ВЕСТНИК

Сибирского государственного
аэрокосмического университета
имени академика М. Ф. Решетнева

Выпуск 5 (31)

(по материалам XII Международного симпозиума
по непараметрическим методам
в кибернетике и системному анализу)

Красноярск 2010

ВЕСТНИК

Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева

Выпуск 5 (31)

Грант РФФИ № 10-01-06090-2 (дополнительное соглашение № 10-0839/22 от 25 марта 2010 г.)

Главный редактор
доктор экономических наук
профессор
Г. П. Беляков

Заместители главного редактора:
доктор технических наук
профессор
И. В. Ковалев

доктор технических наук
профессор
Л. Г. Шаймарданов

Редакционная коллегия:
С. С. Аплеснин
Е. Н. Головенкин
Ю. В. Ерыгин
М. В. Краев
В. Д. Лаптенок
А. Н. Ловчиков
А. В. Медведев
В. В. Москвичев
В. А. Охорзин
К. Г. Охоткин (отв. секретарь)
С. И. Сенашов
В. Г. Сомов

Учредитель
Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева

© Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева, 2010
© Коллектив авторов, 2010

К СВЕДЕНИЮ ЧИТАТЕЛЕЙ

«Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева» (Вестник СибГАУ) – межрегиональный научный журнал, содержащий результаты научных исследований в области естественных, технических и прикладных наук по вопросам проектирования, производства и эксплуатации авиационной и ракетно-космической техники, а также экономики

Представленные в журнал статьи публикуются после обязательного рецензирования и при оформлении их в соответствии с требованиями редакции (www.vestnik.sibsau.ru)

Периодичность – 4 раза в год
Специальный выпуск – 1 раз в год

Электронная версия журнала представлена на сайте Научной электронной библиотеки (http: //www.elibrary.ru) и сайте журнала (www.vestnik.sibsau.ru)

Журнал включен в каталоги:
российской прессы «Роспечать» (МАП) – 11399;
объединенный «Пресса России» – 39263

АДРЕС РЕДАКЦИИ:

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева
Россия, 660014, Красноярск, проспект имени газеты «Красноярский рабочий, 31, П-415
Тел./факс (391)291-90-19
E-mail: vestnik@sibsau.ru

Журнал зарегистрирован Федеральной службой по надзору за соблюдением законодательства в сфере массовых коммуникаций и охраны культурного наследия, свидетельство ПИ № ФС77-22189 от 27.10.2005 г.

Журнал включен ВАК России в «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора наук»

При перепечатке или цитировании материалов из журнала «Вестник СибГАУ» ссылка обязательна

Редакторы Т. Е. Ильющенко, К. С. Мирошникова, Е. Г. Некрасова
Редактор английского текста Е. А. Гончаров
Компьютерная верстка М. А. Белоусовой
Подп. в печать 30.12.2010. Формат 70×108/16.
Бумага офсетная. Печать плоская. Усл. печ. л. 24,0.
Уч.-изд. л. 52,0. Тираж 1000 экз. С 168/10. Заказ 87/80.
Редакционно-издательский отдел
Сиб. гос. аэрокосмич. ун-та.
660014, г. Красноярск,
просп. им. газ. «Красноярский рабочий», 31.
Отпечатано в типографии ИП Суховольской Ю. П.
660125, г. Красноярск, ул. Водопьянова, 2.

ВСТУПИТЕЛЬНОЕ СЛОВО

В сентябре 2010 года Красноярск принимал участников XII Международного симпозиума по непараметрическим методам в кибернетике и системному анализу. Его организаторами были Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева, Национальный исследовательский Томский государственный университет, Институт вычислительного моделирования СО РАН. В настоящий «Вестник» вошли некоторые работы, представленные на симпозиуме.

Симпозиум был посвящен развитию современных математических методов для создания интеллектуальных компьютерных систем различного назначения, функционирующих в условиях неполных данных об исследуемом процессе. Обсуждались также проблемы системного анализа.

Тематика симпозиума представлена следующими секциями:

– непараметрическая и робастная статистика;

– адаптивные и обучающиеся системы;

– непараметрические системы управления;

– оптимизация и принятие решений;

– идентификация и моделирование;

– распознавание образов, классификация и обработка изображений;

– проблемы системного анализа;

– управление экономическими и организационными системами;

– искусственный интеллект и экспертные системы;

– разработка и применение компьютерных систем различного назначения и систем автоматизации.

Красноярск не впервые становится местом проведения симпозиума; проходил он также и в Томске, Иркутске, Железногорске, Дивногорске, Шушенском. Традиционно в первой половине дня заслушивались два пленарных доклада, во второй половине дня – секционные доклады. Результаты научных исследований широко обсуждались участниками симпозиума в течение всего времени его проведения.

Во время работы симпозиума спонтанно возникали дискуссии по различным проблемам кибернетики, информатики, системного анализа, менеджмента, образования. Материалы одной из дискуссий, посвященной образовательному процессу в нашей стране, приведены в данном «Вестнике».

Считаем необходимым отметить большую и эффективную работу доцентов кафедры системного анализа и исследования операций СибГАУ Е. Д. Агафонова и О. В. Шестерневой, аспиранта кафедры А. С. Орловой, студентов кафедры Л. Н. Голуб и Д. В. Безмен по организации и проведению симпозиума и подготовке материалов к опубликованию.

Профессор А. В. МЕДВЕДЕВ

Профессор Ф. П. ТАРАСЕНКО

Раздел
1

КИБЕРНЕТИКА, СИСТЕМНЫЙ АНАЛИЗ, ПРИЛОЖЕНИЯ

УДК 62.501

А. В. Медведев

К ТЕОРИИ НЕПАРАМЕТРИЧЕСКИХ СИСТЕМ

Приводятся краткие сведения о параметрической теории управления дискретно-непрерывными процессами в частности, теории дуального управления и параметрической теории адаптивных систем. Обсуждается вопрос о месте теории непараметрических систем в общей теории управления. Рассматриваются некоторые модели и алгоритмы управления в условиях непараметрической неопределенности.

Ключевые слова: дискретно-непрерывный процесс, параметрическая теория, дуальное управление, непараметрические методы, адаптивное управление, априорная информация, неопределенность, случайные помехи.

Теория – в виду практики.

Девиз конгрессов IFAC

Истинная теория должна заключаться о одном простом, единственном начале, откуда явление берется как необходимое следствие, со всем своим разнообразием.

Н. И. Лобачевский

Введение. Ранее [1] были рассмотрены различные уровни априорной информации, возникающие в различных задачах кибернетики. Ниже мы проанализируем два уровня априорной информации: параметрический и непараметрический.

Пусть $u(t)$ – управляющее воздействие, поступающее на объект, а $x(t)$ – выход объекта, задающее воздействие обозначим $x^*(t)$, (t) – непрерывное время, $\{u_t, x_t, t = \overline{1,s}\}$ – выборка наблюдений $u(t)$ и $x(t)$ в дискретное время t объемом s. Обычно в качестве критерия в теории управления берется функционал:

$$R = \int_0^T (x(t) - x^*(t))^2 dt, \qquad (1)$$

где T – фиксированная величина, $0 \le t \le T$.

Движение динамического процесса, например, может быть задано уравнением

$$a_0 \frac{d^n x(t)}{dt^n} + a_1 \frac{d^{n-1} x(t)}{dt^{n-1}} + \ldots + a_n x(t) = u(t), \quad (2)$$

где a – вектор параметров.

В задаче об оптимальном быстродействии требуется перевести объект из одного состояния в другое за минимальное время T при любых начальных условиях и любой функции $x^*(t)$. При этом $|u(t)| < U^0$, где U^0 – ограничение на управление [2].

Теория дуального управления. Феномен дуализма в системах управления был открыт в 1962 г. А. А. Фельдбуамом и в последующем существенно развит им и его последователями. Сущность дуализма состоит в том, что управляющие воздействия носят двойственный характер. Они, как замечает А. А. Фельдбаум, «должны быть в известной мере изучающими, но в известной мере направляющими» [2].

Предметом изучения является дискретно-непрерывная система (рис. 1).

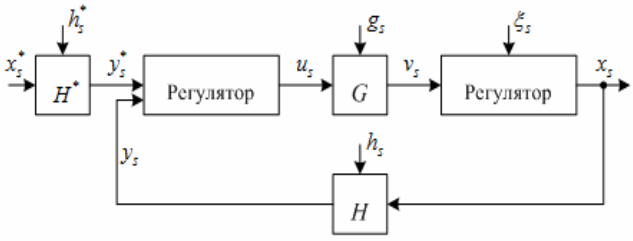


Рис. 1

Приняты следующие обозначения: x_s^* – задающее воздействие, которое через канал H^* смешивается с шумом h_s^* и поступает в качестве y_s^* в регулятор; выход объекта x_s проходя через канал H и смешиваясь с шумом h_s в виде y_s также поступает в регулятор; управляющее воздействие u_s проходя через канал G, смешиваясь с помехой g_s поступает в виде v_s на объект, который находится под воздействием помехи ξ_s; s – дискретное время.

Далее предполагается следующее:

– рассматриваемая задача – байесова, h_s^*, h_s, g_s – последовательности независимых случайных величин с неизменными плотностями вероятности $P(h_s^*)$, $P(h_s)$, $P(g_s)$; $\xi_s = \xi(s, \mu)$, где μ – случайный вектор с известной априорной плотностью вероятности $P(\mu)$. Аналогично полагаем, что $x_s = x(\lambda, s)$, где λ – случайный вектор с заданной плотностью вероятности $P(\lambda)$ и все внешние воздействия – ξ_s, h_s^*, h_s, g_s, x_s^* – статистически независимы;

– объект не имеет памяти и описывается уравнением $x_s = F(\xi_s, v_s)$, где F – ограничена, однозначна и дифференцируема;

– способы комбинации сигнала и шума в блоках H^*, H, G считаются известными и неизменными, а са-

ми блоки не имеющими памяти, т. е. $y_s^* = y^*(x_s^*, h_s)$, $v_s = v(u_s, g_s)$, $y_s^* = y^*(x_s^*, h_s)$ вместо которых и вероятностных характеристик шумов можно сразу задать условные плотности вероятности $P(v_s / u_s)$, $P(y_s / x_s)$, $P(y_s^* / x_s^*)$.

Задача состоит в определении оптимальной стратегии регулятора.

Постановка задачи. Введем удельную функцию потерь $W_s = W(s, x_s, x_s^*)$, тогда общая функция потерь W примет вид

$$W_s = \sum_{s=0}^{n} W(s, x_s, x_s^*). \tag{3}$$

Назовем оптимальной систему, для которой полный риск минимален

$$R = M\{W\} = \sum_{s=0}^{n} M\{W(s, x_s, x_s^*)\} = \sum_{s=0}^{n} R_s, \tag{4}$$

где R_s – удельный риск.

Будем считать, что регулятор в общем случае обладает памятью и характеризуется случайной стратегией. Введем временные векторы $\vec{u}_s = (u_0, ..., u_s)$, $\vec{x}_s = (x_0, ..., x_s)$ и по аналогии $\vec{v}_s, \vec{x}_s^*, \vec{y}_s^*, \vec{y}_s$, $0 \le s \le n$.

Теперь поставим задачу отыскания оптимальной случайной стратегии регулятора, т. е. оптимальных плотностей вероятности

$$\vec{v}_s, \vec{x}_s^*, \vec{y}_s^*, \vec{y}_s, 0 \le s \le n, \tag{5}$$

при которых полный риск R минимален.

Поскольку Γ_s – суть плотность вероятности, то

$$\Gamma_s \ge 0, \quad \int_{\Omega(u_s)} \Gamma_s(u_s) d\Omega = 1, \tag{6}$$

где $\Omega(u_s)$ – область возможных значений u_s; $d\Omega$ – ее бесконечно малый элемент. Γ_s, $s = 0, ..., n$ называется удельными стратегиями, а их совокупность – полной стратегией.

Синтез оптимальной стратегии. Первый этап задачи состоит в выводе формулы полного риска. Сначала найдем условный удельный риск r_s. Последний в теории дуального определяется следующим образом:

$$r_s = M\{W_s / \vec{y}_s^*, \vec{u}_{s-1}^*, \vec{y}_{s-1}\} = \int_{\Omega(\lambda, x_s)} W_s(s, x_s^*(s, \lambda), x_s) P(\lambda, x_s / \vec{y}_s^*, \vec{y}_{s-1}, \vec{x}_{s-1}) d\Omega. \tag{7}$$

По теореме умножения вероятностей (7) можно записать

$$P(\lambda, x_s / \vec{y}_s^*, \vec{y}_{s-1}, \vec{x}_{s-1}) = P(\lambda / \vec{y}_s^*, \vec{y}_{s-1}, \vec{x}_{s-1}) P(x_s / \vec{y}_s^*, \vec{y}_{s-1}, \vec{x}_{s-1}). \tag{8}$$

Учитывая, что λ в (8) зависит только от $\vec{y}_s^*$, то

$$P(\lambda / \vec{y}_s^*, \vec{y}_{s-1}, \vec{x}_{s-1}) = P(\lambda / \vec{y}_s^*) = P_s(\lambda), \tag{9}$$

где $P_s(\lambda)$ обозначает апостериорную плотность вероятности λ в момент времени s на основе наблюдений $\vec{y}_s^*$. Поскольку во втором сомножителе (8) фиксация λ ничего не меняет, раз уже зафиксировано $\vec{y}_s^*$, то

$$P(x_s / \lambda, \vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}) = P(x_s / \vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}). \tag{10}$$

Подставляя (8) в (7), с учетом (9) и (10) получим

$$r_s = \int_{\Omega(\lambda, x_s)} W_s(s, x_s^*(\lambda, s), x_s) P(\lambda) \times \times P(x_s / \vec{y}_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}) d\Omega. \tag{11}$$

Следуя [2], выражение (11) преобразуется к виду

$$r_s = \int_{\Omega(\lambda, \mu, x_s, u_s)} W_s(s, x_s^*(s, \lambda), x_s) \frac{P(\lambda) P(\vec{y}_s^* / \lambda)}{P(\vec{y}_s^*)} \times \times P(x_s / \mu, s, u_s) \frac{P(\lambda) \prod_{i=0}^{S-1} P(y_i / \mu, i, u_i)}{P(\vec{y}_{s-1}, \vec{u}_{s-1} / \vec{y}_s^*)} \times \times \prod_{i=0}^{S} \Gamma(u_i / \vec{y}_i^*, \vec{u}_{i-1}, \vec{y}_{i-1}) d\Omega. \tag{12}$$

Усреднение r_s по $\vec{y}_s^*$, $\vec{y}_s^*$, $\vec{u}_{s-1}$ дает

$$R_s = \int_{\Omega(\vec{y}_s^*, \vec{y}_{s-1}, \vec{u}_{s-1})} r_S P(\vec{y}_s^*, \vec{y}_{s-1}, \vec{u}_{s-1}) d\Omega, \tag{13}$$

и с учетом

$$P(\vec{y}_s^*, \vec{y}_{s-1}, \vec{u}_{s-1}) = P(\vec{u}_{s-1}, \vec{y}_{s-1} / \vec{y}_s^*) P(\vec{y}_s^*) \tag{14}$$

получим средний удельный риск

$$R_s = \int_{\Omega(\lambda, \mu, x_s, \vec{y}_s^*, \vec{u}_s, \vec{y}_{s-1})} W_s(s, x_s^*(s, \lambda), x_s) P(\lambda) \times \times P(x_s / \mu, s, u_s) P(\mu) \prod_{i=0}^{S} P(\vec{y}_s^* / i, \lambda) \times \times \prod_{i=0}^{S-1} P(y_i / \mu, i, u_i) \prod_{i=0}^{S} \Gamma(u_i / \vec{y}_i^*, \vec{u}_{i-1}, \vec{y}_{i-1}) d\Omega. \tag{15}$$

Проанализируем формулу среднего удельного риска (15), характеризующего качество работы системы, представленной на рис. 1. Хотя объект не имеет памяти, но R_s зависит от всей последовательности Γ_i, $i = 1, ..., s$, т. е. выбор Γ_i, $i = 1, ..., s$ влияет на величину R_s к s-му такту времени. В этом факте и содержится феномен дуализма. Рассмотрим k-ый момент времени $k < s$. Запишем полный риск в виде

$$R = \sum_{i=1}^{S} R_i = \sum_{i=1}^{k-1} R_i + R_k + \sum_{i=k+1}^{S} R_i. \tag{16}$$

Анализ (16) показывает, что на k-ом шаге первая составляющая дает ошибку при прохождении $(k-1)$-го такта управления; второе слагаемое названо А. А. Фельдбаумом риском действия – раз мы находимся на k-ом шагу, а третье слагаемое представляет собой риск изучения. Поэтому, чтобы управление было оптимальным для s шагов, необходимо на каждом такте управления минимизировать сумму рисков действия и изучения, т. е. для $k = 1, 2, ..., s$.

Для определения оптимальной стратегии дуального управления А. А. Фельдбаум использует метод ди-

намического программирования. Введем вспомогательные функции:

$$\alpha_s = \int\limits_{\Omega(\lambda,\mu,x_s)} W(s, x_s^*(s,\alpha), x_s)P(\lambda)\prod_{i=0}^{s} P(y_i^*/i,\lambda)\times \\ \times P(x_s/\mu, s, u_s)P(\mu)\prod_{i=0}^{s-1} P(y_i/\mu, i, u_i)d\Omega \quad (17)$$

и

$$\beta_s = \prod_{i=0}^{s}\Gamma_i = \prod_{i=0}^{s}\Gamma_i(u_i/y_i^*, \vec{u}_{i-1}, \vec{y}_{i-1}). \quad (18)$$

Выпишем с учетом последних формул выражение для R_s

$$R_s = \int\limits_{\Omega(\vec{u}_s, \vec{y}_{s-1}, y_s^*)} \alpha_s(\vec{u}_s, \vec{y}_{s-1}, y_s^*)\beta_{s-1}(\Gamma_s)d\Omega = \\ = \int\limits_{\Omega(\vec{u}_s, \vec{y}_{s-1}, y_s^*)} \beta_{s-1}\chi_s(\vec{u}_s, \vec{y}_{s-1}, y_s^*)d\Omega, \quad (19)$$

где

$$\chi_s(\vec{u}_s, \vec{y}_{s-1}, y_s^*) = \\ = \int\limits_{\Omega(u_s)} \alpha_s(u_s, \vec{u}_{s-1}, \vec{y}_{s-1}, y_s^*)\Gamma_s(u_s/\vec{u}_{s-1}, \vec{y}_{s-1}, y_s^*)d\Omega. \quad (20)$$

На основании теоремы о среднем и с учетом (6) можно записать

$$\chi_s = (\alpha_s)_{\text{ср}} \int\limits_{\Omega(u_s)} \Gamma_s\Omega = (\alpha_s)_{\text{ср}} \geq (\alpha_s)_{\min}. \quad (21)$$

Далее необходимо подобрать Γ_s так, чтобы минимизировать R_s, а это возможно, если для любых $\vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s^*$ найти Γ_s так, чтобы χ_s была минимальной. Пусть $u_s = u_s^*$ такое, что $\alpha_s = \alpha_s^* = \alpha_s^{\min}$:

$$\alpha_s^* = \alpha_s(u_s^*, \vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s^*) = \\ = \min_{u_s\in\Omega(u_s)} \alpha_s(u_s, \vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s^*). \quad (22)$$

Величина u_s^* является, очевидно, функцией от $\vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s^*$

$$u_s^* = u_s^*(\vec{u}_{s-1}, \vec{y}_{s-1}, \vec{y}_s^*), \quad (23)$$

но это возможно, если

$$\Gamma_s^* = \delta(u_s - u_s^*), \quad (24)$$

где $\delta(u_s - u_s^*)$ – дельта-функция Дирака, т. е. Γ_s^* является регулярной (нерандомизируемой) стратегией. Если подставить (24) в (20), то получим

$$\chi_s = (\alpha_s)_{u_s=u_s^*} = (\alpha_s)_{\min}, \quad (25)$$

но согласно (21), это и есть минимальное значение для χ_s, следовательно Γ_s^* представляет собой оптимальную стратегию.

Чтобы найти оптимальную стратегию Γ_i^*, $i < s$ нужно последовательно перемещаться от конечно-малого момента времени $i = s$ к началу. Можно показать, что оптимальная стратегия для k-го от последнего такта s равна

$$\Gamma_{s-k}^* = \delta(u_{s-k} - u_{s-k}^*), \quad (26)$$

а

$$u_{s-k}^* = u_{s-k}^*(\vec{u}_{s-k-1}, \vec{y}_{s-k-1}, \vec{y}_{s-k}^*). \quad (27)$$

Отсюда следует важное с практической точки зрения заключение: оптимальное управляющее устройство физически реализуемо, так как к $(s - k)$-му такту для определения u_{s-k}^* необходимы наблюдаемые значения $\vec{u}_{s-k-1}, \vec{y}_{s-k-1}, \vec{y}_{s-k}^*$.

Приведем оптимальный закон управления для простейшего безынерционного объекта при аддитивных нормально-распределенных помехах g и ξ

$$u_s^* = x_s^* - \sum_{i=0}^{s-1}(x_i - u_i)\Bigg/ s + \left(\frac{\sigma_g}{\sigma_\zeta}\right)^2, \quad (28)$$

где σ_g^2, σ_ξ^2 – дисперсии помех g и ξ.

Адаптивное дуальное управление. Проанализируем основную идею адаптивного управления на примере [3]. Пусть линейный объект описывается разностным уравнением

$$x_t = \sum_{i=1}^{n} a_i x_{t-i} + a_0 u_{t-1}, \quad (29)$$

где вектор коэффициентов $c = (a_0, a_1, ..., a_n)$, а вектор ситуаций $Z_t = (x_{t-1}, ..., x_{t-n}, u_{t-1})$, тогда

$$x_t = c^T Z_t, \quad (30)$$

где T – знак транспонирования.

Закон управления примем линейным

$$u_{t-1} = \sum_{j=1}^{m} b_j x_{t-i} \text{ либо } u_{t-1} = b^T Y_t, \quad (31)$$

где вектор параметров $b = (b_1, ..., b_m)$, а $Y_t = (x_{t-1}, ..., x_{t-m})$ – вектор входных переменных управляющего устройства.

Идентификация объекта осуществляется с помощью алгоритма «изучения»

$$c_t = c_{t-1} + \gamma_t^c(x_t - c_{t-1}^T z_t)z_t \quad (32)$$

при квадратичном критерии качества $M\left\{\left(x_t - c^T z_t\right)^2\right\}$ идентификации.

Алгоритм управления может быть представлен с учетом (29)–(32) в следующем виде:

$$b_t = b_{t-1} + \gamma_t^b(x^*_t - c_{t-1}^T z_t)V_t c_t, \quad (33)$$

где x^*_t – задающее воздействие, а V_t имеет вид матрицы функций чувствительности размера $m\times(m+1)$.

$$V_t = \begin{Vmatrix} \dfrac{\partial x_{t-1}}{\partial b_{t-1}^1} & \cdots & \dfrac{\partial x_{t-m}}{\partial b_{t-1}^1} & \dfrac{\partial u_{t-1}}{\partial b_{t-1}^1} \\ \cdots & & \cdots & \cdots \\ \dfrac{\partial x_{t-1}}{\partial b_{t-1}^m} & \cdots & \dfrac{\partial x_{t-m}}{\partial b_{t-1}^m} & \dfrac{\partial u_{t-1}}{\partial b_{t-1}^m} \end{Vmatrix}. \quad (34)$$

Таким образом, адаптивные алгоритмы дуального управления представлены формулами (30)–(33), а структурная схема на рис. 2.

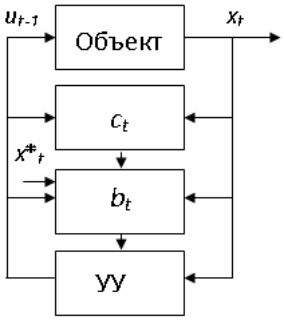


Рис. 2

Видно, что она представляет собой объединение блоков идентификации (оценка параметров c) и управление (оценка параметров b). В этой адаптивной системе каждое новое значение состояния объекта x_t вызывает изменение параметров c_t и b_t. Возможны и иные стратегии «изучения» и «управления» [3]. Блок УУ реализует закон управления (31).

Заметим, что математическая технология синтеза рекуррентных параметрических адаптивных алгоритмов дуального управления существенно отличается от рассмотренных выше, которые основаны на Байесовой постановке задачи синтеза оптимальной стратегии управления.

Непараметрическое дуальное управление. Вернемся к объекту, показанному на рис. 1. В теории дуального управления [2] и в теории адаптивных систем [3] предполагается математическое описание объекта с точностью до вектора параметров (2), (29). В большинстве случаев, априорной информации недостаточно, чтобы обосновано выбрать управление исследуемого процесса. Поэтому приходится проводить серию экспериментов на объекте (часто длительных и дорогостоящих), чтобы качественно, с практической точки зрения, решить задачу идентификации.

В условиях непараметрической неопределенности [4] уравнение процесса с точностью до вектора параметров неизвестно, но известны свойства объекта качественного характера, например, однозначность характеристик или неоднозначность для безынерционных процессов; линейность или тип нелинейности – для динамических. Если вид уравнения, описывающего процесс, неизвестен, то известные параметрические методы теории управления [5; 6] не применяются для решения задач идентификации и управления.

Из соображений простоты рассмотрим схему, представленную на рис. 3.

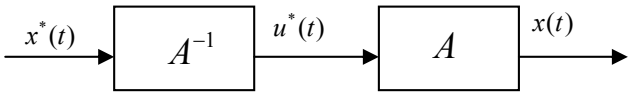


Рис. 3

На схеме A является неизвестным оператором, описывающим процесс, т. е.

$$x(t) = A < u(t) > . \qquad (35)$$

Если существует оператор, обратный A, т. е. A^{-1}, $AA^{-1} = I$ – единичный оператор, то

$$u(t) = A^{-1} < x(t) > . \qquad (36)$$

Задавая теперь траекторию $x^*(t)$, находим из (36) идеальное значение $u^*(t)$. Таким образом, (36) может быть отнесен к категории идеальных регуляторов. В дальнейшем будем его называть u-регулятор, чтобы отличить от многих известных. Однако проблема состоит в том, что в большинстве случаев его построить нельзя, тем более, что оператор A – неизвестен. Попытки, как то, хотя бы частично, решить эту проблему введением в УУ корректирующих цепочек, компенсирующих звеньев и т. п. предпринимались. В некоторых технических системах это приводило к успеху [6].

В 1950-х гг. прошлого столетия академиком В. С. Кулебакиным был предложен и существенно развит метод $K(D)$-изображений, который привел к появлению теории инвариантности автоматически регулируемых и управляемых систем. Но в этом случае, необходима высокая точность описания исследуемых процессов дифференциальными уравнениями. Если вид уравнения, оценивающего исследуемый процесс, неизвестен, то классические методы теории управления не применяются.

Рассмотрим частный случай. Пусть объект описывается линейным дифференциальным уравнением неизвестного порядка, например n, n – неизвестно. В этом случае при нулевых начальных условиях $x(t)$ равен [7]

$$x(t) = \int_0^t h(t-\tau)u(\tau)d\tau, \qquad (37)$$

где $h(t - \tau)$ – весовая функция системы, являющаяся производной переходной функции $k(t)$, т. е. $h(t) = k'(t)$. Известно, что обратным оператором (37) является оператор [7]

$$u(t) = \int_0^t v(t-\tau)x(\tau)d\tau, \qquad (38)$$

где $v(t)$ – весовая функция объекта в направлении «выход–вход» и $v(t) = w'$ (t), где $w(t)$ – переходная функция системы в том же направлении. В этом случае (см. рис. 3), A представлен оператором (37), а A^{-1} – выражением (38). Следовательно, теперь проблема состоит в отыскании весовой функции $h(t)$. Один из возможных путей решения этого вопроса состоит в решении уравнения Винера–Хопфа. Другой – в снятии переходной характеристики на реальном объекте с последующей оценкой его весовой функции по результатам измерений $\{x_i = k_i, t_i, \ i = \overline{1, s}\}$.

Непараметрическая модель (37) будет иметь вид

$$x_s(t) = \int_0^t h_s(t-\tau, \vec{k}_s, \vec{t}_s)u(\tau)d\tau, \qquad (39)$$

где $\vec{k}_s$, $\vec{t}_s$ – временные векторы $\vec{k}_s = (k_1, ..., k_s)$, $\vec{t}_s = (t_1, ..., t_s)$, а $h_s(\cdot)$ равна

$$h_s(t) = \frac{1}{sc_s}\sum_{i=1}^{s} k_i H'\left(\frac{t-t_i}{c_s}\right), \qquad (40)$$

$H(\cdot)$ – колоколообразные (ядерные) функции; c_s – параметр размытости, удовлетворяющий условиям сходимости [4].

Весовую функцию $v(t)$ в направлении «выход–вход», а также переходную $w(t)$ на объекте «снять» нельзя. Было предложено переходную функцию «вспять» снять на модели. По-видимому, впервые это было сделано в [8]. Таким образом, из соотношения

$$x_s(t) = 1(t) = \int_0^t h_s(t-\tau, \vec{k}_s, \vec{t}_s)u(\tau)d\tau, \qquad (41)$$

можно получить выборки $\{u_j, t_j,\ j = \overline{1, s}\}$. Тогда непараметрический алгоритм управления линейной динамической системы примет вид

$$u_s^*(t) = \int_0^t \left(\frac{1}{sc_s}\sum_{j=1}^{s} w_j H'\left(\frac{t-\tau-t_i}{c_s}\right)\right)x^*(\tau)d\tau. \qquad (42)$$

Ясно, что объемы выборок при «снятии» переходных характеристик на реальном объекте и модели могут не совпадать. Фрагмент работы алгоритма (42) будет представлен ниже. Отметим лишь, что была реализована схема, соответствующая рис. 3, неизвестные операторы A и A^{-1} оценивались по исходным переходным характеристикам процесса (уравнение процесса было неизвестно) в классе непараметрических статистик.

Поскольку операторы A и A^{-1} по реальным данным будут оценены неточно, то возникает необходимость несколько изменить схему, представленную на рис. 3 (рис. 4).

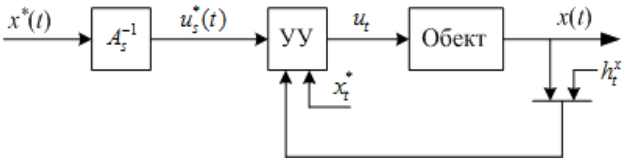


Рис. 4

Непараметрической оценкой обратного оператора объекта в данном случае является A_s^{-1}, u_s^* – выход (оценка A^{-1}), помеха h_t^x действует в канале обратной связи (см. рис. 4). Непараметрический алгоритм дуального управления имеет вид

$$u_{s+1,t} = u_{s,t}^* + \Delta u_{s+1,t}, \qquad (43)$$

где $u_{s,t}^*$ определяется по формуле (42), а $\Delta u_{s+1,t} = \varepsilon(x_t^* - x_t, s)$ – поисковые шаги.

Таким образом, в $u_{s,t}^*$ сосредоточены «знания» об объекте, а $\Delta u_{s+1,t}$ – «изучающие» поисковые шаги. В этом и состоит дуализм алгоритма (43).

Поясним его на примере безынерционного объекта $x = f(u)$, в качестве оценки которого примем непараметрическую оценку функции регрессии по наблюдениям $\{x_i, u_i,\ i = \overline{1, s}\}$ [4]:

$$x_s(u) = \sum_{i=1}^{s} x_i \Phi\left(\frac{u-u_i}{c_s}\right) \Big/ \sum_{i=1}^{s} \Phi\left(\frac{u-u_i}{c_s}\right), \qquad (44)$$

где колоколообразные функции $\Phi(\cdot)$ и параметр размытости c_s удовлетворяют некоторым условиям сходимости [4]. Аналогом выражения (42) в этом случае будет $u = f^{-1}(x)$, а u_s^* из (43) будет равно

$$u_{s,t}^* = \sum_{i=1}^{s} u_i \Phi\left(\frac{x_t^* - x_i}{c_s}\right) \Big/ \sum_{i=1}^{s} \Phi\left(\frac{x_t^* - x_i}{c_s}\right). \qquad (45)$$

Ясно, что класс функций $x = f(u)$ – взаимнооднозначные и непрерывные. В дальнейшем мы увидим, что он может быть расширен.

Проанализируем характер дуализма алгоритма (43). На начальной стадии управления основная роль принадлежит второму слагаемому $\Delta u_{s+1,t}$ формулы (43). Это случай активного накопления информации в системе дуального управления, который начинается с появления первого наблюдения входной и выходной переменной объекта. По мере процесса обучения (накопления информации) все возрастающую роль при формировании управляющего воздействия $u_{s+1,t}$ начинает играть первое слагаемое, т. е. $u_{s,t}^*$. Таким образом, в процессе дуального управления объектом фигурируют как этап изучения объекта, так и этап приведения его к цели.

Численные эксперименты. Ниже приведем некоторые результаты расчетов, иллюстрирующих частные задачи при исследовании динамических и безынерционных объектов.

Первый эксперимент иллюстрирует «снятие» весовых функций линейной динамической системы. На рис. 5 показаны аналитическая весовая функция полученная при решении дифференциального уравнения типа (2), когда в правой части принята дельта-функция, здесь же приведены весовые функции для различных аналогов дельта-функций.

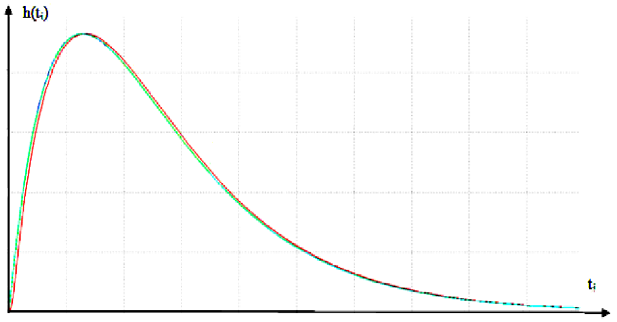


Рис. 5

Весовые функции, показанные на рис. 5, получены при различных дельтаобразных ступенчатых функциях следующего вида:

$$\delta_1(t) = \begin{cases} \Delta t^{-1}, \text{если } \Delta t \in [0;\, 0,1]; \\ 0, \text{если } \Delta t \notin [0;\, 0,1]; \end{cases}$$

$$\delta_2(t) = \begin{cases} \Delta t^{-1}, \text{если } \Delta t \in [0;\, 0,01]; \\ 0, \text{если } \Delta t \notin [0;\, 0,01]; \end{cases}$$

$$\delta_3(t) = \begin{cases} \Delta t^{-1}, \text{если } \Delta t \in [0;\, 0,001]; \\ 0, \text{если } \Delta t \notin [0;\, 0,001]. \end{cases}$$

Как видно, весовая функция, полученная аналитическим путем, а также весовые функции, полученные при воздействии на объект, существенно отличающихся дельтаобразных функций практически совпадают. Это приводит к выводу, что на практике в эксперименте по наблюдению весовых функций возможна подача на вход объектов, существенно отличающихся дельтаобразных управляющих воздействий. Существенным здесь, однако, является связь (соотношение) двух элементов: «динамический объект (процесс) – дифференциальное уравнение».

Результаты управления линейным динамическим объектом, соответствующие схеме представленной на рис. 3, показаны на рис. 6, при условии, что x_t^* – случайная величина, генерируемая датчиком равномерно распределенных случайных чисел.

Эксперимент был проведен по следующей схеме: сначала на объекте снимались переходные характеристики и с их использованием оценивался оператор A по формуле (39) и обратный оператор A^{-1} по формуле (42). Благодаря рис. 6, можно увидеть хорошее качество управления даже в таком «экзотическом» случае. С подобной задачей не справится ни один из известных регуляторов.

В случае безынерционного объекта рассмотрим следующую задачу: пусть имеется выборка наблюдений $\{x_i,\ u_i,\ i = \overline{1, s}\}$. В качестве оценки $M\{x|u\}$ и $M\{u|x\}$ примем оценки функции регрессии вида (44) и (45) (рис. 7).

Выборка наблюдений и непараметрическая оценка $M\{x|u\}$ (44) показаны на рис. 7, *а*, а непараметрическая оценка $M\{x \,|\, M\{u \,|\, x^*\}\}$ – на рис. 7, *б*.

В качестве характеристики объекта был взят отрезок параболы (рис. 7, *а*). Результаты функционирования схемы, представленной на рис. 3, приведены на рис. 7, *б*.

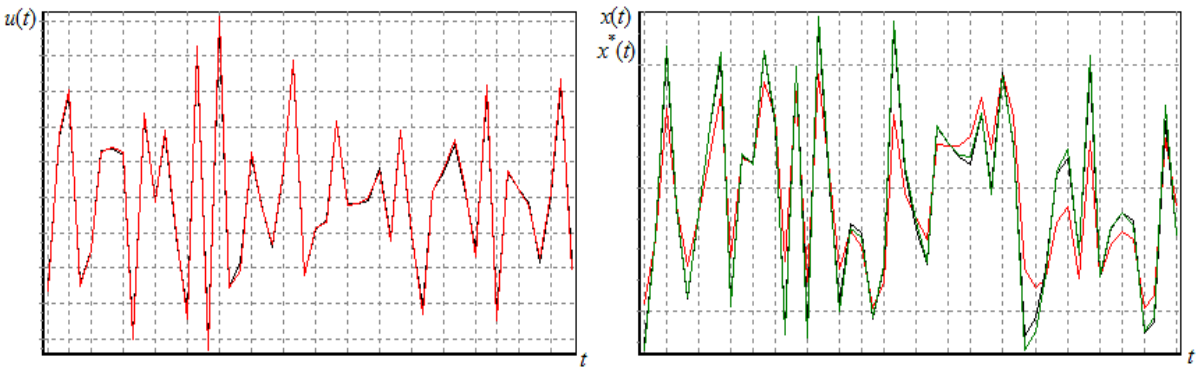
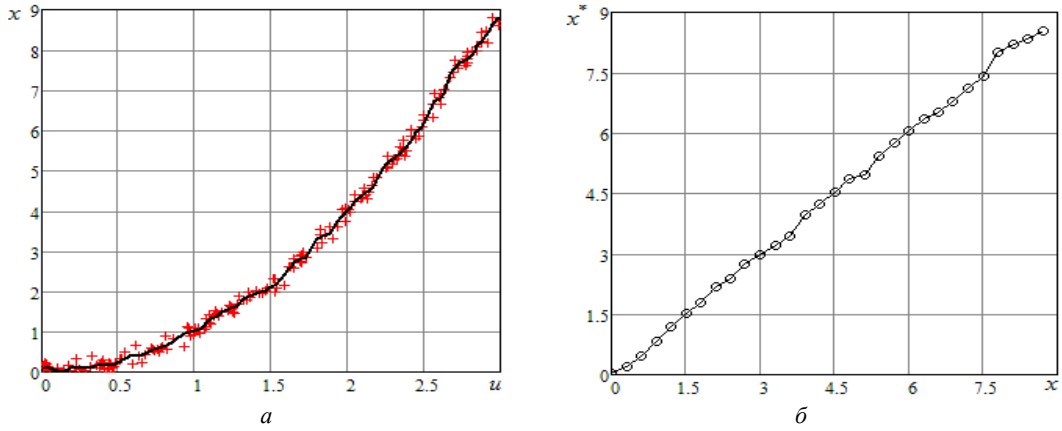


Рис. 6



а *б*

Рис. 7

Контроль переменных, измерения. Здесь мы подчеркнем важность проблемы измерения «входных–выходных» переменных исследуемого объекта, процесса. Ранее [9; 10] уже отмечалось, что отличающиеся средства контроля даже для одних и тех же процессов приводят к различным формулировкам задач идентификации. Главное, что следует выделить в этой проблеме, состоит в том, что нередко динамический объект мы вынуждены рассматривать как статический с запаздыванием из-за длительной процедуры контроля (измерения, анализа) некоторых переменных, существенно превышающей постоянную времени объекта.

Безусловно, при моделировании и управлении дискретно-непрерывными процессами, целесообразно использовать сигналы или аналогичные им, но это требует тщательного анализа не только самого конкретного объекта, но и средств и технологии контроля всех доступных переменных, а также априорной информации, которая одновременно по различным каналам измерения переменных многомерной системы объекта может соответствовать различным уровням априорной информации. Неучет тех или иных переменных, параметров, характера измерения и контроля, априорной информации, а также некоторая «вольность» при принятии тех или иных допущений, неизбежных при математической постановке задачи, может привести, в конечном счете, к негативным последствиям. Вся эта сумма вопросов часто обходится при исследовании проблемы моделирования с теоретической точки зрения [11; 12]. При решении же прикладных задач, построении моделей конкретных процессов, это просто невозможно, ибо истина ничуть не страдает от того, если кто-либо ее не признает (И. Ф. Шиллер). Представляется уместным еще раз акцентировать внимание исследователя на формулировку проблемы идентификации реального процесса на самой начальной стадии: гораздо труднее увидеть проблему, чем найти ее решение. Для первого требуется воображение, а для второго – только умение (Д. Д. Бернал).

Отметим еще одну важную черту, которая сопутствует измерению многих переменных. Это непредставительность пробы, предназначенной для контроля. Проблема здесь состоит в том, что результат измерения (анализа) тех или иных переменных присваивается целой партии продукта (изделия). При этом, для анализа берутся десятки грамм продукта, а результат анализа присваивается многотонной партии. Следует заметить, что сам анализ проводится с высокой точностью, например, химический, физико-химический, физико-механический и др. Существенно другое: где и как брать пробу? В некоторых отраслях это регламентируется ГОСТом, в других случаях приняты те или иные рекомендации. Одним словом, вопрос очень серьезный и требует тщательного анализа в каждом конкретном случае. Неточности на этой стадии часто приводят к очень «грубым» моделям процесса, а, следовательно, и к неудовлетворительным системам управления. Мы здесь не будем обсуждать проблемы разрушающего контроля. Это отдельный, самостоятельный вопрос, требующий специального исследования.

Математические постановки задач моделирования и управления. Совершенно очевидным является факт наличия существенно различной априорной информации об исследуемом процессе [10]. Как следствие этого – различные математические постановки задач, с точки зрения математической строгости. Одним из основных «камней преткновения» на этом пути является несоответствие наших предположений об исследуемом объекте самому объекту. После традиционно произносимого «Пусть процесс…», следуют такие предположения, гипотезы, которые, к сожалению, часто имеют отдаленное отношение к реальности. Трудно представить себе процесс, объект, характеристики которого были бы неизменными или менялись бы по известному закону с течением времени. Мы имеем в виду процессы, описанные в [9], средства и технологии измерения переменных объектов, которые представляют интерес в теории автоматического управления. Основные их черты состоят в недостатке априорной информации, воздействии случайных факторов, характеристики которых нам неизвестны, недостаток и несовершенства средств контроля переменных, непредставительность отбора проб для измерений и др. Наше незнание приходится, к сожалению, заменять, говоря «Пусть…». Ясно, что если наши допущения достаточно близки к реальности, то в итоге можно рассчитывать на успех при решении той или иной задачи, если же нет, то неудача неизбежна. Действительно, многие процессы и объекты, в основе функционирования которых лежат фундаментальные законы физических, химических, электрических, механических и других явлений, могут быть описаны с высокой степенью точности. Соответственно, для них могут создаваться и модели и системы управления достаточно высокого качества, что во многих случаях имеет место [5; 6].

Если же допущения слишком «грубые», то, видимо, есть два пути. Первый – восполнение нашего «незнания» о процессе, когда можно будет сделать аккуратную, с математической точки зрения, постановку задачи. Второй путь состоит в развитии математического подхода, адекватного тому уровню априорной информации, которым мы реально располагаем.

В этой связи, хотелось бы напомнить некоторые факты, известные, например из статьи Р. Калмана [13]. Приведем некоторые выдержки из этой статьи: «…классический (колмогоровский) вероятностный подход не может работать в реальных задачах с недостоверными данными. Для того, чтобы моделировать неопределенность при помощи вероятностного механизма необходимо иметь чересчур много информации, которая не может быть извлечена из доступных данных в большей массе практических задач». И еще, Л. С. Понтрягин: «Математики не верят в вероятность»; А. Н. Колмогоров: «…со статистикой что-то не в порядке».

Несколько отличающаяся аксиоматика теории вероятностей изложена в [14].

Нам предстоит в будущем моделировать и управлять реальными процессами, описанными в [9], включая организационные, потому что этого требует реальность, практика. В частности многие экономические процессы могут быть отнесены к организационным. Еще в середине прошлого столетия по поводу применения математики в экономике Дж. фон Нейман и О. Моргенштерн писали [15]: «Прежде всего отдадим себе отчет в том, что в настоящее время в экономической теории не существует универсальной системы и, что если она и будет создана, то едва ли это произойдет в ближайшее время. Причина этого кроется в том, что экономика является слишком сложной наукой…». И далее: «Часто аргументация против применения математики состоит из ссылок на субъективные элементы, психологические факторы и т. п. …». «Важно осознать, что экономисты не могут надеяться на более легкую судьбу, чем та, которая постигла ученых других специальностей».

Прошло более полувека, но математики для экономической науки, а также для моделирования и управления организационными процессами не появилось, хотя некоторые продвижения в этом направлении есть: теория размытых множеств, теория принятия решений, системный анализ и теория систем и др.

О теории непараметрических систем. Термины «непараметрическая идентификация», «непараметрические методы обработки данных» встречаются в монографиях по идентификации [11; 12], но непараметрических алгоритмов идентификации не приведено. Обычно непараметрическую идентификацию линейных динамических процессов связывают с отысканием весовых или переходных функций системы в результате решения интегральных уравнений Фредгольма 1-го рода, в частности уравнения Винера-Хопфа [11; 12; 16].

Выше мы говорили о моделях и *и*-регуляторах, которые свободны от выбора с точностью до вектора параметров моделей исследуемого процесса или параметрической структуры управляющих устройств, а также параметрической структуры других характеристик процесса, например, корреляционных функций, спектральных плотностей и др. Таким образом, речь идет об идентификации и управлении в условиях непараметрической неопределенности [4]. Представляется, что это наименьший уровень априорной информации об исследуемом объекте, когда возможно решение широкого класса задач кибернетики, наиболее адекватных реальным процессам. Заметим также, что первые исследования по непараметрическому управлению безынерционными объектами относятся к началу 1970-х гг. прошлого столетия. Можно считать, что теория непараметрических систем охватывает различные задачи кибернетики, ориентированные на непараметрический уровень априорной информации.

Вышеизложенное охватывает некоторые задачи идентификации и управления на уровне параметрической и непараметрической априорной информации. В отличие от хорошо развитой параметрической теории, непараметрическая ориентирована на уровень меньшей априорной информации об исследуемых объектах и процессах. Обращается специальное внимание на системы дуального управления байесового типа, адаптивного дуального управления и непараметрического дуального управления. Приводятся некоторые непараметрические модели и алгоритмы дуального управления, а также частные результаты численных расчетов, имеющих иллюстративный характер.

Библиографические ссылки

1. Медведев А. В. Теория непараметрических систем. Общий подход // Вестник СибГАУ. Вып. 3. Красноярск, 2008.

2. Фельдбаум А. А. Основы теории оптимальных автоматических систем. М. : Физматгиз, 1963.

3. Цыпкин Я. З. Адаптация и обучение в автоматических системах. М. : Наука, 1968.

4. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука, 1983.

5. Методы классической и современной теории автоматического управления. Статистическая динамика и идентификация систем автоматического управления / под ред. К. А. Пупкова, Н. Д. Егупова. Т. 2. М. : Изд-во МГТУ им. Н. Э. Баумана, 2004.

6. Методы классической и современной теории автоматического управления. Синтез регуляторов систем автоматического управления / под ред. К. А. Пупкова, Н. Д. Егупова. Т. 3. М. : Изд-во МГТУ им. Н. Э. Баумана, 2004.

7. Куликовский Р. Оптимальные и адаптивные процессы в системах автоматического регулирования. М. : Наука, 1967.

8. Medvedev A. V. Identification and control for linear dynamic System of unknown order / A. V. Medvedev // Optimization techniques IFIP Technical Conference. Berlin–Heidelberg–New York : Springer-Verlag, 1975. P. 48–56.

9. Медведев А. В. Теория непараметрических систем. Процессы // Вестник СибГАУ. Вып. 3. Красноярск, 2010.

10. Медведев А. В. Теория непараметрических систем. Моделирование // Вестник СибГАУ. Вып. 4. Красноярск, 2010.

11. Эйкхофф П. Основы идентификации систем управления. М. : Мир, 1975.

12. Льюнг Л. Идентификация систем. М. : Наука, 1991.

13. Калман Р. Е. Идентификация систем с шумами // Успехи математических наук. Вып. 4. Т. 40. 1985.

14. Уиттл П. Вероятность. М. : Наука, 1982.

15. Нейман Дж. фон, Моргенштерн О. Теория игр и экономическое поведение. М. : Наука, 1970.

16. Заде Л., Дезоер Ч. Теория линейных систем. М. : Наука, 1970.

A. V. Medvedev

TO THE THEOTY OF NONPARAMETRIC SYSTEMS

In the article the author gives a brief information about parametric theory of control of discrete-continuous processes, in particular about the theory of adaptive systems, and considers a question about the theory of nonparametric systems and its place in the general theory of control in conditions of nonparametric uncertainty.

Keywords: discrete-continuous process, parametric theory, dual control, nonparametric methods, adaptive control, a priori information, uncertainty, random noise.

© Медведев А. В., 2010

УДК 517

Ф. П. Тарасенко

КУЛЬТУРА СУБЬЕКТА И МОДЕЛИРОВАНИЕ

Обсуждается термин «культура», который во многих областях деятельности человека наполнен различным содержанием. Анализируется связь моделей, моделирования и культуры субъекта. Приводятся некоторые аспекты «погружения» субъекта во взаимодействия культуры, моделей и среды.

Ключевые слова: *культура, модели, субъект, система моделей, язык, саморазвивающаяся система, образование, знания, преподавание, математические модели.*

Как и практически любое слово в обыденном языке, термин «культура» используется в разных смыслах. Одни употребляют его в весьма узко специализированном смысле (подготовка бактериальных, микробных или тканевых культур для микроскопического исследования, выведение различных культур растений, окультуривание почв и ландшафтов); другие считают его характеристикой поведения и убеждений конкретных этнических, социальных, возрастных групп (национальная культура, культура конкретной прослойки общества); третьи рассматривают культуру как специфичную часть человеческой деятельности (культура земледелия; например, в миссию университета кроме образовательной функции входит участие в развитии науки и культуры общества), как продукт интеллектуальных или эстетических устремлений: иногда под культурой подразумевают конкретную форму организации жизнедеятельности цивилизации, нации или их исторического периода («западная» и «восточная», греческая культуры, культура средневековья и т. д.).

Обсудим понятие культуры в самом обобщенном смысле, охватывающем все остальные толкования как частные случаи. В таком качестве культура понимается как феномен, присущий объектам только одного типа, которые именуются субъектами. Не совсем ясно, в какой мере и какая часть всех живых существ может быть отнесена к субъектам, но понятие субъект определенно связывается с человеком: оно и отражает в языке кардинальное, качественное отличие человека от остальных объектов материального мира.

Синтетический подход к культуре позволяет сказать, что ее главная внешняя роль состоит в обеспечении взаимодействий субъекта с окружающей средой, в качестве которой выступают природа и общество, частью которых является субъект. Все возможные действия субъекта в этой среде можно разделить на два типа: познание окружающего мира и преобразование его. Покажем, что ни то, ни другое невозможно без моделирования.

Начнем с преобразования действительности. Для этой активности субъекта характерна ее целенаправленность: субъект вносит изменения в неудовлетворяющую его реальность, стараясь приблизить ее к желаемому состоянию – конечной цели, которая должна быть определена до того, как субъект начнет действовать. Но это не само реальное состояние, а его описание, т. е. его модель. Кроме того, поскольку переход из начального, неудовлетворительного состояния в желаемое конечное состояние может происходить по различным траекториям, необходимо определить заранее, по какой именно траектории следует идти, т. е. указать, какие действия и в какой последовательности следует выполнять. Но описание этих действий (план, алгоритм) – это не сами действия, а их модель. Таким образом, любая целенаправленная деятельность невозможна без моделирования, основана на моделировании.

То же относится и к познанию реальности. Во-первых, результаты познания должны быть представлены, зафиксированы, описаны (в виде протоколов наблюдений, таблиц измерений, фото- или аудиозаписей и т. п.), это есть не сами наблюдения, а их описания, т. е. модели (в этом смысле можно сказать, что модели есть форма существования знаний). Во-вторых, без моделей не происходит и сам процесс получения субъектом информации о внешней среде. Это нам только кажется, что – открыл глаза, и информа-

ция сама попадет вам в глаза, потрогай предмет – и узнаешь, холодный он или горячий. На деле же субъект получает информацию о среде посредством моделей, с помощью моделей. Во всех языках есть глаголы смотреть и видеть, и это разные «вещи»: можно смотреть и не видеть, можно видеть то, чего нет (гипноз, иллюзии зрения и т. п.), можно заблокировать осязательные модели (например, инъекцией новокаина), и вы перестанете отличать холодное от горячего. Более того, известны психологические опыты, когда от прикосновения холодного утюга, у человека, которого убедили, что утюг горячий, возникали ожоговые волдыри. Все это говорит о том, что не только финал познания требует моделирования, но и сам процесс восприятия информации происходит посредством моделей.

Откуда же берутся модели, обеспечивающие активность субъекта, если без моделей она невозможна? Для субъекта-личности исходными, первичными моделями являются гены, доставшиеся ему по наследству от родителей. Эти модели уже в нужной форме встроены в его организм при рождении: никто нас не учил дышать, кричать, сосать материнское молоко и т. д. Мы от рождения наделены моделями, обеспечивающими наши каналы связи с внешним миром, – органами чувств и моделями мозга, позволяющими «переводить» наши ощущения в приобретаемые модели действительности.

Врожденные и благоприобретенные модели субъекта и образуют ту среду, посредством которой он общается с внешним миром. Этот «буферный» мир моделей между субъектом и физической окружающей его средой, обеспечивающий его любые взаимодействия с нею, и есть культура субъекта.

Признание теснейшей связи культуры и моделирования имеет ряд далеко идущих последствий.

Первое следствие. Культура является для субъекта «второй природой», в которой он живет и через которую он взаимодействует с «первой природой», реальной, как субъект – по законам культуры, в отличие от его взаимодействий с нею в качестве одного из остальных объектов, по общим для всех объектов законам природы: физическим, химическим, индивидуально-биологическим. Но одновременно субъект находится в культурной среде и, как субъект, может взаимодействовать с «первой природой» только через посредство культуры. Культура является специфичной системой, со своими дополнительными условиями и ограничениями, которые являются столь же обязательными и нерушимыми, как и законы «первой природы». Эти законы – системные правила моделирования, закономерности культуры. При попытках нарушить эти законы, субъекта настигает неудача, а то и гибель, как и при попытках обойти законы природы.

Второе следствие. Культура сугубо индивидуальна – нет ни одного субъекта, тождественного другому. Для субъектов-индивидов это очевидно, так как исходный слой врожденных моделей у каждого свой: кто не наделен абсолютным слухом, тому при всем старании не стать выдающимся музыкантом; от рождения лишенные слуха и зрения люди поневоле будут надстраивать свой мир моделей только на доступной им информации. У некоторых особей от рождения существуют особо замечательно выраженные способности к тем или иным видам деятельности – такие особенности индивида называются талантами. Однако раскрытие их потенциала нуждается в том, чтобы их заметить и поддержать. Но даже при (предположительно) одинаковых генах (как у однояйцовых близнецов) субъект будет наращивать свою культуру из собственного опыта – информации, получаемой из непосредственно окружающей его среды, которая у каждого сугубо индивидуальна. Отсюда и уникальность культуры субъекта.

Индивидуальность культуры характерна и для «коллективных» субъектов, и по аналогичным причинам.

Уникальность культуры каждого субъекта является причиной того, что разные субъекты по-разному воспринимают и оценивают одну и ту же ситуацию и по-разному ведут себя в ней. При необходимости обеспечить единообразие поведения субъектов в определенной ситуации (например, солдат в бою или рабочих в цеху), требуется сформировать у них достаточно близкие фрагменты культуры, обеспечивающие нужное поведение – путем специальной профессиональной подготовки.

В процессах формирования культуры многое познано (и включено в технологии воспитания, образования, управления), но их понимание еще далеко от удовлетворительного. Например, Р. Акофф и Д. Гринберг [1] ставят вопрос о том, что обучение «вообще следует перевернуть с головы на ноги» и жестко критикуют существующую в США систему образования. Внимание других исследователей привлечено к той роли, которую гены играют не только в физическом развитии человека, но и в его поведении в течение всей жизни. В современном менеджменте все большее значение придается вопросам изучения и формирования корпоративной культуры каждой организации.

Третье следствие. Поскольку субъект «погружен» в культуру, взаимодействует со средой «сквозь» нее посредством составляющих ее моделей, субъекту часто представляется, будто его модели реальности и есть сама реальность. Это проявляется практически во всех сферах деятельности субъекта: этнических стереотипах о других нациях, всевозможных других мифах, в том числе эгоцентризме и ксенофобии, да и в повседневных стереотипах, вызывающих межкультуральные конфликты (известны случаи, когда «дикие» африканские племена убивали туристов за то, что те стреляли по птицам, считавшимися священными, или за осквернение священного камня присаживанием на него). У разных народов различны представления о женской красоте, правилах хорошего тона. Эта специфичность индивидуальных моделей открыто, в явной форме подчеркивается и используется в искусствах, особенно живописи, литературе и музыке. Пожалуй, единственным видом деятельности, где внимание постоянно сосредоточено на различиях между моделями и реальностью, является наука. В остальных де-

лах субъекты исходят только из своих представлений о реальности. Например, профессионалы-управленцы редко отдают себе отчет в том, что причиной их неудач являются не «вражьи происки», неумелость исполнителей или нехватка ресурсов, а неадекватность их собственных представлений о реальности, основываясь на которых они принимали свои решения.

Роль априорных предположений в работе с любой системой чрезвычайно велика. Это должно быть предметом особого внимания в управлении социальными системами.

Четвертое следствие. Деление культуры на духовную и материальную вполне соответствует классификации моделей на абстрактные и реальные. Абстрактные модели создаются средствами мышления и психики; реальные модели являются материальными предметами, на которых одним из трех способов запечатлена информация о моделируемом оригинале (модели прямого, косвенного и условного подобия [2]).

Пятое следствие. Язык субъекта сугубо индивидуален. Из богатого разнообразия проявлений моделирования выделим лишь один, но очень существенный аспект культуры субъекта – язык. Роль языка в культуре столь велика, что П. Флоренский даже отождествил их: «Культура – это язык, объединяющий человечество» [3].

Язык оказывается объектом, на котором проявляется грубость классификации моделей на абстрактные и реальные: пока человек мыслит в языке, он строит абстрактные модели, а когда он облекает эти модели в речевую или письменную форму, они становятся реальными моделями, которые можно обмерять и обсчитывать. Языкознание является одной из сложнейших естественных наук, которую почему-то относят к гуманитарным. (Это еще один пример упрощенности четкой классификации вообще и, в частности, искусственности деления наук на точные, естественные и гуманитарные.)

Язык является универсальным средством моделирования: на языке можно говорить о чем угодно. Возможность говорить конечными фразами о бесконечно разнообразной реальности приобретается ценой того, что модель отображает реальность не в ее бесконечной неисчерпаемости, а лишь в нужном субъекту отношении (целевая предназначенность моделей), и существенных подробностях, приближенно, с конечной достаточной точностью. При этом бесконечное разнообразие рассматриваемого множества приближенно описывается разнообразием конечного числа подмножеств («классов»), каждому из которых присваивается обозначение, – в разговорном языке это слова. В зависимости от способа классификации («искусственная» или «естественная», четкая, расплывчатая или статистическая [2]) и способа присвоения имен классам, возникают многозначность и расплывчатость смысла слов, характерные для всех естественных языков.

Из уникальности культуры каждого субъекта следует, что у каждого субъекта свой язык. Национальный язык принадлежит этносу; каждая группа имеет свой язык (на своих «профессиональных» языках говорят юристы, философы, врачи, физики и т. д.), известны разные диалекты, различается английский язык у британцев и американцев; собственные языки у уголовников («феня») и у лондонских низов («кокни»). Да и два индивида одной национальности говорят по-разному (в силу различия их личных культур), и одно и то же слово понимают по-разному (в силу размытости смысла слов). Англичане особенно чувствительны к произношению индивида (это хорошо описано в «Пигмалионе» Б. Шоу). Огромное значение имеет введение единого государственного языка в многонациональном государстве.

Расплывчатость смысла слов в разговорном языке, столь полезная в бытовом общении, становится помехой для обеспечения деятельности в «профессиональной» сфере. Необходимое уточнение (уменьшение неопределенности) возможно только путем внесения в язык новой специфической информации. Это делается либо за счет придания нового смысла старым словам (примером может служить разный смысл слова «поле» у физиков, геологов, агрономов, печатников), либо введения в язык новых терминов (изобретенных или заимствованных из других языков). По мере добывания новых знаний специальный язык развивается, точность его языковых моделей взрастает. Предельной точности язык достигает, включая в себя язык математики. Но достичь этого можно только собрав достаточно информации для построения адекватной математической модели. По-видимому, именно это имел в виду Кант, говоря, что некая область знаний с тем большим правом может называться наукой, чем больше она использует математику в своих целях. Но если какая-то деятельность не применяет математические модели, то либо ее цели достигаются без них, либо она имеет дело со столь сложным и столь мало изученным явлением, что до математических моделей дело пока не дошло. Иногда некоторые нетерпеливые специалисты пытаются насильственно применить к сложному явлению красивое уравнение и предметно толковать его решение. Такие попытки редко бывают удачными.

Шестое следствие. Культура – саморазвивающаяся система моделей. Культуры, как и все феномены, со временем изменяются. Как и остальные системы, они проходят свой жизненный цикл, в котором различные тенденции друг друга сменяют, протекают с разной скоростью; и любые системы, возникнув однажды, когда-то исчезают. Неудивительно, что из пяти вариантов изменений (развитие, рост, стабильность, спад, деградация) наиболее желательными для субъекта являются рост и развитие.

Рост и развитие, хотя и могут протекать одновременно, но по существу это качественно разные вещи, и могут происходить независимо друг от друга. Кладбища и свалки растут, но об их развитии можно говорить лишь при придании им новых качеств. Человек растет до 25 лет, но развиваться продолжает по мере способностей и желания (Эйнштейн с возрастом даже стал худеть). Акофф развивал эту тему в ряде публи-

каций; наиболее сжато это выражено в [4]: «Развитие – это увеличение компетентности. Объемность – результат роста; компетентность – результат развития. Рост имеет количественный характер, развитие – качественный. Рост состоит в приобретении, развитие состоит в обучении. Развитие заключается не столько в том, как много организация имеет, сколько в том, как много организация может сделать с тем, что имеет».

Из многообразия аспектов, составляющих проблематику жизненного цикла культуры, для нашей темы – связь культуры и моделирования – особый интерес представляет тот факт, что развитие является следствием приобретения субъектом новой информации, включения ее в свою культуру и последующего использования с своей деятельности, т. е. результатом обучения. А обучение – это, пожалуй, единственная функция субъекта, которую нельзя выполнить для субъекта вместо него. Роль учителей и воспитателей – в том, чтобы помогать субъекту обучаться. Но, если субъект не способен или не желает обучаться, то никакие усилия преподавателей не будут иметь успеха.

Таким образом, развитие невозможно без обучения, обучение – активная функция самого обучающегося, и, следовательно, развитие возможно только как саморазвитие. Значит, культура не планируется извне, не надстраивается извне, культура есть саморазвивающаяся система, взаимодействующая с изменчивой окружающей средой.

Отметим несколько проблем, рассмотрение которых может углубить наше понимание процессов развития культур.

Во-первых, для саморазвития системы должны быть внутренние механизмы в ней, какие-то собственные прямые и обратные (положительные и отрицательные) связи. Предстоит выявить их и их взаимосвязи. Пока лишь можно дать перечень некоторых кандидатов в элементы этих механизмов, обеспечивающих рост и развитие культуры: любознательность, творчество, природная мотивация, обучаемость, адаптация, сочетание стабильности и изменчивости, эволюционный тип развития, критицизм, мораль, этика…

Во-вторых, интересно уточнить соотношение между искусственным (субъективным) и естественным (объективным) в культуре. В связи с трудовой, преобразовательной активностью субъектов различают два класса систем – естественные и искусственные. К естественным системам относятся природно организованные неоднородности – Солнечная система, живые организмы, бассейны рек, галактики, месторождения полезных ископаемых и т. д. и т. п. Искусственными называются системы, возникшие по воле и стараниями субъекта, – здания, сооружения, машины, приборы, инструменты и т. д. и т. п. Но все реальные искусственные системы состоят из преобразованных или приспособленных естественных частей (лошадь в упряжке, лесная деляна в леспромхозе, лесопарк в городе, природные стройматериалы в здании и т. п.). И по своим характеристикам многие искусственные системы в процессе технического совершенствования приближаются к параметрам своих естественных аналогов (обводы подводных лодок – к формам китов, профили самолетов – к контурам птиц и т. д.). Чисто искусственными, субъективными являются (по своей субстанции) лишь абстрактные модели, составляющие мышление и психику, и то, что называется духовной культурой. Проблему соотношения искусственного и естественного в культуре можно считать современной версией проблемы первичности сознания и/или бытия.

Еще один аспект саморазвиваемости культуры – ее автономное, неосознаваемое как бы биологическое, нейрофизиологическое, интуитивное, эмоциональное развитие. Этот феномен замечен многими исследователями. Например, Хайек пишет: «Нельзя считать, что наше логическое обоснование является высшим суждением, и только те законы имеют моральную ценность, которые поддерживаются нашей логикой» [5, с. 21]. Наглядный пример такой ситуации приведен в [6] – антизакон № 10: «Степень иррациональности, которую руководитель приписывает другим, прямо пропорциональна его собственной неразумности». И еще: «Если бы мы не стали делать ничего без определенного резонного обоснования, мы скорее всего погибли бы» [5, с. 68]. Еще один пример – то, как мы познаем правила языка и поведения в обществе: ведь нас никто не инструктирует, мы просто открываемся окружающей среде – семье, ее социальным связям, и – обучаемся. Нормы поведения, традиции, мораль возникают в процессе человеческих взаимодействий, без их преднамеренного проектирования и введения в нашу культуру.

В-третьих, интересной проблемой является сам механизм развития, т. е. процесс получения, восприятия новой информации и присоединения ее к существующей культуре. В свете нашей гипотезы о тождественности культуры и моделирования, вопрос о включении новых моделей в состав культуры существенно связывается со степенью согласованности между ними (ингерентностью модели культуре [2]). Культура субъекта воспримет только те новые идеи, которые совместимы с нею. Не зря говорят, что некоторые идеи «родились слишком рано». Идеи кибернетики были предложены (и даже под тем же названием!) Ампером и Трентовским в середине XIX в., затем Богдановым (под названием «тектологии») на рубеже XIX и XX вв., но были восприняты обществом только в середине XX в., благодаря тому, что Винер преподнес их как отклик на потребность самой культуры в совершенствовании управления и связи – того, что теперь называется информационными технологиями. Другой пример – процесс преподавания. Если первокурсник попадет на лекцию для старшекурсников, он мало что (если вообще) поймет: сообщаемые модели несовместимы с его культурой. Одна из сложнейших проблем преподавания – постепенное наращивание культуры обучающихся для последующего восприятия все более сложных моделей (отчасти поэтому столь высока корреляция между успеваемостью и посещаемостью студентов в вузах).

К этому же кругу вопросов относится проблема понимания в ходе общения двух или более субъектов. Полное понимание возможно лишь в рамках общего

языка. В силу различия индивидуальных культур, понимание может быть лишь в пределах их совпадения. Добиваться понимания значит добиваться расширения поля совпадения культур сторон общения. Наглядные примеры – практика переговоров сторон с различающимися интересами; период выработки общего языка в группе экспертов разных специальностей, работающих над решением единой сложной проблемы; проблема доведения смысла распоряжения менеджера до всех исполнителей; гипотетическое общение с инопланетянами; укрепление целостности многонационального государства посредством единого государственного языка.

Седьмое следствие. Специфика внутренних функций культуры. Если внешние функции культуры обеспечивают активные действия субъекта в окружающей среде, то внутренние обеспечивают самореализацию субъекта, раскрытие его врожденного потенциала. Разнообразие проявлений активности субъекта мы описываем с помощью классификаций. Можно поразному различать виды поведения субъекта; однако, в разных классификациях просматривается и некоторая общность:

– работа, игра, развлечения, вдохновение [4];

– труд, познание, получение удовольствия, поиск смысла существования (самоопределение) [6];

– преобразовательная, познавательная, проектная деятельности (целеполагание), достижение удовлетворения [2].

Каждый из типов деятельности обеспечивается соответствующими моделями (культурой), каждый из них несет внешнюю и внутреннюю функции – в разных пропорциях. К внутренним функциям относится все, что связано с развитием (саморазвитием).

И здесь на первый план для субъекта-индивида выходит образование. Центральная идея критики современной образовательной системы [6] заключается в том, чтобы считать главным в образовательном процессе не работу преподавателя, а работу обучающегося. Следующий момент – обучение должно продолжаться всю жизнь. Сегодня все громче говорят о «послевузовском» образовании. В [6] подробно обсуждаются вопросы образования в разные периоды жизни человека, начиная с дошкольного возраста (3–4 года), и включая образование для пенсионеров. Но вполне логично пойти дальше – вплоть до перинатального образования: ведь известно, что ребенок уже во чреве матери воспринимает и различает звуки и другие сигналы извне.

Общим для всех индивидов фактом образования является любопытство. Еще Аристотель во введении к трактату «Основы натуральной философии» (в другом переводе – «Метафизика») подчеркнул, что «люди по природе своей стремятся к знаниям». Любознательность признается постоянно действующей активной силой, заставляющей людей от рождения до смерти активно действовать.

Другой, общий для всех индивидов фактор, – врожденная способность к игре как средству обучения. Если игру понимать как любую деятельность при неизвестности ее исхода, то становится явной ее познавательная роль.

Однако, направленность действий и выбор способов движения определяются характером личности, – особой уникальностью субъекта. Происхождение характера остается тайной природы: люди, выросшие в одной семье и одинаковом окружении, всегда имеют уникальные и различные характеры. Гринберг [6] сделал попытку проанализировать компоненты характера: страстность, решительность, уверенность, способность решать проблемы, способность к общению. Разная выраженность каждого компонента и их взаимосвязей является, по мнению Гринберга, причиной различия характеров.

Развитие субъекта-группы выражается в различных формах придания его культуре новых качеств. Например, для человеческих сообществ этот процесс проявился в виде исторических переходов от одной общественной формации к другой. Попытка спрогнозировать этот процесс в будущем была сделана Ф. Фукуямой [7]; он считает, что либеральная демократия – неулучшаемая стадия развития человечества.

Библиографические ссылки

1. Акофф Р. Л., Гринберг Д. Преобразование образования : пер. с англ. Томск : Изд-во ТГУ, 2009.

2. Тарасенко Ф. П. Прикладной системный анализ. Томск : Изд-во ТГУ, 2004.

3. Зинченко В. П. Наука – неотъемлемая часть культуры? // Вопросы философии. 1990. № 1. С. 33–50.

4. Акофф Р. Л. Менеджмент в XX в. Преобразование корпорации : пер. с англ. Томск : Изд-во ТГУ, 2006.

5. Hayek F. The Fatal Conceit. University of Chicago press, 1988.

6. Акофф Р. Л., Гринберг Д., Бибб С. Анти-законы менеджмента. Томск : Изд-во ТГУ. 2008.

7. Фукуяма Ф. Конец истории, или последний человек. М. : АСТ. 2007.

F. P. Tarasenko

CULTURE OF AN INDIVIDUAL AND MODELING

In this article the author considers the term «culture» which has different content in different areas of human activity and gives analysis to interconnection of models, modeling and culture of an individual. Some aspects of immersion of individuals in culture, models and environment interaction are given as well.

Keywords: culture, models, individual, models system, language, self-developing system, education, knowledge, teaching, mathematical models.

© Тарасенко Ф. П., 2010

УДК 519.95

Н. Г. Загоруйко, И. А. Борисова, В. В. Дюбанов, О. А. Кутненко

ФУНКЦИИ КОНКУРЕНТНОГО СХОДСТВА В АЛГОРИТМАХ РАСПОЗНАВАНИЯ КОМБИНИРОВАННОГО ТИПА*

Для решения комбинированных задач распознавания образов предлагаются методы, основанные на функции конкурентного сходства (FRiS-функции). С ее помощью можно оценивать сходство между объектами, строить решающие правила, оценивать компактность образов и информативность признаков. Приводятся примеры использования предлагаемых методов для решения задач распознавания и прогнозирования.

Ключевые слова: функция конкурентного сходства, распознавание, компактность, информативность.

В задачах распознавании для принятия решения о принадлежности контрольного объекта z к образу A недостаточно знать расстояние $R(z, A)$. Нужно знать также расстояние $R(z, B)$ до конкурирующего образа B и определить, что $R(z, A)$ меньше $R(z, B)$. В методе k ближайших соседей (*kNN*) эти расстояния сравниваются в шкале порядка. Но оценками сходства можно распорядиться более эффективно, если измерять сходство в более сильной абсолютной шкале. Сформулируем следующие требования, которым должна удовлетворять мера $F(a, b|z)$ сходства объекта z с объектом a в конкуренции с объектом b.

Свойство локальности. Мера сходства должна зависеть не от характера распределения всего множества объектов, а от особенностей распреде- ления объектов в окрестности объекта z.

Свойство нормированности. Если оценивается мера сходства объекта z с объектом a, и ближайшим соседом z является объект b, $b \neq a$, то при совпадении объектов z и a мера $F(a, b|z)$ должна иметь максимальное значение, равное $+1$, а при совпадении z с b – минимальное значение, равное -1. Во всех остальных случаях мера конкурентного сходства принимает значения от -1 до 1. При одинаковых расстояниях $R(z, a)$ и $R(z, b)$ объект z в равной степени будет похожим на объекты a и b, и $F(a, b|z) = F(b, a|z)) = 0$.

Свойство инвариантности. Значения $F_{a/b}(z)$ и $F_{b/a}(z)$ должны сохраняться при сдвиге начала и повороте координат, а также при умножении всех координат на одно и то же положительное число.

Предлагаемая нами функция конкурентного сходства FRiS (от Function of Rival Similarity) [1] удовлетворяет всем этим требованиям:

$$F(a, b|z) = (r_2 - r_1)/(r_2 + r_1), \qquad (1)$$

где r_1 r_2 – расстояния $R(z, a)$ и $R(z, b)$ соответственно. Опыт показал, что использование этой меры сходства позволяет создавать эффективные алгоритмы решения задач распознавания, легко объединяемые в алгоритмы комбинированного типа. Опишем некоторые из этих алгоритмов.

Оценка компактности и информативности. Практически все алгоритмы распознавания основаны на использовании гипотезы компактности, имеющиеся определения которой не позволяют найти ее количественной оценки. Мы предлагаем использовать для этой цели FRiS-функцию. Для каждого объекта a_i образа A, $i = 1, 2, \ldots, M_A$ найдем оценки сходства с ним всех M_A объектов a_j этого образа и оценки отличия от a_i всех M_B объектов b_s конкурирующего образа B. При определении сходства объектов a_j с a_i конкурентом будет служить объект b_j образа B, который является ближайшим соседом объекта a_j. Для оценки отличия от a_i объекта b_s образа B нужно найти объект b_q, который является ближайшим соседом объекта b_s. Расстояние от b_s до b_q принимается равным r_1, от b_s до a_i – равным r_2, по ним вычисляется мера сходства объекта b_s с объектом b_q в конкуренции с объектом a_i, и эта величина принимается в качестве меры отличия b_s от a_i.

Просуммировав оценки сходства объекта a_i со всеми своими объектами и оценки отличия от всех чужих объектов и разделив полученную сумму на величину $(M_A + M_B)$, мы получим среднюю оценку C_i его сходства со своими и отличия от чужих объектов. Повторив эти процедуры для всех объектов образа A, мы можем найти оценку компактности C_A этого образа:

$$C_A = (1/M_A)\sum_{i=1}^{M_A} C_i. \qquad (2)$$

Общая оценка C FRiS-компактности K образов может быть получена путем геометрического усреднения оценок C_t:

$$C = \sqrt[K]{\prod_{t=1}^{K} C_t}. \qquad (3)$$

Описанная мера компактности тем больше, чем выше плотность объектов внутри образов и дальше образы отстоят друг от друга. Таким же свойством обладает и мера, предложенная Фишером для оценки информативности признаков при нормальном распределении образов. Вполне естественным является использование FRiS-компактности в качестве *критерия информативности* признаков при произвольных распределениях образов. Она используется в этом качестве в алгоритме GRAD [2].

* Работа выполнена при финансовой поддержке РФФИ, проект № 08-01-00040.

Наши эксперименты с этим критерием показали его существенное преимущество по сравнению с широко используемым критерием минимума ошибок при распознавании тестовой выборки методами Cross Validation или One Leave Out.

Выбор эталонов (алгоритм FRiS-Stolp). Для распознавания контрольного объекта используются меры его сходства с эталонными представителями («столпами») каждого образа. Алгоритм FRiS-Stolp [1] выбирает эталоны, которые обладают высокими значениями двух свойств: *«обороноспособности»* – высокая степень сходства с ним других объектов этого образа позволяет надежно распознавать свои объекты; *«толерантности»* – низкая степень сходства с ним объектов других образов предотвращает распознавание чужих объектов в качестве своих.

Алгоритм FRiS-Stolp состоит из следующих шагов.

1. Оценивается объект a_i в роли столпа. Для этого, как и в предыдущем пункте, вычисляем сходство с ним всех M_A объектов образа A и несходство с ним (отличие) всех M_B объектов образа B. Находим среднее значение этих величин C_i'.

2. После того, как все объекты образа A побывают в роли столпа, выбирается тот объект a_i, который набрал наибольшую величину C_i'. Он объявляется столпом первого кластера. В состав этого кластера включаются m_1 объектов образа A, сходство которых со столпом превышает заданный порог F^*, например, $F^* = 0$. Среднее значение сходства m_1 объектов со своим столпом (C_1) служит мерой компактности полученного кластера.

3. Если $m_1 < M_A$, то для остальных объектов повторяются шаги 1–2. В итоге получаем список k_A столпов образа A с указанием величин $C_j, j = 1, 2, \ldots, k_A$, накопленных каждым столпом.

4. Повторяем шаги 1–4 для объектов образа B и получаем список k_B столпов этого образа и оценки их компактности $C_q, q = 1, 2, \ldots, k_B$.

5. Если образов больше, чем два, то при определении компактности каждого t-го образа, $t = 1, 2, \ldots, K$, будем считать его образом A, а объекты всех остальных образов объединим в образ B.

Алгоритм FRiS-Stolp при нормальных распределениях, в первую очередь, выбирает столпы, расположенные в районе математического ожидания. Если распределения полимодальны и образы линейно неразделимы, то столпы будут стоять в центрах мод. В процессе распознавания, решение принимается в пользу того образа, на столп которого контрольный объект похож больше всего, а значение функции сходства объекта с выбранным образом позволяет судить о достоверности принятого решения.

Построение решающих правил с одновременным выбором признаков. Рассмотрим следующие алгоритмы.

Алгоритм AdDel [2]. Пусть из N признаков выбирается наиболее информативная подсистема, состоящая из n признаков. Хорошо известны два жадных полиномиальных алгоритма решения этой задачи. Алгоритм Deletion (Del) начинает работать со всеми N признаками и последовательно на каждом шаге вычеркивает из системы наименее полезный признак, пока в системе не останется n признаков. Алгоритм Addition (Ad), напротив, вначале находит самый информативный признак и на каждом из n шагов добавляет в систему тот признак, участие которого в системе делает ее наиболее информативной. Оба описанных алгоритма дают оптимальное решение на каждом шаге, но это не обеспечивает глобального оптимального решения. В алгоритме AdDel методом Ad набирается некоторое количество n_1 информативных признаков, затем n_2 из них ($n_2 < n_1$) исключается методом Del. Такое чередование алгоритмов Ad и Del продолжается до достижения заданного количества признаков n. Наблюдения показывают, что по мере увеличения числа признаков компактность вначале растет, потом рост прекращается и начинается ее снижение за счет добавления малоинформативных, шумящих признаков. Перегиб кривой качества позволяет автоматически выбрать *оптимальное количество признаков*. Это очень важное свойство алгоритмов семейства AdDel, которым не обладают другие алгоритмы выбора информативных признаков.

Алгоритм GRAD [2]. Добавлять и исключать признаки можно как по одному, так и группами (гранулами), состоящими из нескольких признаков. В наших экспериментах было обнаружено, что если все признаки упорядочить по убыванию их индивидуальной информативности, то в составе наиболее информативных пар преобладают признаки с малыми порядковыми номерами. В алгоритме GRAD (GRanulated AdDel) алгоритм AdDel работает на множестве G наиболее информативных гранул, состоящих из w признаков, $w = 1, 2, 3$. Гранулы мощности 1 – это признаки, занимающие первые m мест по индивидуальной информативности. Из них методом полного перебора выбираются гранулы мощности 2 и 3. Затем весь список из m самых информативных гранул подается на вход алгоритма AdDel. Сравнительные эксперименты показали, что алгоритм GRAD по качеству получаемых решений значительно лучше алгоритма AdDel.

Алгоритм FRiS-GRAD [2]. Этот алгоритм одновременно выбирает информативные признаки и в пространстве этих признаков строит решающее правило. Признаки выбираются алгоритмом GRAD, а решающие правила – алгоритмом FRiS. Эффективность этого алгоритма комбинированного типа подтверждена опытом решения большого количества реальных задач.

Примеры приложений. Рассмотрим примеры приложний.

Распознавание двух видов лейкемии (*ALL* и *AML*). Задача распознавания двух типов лейкемии интересна тем, что в литературе представлены результаты ее решения разными группами исследователей. Это дает возможность сравнить наши результаты с лучшими прежними результатами. Анализируемые данные представлены матрицей векторов экспрессии генов,

полученных с помощью биочипов для пациентов с двумя типами лейкемии – ALL и AML [3]. Обучающая выборка содержит 38 объектов, тестовая выборка – 34 объекта. Исходное количество признаков (генов) $N = 7\,129$.

Рассмотрим результаты решения этой задачи, описанные в работе [3]. Информативное подмножество признаков выбиралось методом RFE (разновидностью алгоритма Deletion), решающие правила основаны на методе SVM. Были найдены наилучшие подсистемы, размерность которых кратна степени числа 2: 4 096, 2 048, …, 4, 2 и 1. По двум лучшим признакам, которые можно выбрать по результатам обучения, правильно распознано 30 объектов, по 4 лучшим признакам – 31, по 128 признакам – 33. Нами на тех же данных получены следующие результаты. Информативное подмножество признаков выбиралось с помощью алгоритма FRiS-GRAD. Информативность признаков оценивалась по критерию FRiS-компактности. Из 7 129 признаков было выбрано 18, из которых программа FRiS-Stolp построила 30 вариантов решающих правил. В состав каждого правила входит с разными весами от 4 до 6 признаков. Первые 27 правил дают результат 34 из 34.

Различие между приведенными результатами могут зависеть как от метода выбора признаков, так и от типа решающих правил. Для сравнения решающих правил SVM и FRiS был проведен такой эксперимент. В подпространстве двух признаков (генов 803 и 4 846), выбранных методом FRE, по правилу SVM было получено 30 правильных ответов, а FRiS-методом – 33. По лучшему одному гену (4 846), выбранному методом FRE, результат SVM равен 27, а результат FRiS – 30. Отсюда можно сделать вывод, что как метод выбора признаков, так и решающие правила, основанные на FRiS-функции, обладают достаточно высокими конкурентными качествами.

Прогнозирование покупательского спроса. Достаточно успешным оказалось применение FRiS-функции при решении задачи на международном конкурсе Data Mining Cup 2009. Задача состояла в предсказании значений переменных, измеренных в абсолютной шкале, и заключалась в следующем. Анализировались данные о том, сколько книг того или иного жанра было продано в разных магазинах в течение года. Эти данные представляли собой очень разреженную таблицу (84 % клеток таблицы имели значение 0), в которой *M*-строками (объектами) являлись магазины ($M = 4\,812$), а *N*-столбцами (признаками) – жанры книг ($N = 1\,864$). На пересечении строк и столбцов указывалось количество книг данного жанра, проданных в течение года в том или ином магазине. Последние 8 признаков являлись целевыми. Таблица была разделена по горизонтали на два слоя. В первом (обучающем) слое содержалось $M_o = 2\,394$ магазина. Для 2 418 контрольных магазинов требовалось предсказать, сколько и каких из 8 жанров книг было продано в каждом из них. В конкурсе изъявили желание участвовать 618 команд из 164 организаций 42 стран. Эту задачу решила 321 команда и прислала свои результаты. Наша команда, используя алгоритм FRiS-Pro [4], заняла 4 место. Полученные результаты подтверждают возможность использования FRiS-функции в алгоритмах решения задач прогнозирования количественных переменных.

Таким образом, использование относительной меры сходства, учитывающей конкурентную обстановку, позволяет строить эффективные методы решения всех основных задач Data Mining, в том числе задач распознавания образов. Методы инвариантны к количеству образов, характеру их распределений и обусловленности обучающей выборки (соотношению между M и N).

Библиографические ссылки

1. Methods of Recognition Based on the Function of Rival Similarity / N. G. Zagoruiko [et al.] // Pattern Recognition and Image Analysis. 2008. Vol. 18. P. 1–6.

2. Attribute selection through decision rules construction (algorithm FRiS-GRAD) / N. G. Zagoruiko [et al.] // Pattern Recognition and Image Analysis : New Information Technologies : Proc. of 9th Intern Conf. Nizhni Novgorod. 2008. Vol. 2. P. 335–338.

3. Gene Selection for Cancer Classification using Support Vector Machines / I. Guyon [et al.] // Machine Learning. 2002. Vol. 46 (1–3). P. 389–395.

4. Дюбанов В. В. Использование FRiS-функции в алгоритмах предсказания количественных переменных (Алгоритм FRiS-Pro) // Знания-Онтологии-Теории (ЗОНТ-09) : тр. Всерос. конф. Новосибирск, 2009. Т. 2. С. 258–260.

N. G. Zagoruiko, I. A. Borisova, V. V. Dyubanov, O. A. Kutnenko

FUNCTIONS OF RIVAL SIMILARITY IN ALGORITHMS OF RECOGNITION OF COMBINED TYPE

For decision of combined tasks of pattern recognition we offer methods based on function of rival similarity (FRiS-function). With the help of the function we can estimate similarity between objects, build decision rules, estimate compactness of patterns and informativeness of attributes. Examples of use of the offered methods for decision of tasks of recognition and forecasting are given in the article.

Keywords: function of rival similarity, recognition, compactness, informativeness.

© Загоруйко Н. Г., Борисова И. А., Дюбанов В. В., Кутненко О. А., 2010

УДК 519.2

И. А. Бодягин, Ю. С. Харин

ПРОГНОЗИРОВАНИЕ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ ПРИ НАЛИЧИИ ЦЕНЗУРИРОВАНИЯ

Рассмотрена задача статистического прогнозирования авторегрессионных временных рядов при наличии интервального цензурирования. Построена оптимальная прогнозирующая статистика, для нее вычислен условный среднеквадратический риск. Для авторегрессии первого порядка проведено сравнение оптимальной прогнозирующей статистики с прогнозирующими статистиками, часто используемыми на практике. Представлены численные результаты.

Ключевые слова: авторегрессия, прогнозирование, цензурирование, риск.

Задача статистического прогнозирования возникает во многих приложениях: в медицине, экономике, метеорологии, технике, астрономии [1]. Для описания временных рядов с зависимыми наблюдениями и прогнозирования будущих значений широко применяется модель авторегрессии [1]. На практике значения временного ряда часто наблюдаются с искажениями различных типов: выбросы, пропуски, гетероскедастичность [2], цензурирование [3] и др.; обзор типов искажений и их математические описания представлены в работе [2]. Цензурирование временного ряда заключается в том, что часть наблюдений ряда известна точно, а об остальных наблюдениях известно лишь, что они принадлежат некоторым числовым интервалам. Такая ситуация может возникать из-за наличия у приборов конечных пределов измерения, высокой стоимости проведения точных измерений, разладки оборудования и других причин.

Цензурированные выборки независимых наблюдений подробно изучены в математической теории надежности [4]. Однако статистический анализ цензурированных временных рядов остается малоизученным и актуальным направлением исследований.

Математическая модель. Пусть временной ряд x_t описывается моделью АР(p) авторегрессии порядка $p \in N$ [1]:

$$x_t = \sum\nolimits_{i=1}^{p} \theta_i x_{t-i} + u_t,\ t \in Z, \tag{1}$$

где $\{\theta_i\}_{i=1}^{p}$ – коэффициенты авторегрессии такие, что все корни порождающего характеристического многочлена $z^p - \sum\nolimits_{j=1}^{p} \theta_j z^{p-j}$ лежат внутри единичного круга; $\{u_t\}$ – независимые в совокупности одинаково распределенные случайные величины, имеющие нормальный закон распределения вероятностей $L\{u_t\} = N(0, \sigma^2)$.

Пусть вместо значений временного ряда наблюдаются случайные события:

$$A_t^* = \{x_t \in A_t\},\ t \in \{1, \ldots, T\}, \tag{2}$$

где $\{A_t\}$ – заданные борелевские множества; $T > p$ – длительность наблюдения.

При наличии интервального цензурирования возможны два случая:

1) A_t состоит из одного элемента ($A_t = \{x_t\}$), тогда значение x_t известно точно;

2) A_t является числовым интервалом ($A_t = [a_t, b_t)$, $a_t < b_t$), тогда имеет место интервальное цензурирование значения x_t, а интервал $[a_t, b_t)$ называется интервалом цензурирования.

Статистическое прогнозирование будущего значения $x_{T+1} \in \mathrm{R}$ заключается в вычислении оценки $\hat{x}_{T+1} \in \mathrm{R}$ на основе имеющейся информации о наступлении событий $A_1^*, \ldots, A_T^*$:

$$\hat{x}_{T+1} = f\left(A_T^*, A_{T-1}^*, \ldots, A_1^*\right). \tag{3}$$

Погрешность прогнозирования будем характеризовать условным риском прогнозирования

$$r_T(f) = \mathrm{E}\left\{\left(x_{T+1} - \hat{x}_{T+1}\right)^2 \middle| A_T^*, \ldots, A_1^*\right\} \geq 0, \tag{4}$$

т. е. среднеквадратической ошибкой прогнозирования.

Рассмотрим задачу построения оптимальной прогнозирующей статистики (ОПС) $f_0(\cdot)$, минимизирующей условный риск (4), в случае известных параметров модели (1), (2):

$$r_T(f_0) = \inf_{f(\cdot)} r_T(f). \tag{5}$$

Основные результаты. *Теорема 1.* Если временной ряд x_t наблюдается при наличии цензурирования общего вида (2), то среди всех прогнозирующих статистик вида (3) оптимальная по критерию минимума риска (5) прогнозирующая статистика определяется условным математическим ожиданием:

$$\hat{x}_{T+1} = f_0\left(A_T^*, \ldots, A_1^*\right) = \mathrm{E}\left\{x_{T+1} \middle| A_T^*, \ldots, A_1^*\right\},$$
$$r_T(f_0) = \mathrm{D}\left\{x_{T+1} \middle| A_T^*, \ldots, A_1^*\right\}. \tag{6}$$

Доказательство. Преобразуем условный риск (4):

$$r_T(f) = \mathrm{E}\left\{\left(x_{T+1} - \hat{x}_{T+1}\right)^2 \middle| A_T^*, \ldots, A_1^*\right\} =$$
$$= \mathrm{E}\left\{\left(x_{T+1} - f\left(A_T^*, \ldots, A_1^*\right)\right)^2 \middle| A_T^*, \ldots, A_1^*\right\} =$$

$$= \mathrm{E}\left\{\left(\left(x_{T+1} - \mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\}\right) + \left(\mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\} - \right.\right.$$

$$\left.\left. - f\left(A_T^*, \ldots, A_1^*\right)\right)\right)^2 \middle| A_T^*, \ldots, A_1^*\right\}.$$

Заметим, что второе слагаемое зависит только от $A_T^*, \ldots, A_1^*$:

$$r_T(f) = \mathrm{E}\left\{\left(x_{T+1} - \mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\}\right)^2 \middle| A_T^*, \ldots, A_1^*\right\} +$$

$$+\left(\mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\} - f\left(A_T^*, \ldots, A_1^*\right)\right)^2 +$$

$$+2\,\mathrm{E}\left\{\left(x_{T+1} - \mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\}\right)\middle| A_T^*, \ldots, A_1^*\right\} \times$$

$$\times\left(\mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\} - f\left(A_T^*, \ldots, A_1^*\right)\right) = \mathrm{D}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\} +$$

$$+\left(\mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_1^*\right\} - f\left(A_T^*, \ldots, A_1^*\right)\right)^2 \to \min_{f(\cdot)}.$$

Из этого представления следует, что (6) есть решение задачи (5).

Теорема 1 является обобщением известного результата [1] в ситуации, когда цензурирование отсутствует.

Рассмотрим случай, когда цензурированы только последние q ($0 \le q \le T$) значений временного ряда, а остальные $T - q$ наблюдений известны точно. Обозначим

$$\mu(t, m) = \theta_1 x_{t-1} + \ldots + \theta_m x_{t-m} = \sum_{i=1}^{m} \theta_i x_{t-i}, \quad t, m \in N.$$

Теорема 2. Пусть в рамках модели (1), (2) наблюдаются значения $x_1, \ldots, x_{T-q}$ и случайные события $A_{T-q+1}^* = \left\{x_{T-q+1} \in \left[a_{T-q+1}, b_{T-q+1}\right)\right\}, \ldots,$ $A_T^* = \left\{x_T \in \left[a_T, b_T\right)\right\}$. Тогда ОПС имеет вид

$$\hat{x}_{T+1} = \frac{\int_{a_T}^{b_T} \ldots \int_{a_{T-q+1}}^{b_{T-q+1}} \mu(T+1, p)\, p(x_T, \ldots, x_{T-q+1} | x_{T-q}, \ldots, x_1)\, dx_{T-q+1} \ldots dx_T}{\int_{a_T}^{b_T} \ldots \int_{a_{T-q+1}}^{b_{T-q+1}} p(x_T, \ldots, x_{T-q+1} | x_{T-q}, \ldots, x_1)\, dx_{T-q+1} \ldots dx_T}. \quad (7)$$

Доказательство. Оценку (6) в силу (1) можно представить следующим образом:

$$\hat{x}_{T+1} = \mathrm{E}\left\{x_{T+1}\middle|A_T^*, \ldots, A_{T-q+1}^*, x_{T-q}, \ldots, x_1\right\} =$$

$$= \mathrm{E}\left\{\sum_{i=1}^{p} \theta_i x_{T+1-i} + u_{T+1}\middle|A_T^*, \ldots, A_{T-q+1}^*, x_{T-q}, \ldots, x_1\right\} =$$

$$= \mathrm{E}\left\{\mu\left(T+1, p\right)\middle|A_T^*, \ldots, A_{T-q+1}^*, x_{T-q}, \ldots, x_1\right\},$$

так как случайная величина u_{T+1} не зависит от $A_T^*, \ldots, A_{T-q+1}^*$, $x_{T-q}, \ldots, x_1$ и $\mathrm{E}\{u_{T+1}\} = 0$. Вычислив полученное математическое ожидание, получим требуемое равенство (7).

Введем обозначение:

$$\Psi(x, y, m, s, u, v) = \left(u\,\varphi\left((x-m)/s\right) - v\,\varphi\left((y-m)/s\right)\right) \times$$

$$\times\left(\Phi\left((y-m)/s\right) - \Phi\left((x-m)/s\right)\right)^{-1},$$

$$x, y, m, s, u, v \in \mathrm{R},$$

где $\varphi(x) = \left(1/\sqrt{2\pi}\right)\exp(-x^2/2)$, $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$ – соответственно, плотность и функция распределения вероятностей стандартного нормального закона $N(0, 1)$.

Следствие. Если в рамках модели (1), (2) наблюдаются значения $x_1, \ldots, x_{T-1}$ и случайное событие $A_T^* = \left\{x_T \in (a_T, b_T)\right\}$ ($q = 1$), то ОПС (4) имеет вид

$$\hat{x}_{T+1} = \theta_1 \mu(T, p) + \sum_{i=2}^{p} \theta_i x_{T-i+1} + \theta_1 \sigma \Psi \times$$

$$\times\left(a_T, b_T, \mu(T, p), \sigma, 1, 1\right).$$

Доказательство. Воспользовавшись известным соотношением для модели AP(p) [1]

$$p\left(x_T | x_{T-1}, \ldots, x_1\right) = \varphi\left(\left(x_T - \mu\left(T, p\right)\right)/\sigma\right)/\sigma$$

и теоремой 2 для $q = 1$, получим требуемое соотношение.

Если среди последних q значений временного ряда имеются не только цензурированные наблюдения, но и k ($1 \le k \le q$) известных наблюдений $x_{l_1}, \ldots, x_{l_k}$ ($T - q + 1 \le l_1 < \ldots < l_k \le T$), то ОПС может быть получена из (7) предельным переходом $b_{l_1} \to a_{l_1}, \ldots, b_{l_k} \to a_{l_k}$.

Рассмотрим частный случай модели (1) – авторегрессию первого порядка ($p = 1$):

$$x_t = \theta x_{t-1} + u_t, \; t \in \mathrm{Z}, \quad (8)$$

и $q = 1$, причем предполагается, что параметры модели θ и σ известны. Для этого случая исследуем зависимость условного риска прогнозирования от длины интервала цензурирования и проведем сравнение ОПС $f_0(\cdot)$ с прогнозирующими статистиками, часто используемыми на практике [3]. В этом случае последнее значение x_T временного ряда цензурировано интервалом (a_T, b_T), а предпоследнее значение x_{T-1} известно точно. Поскольку в данном случае результат зависит только от одного интервала цензурирования $A_T = (a_T, b_T)$, то для упрощения обозначений вместо a_T и b_T будем писать a и b. Используя теорему 2, можно доказать следующую теорему.

Теорема 3. Пусть для модели (8) наблюдаются значение x_{T-1} и случайное событие $A_T^* = \left\{x_T \in [a, b)\right\}$, тогда ОПС и ее условный риск имеют вид:

$$\hat{x}_{T+1} = f_0\left(A_T^*, x_{T-1}\right) = \theta^2 x_{T-1} + \theta\sigma\Psi\left(a, b, \theta x_{T-1}, \sigma, 1, 1\right), \quad (9)$$

$$r_T(f_0) = (1+\theta^2)\sigma^2 - \left(\theta\sigma\Psi\left(a, b, \theta x_{T-1}, \sigma, 1, 1\right)\right)^2 + \quad (10)$$

$$+\,\theta^2\sigma\Psi\left(a, b, \theta x_{T-1}, \sigma, a - \theta x_{T-1}, b - \theta x_{T-1}\right).$$

Следствие. В условиях теоремы 3 для условного риска прогнозирования справедливо асимптотическое разложение при $\tau = b - a \to 0$:

$$r_T(f_0) = \sigma^2 + \theta^2\tau^2/12 - \theta^2\tau^4 \times$$

$$\times\left(3a^2 - 6a\theta x_{T-1} + 3\theta^2 x_{T-1}^2 + 2\sigma^2\right)/720\sigma^4 + o(\tau^4).$$

Доказательство. Учитывая дифференцируемость функции $\Psi(\cdot)$ в (10) по τ, воспользуемся формулой Тейлора с остаточным членом в форме Пеано и получим требуемое соотношение для условного риска.

Из доказанного следствия получаем, что безусловный риск ОПС имеет следующее асимптотическое разложение при $\tau = b - a \to 0$:

$$\mathrm{E}\{r_T(f_0)\} = \sigma^2 + \theta^2\tau^2/12 - \theta^2\tau^4 \times \\ \times\left(3a^2 + 3\theta^2\sigma^2/\left(1-\theta^2\right) + 2\sigma^2\right)/720\sigma^4 + o(\tau^4). \quad (11)$$

Одной из возможных альтернативных прогнозирующих статистик является следующая [3]:

$$\hat{x}_{T+1} = f_1(A_T^*) = \theta\,\mathrm{E}\left\{x_T \mid A_T^*\right\} = \\ = \theta\,\mathrm{E}\left\{x_T \mid x_T \in [a,b)\right\}. \quad (12)$$

Теорема 4. Если для модели (8) наблюдается случайное событие $A_T^* = \left\{x_T \in [a,b)\right\}$, то прогнозирующая статистика (12) имеет вид

$$\hat{x}_{T+1} = f_1(A_T^*) = \left(\theta\sigma/\sqrt{1-\theta^2}\right)\times \\ \times\Psi\left(a,b,0,\theta\sigma/\sqrt{1-\theta^2},1,1\right), \quad (13)$$

и ее условный риск прогнозирования равен:

$$r_T(f_1) = \frac{\sigma^2}{1-\theta^2} - \frac{\theta^2\sigma^2}{1-\theta^2}\times \\ \times\left(\Psi\left(a,b,0,\frac{\sigma}{\sqrt{1-\theta^2}},1,1\right)\right)^2 + \quad (14) \\ + \frac{\theta^2\sigma}{\sqrt{1-\theta^2}}\Psi\left(a,b,0,\frac{\sigma}{\sqrt{1-\theta^2}},a,b\right).$$

Доказательство. Прогнозирующая статистика (12) выглядит следующим образом:

$$\hat{x}_{T+1} = \theta\,\mathrm{E}\left\{x_T \mid x_T \in [a,b)\right\} = \theta\left(\int_a^b xn\left(x\middle|0,\sigma^2/1-\theta^2\right)dx\right)\times \\ \times\left(\int_a^b n\left(x\middle|0,\sigma^2/1-\theta^2\right)dx\right)^{-1}.$$

Воспользовавшись [5] для вычисления интегралов, получим статистику (13). Аналогично вычисляется условный риск прогнозирования (14).

Следствие. В условиях теоремы 4 для условного риска прогнозирования справедливо асимптотическое разложение при $\tau = b - a \to 0$:

$$r_T(f_1) = \sigma^2 + \theta^2\tau^2/12 - \theta^2\left(1-\theta^2\right)^2\times \\ \times\tau^4\left(3a^2 + 2\sigma^2/(1-\theta^2)\right)/720\sigma^4 + o(\tau^4).$$

Доказательство. Проводится аналогично доказательству следствия теоремы 3.

Сравнивая $\mathrm{E}\{r_T(f_0)\}$ и $r_T(f_1)$ при $\tau \to 0$, замечаем, что усредненный риск ОПС (11) меньше риска прогнозирующей статистики (12).

Рассмотрим еще одну часто используемую прогнозирующую статистику:

$$\hat{x}_{T+1} = f_2\left(A_T^*\right) = \theta(a+b)/2. \quad (15)$$

Теорема 5. Если для модели (8) наблюдается случайное событие $A_T^* = \left\{x_T \in [a,b)\right\}$, то условный риск прогнозирования для статистики (15) равен*:*

$$r_T(f_2) = \sigma^2/1-\theta^2 + \theta^2(a+b)^2/4 - \left(\theta^2\sigma/\sqrt{1-\theta^2}\right)\times \\ \times\Psi\left(a,b,0,\sigma/\sqrt{1-\theta^2},b,a\right). \quad (16)$$

Доказательство. Проводится аналогично доказательству теоремы 4.

Следствие. Пусть выполнены условия теоремы 5 и $\tau = b - a \to 0$. Тогда для условного риска справедливо следующее асимптотическое разложение:

$$r_T(f_2) = \sigma^2 + \theta^2\tau^2/12 + \theta^2\left(1-\theta^2\right)^2\times \\ \times\tau^4\left(a^2 - \sigma^2/(1-\theta^2)\right)/360\sigma^4 + o(\tau^4).$$

Доказательство. Проводится аналогично доказательству следствия теоремы 3.

Легко увидеть, что при τ близких к нулю, риск прогнозирования статистики (15) больше, чем риск прогнозирования для статистик (9) и (12).

Численные результаты. Для сравнения прогнозирующих статистик (9), (12) и (15) проведены компьютерные эксперименты. Применялся метод Монте-Карло с числом прогонов N = 10 000. Для моделирования временного ряда использованы значения параметров: p = 1, θ = 0,8, σ = 1, q = 1, T = 100, $\tau \in \{0,\ 0{,}5,\ \ldots,\ 15\}$, по наблюдению x_T строился интервал цензурирования $(a,\ b)$ длины τ, где $a = x_T - \alpha\tau$ и $b = x_T + \alpha(1-\tau)$, α – случайная величина, равномерно распределенная на $[0,\ 1]$.

На рисунке *а* изображены графики зависимостей экспериментальных значений риска для всех трех прогнозирующих статистик от τ. ОПС (9) имеет наименьший риск, риск статистики (12) принимает большие значения, а риск статистики (15) возрастает очень быстро и уже при малых τ принимает достаточно большие значения.

На рисунках *б–г* изображены усредненные теоретические значения риска прогнозирования для статистик (9), (12) и (15) в зависимости от τ, вычисленные по формулам (10), (14) и (16), соответственно, и 95 %-ные доверительные границы риска.

Таким образом, в настоящей работе найдена ОПС для авторегрессионных временных рядов при наличии цензурирования и ее риск; в случае авторегрессии первого порядка проведено сравнение ОПС с прогнозирующими статистиками, часто используемыми на практике; проведены компьютерные эксперименты, которые показали, что экспериментальные и теоретические значения риска находятся в хорошем согласии.

Библиографические ссылки

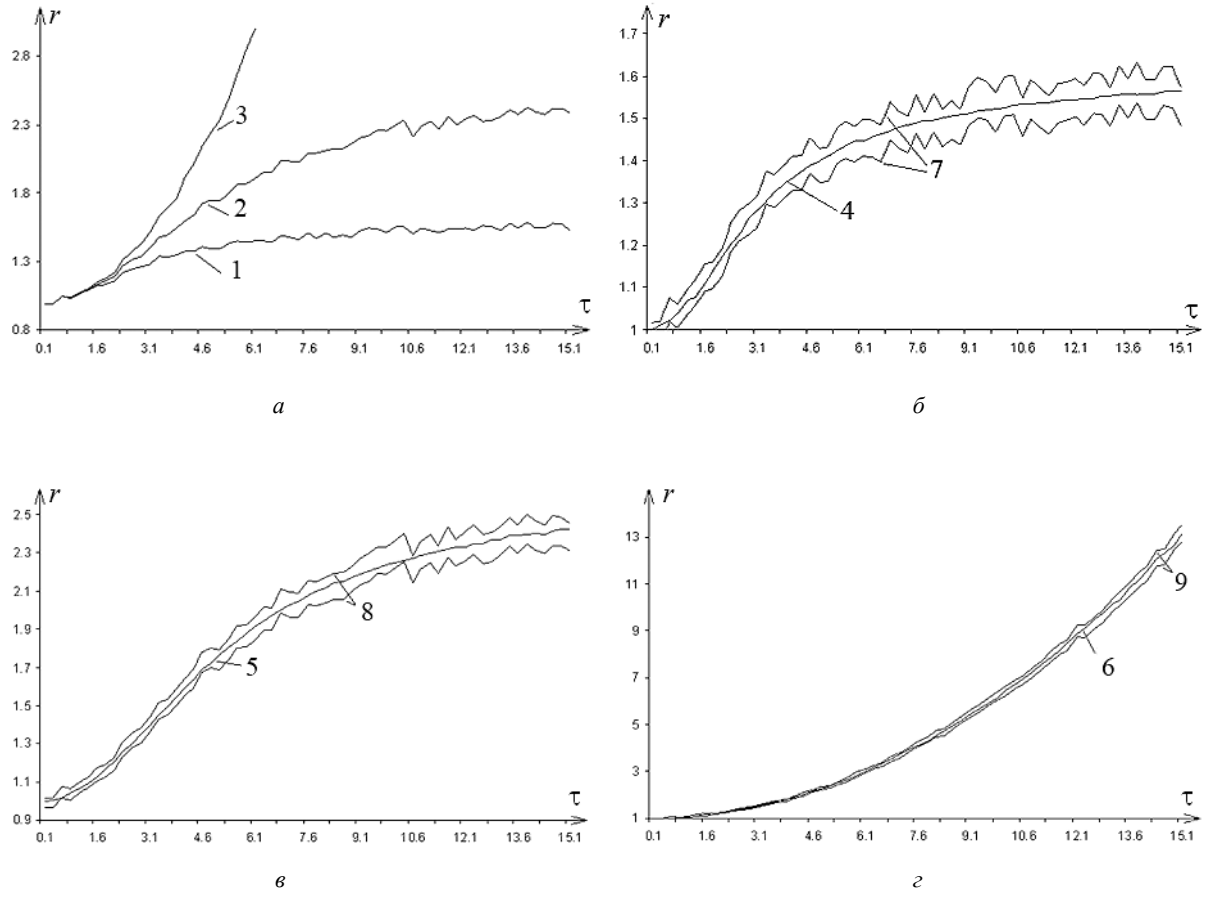
1. Андерсон Т. Статистический анализ временных рядов : монография. М. : Мир, 1976.

2. Харин Ю. С. Оптимальность и робастность в статистическом прогнозировании : монография. Минск : БГУ, 2008.

3. Park J. W., Genton M. G., Ghosh S. K. Censored time series analysis with autoregressive moving average models // The Canadian Journal of Statistics. 2007. Vol. 35, № 1. P. 151–168.

4. Gomez G., Espinal A., Lagakos W. Inference for a linear model with an interval-censored covariate // Statistics in medicine. 2003. № 22. P. 409–425.

5. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М. : Физматгиз, 1963.



Результаты численных экспериментов:
а – сравнение трех прогнозирующих статистик: *1* – риск оптимальной прогнозирующей статистики (9), *2* – риск статистики (12), *3* – риск статистики (15); *б–г* – сравнение теоретических и экспериментальных значений риска прогнозирующих статистик (9), (12) и (15): *4* – теоретическое значение риска статистики (9), *5* – теоретическое значение риска статистики (12), *6* – теоретическое значение риска статистики (15), *7–9* – 95 % доверительные границы соответствующих значений риска

I. A. Badziahin, Yu. S. Kharin

FORECASTING OF AUTOREGRESSIVE TIME SERIES UNDER CENSORING

Problems of statistical forecasting are considered for autoregressive time series observed under interval censoring. Optimal forecasting statistic is proposed, its mean-square risk is evaluated. Comparison of optimal and widely used in practice forecasting statistics is made. Numerical results are given.

Keywords: autoregression, forecasting, censoring, risk.

© Бодягин И. А. , Харин Ю. С., 2010

УДК 62-544

Е. Д. Агафонов

НЕПАРАМЕТРИЧЕСКИЙ АЛГОРИТМ УПРАВЛЕНИЯ НЕЛИНЕЙНЫМИ ДИНАМИЧЕСКИМИ СИСТЕМАМИ С ИСПОЛЬЗОВАНИЕМ СКОЛЬЗЯЩИХ РЕЖИМОВ

Предложен алгоритм управления нелинейными динамическими системами с применением скользящих режимов, в которых объект управления с одним входом и одним выходом представлен конечно-разностной непараметрической моделью в пространстве состояний. Даны рекомендации по настройке и оптимизации алгоритма. Разработанный алгоритм реализован в пакете MATLAB, приведены результаты управления движением обратного маятника.

Ключевые слова: нелинейная динамическая система, разностная модель, непараметрический регулятор, устойчивость по Ляпунову, метод скользящих режимов.

Проблемы, связанные с созданием систем управления динамическими объектами, актуальны по нынешний день в силу сложности математического описания объектов управления и синтеза соответствующих управляющих алгоритмов.

Наиболее развита теория управления линейными стационарными системами. Во-первых, известны подходы к синтезу алгоритмов управления, когда объект управления описывается системой линейных дифференциальных или разностных уравнений [1]. Во-вторых, разработаны алгоритмы управления в случае описания линейной системы в классе обобщенных функций с использованием интеграла Дюамеля или уравнения Винера-Хопфа [2; 3]. В случае недоступности точного описания объекта управления отыскивают его модель, опираясь на априорную информацию о виде и структуре уравнений, законах распределения помех измерений входных и выходных величин и т. д. Принято говорить, что первому и второму вариантам описания объектов в стохастической постановке задачи соответствуют параметрические и непараметрические алгоритмы управления для линейных динамических систем.

Теория управления для нелинейных систем в настоящее время активно развивается. Среди эффективных подходов к разработке алгоритмов управления можно выделить подход с использованием регуляторов со скользящими режимами [4–6], обладающих высокими динамическими показателями и свойством грубости (робастности) по отношению к параметрическим и координатным возмущениям. Управляющее воздействие в методе скользящих режимов синтезируется из условий устойчивости, заданных условиями теоремы устойчивости Ляпунова. Метод имеет следующие особенности:

– применим для нелинейных динамических систем;

– управляющее воздействие отыскивается в классе разрывных функций;

– метод требует знания модели объекта управления (в общем случае модель в пространстве состояний);

– для функционирования регулятора требуется информация о значениях переменных состояния;

– метод устойчив к ошибкам идентификации модели объекта управления и ошибкам измерения переменных состояния.

Постановка задачи. Пусть объект управления задан дифференциальным уравнением, линейным относительно входного воздействия:

$$x^{(n)} = f\left(x, x', \ldots, x^{(n-1)}\right) + \varphi\left(x, x', \ldots, x^{(n-1)}\right) u,$$

где $x = x(t)$ – траектория движения объекта, его реакция на входной процесс $u = u(t)$; функции $\varphi\left(x, x', \ldots, x^{(n-1)}\right)$ и $f\left(x, x', \ldots, x^{(n-1)}\right)$ – в общем случае нелинейные и разрывные, потребуем также выполнения условия $\varphi\left(x, x', \ldots, x^{(n-1)}\right) \neq 0$.

Определим цель управления как $x^* = 0$. В ином случае, если $x^* \neq 0$, корректируем функции φ и f так, чтобы цель управления стала $x^* = 0$.

Вводится понятие поверхности скольжения, изменяющейся со временем поверхности в пространстве состояний, обеспечивающей устойчивость движения системы. Поверхность скольжения задается скалярным уравнением

$$s\left(x, x', \ldots, x^{(n-1)}, t\right) = 0.$$

В общем случае для системы порядка n поверхность скольжения определяется выражением

$$s\left(x, x', \ldots, x^{(n-1)}, t\right) = \left(\frac{d}{dt} + \gamma\right)^{n-1} x, \ \gamma > 0.$$

Управление объектом осуществляется в два этапа: вывод процесса на поверхность скольжения и скользящий режим – движение процесса по поверхности скольжения до состояния равновесия. Для обеспечения устойчивости движения к поверхности скольжения вводится функция Ляпунова

$$V = \frac{1}{2} s^2.$$

Условие устойчивости имеет вид

$$V' \leq 0,$$

где

$$V' = \left(\frac{1}{2}s^2\right)' = ss' = s\left[\left(\frac{d}{dt}+\gamma\right)^{n-1}\frac{d}{dt}x\right] = s\left(f+\varphi u+y\right);$$

$$y = y\left(x',\ldots,x^{(n-1)},\gamma\right) = (n-1)\ \gamma\ x^{(n-1)} +$$

$$+(n-1)(n-2)\gamma^2 x^{(n-2)}/2!+\ldots+\gamma^n x'.$$

Устойчивость движения обеспечивается выбором управляющего воздействия:

$$u\ \begin{cases} <\beta,\ s>0,\\ =\beta,\ s=0,\\ >\beta,\ s<0, \end{cases}$$

где

$$\beta = -\frac{f+y}{\varphi}.$$

Метод скользящих режимов доказал свою эффективность для широкого круга практических приложений. Однако, для его использования необходима модель нелинейного объекта управления. Поэтому синтез алгоритма предваряется решением задачи идентификации объекта управления, в частности, в нашем случае, необходимо идентифицировать функции φ и *f*. На практике не всегда возможно отыскать модель в виде уравнения, известного с точностью до вектора параметров, с последующей его оценкой, так как в теории не существует универсального алгоритма идентификации структуры объекта. В работе предлагается отыскивать модель нелинейного объекта управления в классе непараметрических моделей.

Алгоритм решения задачи. Обобщим описание объекта управления:

$$x^{(n)} = \varphi\left(x,x',\ldots,x^{(n-1)},u\right).$$

Линейность уравнения относительно входного воздействия не требуется. Первая производная функции Ляпунова имеет вид

$$V' = ss' = s\left(\varphi\left(x,x',\ldots,x^{(n-1)},u\right)+y\left(x',\ldots,x^{(n-1)},\gamma\right)\right).$$

Для устойчивости движения управляющее воздействие *u* выбираем так, чтобы обеспечивалось выполнение следующих условий:

$$\varphi\left(x,x',\ldots,x^{(n-1)},u\right)+y\left(x',\ldots,x^{(n-1)},\gamma\right)>0,\ \text{если}\ s<0,$$

$$\varphi\left(x,x',x^{(n-1)},u\right)+y\left(x',\ldots,x^{(n-1)},\gamma\right)=0,\ \text{если}\ s=0,$$

$$\varphi\left(x,x',\ldots,x^{(n-1)},u\right)+y\left(x',\ldots,x^{(n-1)},\gamma\right)<0,\ \text{если}\ s>0.$$

Преобразуем условия устойчивости к следующему виду:

$$\varphi\left(x,x',\ldots,x^{(n-1)},u\right)=a-y\left(x',\ldots,x^{(n-1)},\gamma\right),\ \text{если}\ s<0,$$

$$\varphi\left(x,x',\ldots,x^{(n-1)},u\right)=-y\left(x',\ldots,x^{(n-1)},\gamma\right),\ \text{если}\ s=0,$$

$$\varphi\left(x,x',\ldots,x^{(n-1)},u\right)=-a-y\left(x',\ldots,x^{(n-1)},\gamma\right),\ \text{если}\ s>0.$$

где $a=\text{const}>0$. Решением задачи будет обратная зависимость значений входного процесса от переменных состояния и вспомогательной величины $\pm a-y$:

$$u = q\left(x,x',\ldots,x^{(n-1)},\pm a-y\right).$$

Для синтеза модели перейдем к конечно-разностному представлению:

$$u_d = q_d\left(x(t),x(t-1)\ldots,x(t-n+1),\pm a-y(t)\right).$$

Здесь аргумент *t* представляет собой дискретную величину: $t=1,2,\ldots$ Для перехода к дискретной модели используем аппроксимацию производной по времени в виде конечной разности с первым порядком аппроксимации $\frac{dx(t)}{dt}\approx\frac{\Delta x(t)}{\Delta t}=\frac{x(t)-x(t-\Delta t)}{\Delta t}$ либо с использованием других более сложных схем [7]. Для удобства записи переименуем переменные:

$$z_1 = x(t),\ z_2 = x(t-1),\ z_3 = x(t-2),\ \ldots,$$

$$z_{n-1} = x(t-n+1),\ z_n = \pm a-y(t).$$

В результате получили модель в виде

$$u_s = q_s\left(z_1\ldots,z_n\right).$$

Проблема построения модели динамического процесса сводится к решению проблемы идентификации статического объекта, входными воздействиями которого становятся состояния в текущий и предыдущие моменты времени, а также введенная ранее вспомогательная величина.

Во время функционирования объекта накапливаем выборку $\{z_k[i],u[i]\}$, $k=1,2,\ldots n$, $i=1,2,\ldots s$. Полученная выборка используется для построения непараметрической модели Надарая–Ватсона [2; 3]:

$$u_s = \frac{\sum_{i=1}^{s}u[i]\prod_{k=1}^{n}\Phi\left(\frac{z_k-z_k[i]}{C_{sk}}\right)}{\sum_{i=1}^{s}\prod_{k=1}^{n}\Phi\left(\frac{z_k-z_k[i]}{C_{sk}}\right)}.$$

В случае равенства нулю знаменателя дополняем модель:

$$u_s = 0,\ \text{если}\ \sum_{i=1}^{s}\prod_{k=1}^{n}\Phi\left(\frac{z_k-z_k[i]}{C_{sk}}\right)=0.$$

Параметры C_s удовлетворяют условиям сходимости:

$$C_s>0,\ \lim_{s\to\infty}C_s=0,\ \lim_{s\to\infty}sC_s=\infty,$$

а ядерная функция Ф приобретает вид

$$\frac{1}{C_s}\int_{-\infty}^{\infty}\Phi\left(\frac{t-T}{C_s}\right)dt=1,$$

$$\lim_{C_s\to 0}\left(\frac{1}{C_s}\int_{-\infty}^{\infty}\varphi(t)\ \Phi\left(\frac{t-T}{C_s}\right)dt\right)=\varphi(T).$$

Перед использованием модели необходимо провести процедуру оптимизации относительно вектора параметров $\overline{C}_s = \{C_{s1}, C_{s2}, \ldots, C_{sn}\}$ совместно с параметром α. Предлагается использовать критерий оптимизации, минимизирующий средний квадрат отклонения выходного процесса от желаемого. Критерий оптимизации принимает вид

$$\sum_{t=0}^{T}\left[x\left(u_s\left(\overline{C}_s, a\right), t\right)\right]^2 \to \min_{\overline{C}_s, a}.$$

Указанная форма критерия обусловлена равенством нулю желаемого выхода процесса.

Результаты моделирования. Тестирование алгоритма проводилось на задаче управления положением обратного маятника, совершающего колебания на плоскости, т. е. имеющего одну степень свободы. Требовалось привести маятник в вертикальное положение из некоторого начального состояния, заданного пользователем, а в дальнейшем сохранять это положение. Была взята физическая модель маятника – тонкого невесомого стержня с заданной массой, закрепленной на его верхнем конце. Во время перемещения в пространстве груз испытывает сопротивление окружающей среды. Управляющее воздействие представляло собой момент сил, приложенный к маятнику в точке его опоры – «нижнем шарнире». Движение маятника описывается системой нелинейных дифференциальных уравнений:

$$\begin{cases} \alpha' = \omega, \\ \omega' = \dfrac{g}{l}\sin\alpha - \dfrac{k_F l}{m}\omega^2 \operatorname{sign}\omega + \dfrac{M}{ml^2}, \end{cases}$$

где l – длина маятника; m – масса груза; g – ускорение свободного падения; M – момент сил, действующих на маятник, – управляющее воздействие; k_F – коэффициент сопротивления окружающей среды; α – угол отклонения маятника от вертикального положения; ω – угловая скорость маятника.

Управляющее устройство включено в цепь обратной связи, оно получает информацию о состоянии объекта управления и генерирует управляющее воздействие. Физическая модель объекта управления устройству управления недоступна. Информация об объекте заключена в выборке измерений переменных состояния и вспомогательной величины, накопленных заранее в процессе функционирования объекта управления. Необходимо отметить, что во время накопления выборки объект управляется с использованием каких-либо других регуляторов, вручную оператором либо, если это допустимо, претерпевает «неудачные» попытки управления в разомкнутой схеме, в том числе, например, реагирует на случайное, хаотическое управляющее воздействие.

Алгоритм управления реализован в пакете MATLAB версии 7 с применением визуальной системы моделирования динамических процессов Simulink (рис. 1).

Блок *Inverted Pendulum* содержит физическую модель обратного маятника. На вход блока подается значение начального угла отклонения маятника от вертикального положения *Initial Angle*, а также входное воздействие $u(t)$, которое генерируется управляющим элементом *Controller*. Входная информация для элемента управления: переменные состояния $x(t)$ и $x'(t)$, вспомогательная величина $\pm a - y(t)$, параметры размытости c_1, c_2 и c_3, коэффициенты α и γ. Блок *Supplement* вычисляет значение вспомогательной величины на каждом такте управления.

Как отмечалось ранее, перед использованием алгоритма должна осуществляться процедура оптимизации с применением среднеквадратического критерия. Оптимизация проводилась методом сопряженных градиентов с мультистартом по параметрам, обозначенным на схеме k_1, k_2 и k_3. Из практических соображений эти параметры были выбраны следующим образом: $k_1 = c_1$, $k_2 = c_2 = c_3$, $k_3 = \alpha$.

Результаты одного из экспериментов приведены ниже (рис. 2).

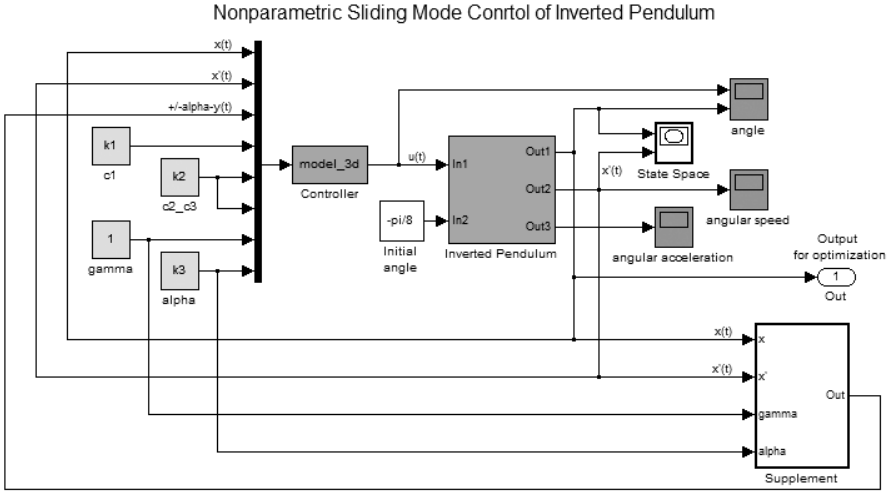


Рис. 1. Модель Simulink, реализующая алгоритм управления

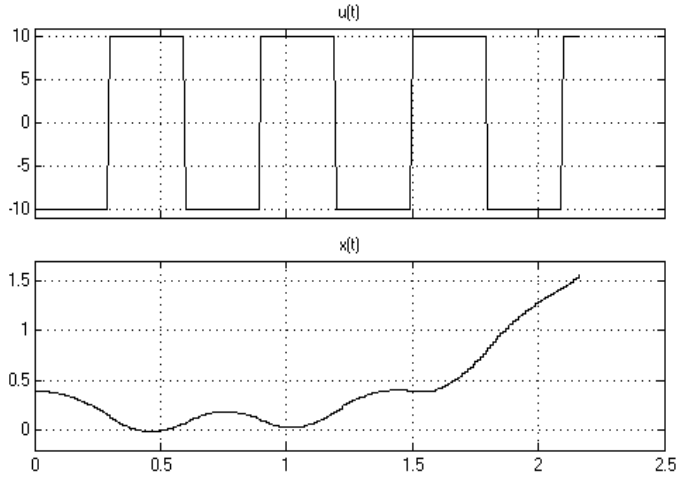


Рис. 2. Тестовое управляющее воздействие и реакция маятника в разомкнутой схеме

Момент времени, когда процессы прерываются, соответствует падению маятника на условную горизонтальную плоскость, т. е. $x(t) = \pi/2$. Объем обучающей выборки $s = 218$, выборочные значения поступают через равный временной интервал $\Delta t = 0,01$ и не содержат случайных ошибок.

Начальное положение маятника было принято $x(0) = -\pi/8 \approx -0,39$. Это значение лежит вне диапазона обучающей выборки, в этом нетрудно убедиться, взглянув на рис. 2. Остальные параметры $l = 1$ м, $m = 1$ кг, $k_F = 0$, $\gamma = 1$. В результате оптимизации были вычислены следующие значения параметров устройства управления: $c_1 = 0,8$, $c_2 = c_3 = 1,7$, $\alpha = 0,7$. Результат управления маятником изображен на рис. 3.

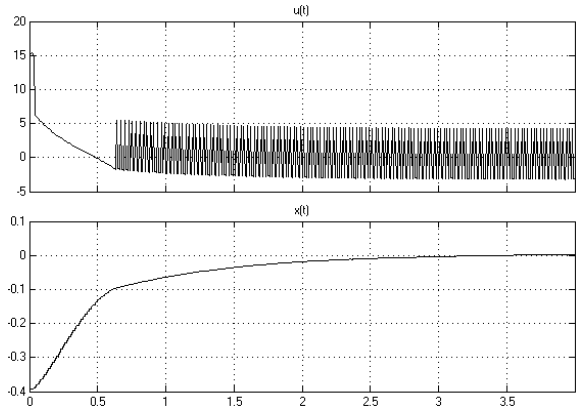


Рис. 3. Входное воздействие $u(t)$ и результирующий отклик маятника $x(t)$

Нетрудно заметить, что входной процесс $u(t)$ состоит из двух составляющих: гладкой и осциллирующей. Они соответствуют двум стадиям работы регулятора: выводу на скользящую поверхность и режиму скольжения (рис. 4).

Выход на поверхность скольжения $s = x(t) + x'(t) = 0$ происходит со второй попытки. Осцилляция во время процесса скольжения наблюдается в области значений производной выходного процесса.

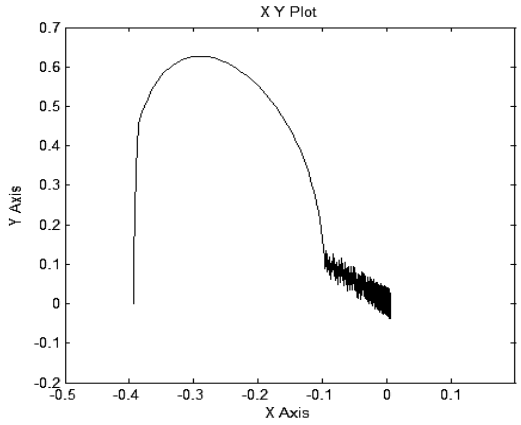


Рис. 4. Траектория управляемого движения маятника в пространстве состояний

Таким образом, использование непараметрического подхода к идентификации позволяет осуществить метод скользящих режимов в управлении нелинейным динамическим объектом, модель которого невозможно получить с точностью до набора параметров. Генерация управляющего сигнала в классе разрывных функций позволяет свести к минимуму негативный эффект ошибок (как ошибок идентификации, так и влияния случайных ошибок измерений) в осуществлении процесса управления.

В настоящей работе была предпринята попытка построения эффективного непараметрического алгоритма управления, информация для функционирования которого заключена в обучающей выборке. Выборка должна быть известна заранее, до начала процесса управления. В дальнейшем необходимо совершенствование предложенного алгоритма управления и создание его адаптивной модификации.

Библиографические ссылки

1. Воронов А. А., Титов В. К., Новогранов Б. Н. Основы теории автоматического регулирования. М. : Высш. шк., 1977.

2. Медведев А. В. Адаптация в условиях непараметрической неопределенности // Адаптивные системы и их приложения : сб. науч. тр. / Новосибирск : Наука, 1978. С. 4–34.

3. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука, 1983.

4. Уткин В. И. Скользящие режимы и их применения в системах с переменной структурой. М. : Наука, 1974.

5. Kjaer M. A. Sliding Mode Control. Dept. of Autom. Control // Lund Institute of Technology, Sweden, 2004.

6. Slotine J.-J. E., Li W. Applied Nonlinear Control. Englewood Cliffs : Prentice-Hall, 1991.

7. Самарский А. А. Теория разностных схем. М. : Наука, 1989.

E. D. Agafonov

NONPARAMETRIC CONTROL ALGORITHM FOR NONLINEAR DYNAMIC SYSTEMS USING SLIDING MODES

The paper deals with a new control algorithm designed for nonlinear dynamic systems with the use of Sliding Mode control approach. The SISO object is represented by its nonparametric finite differences model in state space. The paper gives recommendations for tuning and optimization of the control algorithm. The proposed control algorithm is implemented in MATLAB/Simulink technical computing software. As an illustrative example we presrnt results of control process of inverted pendulum.

Keywords: nonlinear dynamic system, finite difference model, nonparametric control, Lypunov stability, sliding mode control.

© Агафонов Е. Д., 2010

UDC 519.234

K. Zablotskaya, S. Walter, S. Zablotskiy, W. Minker

DETERMINATION OF VARIABLES SIGNIFICANCE USING ESTIMATIONS OF THE FIRST-ORDER PARTIAL DERIVATIVE

In this paper we describe and investigate a method which allows us to detect the most informative features out of all data extracted from a certain data corpus. The significance of input features is estimated as an average absolute value of the first-order partial derivative. The method requires the values of the objective function at the certain assigned points. If there is no possibility to calculate these values (the object is not available for experiments), we use non-parametric kernel regression to approximate them. The algorithm is tested on different simulated objects and is used for investigation of the dependency between linguistic features of spoken utterances and speakers' capabilities.

Keywords: non-parametric kernel regression, first-order partial derivative.

In our research we try to investigate if there is a dependency between spoken utterances of a person and his capabilities. For this purpose we collected a corpus of monologues and dialogues of different speakers [1]. Their verbal intelligence was measured with an intelligence test [2]. Dealing with the corpus, we try to extract relevant information enough for clustering, classification, regression, or other data mining tasks. There are normally lots of different features which could be extracted from the monologues and dialogues, but their importance or relevance is not always obvious. Most of them are noise fields, which make the analysis of data increasingly difficult. When working with high dimensional spaces, the computational effort required by data analysis tools may be tremendous. It is therefore essential to detect irrelevant or weakly correlated features and exclude them out of consideration.

There exist different solutions to this problem. One of them is the use of Pearson's coefficient or the coefficient of multiple correlation. However, if Pearson's coefficient is close to 0, it does not mean that the output and input variables are not correlated. It just shows that there is no linear dependency between them. Such features should not be excluded out of consideration without additional analysis. Another approach to decrease the number of features is Principal Component Analysis. This method involves a mathematical procedure that transforms correlated variables into a smaller number of uncorrelated ones called principal components. But it does not determine the contribution of a certain feature to an objective function.

In this paper we describe a method which determines the most informative features even if the dependency between input variables and the output is not linear.

Determination of the Most Informative Features. To determine if there is a dependency between input features (or extracted features) and the output, we make a series of experiments on the object (if it is available) or create a model using non-parametric kernel regression and estimate the average first-order partial derivative with respect to each input feature. The feature with the largest average partial derivative is the most important. This algorithm may be described in the following way.

Non-parametric kernel regression (NPR) allows us to create a model using the data set $x_1[t], ..., x_n[t], y[t], t = \overline{1,s}$ without additional knowledge about the dependency structure [3; 4]. NPR estimates the dependency between inputs and outputs using a weighted average of the observations $y[t]$:

$$\hat{M}\{Y \mid x\} \equiv \hat{y}(\bar{x}) = \frac{\sum_{t=1}^{s} y[t] \prod_{i=1}^{n} \Phi\left(\frac{x_i - x_i[t]}{C_i}\right)}{\sum_{t=1}^{s} \prod_{i=1}^{n} \Phi\left(\frac{x_i - x_i[t]}{C_i}\right)},$$

where C_i – is the bandwidth or smoothing parameter, $\Phi(z)$ – is a kernel function.

The kernel function assigns weights for each observation. The weighted sum of $y[t]$ estimates the output at any point $\bar{x}$. The parameters C_i determine how many points from the training data set are used for calculating $\hat{y}(\bar{x})$. The observations which are nearer to $\bar{x}$ have larger weights and are more significant for $\hat{y}(\bar{x})$. If C_i are large, a lot of observations are taken into account and the model is not precise. These parameters should be trained on the existing data set, and C_i providing the smallest mean square error (MSE) are used for other investigations.

Let an object have an input vector $\vec{x} = (x_1, x_2, ..., x_n)$ and an output $y = f(\vec{x})$. A feature x_i is informative if its average influence on the output is significant, given the other $n-1$ features fixed. We estimate this significance as an average absolute value of the first-order partial derivative with respect to this variable.

Let the variables $\vec{x} = (x_1, x_2, ..., x_n)$ belong to the intervals $[a_1; b_1]$, $[a_2; b_2]$, …, $[a_n; b_n]$. We generate random values $\{\tilde{x}_1[1], ..., \tilde{x}_1[m], \tilde{x}_2[1], ..., \tilde{x}_2[m], ...\}$ in corresponding intervals, m is a predefined value. To get a precise estimation of the average first-order partial derivative, we generate these random values near one observation value $\bar{x}[l]$, $l = \overline{1,s}$:

$$\sum_{t=1}^{s} \prod_{i=1}^{n} \Phi\left(\frac{\tilde{x}_i[k] - x_i[t]}{C_i}\right) > 0,$$

for all $k = \overline{1,m}$. Then we fix the features $(x_2, ..., x_n)$ at some points, for example at $(\tilde{x}_2[1], \tilde{x}_3[1], ..., \tilde{x}_n[1])$. The outputs of the goal function are estimated at the following points:

$$\hat{y}_1^+ = f(\tilde{x}_1[1] + h_1^+, \tilde{x}_2[1], ..., \tilde{x}_n[1]),$$
$$\hat{y}_1^- = f(\tilde{x}_1[1] - h_1^-, \tilde{x}_2[1], ..., \tilde{x}_n[1]),$$
$$\hat{y}_2^+ = f(\tilde{x}_1[2] + h_2^+, \tilde{x}_2[1], ..., \tilde{x}_n[1]),$$
$$\hat{y}_2^- = f(\tilde{x}_1[2] - h_2^-, \tilde{x}_2[1], ..., \tilde{x}_n[1]), \ldots,$$

where h_1^+ and h_1^- are random values from a small interval (for example, $h \in [0,01; 0,5]$).

The first component of the average first-order partial derivative with respect to x_1 is estimated as:

$$\hat{f}'_{x_1}[1] = \frac{\frac{|\hat{y}_1^+ - \hat{y}_1^-|}{h_1^+ - h_1^-} + \frac{|\hat{y}_2^+ - \hat{y}_2^-|}{h_2^+ - h_2^-} + ... + \frac{|\hat{y}_m^+ - \hat{y}_m^-|}{h_m^+ - h_m^-}}{m}.$$

Then the features $(x_2, ..., x_n)$ are fixed at other points, for example at $(\tilde{x}_2[2], \tilde{x}_3[1], ..., \tilde{x}_n[1])$, and the same procedure is repeated for x_1. The average absolute value of the partial derivative in the neighborhood of $\bar{x}[l]$ is estimated as:

$$\hat{f}'_{x_1}(\bar{x}[l]) = \frac{\hat{f}'_{x_1}[1] + \hat{f}'_{x_1}[2] + ... + \hat{f}'_{x_1}[M]}{M},$$

where M – the number of all possible combinations, $M = m^{(n-1)}$. As these random values have been generated in the neighborhood of one observation values, only small space is investigated. We generate $\{\tilde{x}_1[1], ..., \tilde{x}_1[m], \tilde{x}_2[1], ..., \tilde{x}_2[m], ...\}$ next to another observation point $\bar{x}[l']$ and find $\hat{f}'_{x_1}(\bar{x}[l'])$ in the same way. This procedure is repeated K times. The average absolute value of the partial derivative $\hat{f}'_{x_1}$ – is estimated as: $\hat{f}'_{x_1} = \sum_{l=1}^{K} \hat{f}'_{x_1}(\bar{x}[l]) \Big/ K$, where K is a predefined value. In the same way the average absolute values of $\hat{f}'_{x_i}$, $i = \overline{1,n}$ are estimated.

Investigation of the Algorithm. In this section we show the results of the algorithm's work when the object is not available for experiments, i. e. there is only collected data. In the following experiments the average absolute value of the partial derivative is estimated with $M = 2$ and $K = 20$. The function for simulating the object is given by: $f(\bar{x}) = 5x_1 + 0,5x_2 - 10x_3 + 0,1x_4 + 2x_5$.

In our first experiment, the non-parametric regression model is trained using all the input variables ($C_i = [0,4; 1,7; 0,3; 1,8; 0,8]$, $MSE = 0,08$).

Then we take away the first feature x_1 from the data set. In this case the situation with incomplete data is simulated, but the most informative features should nevertheless be found ($C_i = [1,9; 0,3; 1,9; 0,9]$, $MSE = 0,39$). The results of the algorithm are shown in Table 1. As we can see, the algorithm was able to find the most important features in both cases.

Table 1

Results of the algorithm's work

Features	5 inputs are used		4 inputs are used	
	$\hat{f}'_x$ (Real ranks)	Algorithm's $\hat{f}'_x$ (Ranks)	$\hat{f}'_x$ (Real ranks)	Algorithm's $\hat{f}'_x$ (Ranks)
x_1	5,0 (2)	3,42 (2)	–	–
x_2	0,5 (4)	1,40 (4)	0,5 (3)	0,39 (3)
x_3	10,0 (1)	6,51 (1)	10,0 (1)	7,97 (1)
x_4	0,1 (5)	1,20 (5)	0,1 (4)	0,28 (4)
x_5	2,0 (3)	2,10 (3)	2,0 (2)	1,39 (2)

Table 2

Results of the algorithm's work

Features	5 inputs are used		8 inputs are used	
	$\hat{f}'_x$ (Real ranks)	Algorithm's $\hat{f}'_x$ (Ranks)	$\hat{f}'_x$ (Real ranks)	Algorithm's $\hat{f}'_x$ (Ranks)
x_1	3,83 (5)	3,15 (5)	3,83 (5)	2,71 (5)
x_2	4,23 (4)	3,29 (4)	4,23 (4)	3,24 (4)
x_3	4,38 (3)	3,50 (3)	4,38 (3)	4,38 (3)
x_4	7,05 (2)	4,24 (2)	7,05 (2)	4,43 (2)
x_5	30,0 (1)	31,48 (1)	30,0 (1)	26,23 (1)
x_6	–	–	0,027 (7)	0,43 (7)
x_7	–	–	0,021 (8)	0,07 (8)
x_8	–	–	0,06 (6)	0,74 (6)

Now let's use the following function for generating the input and output data: $f(\bar{x}) = 7\sin(x_1) + 6\cos(x_2) - -8\sin(x_3) - 10\cos(x_4) + 5x_5^2$. In this dependency there are no features which influence on the output is linear. This is a more complex situation for the algorithm. However, if the model is trained well ($C_i = [1,0;\ 0,6;\ 0,7;\ 0,5;\ 0,1]$, $MSE = 0,32$), the algorithm gives us good results (see Table 2).

Let's use the same function for simulating the data set and add three more features to the input variables. We simulate the situation when the data set is large and not all features influence on the output. These additional input features are: $x_6 = 0,05\sin(t)$, $x_7 = 0,03\cos(t)$, $x_8 = 0,01t^2$. The coefficients with these features are small, so that x_6, x_7 and x_8 are noises for the output. In this case we use all the features to train the model. The results are shown in Table 2. The algorithm could find the most informative and the least informative features ($C_i = [1,0;\ 0,6;\ 0,7;\ 0,5;\ 0,1;\ 1,9;\ 1,5;\ 1,4]$, $MSE = 0,32$).

Let's use for simulating the function:

$$f(\bar{x}) = 0,2\sin(2x_1) + 2\cos(8x_2) + 5\sin(x_3) + \\ + 0,1x_4 + 0,5x_5 + x_6 + 2x_7 + 3x_8 + 4x_9 + 5x_{10},$$

generate the data set and take away the features x_1, x_2 and x_3. The results of the algorithm ($C_i = [1,5;\ 1,4;\ 1,0;\ 0,9;\ 0,5;\ 0,5;\ 0,5]$, $MSE = 0,3$) are given in Table 3.

Analyzing the results in the tables, we may say that the algorithm with the non-parametric model can find the most informative features. This method can be used for analyzing a high dimensional data set. It allows us to exclude the least informative features from consideration.

Table 3

Results of the algorithm's work

Features	$\hat{f}'_x$ (Real ranks)	Algorithm's $\hat{f}'_x$ (Ranks)
x_1	–	–
x_2	–	–
x_3	–	–
x_4	0,1 (7)	0,61 (7)
x_5	0,5 (6)	0,71 (6)
x_6	1 (5)	1,26 (5)
x_7	2 (4)	1,80 (4)
x_8	3 (3)	3,79 (3)
x_9	4 (2)	4,15 (2)
x_{10}	5 (1)	4,83 (1)

Experiments with the Corpus. We decided to analyze different features extracted from monologues of German native speakers using the algorithm described above. The corpus consists of transcribed descriptions of a short film by different candidates. German native speakers of different ages and educational levels were asked to watch a short film and to describe it with their own words. The film was about an experiment on how long people could be without sleep. The participants were also asked to take an intelligence test. The verbal part of the test consists of 6 subtests. The first subtest is «Information». With this sub-test the general knowledge is measured; 25 questions come from a particular culture. For example, «What is the capital of Russia?» Overall, 56 candidates were tested; 3 hours 30 minutes of audio data were collected.

To extract features from the monologues, all the words from the descriptions were compared with a special dictionary [5]. The dictionary consists of different words sorted by 64 categories. For example, the category «Articles» contains words die, das, der, ein, eine, einen, etc. Each word from the dictionary may refer to several

categories. For example, the word traurig (sad) refers to the categories «Affect», «Negative emotion» and «Sadness». We analyzed all the monologues, calculated the number of words for each category and divided them by the total amount of words in each monologue. By this way we got 64 characteristics of 56 monologues. Our task was to investigate the dependency between these 64 features and the results of the subtest «Information», and to find several informative features out of 64 characteristics.

We combined 4 or 5 features together, trained the non-parametric model and applied our method. As a result, the category «Affection» had the largest value of the first-order partial derivative and was estimated as a more informative feature. «Positive emotions» and «Negative emotions» are subcategories of «Affection» and are also relevant according to our algorithm. However, «Anger» and «Optimism» do not have large values of $\hat{f}'_{x_i}$. The category «Cognitive mechanism» is estimated as irrelevant, however, the category «Cause» which is a subcategory of «Cognitive mechanism» is more important.

Discussion and Future work. The goal of this work was to apply the method to the corpora. In each combination of the features the category of emotions was determined as the most informative feature. It means that there is a dependency between speaker's general knowledge and the amount of emotional words which he uses in his speech. We could not find any references describing this dependency. Only in LEAS [6] emotional intelligence is measured linguistically, however, the correlation between them was not found. For our algorithm we used a small data set that also influenced the results. Also, these emotional words may be a subcategory of another category which was not analyzed. For example, they may create a group of frequently-used words, or they are formed from abstract words which show the level of intelligence in spoken utterances. This research and the results are preliminary; in our future work we are going to further investigate this phenomenon, to find other linguistic features which reflect verbal intelligence and to collect more data for more precise estimations.

References

1. Zablotskaya K., Walter S., Minker W. Speech data corpus for verbal intelligence estimation. In Proceedings of International Conference on Language Resources (LREC) 2010.

2. Wechsler D. Handanweisung zum Hamburg-Wechsler-Intelligenztest fuer Erwachsene (HAWIE). Separatdr., Bern; Stuttgart; Wien, Huber. 1982.

3. Nadaraya E. On estimating regression. Theory of Probability and its Applications. 1964. Vol. 10. P. 186–190.

4. Watson G. Smooth regression analysis. Sankhya – The Indian Journal of Statistics. 1964. Vol. 26. P. 359–372.

5. Computergestuetzte quantitative textanalyse / M. Wolf, A. B. Horn, M. R. Mehl et al. // Diagnostica, 54. 2008. Heft 2. P. 85–98.

6. Lane R. D, Schwartz G. E. Levels of emotional awareness: a cognitive-developmental theory and its application to psychopathology. Am J Psychiatry. 1987. P. 113–143.

© Zablotskaya K., Walter S., Zablotskiy S., Minker W., 2010

UDC 519.234

S. Zablotskiy, T. Müller, W. Minker

ESTIMATION OF RADIO SIGNAL QUALITY DEGRADATION BY MEANS OF NEURAL NETWORK AND NON-PARAMETRIC REGRESSION MODEL

In this paper we present an approach which allows us to avoid expansive and time consuming subjective assessments of audio quality degradation caused by different nature distortions while transmitting and receiving of stereo audio signal through the radio channel. This approach is based on the basic version of PEAQ (Perceptual Evaluation of Audio Quality) originally developed mainly for audio codec estimation. The MOV (Model Output Variables) vector of the PEAQ method is mapped to the audio quality degradation scale using two different models: neural networks and non-parametric regression. The results of two independent approaches are compared.

Keywords: PEAQ, audio quality degradation, neural network, non-parametric regression.

The manufacturers of radio receivers and other radio equipment have to estimate the quality of the new product comparing to the existent equipment. Among other things the common listening comprehension of the perceived degradation with respect to the original (reference) audio signal has to be taken into consideration because humans with their own listening comprehension are supposed to be the main end-users of the designed equipment. Since any high-quality reliable subjective assessments are very expensive and time consuming it is strongly desired to have a tool for automatic perceptual evaluation of the audio quality degradation. This is a fundamental idea behind the PEAQ method, as specified in ITU-R BS.1387 recommendation [1]. According to this recommendation the PEAQ measurement method is applicable to the most types of audio signal processing equipment, both digital

and analog. However, it is expected that many applications will focus on audio codecs. Test results obtained with the help of the PQevalAudio software, compliant with the basic model of PEAQ [2], disagree with the Mean Opinion Score (MOS) of the test persons. The Objective Difference Grade (ODG) of this software after appropriate linear rescaling is always significantly smaller than MOS. This implies that the PQevalAudio estimates the quality of the received (degraded) audio signal to be permanently lower than it seems to be for subjects from a test group. One of the obvious reasons is that the human requirements for audio files (signals) degraded by audio coding (recoding) are significantly higher than the requirements for files degraded by the entire radio transmitting-receiving process.

The output of the software is not only the final ODG value, but also 10-dimensional Model Output Variables (MOV's) combined using a neural network to give the ODG value [2]. Our goal is to create a new model having MOV's as the input variables and to train it on the basis of our test group subjective estimations to be able to use the whole algorithm for automatic perceptual evaluation of transmitted through the radio channel audio quality degradation. We have chosen two different approaches for MOS approximation: neural network [3] and non-parametric regression model [4]. The performance of both algorithms on the test set is compared and the conclusions are presented in the corresponding section.

Alignment of Reference and Received Audio Signals. All test audio files are divided into two unequal parts. The smaller one consists of artificially simulated audio signals degraded by the additionally applied noise signals of different types, which sound more or less real comparing to the radio noise. The second part of the test audio material was obtained under conditions maximally close to the real situation – the reference short audio files were captured from various Internet radio stations with acceptable quality which were aired at the same time (with a small time delay) in our local area. Degraded audio files were captured from the corresponding radio stations in different conditions: inside the buildings and outside, in the driving car.

PQevalAudio software assumes two input audio signals (reference and degraded) to be time and gain aligned. Therefore, all the files are aligned even before the subjective assessment stage to be more accurate, when finding the dependency between MOS and MOV's.

Variable Delay Compensation. The reference and degraded files transmitted through the radio channel have not only a constant delay between each other, but also a variable delay changing through the entire audio signal. The constant delay is calculated as a maximum value of the cross-correlation function. The variable delay compensation is an increasingly complex process. Both reference and degraded stereo files are first converted into monaural ones and then divided into blocks of chosen length. For each pair of blocks the mean delay is found. This vector is then digitally filtered by the use of the normalized symmetric Hanning window. The result delay vector is then linearly interpolated for each sample and the final dynamic shift is performed by the cubic interpolation.

Usually the variable delay seems to have close to linear dependency from the time. However, sometimes due to the strong sudden noise or other distortions there are areas of the variable delay graph, where the curve seems to be almost randomly changing. In this case it is recommended to approximate linearly such «unreliable» areas using the left «trustful» data (see Fig. 1).

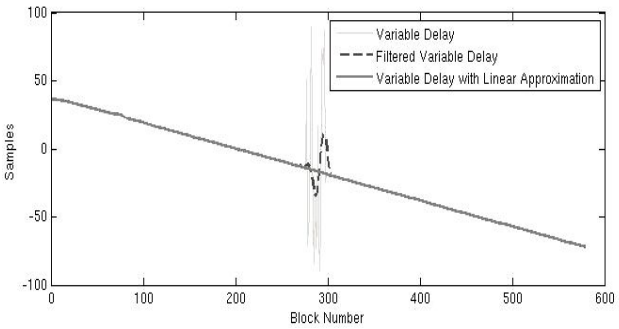


Fig. 1. Variable delay between reference and degraded files with and without approximation

In some specific cases the graph of the variable delay seems to change randomly. It is statistically estimated that in this case it is preferable to ignore the variable delay vector. The recognition of such case is automatically performed according to the acceptance threshold 0.1, i. e. the number of «reliable» data should not be less than 10 % of all block delays data. The smaller threshold would lead to the increasing of the average difference between files.

Gain Alignment. As a first step the reference and degraded stereo audio signals are converted into the monaural files. This should take into account, that the most of the receivers switch the audio to mono in the case of poor radio signal. Therefore, processing on the stereo signal may lead to a wrong gain calculation because of the reference and degraded files difference, caused even only by the switching to monaural signal, which can be quite pure itself.

The gain itself is calculated as the absolute value average of all samples of the monaural files. The sample values of the file yielding a higher gain are linearly reduced according to the gain difference. Then, the proportional normalization is applied for both files.

Subjective Assessments. Aligned files have to be estimated by the candidates from our test group. They should listen at first the reference and corresponding degraded audio files and then estimate the degradation of the second file with respect to the reference one. In our case each candidate has estimated 58 pairs of files, giving for every pair one value from the interval [0, 10] according to Table.

All pairs of files and appropriate MOS values are divided into two sets: training set used for model creation and the test set (10 pairs from 58). The data from the test set are not used for modeling to be honest in evaluating of approximation results.

Subjective estimation scale

Grade	Subjective impression
0	Program content is not/almost not understandable
2	Program content is in common understandable, irritative noise
4	Program content is mostly good understandable, especially when concentrating
6	Noise is (always) audible, also at the usual loudness
8	Reception distortion is audible when concentrating or at the high loudness
10	Reception distortion is inaudible even concentrating or at the high loudness

Neural Network Training. The modified approach for neural network training [5] by means of the genetic algorithm [6] is used. Genetic algorithms are global optimization stochastic algorithms which do not need the knowledge about the function and its derivatives. Only the values of the optimized function in the generated points are required for optimization [7; 8].

In our approach the genetic algorithm is applied two times within each iteration. Firstly it is used to optimize the weights of all neural nets in the current generation. The optimum in our case is the minimal value of the mean squared error (MSE) between the neural network output and the corresponding MOS values from a training set. Secondly the genetic algorithm is used to determine the new generation of the neural nets. In order to apply the optimization algorithm to a network structure its binary representation is used. The training is iterated until the stop conditions are satisfied or the break is performed manually.

Non-parametric regression model. Non-parametric kernel regression of a static object allows us to build a model of this object without any knowledge about the structure of the dependency between input (MOV's) and output (MOS) variables. This model can be presented in the form of a next function:

$$\hat{y}(\vec{x}) = \frac{\sum_{i=1}^{s} y[i] \prod_{p=1}^{n} \Phi\left(\frac{x_p - x_p[i]}{c_p}\right)}{\sum_{i=1}^{s} \prod_{p=1}^{n} \Phi\left(\frac{x_p - x_p[i]}{c_p}\right)}, \qquad (1)$$

where $\Phi(\cdot)$ is a kernel function; c_p – smoothness parameters, $p = \overline{1,n}$; $\vec{x}[i], y[i]$ – input and output variables from the training set, $i = \overline{1,s}$.

The kernel function should satisfy certain requirements [9] and in our case is chosen to be the following function:

$$\Phi(z) = \begin{cases} (3/4)(1-z^2), & if\ |z| \le 1; \\ 0, & if\ |z| > 1. \end{cases} \qquad (2)$$

The only unknown parameters of the algorithm are smoothness parameters c_p which should be trained on the base of the available data, minimizing Mean Square Error (MSE):

$$MSE(c_p) = \left(\frac{1}{s} \sum_{i=1}^{s} \left(\frac{y[i] - \sum_{\substack{j=1, \\ j \ne i}}^{s} y[j] \prod_{p=1}^{n} \Phi\left(\frac{x_p[i] - x_p[j]}{c_p}\right)}{\sum_{\substack{j=1, \\ j \ne i}}^{s} \prod_{p=1}^{n} \Phi\left(\frac{x_p[i] - x_p[j]}{c_p}\right)} \right)^2 \right)^{\frac{1}{2}}. \qquad (3)$$

To minimize this error one could use any optimization methods, especially stochastic methods aiming to find a global minimum, e. g. genetic algorithm. It is also possible to chose one single value of c and calculate all the smoothness parameters in the form of $c_p = c\,(\max_i x_p[i] - \min_i x_p[i])$. The parameter c should be decreased from a fixed value (for instance, $c = 0{,}5$) with a fixed small step until the value of the MSE starts to increase. To get a better result the found values of c_p should be adapted. We did a post-adaptation in our work manually; however, one could prefer a certain optimization method.

Experimental results. As it was previously mentioned, linearly mapped ODG value to the interval [0, 10] tends to be significantly smaller than MOS of the candidates from a test group (the MSE = 5,428).

The MSE of the difference between the output of the found neural network and MOS values calculated on the training data set is equal to MSE = 1,040 and calculated on the test data set is even smaller MSE = 0,971. The result approximation is at least much closer to the MOS, than linearly rescaled ODG software output.

Non-parametric kernel regression model on the same training data has achieved a little bit worse results MSE = 1,091 comparing to neural network. However, on the not-overlapping test data set the non-parametric model performed even better (MSE = 0,910) than our neural network. Moreover, it was found out that 6 out of the 10 sub-optimal smoothness parameters (according to 10-dimensional MOV's) are much larger than the scale of the data in the corresponding dimensions. This means that these 6 input variables do not play any role in the non-parametric model and therefore could be omitted. We created a new non-parametric model on the base of the reduced data size (6 variables were ignored) with the same left smoothness parameters values (re-training of the parameters did not bring any improvement). The result model achieved MSE = 1,092 on the training set and MSE = 0,906 on the test set that is the smallest achieved error on this test set.

The results of both algorithms, rescaled ODG value and the perceptual estimations of the candidates (MOS) for 10 values from a test set are presented in Fig. 2. The closer the approximation to MOS is, the better the model is. Non-parametric regression presented on the graph uses only 4-dimensional input data.

Conclusion and Future Directions. Either the neural network or non-parametric regression model in combination with the PQevalAudio software can be used for the perceptual evaluation of the audio quality degradation, transmitted through the radio channel. The both approaches are obviously not as precise as the

subjective assessments. However, they could be used for the mean opinion score approximation.

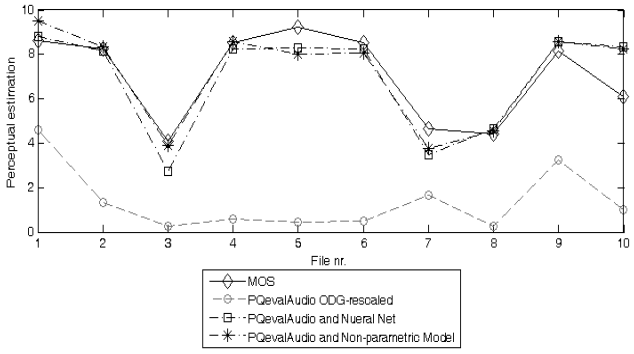


Fig. 2. The algorithms comparison

The non-parametric model has achieved slightly better results on the test set. The main advantage of this model is that we were able to exclude the insignificant features out of the model due to the huge values of the smoothness parameters. The input parameters left in the model are: «average block distortion», «distortion loudness», «noise-to-mask ratio» and «windowed modulation difference». The input parameters which were excluded from a model: «average modulation differences», «bandwidths of the reference and test signal», «harmonic structure of the error» and «relatively disturbed frames». The accuracy of the MOS approximation can be improved by the use of described algorithms trained on the considerably greater amount of data and estimated by a higher number of people.

References

1. ITU-R Recommendation BS.1387, Method for Objective Measurements of Perceived Audio Quality, Dec. 1998.

2. Kabal P. An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality // Department of Electrical & Computer Engineering, McGill University, Canada, 2002.

3. Anderson J. A. An Introduction to Neural Network, The MIT Press, 1995.

4. Nadaraya E. On estimating regression, Theory of Probability and its Applications. 1964. Vol. 10. P. 186–190.

5. Hagan M. T. Neural Network Design (Electrical Engineering). PWS Pub. Co. 1995.

6. Goldberg D. E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional. 1989.

7. Holland J. H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. The MIT Press. 1992.

8. Haupt R. L. Practical Genetic Algorithms, Wiley-Interscience. 1997.

9. Wand M., Jones M. Kernel Smoothing. New York : Chapman and Hall. 1995.

© Zablotskiy S., Müller T., Minker W., 2010

УДК 681.513.5; 517.977

А. В. Кузнецов, А. И. Рубан

АЛГОРИТМЫ МЕТОДА УСРЕДНЕНИЯ КООРДИНАТ ПРИ ПОИСКЕ ГЛАВНЫХ МИНИМУМОВ МНОГОЭКСТРЕМАЛЬНЫХ ФУНКЦИЙ

Построены алгоритмы поиска заданного количества главных минимумов многоэкстремальных функций многих непрерывных переменных при активном учете ограничений неравенств. В основе алгоритмов лежит разбиение заданной области поиска на подобласти, тяготеющие к требуемым главным минимумам, и последующий поиск в каждой найденной подобласти условного глобального экстремума на основе алгоритмов метода усреднения координат. Разбиение на подобласти производится также на основе алгоритмов усреднения координат путем их последовательных запусков и исключением уже найденных подобластей с помощью дополнительных ограничений неравенств. На численных примерах продемонстрирована эффективность работы алгоритмов.

Ключевые слова: глобальная оптимизация, главные минимумы.

При решении проблемы многокритериальной оптимизации часто основываются на одном из показателей эффективности и минимизируют (или максимизируют) его, а остальные показатели выдерживают не хуже заданных (так появляются дополнительные ограничения неравенства). В ограничения неравенства входят также ресурсные и конструктивные ограничения. При такой оптимизации исследователя интересует не только положение глобального минимума выбранного показателя эффективности, но и положение близлежащих (по величине минимизируемой функции) минимумов. Указанные минимумы назовем главными.

В работе [1] предложен алгоритм поиска главных минимумов без ограничений неравенств, в [2, с. 139–143] были проведены испытания алгоритма на простейших

численных. В работе [3] предложено три эвристических способа учета ограничений, основанных на последовательном выборе наилучшей точки в области поиска и выделении окрестности вокруг нее.

В предлагаемой работе рассматривается еще один способ выделения подобластей, соответствующих главным минимумам, путем многократного запуска алгоритмов, основанных на методе усреднения координат, по всей области поиска с исключением найденных областей с помощью добавления дополнительных ограничений неравенств.

Постановка задачи. Необходимо найти заданное количество главных минимумов функции многих переменных:

$$I(x) = \text{гл}\min_{x \in X}, \quad x = (x_1, ..., x_m)^T. \qquad (1)$$

Допустимая область X определяется системой ограничений неравенств

$$\varphi_j(x) \le 0, \; j = \overline{1, m_1}. \qquad (2)$$

Минимизируемая функция качества $I(x)$ имеет широкий диапазон свойств. Она может быть многоэкстремальной, разрывной, недифференцируемой, искажена помехой. Ограничения (2) сужают область поиска экстремума. Функции ограничений $\varphi_j(x), \; j = \overline{1, m_1}$ также могут быть невыпуклыми и недифференцируемыми. Допустимая область может иметь вид «узкого жгута». Поиск минимумов осуществляется только на основе измерений или вычислений указанных функций: $I(x); \varphi_j(x), \; j = \overline{1, m_1}$.

Алгоритмы. Три представленных ниже алгоритма включают в себя два этапа: первый этап – это разбиение исходной области на подобласти, второй – поиск глобального минимума в каждой из найденных подобластей. Особенность предлагаемых алгоритмов заключается в том, что на первом и втором этапе используются одни и те же алгоритмы усреднения координат, но запускаемые в разных областях и с разными наборами ограничений. Учет ограничений производится так же, как и в работе [3].

Исследователь задает:

– исходную гиперпрямоугольную область Π^0 (с центром в x^0 и интервалами Δx^0), которая должна содержать в себе искомые главные экстремумы;

– количество исходных пробных точек n^0, равномерно распределенных внутри области Π^0;

– количество искомых главных минимумов $n_{\text{гл мин}}$;

– количество пробных точек n, вид ядра и степень s его селективности, используемых при поиске глобального минимума в каждой подобласти.

Первый алгоритм (с использованием пробных точек внутри допустимой области).

В заданной прямоугольной области поиска Π^0 последовательно равномерно размещаются пробные точки, и из них формируется n^0 точек, попадающих в допустимую область (2). Разбиение пробных точек $x^{(i)}, \; i = \overline{1, n^0}$ на группы точек, тяготеющих к главным минимумам, производится следующим образом. В пробных точках вычисляется функция качества $I(x^{(i)}) = I^{(i)}, \; i = \overline{1, n^0}$, и на основе алгоритма глобальной оптимизации (описанного ниже) в области $\Pi^1 = \Pi^0 \times X$ отыскивается такая точка $x^1_{\text{мин}}$, которая соответствует наименьшей величине исходной функции качества $I_{\text{мин}} = \min\limits_{x \in \{\Pi^0 \times X\}} I(x)$. Затем вокруг этой точки выделяется подобласть некоторого размера, например $\Delta x^0_v / c$, $v = \overline{1, m}$. Обозначим эту подобласть как $\Pi^1_{\text{пр}}$ и добавим ее к списку ограничений (2), тем самым исключив ее из области Π^0. Далее, аналогичным способом, ведется поиск в области $\Pi^2 = \Pi^1 \setminus \Pi^1_{\text{пр}}$ и получается область $\Pi^2_{\text{пр}}$ с центром в $x^2_{\text{мин}}$ и т. д. Выделение подобластей осуществляем до тех пор, пока их количество не станет равным заданному количеству $n_{\text{гл мин}}$ главных минимумов или алгоритм не сможет выполнить поиск $x_{\text{мин}}$ для очередной подобласти.

Параметр $c \ge 1$ определяет размеры подобластей и задается исследователем пока интуитивно на основе размеров исходной области Π^0, априорной информации о функции качества и требуемого количества главных минимумов $n_{\text{гл мин}}$.

На этом первый этап – разбиение исходной области на подобласти – считается завершенным.

На втором этапе производится отыскание экстремумов в каждой найденной подобласти. При этом к неравенствам (2), которые входят в исходную постановку задачи, добавляются неравенства, определяющие прямоугольные подобласти, найденные на этапе разбиения исходной области на подобласти. Добавление таких неравенств не дает алгоритму выходить за границы этих подобластей и исключает многократные попадания в один и тот же экстремум.

Опишем алгоритм движения к минимуму на $(l + 1)$-м шаге, если на предыдущем шаге был получен вектор искомых переменных x^l:

$$x^{l+1}_v = x^l_v + \Delta x^l_v \, u_{v,\text{мин}}; \; u_{v,\text{мин}} = \sum_{i=1}^{n} u^{(i)}_v \overline{p}^{(i)}_{s,\text{мин}}, \; v = \overline{1, m};$$

$$\overline{p}^{(i)}_{s,\text{мин}} = \frac{p_s(g^{(i)}_{\text{мин}})}{\sum\limits_{j=1}^{n} p_s(g^{(j)}_{\text{мин}})}; \; g^{(i)}_{\text{мин}} = \frac{I^{(i)} - \overline{I}_{\text{мин}}}{\overline{I}_{\text{макс}} - \overline{I}_{\text{мин}}}; \qquad (3)$$

$$\Delta x^{l+1}_v = \gamma_q \cdot \Delta x^l_v \left(\sum_{i=1}^{n} | u^{(i)}_v |^q \overline{p}^{(i)}_{s,\text{мин}} \right)^{1/q}, \; v = \overline{1, m},$$

$$l = 0, \; 1, \; 2, \; \cdots;$$

$$x^0 \equiv x^0_{\text{подобл}}, \; \Delta x^0 \equiv \Delta x^0_{\text{подобл}}; \; [0 < \gamma_q, \; q \in \{1, 2, \cdots\}, \; 0 < s],$$

где $\overline{I}_{\text{мин}} = \text{мин}\{I^{(i)}, i = \overline{1, n}\}$, $\overline{I}_{\text{макс}} = \text{макс}\{I^{(i)}, i = \overline{1, n}\}$; γ_q – коэффициент растяжения области варьирования; s – коэффициент селективности ядра $p_s(\cdot)$; в пере-

менных $u_\nu^{(i)}$, $\overline{p}_{s,\text{мин}}^{(i)}$, $I^{(i)}$ для упрощения записи опущен номер итерации l; всегда $0 \le g_{\text{мин}}^{(i)} \le 1$. Ядра $\overline{p}_{s,\text{мин}}^{(i)}$ нормированы на системе n пробных точек: $\sum_{i=1}^{n} \overline{P}_{s,\text{мин}}^{(i)} = 1$.

Перед совершением $(l+1)$-го рабочего шага для получения пробных точек $x^{(i)}$, $i = \overline{1, n}$, последовательно генерируются точки (равномерно распределенные) в прямоугольной области Δx^l с центром в точке x^l:

$$x_\nu^{(i)} = x_\nu^l + \Delta x_\nu^l \cdot u_\nu^{(i)},\ u_\nu \in [-1; 1],\ \nu = \overline{1, m},\ i = 1, 2, \ldots$$

Из них оставляется n точек, попадающих в допустимую область, задаваемую ограничениями (2): $x^0 = x_{\text{мин}}^j$, $\Delta x^0 = \Pi_{\text{пр}}^j$, $j = \overline{1, n_{\text{гл мин}}}$. В пробных точках вычисляется минимизируемая функция $I^{(i)}$, $i = \overline{1, n}$, и на основе этих экспериментальных данных [$x_\nu^{(i)}$, $\nu = \overline{1, m}$, $I^{(i)}$, $i = \overline{1, n}$] совершается вышеприведенное рабочее движение и пересчитываются размеры области поиска.

Критерием останова процесса поиска обычно служит условие уменьшение размера области варьирования Δx^l до заданной величины: $\max\{|\Delta x_\nu^l|,\ \nu = \overline{1, m}\} \le \varepsilon$.

Значения $p_s(g)$ на интервале изменения $0 \le g \le 1$ являются неотрицательными, убывающими ядрами. Выбор формы и коэффициента s селективности ядра существенно влияет на характеристики поиска глобального минимума. Чаще всего используется степенное ядро $p_s(g) = (1 - g^r)^s$, $r = 1, 2, 3, \ldots$. Коэффициент s для степенных ядер лежит в широком диапазоне целых положительных чисел и подбирается сравнительно просто.

Второй алгоритм (с переходом к штрафной функции).

Введем вместо функции качества (1) и ограничений (2) штрафную функцию:

$$\tilde{I}(x^{(i)}) = \frac{I^{(i)} - \overline{I}_{\text{мин}}}{\overline{I}_{\text{макс}} - \overline{I}_{\text{мин}}} + \alpha \cdot \max\left\{\frac{\varphi_j(x^{(i)}) - \varphi_{j,\text{мин}}}{\varphi_{j,\text{макс}} - \varphi_{j,\text{мин}}},\ j \in J^{(i)} \in \overline{1, m_1}\right\},$$
$$i = \overline{1, n^0},\ \alpha > 0. \qquad (4)$$

Здесь множество $J^{(i)}$ включает только номера тех функций ограничений, для которых ограничения нарушены: при $j \in J^{(i)} \in \overline{1, m_1}$ всегда $0 < \varphi_j(x^{(i)})$; $\overline{I}_{\text{мин}} = \text{мин}\{I^{(i)},\ i = \overline{1, n}\}$, $\overline{I}_{\text{макс}} = \text{макс}\{I^{(i)},\ i = \overline{1, n}\}$; для каждого фиксированного j:

$$\varphi_{j,\text{мин}} = \text{мин}\{\varphi_j(x^{(i)}),\ i : 0 < \varphi_j(x^{(i)})\},$$
$$\varphi_{j,\text{макс}} = \text{макс}\{\varphi_j(x^{(i)}),\ i : 0 < \varphi_j(x^{(i)})\}.$$

Если при некотором $j = l$ множество $\{i : 0 < \varphi_l(x^{(i)})\}$ пустое, то этот номер l не входит во множество $J^{(i)}$ (для каждого фиксированного i). Для пробных точек, в которых ни одно неравенство не нарушено (т. е. множество $J^{(i)}$ пустое) штрафная добавка берется равной нулю.

Если при некотором $j = l$ множество $\{i : 0 < \varphi_l(x^{(i)})\}$ состоит из одного элемента, то соответствующий элемент в функции (4), входящий под оператор $\max\{\}$, полагаем равным 1: $[\varphi_l(x^{(i)}) - \varphi_{l,\text{мин}}] \big/ [\varphi_{l,\text{макс}} - \varphi_{l,\text{мин}}] = 1$.

Далее, разбиение на подобласти происходит как для первого алгоритма, но пробные точки размещаются равномерно по всей подобласти Π^0 (а не в ее части, как в предыдущем алгоритме) и вместо исходной функции $I^{(i)}$ и ограничений $\varphi_j(x^{(i)})$ используется штрафная функция $\tilde{I}^{(i)}$ (4).

После разбиения Π^0 на подобласти, поиск условного глобального минимума в каждой из подобластей (второй этап) ведется так же, как и в первом алгоритме, но пробные точки размещаются равномерно в каждой подобласти $\Pi_{\text{пр}}^j$, $j = \overline{1, n_{\text{гл мин}}}$ без учета ограничений (2) и вместо исходной функции $I^{(i)}$ используется штрафная функция $\tilde{I}^{(i)}$ (4).

Третий алгоритм (с переходом к многомерным ядрам).

Будем учитывать ограничения путем расширения вида нормированные ядра $\overline{p}_s^{(i)}$:

$$\overline{p}_s^{(i)} = \frac{p_{1,s}(g^{(i)}) \prod\limits_{j \in J^{(i)}} p_{2,s}(g_j^{(i)})}{\sum\limits_{\mu=1}^{n} p_{1,s}(g^{(\mu)}) \prod\limits_{j \in J^{(\mu)}} p_{2,s}(g_j^{(\mu)})}, \qquad (5)$$

где $g^{(i)} = \dfrac{I^{(i)} - \overline{I}_{\text{мин}}}{\overline{I}_{\text{макс}} - \overline{I}_{\text{мин}}}$; $g_j^{(i)} = \dfrac{\varphi_j(x^{(i)}) - \varphi_{j,\text{мин}}}{\varphi_{j,\text{макс}} - \varphi_{j,\text{мин}}}$, $j = \overline{1, m_1}$; $p_{1,s}$ – ядро по оптимизируемой функции; $p_{2,s}$ – ядро по ограничениям.

Значения $\overline{I}_{\text{макс}}$, $\overline{I}_{\text{мин}}$, $\varphi_{j,\text{макс}}$, $\varphi_{j,\text{мин}}$ для каждого фиксированного j, а также множество $J^{(i)}$ вычисляются так же, как и для второго алгоритма.

Если при некотором $j = l$ множество $J^{(i)}$ состоит из одного элемента, то аргумент ядра $p_{2,s}$ полагаем равным 0,75. Это значение было получено экспертным путем. Если во всех пробных точках неравенства (2) не нарушены, то ядра становятся одномерными и зависят только от оптимизируемой функции:

$$\overline{p}_s^{(i)} = p_{1,s}(g^{(i)}) \Big/ \sum_{\mu=1}^{n} p_{1,s}(g^{(\mu)}).$$

Таким образом, базовый алгоритм (3) трансформируется следующим образом:

$$x_{\nu}^{l+1} = x_{\nu}^{l} + \Delta x_{\nu}^{l}\, u_{\nu,\text{мин}},\quad u_{\nu,\text{мин}} = \sum_{i=1}^{n} u_{\nu}^{(i)}\, \overline{p}_{s,\text{мин}}^{(i)},\quad \nu = \overline{1,\, m};$$

$$\overline{p}_{s,\text{мин}}^{(i)} = \frac{p_{1,s}(g_{\text{мин}}^{(i)}) \prod\limits_{j\in J^{(i)}} p_{2,s}(g_j^{(i)})}{\sum\limits_{\mu=1}^{n} p_{1,s}(g_{\text{мин}}^{(\mu)}) \prod\limits_{j\in J^{(\mu)}} p_{2,s}(g_j^{(\mu)})}; \qquad (6)$$

$$g_{\text{мин}}^{(i)} = \frac{I^{(i)} - \overline{I}_{\text{мин}}}{\overline{I}_{\text{макс}} - \overline{I}_{\text{мин}}};$$

$$\Delta x_{\nu}^{l+1} = \gamma_q \cdot \Delta x_{\nu}^{l} \left(\sum_{i=1}^{n} |\, u_{\nu}^{(i)}\, |^{q}\, \overline{p}_{s,\text{мин}}^{(i)} \right)^{1/q},\quad \nu = \overline{1,\, m},$$

$$l = 0,\ 1,\ 2, \ldots; \qquad 0 < \gamma_q;\ q \in \{1,\, 2, \cdots\};\ 0 < s.$$

Разбиение на подобласти производится так же, как в первом алгоритме, но пробные точки равномерно размещаются по всей области Π^0, а точки $x_{\text{мин}}^{j}$ отыскиваются на основе алгоритма (6).

После разбиения Π^0 на подобласти поиск минимума в каждой из подобластей ведется также, как и в первом алгоритме, но пробные точки размещаются равномерно в каждой подобласти $\Pi_{\text{пр}}^{j}$, $j = \overline{1, n_{\text{гл мин}}}$ без учета ограничений (2), и поиск условного глобального минимума в каждой подобласти производится на основе алгоритма (6).

Численный пример. Исследуем поведение алгоритмов при наличии аддитивной помехи и допустимой области в виде «узкого жгута». Как и в работе [2, с. 93], в качестве помехи будем использовать аддитивную равномерно распределенную помеху $R[-\theta;\, \theta] \equiv \theta \cdot R[-1;\, 1]$, принадлежащую интервалу $[-\theta;\, \theta]$ (здесь θ определяет уровень помехи). Выбор равномерного закона распределения помехи соответствует наихудшей ситуации, которая, в принципе, может встретиться при реальном искажении оптимизируемой функции. При этом помеха имеет максимальную приведенную энтропию среди всех непрерывных случайных величин (с любым законом распределения), изменяющихся внутри конечного интервала $[-\theta;\, \theta]$. Рассмотрим четырехэкстремальную потенциальную функцию, построенную за счет применения операции min к четырем элементарным экспоненциальным потенциалам:

$$I(x) = \min\Big\{ -3\exp\{-3[|\, x_1 - 3\,|^{1,5} + |\, x_2\,|^{1,5}]\},$$
$$-5\exp\{-2,5[|\, x_1 + 3\,|^{2,5} + |\, x_2\,|^{2,5}]\},$$
$$-7\exp\{-[|\, x_1\,|^{1,2} + |\, x_2 - 3\,|^{1,2}]\},$$
$$-10\exp\{-2[|\, x_1\,|^{2} + |\, x_2 + 3\,|^{2}]\}\Big\}. \qquad (7)$$

Допустимая область имеет вид кольца шириной Δ:

$$x_1^2 + x_2^2 \le (3 + \frac{\Delta}{2})^2 \ \text{ и } \ x_1^2 + x_2^2 \ge (3 - \frac{\Delta}{2})^2. \qquad (8)$$

Пространственный вид функции (7) представлен на рис. 1. Она имеет минимумы в точках (–3; 0), (3; 0), (0; 3), (0; –3) и принимает в них, соответственно, значения (–3), (–5), (–7), (–10). Линии равных уровней функции (7) и допустимая область (8) при $\Delta = 0,4$ приведены на рис. 2. Соответственно, задавая $\theta = 5$, мы получим 100 % помеху.

Все три вышеописанные алгоритма хорошо справляются с поиском условного глобального экстремума при отсутствии помехи и достаточно узкой допустимой области $\Delta = 0,01$.

Изменение по итерациям минимизируемой функции при поиске двух главных минимумов функции (7) при ограничениях (8) и $\Delta = 0,01$ для всех трех алгоритмов отражено на рис. 3.

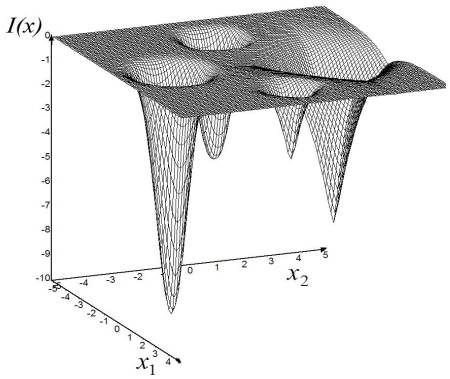


Рис. 1. Потенциальная функция (7)

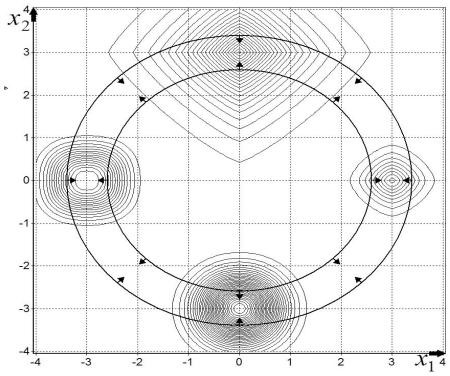


Рис. 2. Равные уровни функции (7) и допустимая область (8) при $\Delta = 0,4$

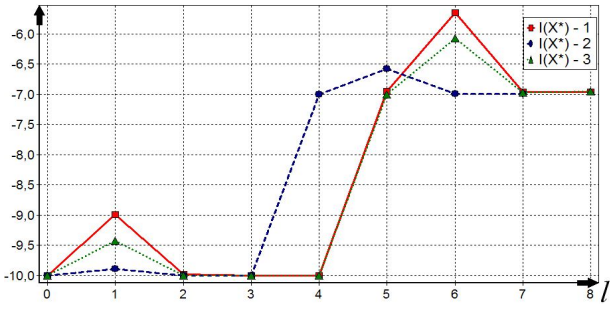


Рис. 3. Процесс поиска двух главных минимумов всеми алгоритмами

Все алгоритмы ведут себя примерно одинаково и отыскивают два главных минимума за 7–8 итераций

(рис. 3). Начальная точка поиска находится в малой окрестности истинного решения и алгоритмы ее лишь уточняют, это обусловлено тем, что еще на первом этапе (выделение подобластей) происходит достаточно точная оценка положения главных минимумов.

Проведем исследование поведения алгоритмов при наличии узкой допустимой области ($\Delta = 0{,}01$) и различных уровнях помехи. Для этого будем менять уровень помехи θ от 0 до 5 (от 0 до 100 % помехи) с шагом 1 и усреднять полученные значения на каждом шаге по 101 запуску. В качестве меры качества работы алгоритмов будем использовать количество попыток размещения пробных точек N и оценки вероятности отыскания «истинных» главных экстремумов $\hat{P}(|x^* - x| \le \varepsilon)$.

Первый алгоритм. Считаем, что $n^0 = 500$, область поиска с центром $x^0 = (0; 0)$ и размером $\Delta x^0 = (4; 4)$, $q = 2$, $\gamma_q = 1{,}2$, ядро по функции качества – параболическое с коэффициентом селективности $s = 300$, размер выборки при поиске экстремума в подобласти $n = 250$. Графики, показывающие оценки вероятности отыскания истинного экстремума и количество пробных точек при разных уровнях помехи, представлены на рис. 4.

Первый алгоритм отыскивает оба экстремума практически с вероятностью 1, но при этом алгоритм размещает очень большое число пробных точек: $1{,}3 \cdot 10^6$ (рис. 4). Такое поведение обусловлено структурой ограничений – узкая область в виде жгута не дает алгоритму размещать точки нигде, кроме как в этой области. Таким образом, алгоритм практически гарантированно отыскивает заданные экстремумы, даже при 100 % помехе (если ограничений станет еще больше, то алгоритм вообще не сможет разместить пробные точки за разумное время).

Второй алгоритм. Зададим такие же параметры, как и для первого алгоритма. Штрафной коэффициент $\alpha = 1{,}1$. Графики, описывающие поведение алгоритма при различных уровнях помехи, представлены на рис. 5.

Второй алгоритм размещает гораздо меньше пробных точек (в 40 раз), чем первый алгоритм (рис. 5). Вероятности отыскания главных минимумов хуже, чем у первого алгоритма, но вполне неплохие при таких уровнях помехи. Этот алгоритм эффективен и прост в настройке.

Третий алгоритм. Зададим такие же параметры алгоритма как и для первого алгоритма. Ядро по функции качества – степенное гиперболическое [2, с. 67] степень ядра $r = 2$, коэффициент селективности $s = 600$, ядро по ограничениям – параболическое с коэффициентом селективности $s = 100$. Графики, описывающие поведение алгоритма представлены на рис. 6.

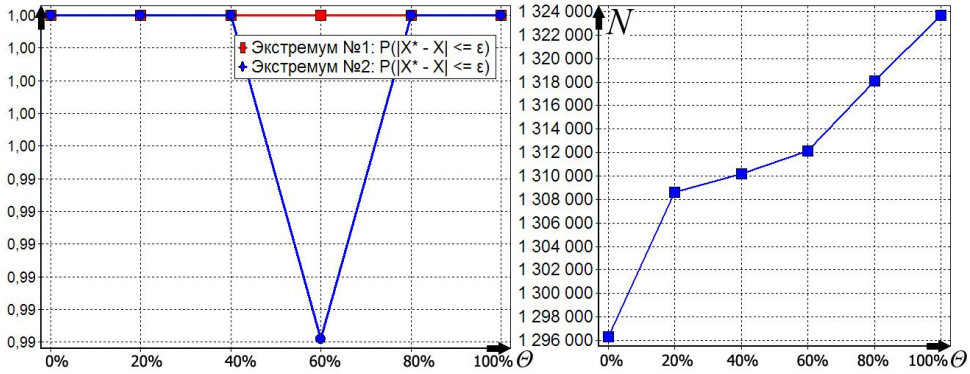


Рис. 4. Зависимость показателей качества работы первого алгоритма от θ

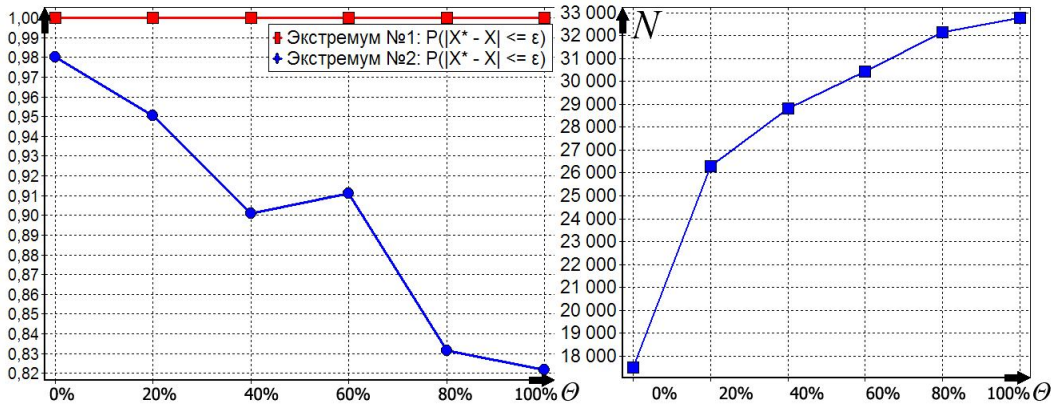


Рис. 5. Зависимость показателей качества работы второго алгоритма от θ

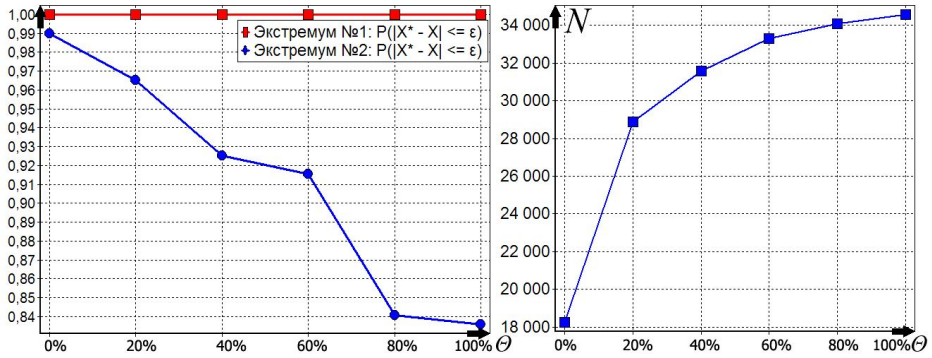


Рис. 6. Зависимость показателей качества работы третьего алгоритма от θ

Третий алгоритм работает чуть лучше, чем второй алгоритм (рис. 6), но он гораздо более сложен в настройке, так как необходимо настраивать два ядра вместо одного (пришлось настраивать не только степень и селективность ядра, но и его вид).

Качество работы второго и третьего алгоритма можно еще повысить, задав $n^0 = 1\ 000$ и $n = 500$, но это приведет к увеличению количества пробных точек.

Программная реализация алгоритмов. Представленные алгоритмы входят в состав пакета глобальной оптимизации Global Optimizer v 2,0, в котором также представлены основные алгоритмы глобальной оптимизации, базирующиеся на методе усреднения координат. Пакет реализован на языке Delphi с использованием среды визуального программирования Delphi 2010 фирмы Embarcadero Technologies и функционирует в операционной системе из семейства Windows 2 000/XP/Vista/7.

При проектировании программных алгоритмов поиска главных минимумов был использован объектно-ориентированный подход. Алгоритмы реализованы как надстройка над алгоритмами поиска глобального экстремума [2]. Этот подход позволяет распараллелить поиск главных экстремумов и тем самым значительно повысить скорость работы алгоритма. Этот вариант требует специальных исследований.

Алгоритмы поиска главных экстремумов привлекают своей простотой, сравнительно небольшими вычислительными затратами, высокими точностными показателями. Проведенные исследования показали, что все три алгоритма примерно одинаково эффективны при наличии сильной помехи. Наилучшие показатели оказались у второго и третьего алгоритмов – они отыскивают главные минимумы с достаточно высокой вероятностью и при этом размещают разумное число пробных точек.

Библиографические ссылки

1. Рубан А. И. Метод непараметрической поисковой оптимизации // Изв. вузов. Физика. 1995. Т. 38, № 9. С. 65–73.

2. Рубан А. И. Глобальная оптимизация методом усреднения координат : монография. Красноярск : ИПЦ КГТУ, 2004.

3. Кузнецов А. В., Рубан А. И. Поиск главных минимумов многоэкстремальных функций при активном учете ограничений неравенств // Техника и технологии : журн. СФУ. 2010. Т. 3, № 3. С. 335–346.

A. V. Kuznetsov, A. I. Rouban

ALGORITHMS OF COORDINATES AVERAGING METHOD FOR SEARCHING OF THE PRINCIPAL MINIMUM OF MULTIEXTREME FUNCTIONS

In the article we propose algorithms for search of the set quantity of the principal minimum of multiextremal functions of continuous variables. The basis of the algorithms is sequential search of global extremum based on coordinates averaging method, with subsequent isolation of the found subdomains with the help of additional inequalities constraints. On numerical examples we demonstrate the overall effectiveness of the algorithms.

Keywords: global optimization, principal minimum.

© Кузнецов А. В., Рубан А. И., 2010

УДК 519.2

Г. С. Лбов, Г. Л. Полякова

МЕТОД ПРОГНОЗИРОВАНИЯ В КЛАССЕ ЛОГИЧЕСКИХ РЕШАЮЩИХ ФУНКЦИЙ*

Рассматривается метод прогнозирования в классе логических решающих функций. Выбор наилучшего разбиения в пространстве переменных осуществляется на основе перебора, использующего идеологию «метода ветвей и границ». Эффективность предложенного метода иллюстрируется результатами решения прикладных задач.

Ключевые слова: анализ эмпирической информации, логические закономерности, прогнозирование.

В статье задача прогнозирования целевой переменной [1–3] решается в классе логических решающих функций. Особенностями указанной задачи являются небольшое число временных отсчетов, необходимость в «наглядности» получаемых результатов. Рассматривается метод прогнозирования в классе логических решающих функций. При этом выбор наилучшего разбиения в многомерном пространстве переменных осуществляется на основе перебора, использующего идеологию «метода ветвей и границ». Предлагаемый метод осуществляет полный перебор различных разбиений многомерного пространства переменных на заданное число прямоугольных областей. Значительное сокращение перебора всевозможных вариантов осуществляется путем выбора только тех вариантов, которые обеспечивают заданную статистическую надежность прогнозирования целевой переменной для любой прямоугольной области.

Отметим, что решение подобного рода задач связано с некоторыми особенностями:

1) отсутствие априорной информации о распределениях в пространстве переменных (характеристик);

2) в трудноформализуемых областях исследователь вынужден включать большое число потенциально полезных переменных (характеристик) из-за сложности изучаемого явления;

3) малое число наблюдений (объектов), как правило, сравнимое с числом переменных;

4) исследователя при изучении сложных объектов нередко интересует не только решение, дающее хорошее качество прогноза, но и сама форма представления такого решения для получения информации о внутренних причинно-следственных связях между характеристиками изучаемых объектов.

При решении статистических задач выявления закономерностей с указанными выше особенностями возникает ряд проблем.

Во-первых, задачи приходится решать в условиях отсутствия априорной информации о виде функций распределения. Любое предположение (например, о нормальности распределения, линейной регрессионной зависимости, независимости переменных, марковости процесса) ставит вопрос о соответствии выбранного предположения истинным ограничениям. Как вводить предположения и какие? В этом состоит первая проблема.

Во-вторых, в условиях малого числа наблюдений и высокой размерности пространства переменных возникает проблема статистической устойчивости получаемых решений. Из теоретических исследований вытекает, что более сложные функциональные зависимости используются для построения решений, больше переменных и меньше число наблюдений (объем выборки), тем больше вероятность получения решения, сильно отличающегося от оптимального. Суть проблемы устойчивости статистических решений заключается в следующем. С одной стороны, сильное ограничение на класс решений ставит вопрос об адекватности наших предположений истинному распределению: чем больше такое несоответствие, тем хуже решение. С другой стороны, чем более сложный класс функций используется при малом объеме выборки, тем выше вероятность получить недостоверное решение. Так, например, может оказаться, что линейная функция, заданная на всем множестве переменных, будет менее информативной по сравнению с линейной функцией, заданной на подмножестве этих переменных. Таким образом, при построении решения необходимо стремиться к максимальной сложности используемого класса решений (для ослабления ограничений на истинное распределение), но при этом сложность класса решений не должна превышать некоторого порога, задаваемого объемом выборки. При малом объеме выборки класс решений должен иметь малую меру сложности. При увеличении объема выборки этот класс должен позволять постепенно увеличивать свою сложность, вплоть до получения оптимального решения при произвольном распределении. Класс решений, обладающий таким свойством, будем называть универсальным классом. Вопрос о соотношении сложности используемого класса решений и объема выборки – наиболее важный и трудный в общих теоретических исследованиях, связанных с построением решений на основе ограниченной эмпирической информации.

По указанным выше причинам, а именно при отсутствии информации о виде распределения, линейной зависимости, наличия малого количества наблюдений (объектов) относительно числа переменных (характеристик), традиционные статистические методы оказываются малоподходящими при решении указанных задач.

* Работа выполнена при финансовой поддержке РФФИ (проект № 10-01-00113-а), проект № 09-07-12087-офи_м. Интеграционного гранта СО РАН (№ 83).

Как показывают теоретические и экспериментальные исследования [1–3], построение логико-вероятностных моделей изучаемых явлений на основе такой информации оказывается достаточно перспективным направлением. Модели представляют собой список логических закономерностей, обладающих достаточно высокой прогнозирующей способностью. Кроме того, в отличие от классических методов статистического анализа результаты в рамках указанной модели представляются на языке, близком к естественному языку логических суждений, что облегчает интерпретацию результатов.

Постулируется, что изучаемое явление характеризуется лишь небольшим числом закономерностей (несколько десятков). Для решения поставленной задачи был использован как алгоритм ТЕМР [3], так и его модификация. Оба алгоритма обнаруживают все закономерности с заданной прогнозирующей способностью. Ниже приводится описание алгоритма ТЕМР и его модификации.

Алгоритм обнаружения логических закономерностей. Как указывалось выше, при решении каждой из 18 задач выявления закономерностей между природными факторами и заболеваемостью клещевым энцефалитом имеем таблицу данных $T = \{x^i, y^i\}$, где $x^i = (x_1^i, ..., x_{12}^i)$, $i = 1, ..., 18$. Так как объем выборки ($N = 20$) является весьма малым, а размерность пространства переменных относительно велика (от 12 до 48), статистически надежные закономерности могут быть получены лишь при огрублении статистического материала.

В данном случае диапазон значений каждой переменной разбивается на ряд интервалов. Интервалы переменной *Y* назовем образами. Из исходной таблицы *T* создаем таблицу $v = \{z^i, u^i\}$, где *z* и *u* – соответствующие интервалы переменных *x* и *y*.

Для каждого сочетания переменных $z_1, ..., z_{12}$ перебираются все наборы интервалов рассматриваемого подмножества переменных (логические высказывания, в данном случае конъюнкции). Пример конъюнкции интервалов: $S = (z_5 = 2) \wedge (z_7 = 1) \wedge (z_{11} = 3)$. Естественно, ее можно записать через границы интервалов указанных переменных, что и делается при описании результатов изучения влияния природных факторов на заболеваемость клещевым энцефалитом.

Логические закономерности определяются отдельно для каждого образа-интервала *Y*. Для этого образ с номером s, $s = 1, ..., k$, назовем первым образом, а объединение всех остальных – вторым образом. Обозначим конъюнкцию через $S(a, E)$, где a – имя объекта (номер переменной); E – некоторая область многомерного пространства переменных. Если набор значений переменных для рассматриваемого объекта принадлежит области E, то будем говорить, что $S(a, E)$ для этого объекта принимает значение «истина», иначе – «ложно». Для любой конъюнкции $S(a, E)$ можно определить по таблице данных v число объектов первого образа $N(1, S)$ и число объектов второго образа $N(2, S)$, на которых указанная конъюнкция истинна.

Конъюнкцию $S(a, E)$ назовем логической закономерностью S^*, характеризующего первый образ, если выполняются неравенства

$$\frac{N(1, S)}{N(1)} \geq \delta, \quad \frac{N(2, S)}{N(2)} \leq \beta,$$

где δ и β – некоторые параметры, $0 \leq \beta < \delta \leq 1$. Чем больше δ и меньше β, тем сильнее логическая закономерность. Множество всех закономерностей обозначим через S^*. Конъюнкцию $S(a, E)$ называем потенциальной логической закономерностью для первого образа (обозначим ее через S'), если выполняются неравенства

$$N(1, S)N(1) \geq \delta, \quad N(2, S)N(2) > \beta.$$

Множество потенциальных закономерностей обозначим через S'. Очевидно, что из $S' \in S'$ можно, вообще говоря, получить закономерность S^* последовательным присоединением предикатов, т. е. $S' \wedge J(a, E_j) \wedge ....$ Если для некоторой конъюнкции $S(a, E)$ выполняется неравенство $N(1, S) / N(1) < \delta$, то конъюнкция S по определению не является закономерностью и присоединение к ней какого-либо предиката не даст закономерности (множество таких конъюнкций обозначим через *S*). Таким образом, любая конъюнкция $S(a, E)$ может быть трех типов: S^*, S' и S.

Назовем конъюнкцию $S(a, E) = J(a, E_{j_1}) \wedge ... \wedge J(a, E_{j_m})$ конъюнкцией длины m, где E_j область значений переменной Z_j.

Алгоритм обнаружения логических закономерностей состоит в последовательном выполнении следующих шагов.

Шаг 1. Рассматриваются всевозможные конъюнкции длины 1, т. е. конъюнкции вида $S(a, E) = J(a, E_j)$, $E_j \in W_j$, $j = 1, ..., n$. Если $S(a, E) \in S^*$, то она включается в список закономерностей и соответствующее подмножество E_j исключается из дальнейшего перебора; если $S(a, E) \in S'$, то соответствующее подмножество E_j оставляется для дальнейшего перебора; если $S(a, E) \in S$, то соответствующее подмножество E_j исключается из дальнейшего перебора. Обозначим через W_j^1 множество подмножеств E_j, оставленных для дальнейшего перебора после выполнения шага 1 алгоритма.

Шаг 2. Рассматриваются всевозможные конъюнкции длины 2, т. е. конъюнкции вида $S(a, E) = J(a, E_j) \wedge J(a, E_l)$, $j \neq l$; $E_j \in W_j^1$, $E_l \in W_l^1$. Если $S \in S^*$ или $S \in S'$, то соответствующие подмно-

жества E_j и E_l исключаются из дальнейшего перебора. Если $S \in S^*$, то соответствующая конъюнкция включается в список закономерностей.

Шаг 3. Рассматриваются всевозможные конъюнкции длины 3, т. е. конъюнкции вида $S(a,E) = J(a,E_j) \wedge J(a,E_l) \wedge J(a,E_m), \quad j \neq l \neq m, \; j \neq m; \; E_j \in W_j^2, \; E_l \in W_l^2, \; E_m \in W_m^2$, где W_j^2 содержит подмножества E_j, оставленные для перебора после шага 2 алгоритма. Далее аналогично рассматриваются конъюнкции длины 4, 5 и т. д.

Предполагается, что реальные таблицы данных таковы, что число элементов множества W_j^m резко уменьшается с увеличением числа шагов m, а также, что уже при небольшом числе шагов происходит останов программы. Из схемы алгоритма следует, что с его помощью обнаруживаются все логические закономерности, характеризующие s-й образ. Последовательно применяя данный алгоритм для каждого из k образов, получаем k списков логических закономерностей: $\{S_1^1, ..., S_{d_1}^1\}, ..., \{S_1^k, ..., S_{d_k}^k\}$. Сделаем сквозную нумерацию полученных закономерностей: $S_1, ..., S_d$, где $d = \sum_{s=1}^{k} d_s$. Заметим, что при малых значениях β ($\beta \simeq 0$) и больших значениях δ ($\delta \simeq 1$) логических закономерностей может не быть; при фиксированном малом β (например, $\beta = 0{,}05$) при уменьшении δ ($\delta \to \beta$) число закономерностей может резко возрастать, а их качество падать. Поэтому при фиксированном значении β необходимо выбрать такую величину δ, чтобы число закономерностей было небольшим (например, десять закономерностей на каждый образ). Этой цели можно достигнуть последовательно уменьшая δ с некоторым шагом $\Delta\delta$. По критерию $F = \delta - \beta$ можно упорядочить полученные закономерности по их вероятностной прогнозирующей способности.

Особенность модифицированного алгоритма ТЕМР состояла в переборе всевозможных конъюнкций сразу заданной длины, указанной в параметрах алгоритма, и нового критерия качества дерева решений.

Метод решения задачи прогнозирования. Рассмотрим задачу прогнозирования [3–6] для каждой переменной Y_l, l = 1, …, m, отдельно, т. е. определим решающую функцию F набором функций $\{\overline{f}_1, ..., \overline{f}_l, ..., \overline{f}_m\}$, где $\overline{f}_l$ – отображение из B в D_{Y_l}. Построение функции $\overline{f}_l$ осуществляется в два этапа [4–6].

Этап I. На основе временного ряда q для фиксированного номера d, $d \in \{1, …, R\}$, и фиксированного номера l организуется обучающая выборка в виде таблицы данных $v_d = \{x^{\mu-d}, y_l^{\mu}\}$, $\mu = R + 1, …, N$. По таблице данных v_d строим логическую решающую функцию $<\alpha_d, r(\alpha_d)>$ с разбиением $\alpha_d = \{E_d^1, E_d^t, ..., E_d^{M_d}\}$ множества $D = \prod_{j=1}^{n} D_{X_j}$ на M_d подмножеств (с помощью алгоритма обнаружения логических закономерностей). Введем новую переменную Z_d с множеством значений $D_{Z_d} = \{1, ..., t, ..., M_d\}$ следующим образом: если предыстория $(x_1^{\mu-d}, ..., x_n^{\mu-d}) \in E_d^t$, то $z_d = t$. Указанные выше вычисления проводятся для всех номеров предысторий (d = 1, …, R). Результатом первого этапа является набор разбиений (α, …, α_d, …, α_R) и соответствующий ему набор переменных (Z_1, …, Z_d, …, Z_R). Это позволяет для любого момента времени t_μ, $\mu = R + 1, …, N$, n-мерную предысторию длины R, представленную набором ($x^{\mu-R}$, …, $x^{\mu-d}$, …, $x^{\mu-1}$), $x^{\mu-d} \in D$, задать в виде одномерной последовательности $z^{\mu} = (z_R^{\mu}, ..., z_d^{\mu}, ..., z_1^{\mu})$, $z_d^{\mu} \in D_{Z_d}$.

Этап II. Временному n-мерному ряду $q = \{x_j^v\}$, j = 1, …, n, v = 1, …, N ставится в соответствие таблица данных $v = \{z^{\mu}, y_l^{\mu}\}$, $\mu = R + 1, …, N$. По таблице v строится логическая решающая функция $\overline{f}_l$, которая позволяет прогнозировать значение переменной Y_l по предыстории длины R, l = 1, …, n. Заметим, что задача прогноза для любого набора $\{Y_{i_1}, ..., Y_{i_m}\} \subseteq \{X_1, ..., X_n\}$, m = 1, …, n решается аналогично.

Из-за сложности статистической задачи (многомерность, нестационарность, малое число наблюдений) возникла необходимость в «огрублении» статистической информации (использовалось небольшое число интервалов значений переменных, включая и показатель заболеваемости).

Совместный анализ климатических и астрофизических факторов в природных очагах клещевого энцефалита проводился на основе следующей статистической информации: температура воздуха (ТВ), относительная влажность воздуха (ОВВ) в приземном слое; количество осадков; данные солнечной радиации: прямой радиации на горизонтальную поверхность, прямой радиации на перпендикулярную поверхность, отраженной радиации; заболеваемость людей клещевым энцефалитом. Исследование влияния взаимозависимых природных факторов на заболеваемость людей клещевым энцефалитом заключалось в совместной статистической обработке 12 временных рядов. Каждый временной ряд представлял собой набор значений среднемесячных показателей перечисленных 6 природных факторов и годовых значений заболеваемости людей в течение 20 лет (с 1991 по 2010 гг.). Таким образом, каждому природному фактору соответствовала таблица из 12 столбцов (переменные X_1, …, X_{12}) и 20 строк (наблюдения или объекты). Для каждой строки указывается количество заболевших за год (переменная Y). Требовалось найти статистические закономерности (взаимосвязи) между всеми X_1, …, X_{12} и Y. В силу того, что природных фак-

торов – 6, а рассматриваемых регионов – 3, возникает 18 задач выявления закономерностей, решение которых и представлено в данной работе.

При решении задачи алгоритмом ТЕМП [4; 5] значения показателя заболеваемости были разделены на 2 диапазона: 1 диапазон соответствовал низкому уровню заболеваемости от 2 до 17 человек на 100 тысяч населения, 2 диапазон – высокому уровню заболеваемости от 17 до 62 заболевших на 100 тысяч населения. Это дало возможность обнаружить логические закономерности, как правило, включая конъюнкции значений не более трех переменных для прогноза уровня заболеваемости. При решении данной задачи модифицированным алгоритмом ТЕМП [6] количество образов (диапазонов значений целевой переменной) задавалось равным k, где $k > 2$.

Приведем некоторые из полученных закономерностей.

Высокий уровень заболеваемости КЭ возникает при совместном (одновременном) выполнении условий: «ТВ в январе – от –20 до –11°С» и «ТВ в августе – от 15 до 18,7 °С» и «ТВ в октябре – от 2,1 до 5,6 °С».

Высокий уровень заболеваемости КЭ определяется совместным выполнением условий: «ОВВ в феврале – от 72,5 до 83,5 %» и «ОВВ в апреле – от 58 до 73 %» и «ОВВ в декабре – от 79,8 до 85 %».

Полученная нами оценка заболеваемости КЭ в НСО в 2009 г. имела значение 7 человек на 100 тысяч населения, реальная заболеваемость составила 6,5 человек на 100 тысяч населения.

Предложен метод прогнозирования целевой переменной на основе обнаружения логических закономерностей на обучающей выборке, который включает перебор различных разбиений многомерного пространства переменных на заданное число прямоугольных областей. Значительное сокращение перебора всевозможных вариантов осуществляется путем выбора только тех вариантов, которые обеспечивают заданную статистическую надежность прогнозирования целевой переменной для любой прямоугольной области.

Эффективность предложенного метода иллюстрируется результатами решения прикладной задачи, состоящей в исследовании влияния природных факторов на заболеваемость клещевым энцефалитом населения Новосибирской, Иркутской областей и Республики Горный Алтай.

Библиографические ссылки

1. Лбов Г. С. Метод анализа многомерных разнотипных временных рядов в классе логических решающих функций // Доклад РАН. 1994. Т. 339. № 6.

2. Лбов Г. С., Методы обработки разнотипных экспериментальных данных. Новосибирск : Наука, 1981.

3. Лбов Г. С., Старцева Н. Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск : Изд-во Ин-та математики, 1999.

4. Исследование влияния природных факторов на заболеваемость клещевым энцефалитом / Г. С. Лбов, Г. Л. Полякова, В. Н. Бахвалова [и др.] // Вестник Новосиб. гос. ун-та. 2010. № 3. С. 31–37. (Серия: Биология. Клиника Медицина).

5. Лбов Г. С., Полякова Г. Л., Пестунов И. А. Метод прогнозирования на основе анализа коротких временных рядов // AIS-IT'10 : тр. Конгресса по интеллектуальным системам и информац. технол. М. : Физматлит, 2010. Т. 1. С. 264–271.

6. Лбов Г. С., Полякова Г. Л. Решение задачи прогнозирования в классе логических решающих функций // Математическая биология и биоинформатика : докл. III Междунар. науч. конф. (10–15 окт. 2010, Пущино). М. С. 226–227.

G. S. Lbov, G. L. Polyakova

FORECASTING METHOD IN CLASS CLASS OF LOGICAL DECISION FUNCTIONS

In this paper we suggest a method using the class of logical decision functions. A search of optimal division of variable space is based on sort out similar to branch and bound method. Results of statistical analysis are represented in the form of logical rules reflected cause-effect relations of object under investigation. Effectiveness of the suggested method is shown by solving applied problems in the sphere of ecology.

Keywords: analyses of empirical information, logical rules, forecasting

© Лбов Г. С., Полякова Г. Л., 2010

УДК 519.234

А. В. Маер, В. А. Симахин

НЕПАРАМЕТРИЧЕСКИЕ ДАТЧИКИ ДЛЯ СЛУЧАЙНЫХ СТАЦИОНАРНЫХ ПРОЦЕССОВ

Рассмотрены алгоритмы непараметрических датчиков случайных стационарных процессов. На основе непараметрических алгоритмов прогноза находится интервал зависимости случайного процесса.

Ключевые слова: датчик, непараметрический, прогноз, процесс, моделирование.

Методы статистического имитационного моделирования нашли широкое распространение как в научных исследованиях, так и в технических приложениях. Именно с их помощью удается получить выводы при исследовании сложных систем, зависящих от большого числа случайных факторов (переменных, показателей, элементов). Особая роль при таких исследованиях отводится датчикам случайных величин, которые должны адекватно отражать суть физических явлений, характерных для случайных процессов, участвующих в исследовании. Определение функции распределения случайного процесса фактически определяет задание математической модели физического явления и тем самым определяет адекватность получаемых решений. Обычно, исходя из физических или интуитивных соображений, выбирают некоторое стандартное распределение $G(x(t), \vec{\theta})$, а затем по обучающей выборке $x_1, ..., x_N$ оценивают неизвестный параметр $\vec{\theta}$ и проверяют по некоторым критериям адекватность параметрической модели, например, линейной авторегрессии (АР), скользящего среднего (МА), АРМА, ARCH, GARCH и т. д. – параметрический подход. Понятно, что при этом задается вид параметрической модели и распределение случайных помех. Построить датчик случайного процесса в данном случае не представляет трудностей. Но при параметрическом подходе имеется ряд трудноформализуемых моментов (например, выбор вида параметрической модели процесса и распределения $G(x(t), \vec{\theta})$, критерия адекватности, интервала зависимости и т. д.). Но необходимо понимать, что любая ошибка в задании априорной информации о виде модели и распределения $G(x(t), \vec{\theta})$ приводит к неизвестному смещению при окончательных выводах. Именно этот момент и вызывает значительный интерес к непараметрическому подходу, когда вид распределения $F(x(t))$ случайного процесса $X(t)$ априори считается неизвестным. Действительно, выборка $x_1, ..., x_N$ в виде эмпирических функций распределения $F_N(x(t))$ несет информацию о неизвестном распределении $F(x(t))$ случайного процесса $X(t)$. Следовательно, возможно построение датчиков на основе непараметрических оценок $F_N(x(t))$. Такие датчики назовем непараметрическими. Интерес к непараметрическим датчикам случайных величин в математической статистике особенно вырос с появлением такого статистического метода, как бутстреп [1], который получил широкое распространение [2]. Как известно, в математической статистике большинство результатов исследований носит асимптотический характер. Желание узнать свойства статистических процедур при конечных объемах выборки N с помощью метода статистических испытаний и породило бутстреп-метод [1]. В последние десятилетия исследования в области создания бутстреп-процедур сосредоточены на создании непараметрических датчиков случайных процессов [3; 4]. В рамках этого направления был предложен ряд непараметрических датчиков случайных процессов: блочный бутстреп, решетчатый бутстреп, локальный бутстреп. Но то, что хорошо для исследования статистических процедур в задачах математической статистики, в большинстве случаев не совсем приемлемо в задачах статистического моделирования, где требуются большие массивы случайных неповторяющихся чисел. При блочном бутстрепе нарушается структура статистической зависимости, решетчатый бутстреп – это фактически параметрический датчик, локальный бутстреп основан на непараметрической регрессии и создает датчик, имитирующий исходную выборку [3].

В данной работе рассмотрены основные принципы построения непараметрических датчиков случайных процессов на основе непараметрических оценок условных функций распределения.

Постановка задачи и алгоритм датчика. Пусть $(x)t$ – стационарный в узком смысле случайный процесс, удовлетворяющий условию слабой зависимости [5] с функцией распределения $F(x(t))$. В реальных задачах бесконечномерная функция распределения $F(x(t))$ представляется в виде k-мерной $F_k(\vec{t}_k)$ с конечным интервалом зависимости.

Пусть $x_1, ..., x_N$ – выборка из $X(t)$ с неизвестной функцией распределения $F(x(t))$, $\vec{X} = (X_1, ..., X_k)^T$ – k-мерная случайная величина с функцией распределения $F_k(\vec{t}_k)$. Обозначим через $F_j(t_j / \vec{t}_{j-1})$, $j = \overline{1,k}$ условные функции распределения случайной величины $\vec{X}$, $F_1(t_1) = F_j(t_j / \vec{t}_{j-1})$, $j = 1$, $\vec{t}_{j-1} = (t_1, ..., t_{j-1})^T$.

Для построения непараметрических датчиков воспользуемся стандартным методом генерирования случайных векторов с помощью преобразования [6]

$$\begin{cases} U_1 = F_1(X_1); \\ U_2 = F_2(X_2 / X_1); \\ . \ . \ . \ . \ . \ . \ . \ . \ . \\ U_k = F_k(X_k / \vec{X}_{k-1}). \end{cases} \quad (1)$$

Как известно [6], случайные величины $U_1, ..., U_k$ будут независимы и равномерно распределены на [0; 1]. С помощью обратного преобразования системы (1) и получают датчики случайных векторов. Метод получения непараметрических датчиков заключается в следующем. Вместо неизвестных $F_1(X_1), ..., F_k(X_k / \vec{X}_{k-1})$ в систему (1) подставляются их непараметрические оценки $F_{1N}(X_1), ..., F_{kN}(X_k / \vec{X}_{k-1})$ и затем решается система типа (1). Для реализации данного метода необходимо выполнение следующих условий:

– оценки $F_{1N}(X_1), ..., F_{kN}(X_k / \vec{X}_{k-1})$ должны быть из класса непрерывных непараметрических оценок условных функций распределения;

– преобразование типа (1), где вместо неизвестных $F_1(X_1), ..., F_k(X_k / \vec{X}_{k-1})$ стоят их оценки $F_{1N}(X_1), ..., F_{kN}(X_k / \vec{X}_{k-1})$, должно также асимптотически приводить к независимым и равномерным в [0; 1] случайным величинам $U_1, ..., U_k$;

– так как решение системы типа (1) сводится к нахождению непараметрических оценок условных квантилей распределения $F_{jN}(X_j / \vec{X}_{j-1})$, то необходимо, чтобы эти оценки сходились по вероятности к квантилям распределения $F_j(X_j / \vec{X}_{j-1})$.

Для этого воспользуемся непараметрическими оценками условных функций распределения [5]:

$$F_{jN}(t_j / \vec{t}_{j-1}) = \frac{A_{jN}}{B_{jN}}; \quad (2)$$

$$A_{jN} = \frac{1}{N h_N^{j-1}} \sum_{i=1}^{N-j} G\left(\frac{t_j - x_j}{h_N}\right) \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{t_\lambda - x_{i-\lambda}}{h_N}\right);$$

$$B_{jN} = \frac{1}{N h_N^{j-1}} \sum_{i=1}^{N-j} \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{t_\lambda - x_{i-\lambda}}{h_N}\right),$$

где K_λ – ядерные функции [6]; $G(u)$ – симметричная функция распределения; $h_N \to 0$, $N \cdot h_N^k \to \infty$, $N \to \infty$.

При некоторых условиях регулярности на $F_k(\vec{x}_k)$, $K_\lambda(u)$, h_N, $G(u)$ можно доказать [5]:

1) $F_{jN}(t_j / \vec{t}_{j-1}) \xrightarrow{\Pr} F_j(t_j / \vec{t}_{j-1})$;

2) вектор $\sqrt{N}(T_{1N}(t_1), ..., T_{kN}(t_k / \vec{t}_{k-1}))^T$ имеет асимптотически нормальное распределение с нулевым вектором средних и диагональной корреляционной матрицей

$$h_N^j \sigma_j^2 = Q^j \cdot f^{-1}(\vec{t}_{j-1}) F(t_j / \vec{t}_{j-1}) \left[1 - F(t_j / \vec{t}_{j-1})\right],$$

где $T_{jN}(t_j / \vec{t}_{j-1}) = \sqrt{h_N^j}\,(F_{jN}(t_j / \vec{t}_{j-1}) - F_j(t_j / \vec{t}_{j-1}))$, $\int K^2(u)du = Q < \infty$, (т. е. асимптотически $F_{1N}(t_1), ..., F_{kN}(t_k / \vec{t}_{k-1})$ – независимы);

3) пусть ξ_j квантиль уровня p условной функции распределения $F_j(t_j / \vec{t}_{j-1})$, тогда $\xi_{jN} \xrightarrow{\Pr} \xi_j$, где ξ_{jN} – оценка условного квантиля, удовлетворяющего уравнению $F_{jN}(\xi_{jN} / \vec{t}_{j-1}) = p$, $0 < p < 1$.

Данные теоретические результаты и позволяют практически создавать алгоритмы непараметрических датчиков стационарных случайных процессов. Для этого необходимо использовать алгоритм стохастической аппроксимации нахождения корня уравнения $F_{jN}(t_j / \vec{t}_{j-1}) = p$ одной переменной t_j при фиксированном векторе $\vec{t}_{j-1}$. Рассмотрим алгоритм получения одного многомерного случайного числа $\vec{x}^* = (x_1^*, ..., x_k^*)^T$. Пусть $u_1, ..., u_k$ – равномерные на [0; 1] числа. Генерируем u_1 и решаем уравнение $F_{1N}(x_1^*) = u_1 \Rightarrow x_1^* = F_{1N}^{-1}(u_1)$. В результате получаем x_1^*. Данный алгоритм давно используется для получения одномерных случайных чисел. Генерируем u_2 и решаем уравнение $F_{2N}(x_2^* / x_1^*) = u_2 \Rightarrow x_2^* = F_{2N}^{-1}(u_2 / x_1^*)$, генерируем u_k и решаем уравнение $F_{kN}(x_k^* / \vec{x}_{k-1}^*) = u_k \Rightarrow x_k^* = F_{kN}^{-1}(u_k / \vec{x}_{k-1}^*)$. На j такте необходимо решить уравнение $F_{jN}(x_j^* / \vec{x}_{j-1}^*) = u_j$ относительно x_k^*, при условии, что известно число $\vec{x}_{j-1}^* = (x_1^*, ..., x_{j-1}^*)^T$ вида

$$\frac{\sum_{i=1}^{N} G\left(\frac{x_j^* - x_i}{h_N}\right) \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)}{\sum_{i=1}^{N} \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)} = u_j, \ j = \overline{2,k}. \quad (3)$$

В результате, при заданном интервале зависимости k, мы получаем начальные значения $x_1^*, ..., x_k^*$ случайного временного ряда. Для получения следующего значения x_{k+1}^* при заданных $x_1^*, ..., x_k^*$ генерируем u_{k+1} и решаем уравнение

$$\frac{\sum_{i=k+1}^{N} G\left(\frac{x_{k+1}^* - x_i}{h_N}\right) \prod_{\lambda=1}^{k} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)}{\sum_{i=k+1}^{N} \prod_{\lambda=1}^{k} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)} = u_{k+1}. \quad (4)$$

Алгоритмы (3), (4) зависят от выбора ядерных функций $G(u)$, K_λ, параметра размытости h_N, интервала зависимости k. Выбор ядерных функций $G(u)$, K_λ и параметра размытости h_N хорошо исследован в рамках непараметрической статистики (на-

пример, в [5]) и в рамках данной статьи не рассматривается. Выбор интервала зависимости k является достаточно критичным во всех бутстреп-процедурах генерации *re*-выборок случайных процессов (длина блока в блочном бутстрепе, порядок АР-модели) и, несмотря на многочисленные исследования, остается достаточно субъективным. В данной работе предлагается алгоритм нахождения параметра k, основанный на непараметрических алгоритмах прогноза [7], который может быть использован в разных бутстреп-процедурах. Возьмем простой непараметрический алгоритм прогноза регрессионного типа [5; 7]. Алгоритм нахождения параметра k заключается в следующем. На основании первых значений ряда $x_1,...,x_j, 1 \le j \le k$ делаем прогноз следующего значения X_{j+1}. Образуем квадрат невязок прогноза $Q(j) = \sum_{i=j+1}^{N} (X_{j+1} - x_i)^2$. Значение параметра k находим из условия минимума $Q(j)$ (для циклических гетероскедастических случайных процессов наблюдается минимум) или $Q(j) \le l$, где l задается пользователем.

Моделирование. Для теста возьмем случайный стационарный процесс

$$X(t) = A\sin(\varpi t + \varphi) + b \cdot \varepsilon,$$

где φ равномерная на $[0; 2\pi]$ случайная величина; b – константа уровня помех; ε – стандартная нормальная величина. Вначале определялся интервал зависимости k. На рис. 1, 2 приводятся зависимости $Q(j)$ (рис. 1. – без помех; рис. 2. – с помехой b = 0,1; N = 100), из которых следует, что интервал зависимости процесса без помех равен 2, с помехами – 5.

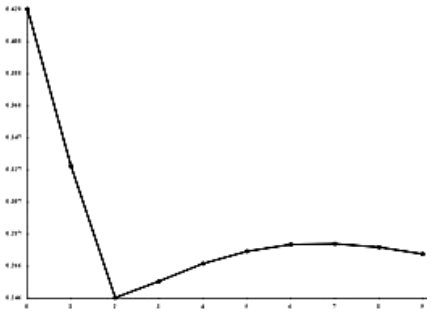


Рис. 1

Без помех: $F_{1N}(t)$ получается равномерным на [–1; 1]; $F_{2N}(t/t_1)$ получается двухточечным; $F_{3N}(t/t_1, t_2)$ вырождается в сингулярное. Моделировались и другие процессы, но наглядность, как в тестовом примере, существенно снижается.

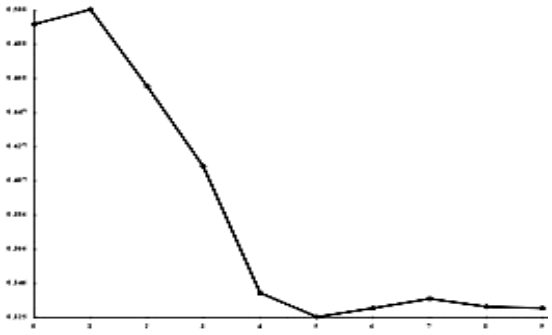


Рис. 2

Таким образом, в работе предложен метод генерации случайных процессов, основанный на классической схеме генерации (3) с использованием непараметрических оценок условных функций распределения.

Предложен метод нахождения интервала зависимости на основе непараметрических оценок прогноза. Данный метод может быть использован в качестве блока адаптивной настройки в классических бутстреп-процедурах (блочный бутстреп, локальный бутстреп).

К недостаткам предложенного метода следует отнести большую вычислительную емкость. Хотя моделирование производилось на кластерах, требуется большая вычислительная производительность. Сейчас рассматривается реализация таких алгоритмов в нейросетевом базисе.

Библиографические ссылки

1. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М. : Финансы и статистика, 1998.

2. Davison A. C., Hinkley D. V. Bootstrap Methods and Their Application. Cambridge : Cambridge University Press, 1997.

3. Бюльман П. Бутстрап-схемы для временных рядов // Квантиль. 2007. № 3. С. 37–56.

4. Hardle W., Horovitz J., Kreiss J-P. Bootstrap methods for time series // International Statist. Review. 2003. № 71. P. 435–459.

5. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М. : Наука, 2004.

6. Ермаков С. М., Михайлов Г. А. Курс статистического моделирования. М. : Наука, 1976.

7. Simakhin V. A. Nonparametric Robust Prediction Algorithms // Book of Abstracts of the International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations managemen. Beer Sheva, 2010. P. 204–218.

A. V. Mayer, V. A. Simakhin

NONPARAMETRIC SENSORS FOR STOHASTIC STATIONARY PROCESS

In the article we consider an algorithm of nonparametric generator's building for stohastic stationary process. Dependency interval of stochastic process is determined with the help of nonparametric algorithms of prognosis.

Keywords: generator, nonparametric, prognosis, process, modeling.

© Маер А. В., Симахин В. А., 2010

УДК 519.224:330.46

А. А. Новоселов

ПОСТРОЕНИЕ МНОГОМЕРНЫХ ДИСКРЕТНЫХ РАСПРЕДЕЛЕНИЙ С ЗАДАННОЙ КОРРЕЛЯЦИОННОЙ СТРУКТУРОЙ

Рассмотрены методы воспроизведения многомерного дискретного распределения с заданной корреляционной структурой и маргинальными распределениями. Для воспроизведения используются смеси базовых распределений и решение некоторых оптимизационных задач.

Ключевые слова: дискретное распределение, корреляция, копула, смесь.

Пусть заданы нормальные распределения со средними значениями $\mu_1, \ldots, \mu_d$ и стандартными отклонениями $\sigma_1, \ldots, \sigma_d$. Для произвольной корреляционной матрицы R существует единственное многомерное нормальное распределение, обладающее такими маргинальными распределениями и корреляционной матрицей. Хорошо известный алгоритм воспроизведения соответствующего случайного вектора $X = (X_1,..., X_d)$ основан на факторизации ковариационной матрицы.

Обозначим

$$\Lambda = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_d \end{pmatrix}$$

диагональную матрицу со стандартными отклонениями на диагонали, тогда $C = \Lambda R \Lambda$ является ковариационной матрицей распределения вектора X. Будучи неотрицательно определенной и симметричной, ковариационная матрица C может быть представлена в виде

$$C = A'A \qquad (1)$$

с некоторой матрицей A, причем последняя определяется не одним образом. Примерами такого представления являются разложение Холецкого и ортогональное разложение.

При наличии разложения (1) вектор X воспроизводится из стандартного нормального случайного вектора Z по формуле

$$X = A'Z. \qquad (2)$$

Действительно, для Z справедливо $\boldsymbol{E}ZZ' = I$, где I – единичная матрица соответствующего размера, поэтому $\boldsymbol{E}XX' = \boldsymbol{E}(A'ZZ'A) = A'(\boldsymbol{E}ZZ')A = A'A = C$, так что X обладает требуемой ковариационной структурой.

В случае, когда компоненты X имеют фиксированные дискретные распределения, аналогичный метод оказывается неприменимым. Во-первых, заданным маргинальным распределениям и ковариационной матрице соответствуют, вообще говоря, многие многомерные дискретные распределения. Может оказаться и так, что подходящее многомерное распределение не существует.

Во-вторых, алгоритм вращения (2) не сохраняет дискретную решетку значений, на которой задано распределение.

В работах [1; 2] анонсированы методы воспроизведения двумерного дискретного распределения с заданными маргинальными распределениями и корреляцией, основанные на смесях некоторых базовых распределений и минимизации уклонения от независимого распределения. В настоящей работе предлагается обоснование этих методов.

Описание двумерного дискретного распределения. Пусть размерность $d = 2$. Обозначим $K = \{1,\ldots,m\} \times \{1,\ldots,n\}$. Дискретное распределение вектора $X = (X_1, X_2)'$ задается на прямоугольной сетке значений $\{x_{11},\ldots, x_{1m}\} \times \{x_{21},\ldots, x_{2n}\}$ в виде $\boldsymbol{P}(X_1 = x_{1i}; X_2 = x_{2j}) = r_{ij}$, $(i, j) \in K$. Обозначим совместное распределение компонент вектора X:

$$r = \{r_{ij}, (i,j) \in K\}. \qquad (3)$$

Здесь $P(X_1 = x_{1i}) = p_i$, $i = 1,\ldots,m$ и $P(X_2 = x_{2j}) = q_j$, $j = 1,\ldots,n$, так что векторы

$$p = (p_1,\ldots,p_m),\ q = (q_1,\ldots,q_n) \qquad (4)$$

описывают маргинальные распределения компонент.

Средние значения

$$a_1 = EX_1 = \sum_{i=1}^{m} x_{1i} p_i \,, \quad a_2 = EX_2 = \sum_{j=1}^{n} x_{2j} q_j \,,$$

$$a_{12} = E(X_1 X_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{1i} x_{2j} r_{ij} \,,$$

стандартные отклонения

$$\sigma_1 = \sqrt{E(X_1 - EX_1)^2} \,, \quad \sigma_2 = \sqrt{E(X_2 - EX_2)^2}$$

и коэффициент корреляции

$$c = \frac{E(X_1 X_2) - EX_1 EX_2}{\sigma_1 \sigma_2} \tag{5}$$

вычисляются как обычно.

Если маргинальные распределения (4) известны, то выполняются следующие соотношения:

$$\sum_{i=1}^{m} r_{ij} = q_j, j = 1, \ldots, n \,; \tag{6}$$

$$\sum_{j=1}^{n} r_{ij} = p_i, i = 1, \ldots, m \,. \tag{7}$$

Отметим, что среди $m + n$ уравнений (6), (7) имеется лишь $m + n - 1$ независимых, поскольку сумма компонент любого распределения равна 1.

Если же дополнительно известен коэффициент корреляции c компонент X, то справедливо и уравнение

$$\sum_{i=1}^{m} \sum_{j=1}^{n} x_{1i} x_{2j} r_{ij} = c\sigma_1 \sigma_2 + a_1 a_2 \,. \tag{8}$$

Обозначим $F(p, q)$ класс всех двумерных распределений (3) с маргинальными распределениями (4) (его еще называют классом Фреше), а $F_c(p, q)$ его подкласс распределений с корреляцией c. Известно [3], что среди всех распределений в $F(p, q)$ наименьшей корреляцией $c_{\min} = c_{\min}(p, q)$ обладает антикомонотонное распределение $R^-(p, q)$, а наибольшей корреляцией $c_{\max} = c_{\max}(p, q)$ обладает комонотонное распределение $R^+(p, q)$, причем $c_{\min} > -1$ и $c_{\max} < 1$. Класс Фреше представим в виде

$$F(p,q) = \bigcup_{c \in [c_{\min}, c_{\max}]} F_c(p,q) \,. \tag{9}$$

Обозначим еще $R^0(p, q)$ независимое распределение из класса Фреше $F(p, q)$.

Пример. Пусть на сетке значений

$$\{0, 1, 2\} \times \{0, 1\} \tag{10}$$

заданы маргинальные распределения

$$p = (1/3, 1/2, 1/6), q = (2/5, 3/5). \tag{11}$$

Основные характеристики маргинальных распределений равны:

$$a_1 = \frac{5}{6}, a_2 = \frac{3}{5}, \sigma_1 = \frac{\sqrt{17}}{6}, \sigma_2 = \frac{\sqrt{6}}{5}.$$

Двухпараметрическое представление класса Фреше имеет вид

$$R = \begin{pmatrix} u & v & 2/5 - u - v \\ 1/3 - u & 1/2 - v & u + v - 7/30 \end{pmatrix}. \tag{12}$$

Область допустимых значений параметров задается неравенствами

$$0 \le u \le 1/3, 0 \le v \le 1/2, 7/30 \le u + v \le 2/5 \tag{13}$$

и представлена на рисунке. Окружностями на нем отмечены антикомонотонное $R^-(p, q)$

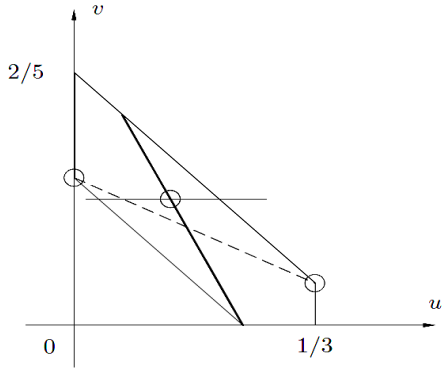
$$R^-(p,q) = \begin{pmatrix} 0 & 7/30 & 1/6 \\ 1/3 & 4/15 & 0 \end{pmatrix}; \tag{14}$$

независимое $R^0(p, q)$

$$R^0(p,q) = \begin{pmatrix} 2/15 & 1/5 & 1/15 \\ 1/5 & 3/10 & 1/10 \end{pmatrix}; \tag{15}$$

комонотонное $R^+(p, q)$ распределение

$$R^-(p,q) = \begin{pmatrix} 1/3 & 1/15 & 0 \\ 0 & 13/30 & 1/6 \end{pmatrix}. \tag{16}$$



Допустимая область в плоскости параметров (u, v):
—— – некоррелированные компоненты; О (слева направо) – антикомонотонное, независимое и комонотонное распределения; —— – решения задачи 1

Класс $F_c(p, q)$ описывается в параметрическом представлении следующим уравнением:

$$v = \frac{7}{15} + c\frac{\sqrt{102}}{30} - 2u \,. \tag{17}$$

Некоррелированные распределения лежат на отрезке прямой, описываемом выражением

$$u + v = \frac{7}{15}, \frac{1}{15} \le u \le \frac{7}{30} \,.$$

Коэффициент корреляции в данном примере заключен в интервале

$$c \in \left[-\frac{7}{\sqrt{102}}; \frac{8}{\sqrt{102}} \right]. \tag{18}$$

Критерий близости к независимому распределению. Теперь сформулируем две задачи выбора из всего множества совместных распределений, удовлетворяющих условиям (6)–(8), единственного распределения, наименее уклоняющегося от независимого распределения в смысле критерия

$$f(r)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_{ij}-p_iq_j\right)^2\to\min_r. \qquad (19)$$

Задача 1. В первой задаче целевая функция (19) минимизируется при ограничениях (6)–(8).

Функция Лагранжа этой задачи имеет вид

$$L(r,\lambda,\mu,\nu)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(r_{ij}-p_iq_j\right)^2+\sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{m}r_{ij}-q_j\right)+$$
$$+\sum_{i=1}^{m}\mu_i\left(\sum_{j=1}^{n}r_{ij}-p_i\right)+\nu\left(\sum_{i=1}^{m}\sum_{j=1}^{n}x_{1i}x_{2j}r_{ij}-c\sigma_1\sigma_2-a_1a_2\right).$$

Дифференцируя функцию Лагранжа по всем переменным и приравнивая производные к нулю, получаем уравнение

$$\partial L/\partial r_{kl}=(r_{kl}-p_kq_l)+\lambda_l+\mu_k+\nu x_{1k}x_{2l}=$$
$$=0,k=1,\dots m,l=1,\dots n, \qquad (20)$$

а также выражения (6)–(8). Отметим, что ввиду соотношения $\sum_{i=1}^{m}p_i=\sum_{j=1}^{n}q_j=1$ в системе (6)–(8) одно из уравнений, например первое, является следствием остальных, и его можно отбросить вместе с соответствующим множителем Лагранжа λ_1. Для решения полученной системы уравнений выразим элементы матрицы *r* из уравнения (20) в виде функции множителей Лагранжа λ,μ,ν:

$$r_{ij}=p_iq_j-\left(\lambda_j+\mu_i+\nu x_{1i}x_{2j}\right)=0,(i,j)\in K. \qquad (21)$$

Подставив полученное выражение в уравнения (6)–(8), имеем:

$$-m\lambda_j-\sum_{i=1}^{m}\mu_i-\nu x_{2j}\sum_{i=1}^{m}x_{1i}=0,j=2,\dots n; \qquad (22)$$

$$-\sum_{j=1}^{n}\lambda_j-n\mu_i-\nu x_{1i}\sum_{j=1}^{n}x_{2j}=0,i=1,\dots m; \qquad (23)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}x_{1i}x_{2j}\left(p_iq_j-\left(\lambda_j+\mu_i+\nu x_{1i}x_{2j}\right)\right)=c\sigma_1\sigma_2+a_1a_2.$$

Обозначив $s_1=\sum_{i=1}^{m}x_{1i},s_2=\sum_{j=1}^{m}x_{2j},s_{12}=\sum_{i=1}^{m}x_{1i}^2\sum_{j=1}^{n}x_{2j}^2$,

преобразуем последнее уравнение к виду

$$-s_1\sum_{j=1}^{n}\lambda_jx_{2j}-s_2\sum_{i=1}^{m}\mu_ix_{1i}-\nu s_{12}=c\sigma_1\sigma_2. \qquad (24)$$

Обозначим I_n единичную матрицу размера $n\times n$, а J_{mn} – прямоугольную матрицу размера $m\times n$, все элементы которой равны 1. Далее обозначим $\bar{x}_2=(x_{22},\dots,x_{2n})',\bar{x}_1=(x_{11},\dots,x_{1m})'$. В системе уравнений (6)–(8), как уже отмечалось, первое уравнение является следствием остальных, поэтому его можно отбросить. Кроме того, в полученной системе $m+n$ уравнений относительно $m+n+1$ неизвестных одну из неизвестных можно выбрать произвольным образом. Мы будем полагать $\lambda_1=0$, что соответствует отбрасыванию первого столбца в матрице системы. После этих операций система уравнений запишется в виде

$$A\gamma=b, \qquad (25)$$

где неизвестные множители Лагранжа обозначены

$$\gamma=(\lambda_2,\dots,\lambda_n,\mu_1,\dots,\mu_n,\nu)', \qquad (26)$$

а матрица системы *A* и вектор правых частей имеют вид

$$A=\begin{pmatrix}-mI_{n-1} & -J_{n-1,m} & -s_1\bar{x}_2\\ -J_{m,n-1} & -nI_m & -s_2\bar{x}_1\\ -s_1\bar{x}_2' & -s_2\bar{x}_1' & -s_{12}\end{pmatrix},\quad b=\left(0,\dots,0,c\sigma_1\sigma_2\right).$$

Квадратная система линейных уравнений (25) имеет единственное решение (26), добавляя в которое значение $\lambda_1=0$, из (21) вычисляем искомое распределение.

Задача 2. Целевая функция (19) минимизируется при ограничениях (6)–(8) и

$$r_{ij}\ge 0,(i,j)\in K. \qquad (27)$$

Аналитическое решение этой задачи в общем виде недоступно, однако численные методы позволяют эффективно решать ее. Приведем решения для рассмотренного примера.

Максимальному значению $c=c_{\max}=8/\sqrt{102}$ соответствует решение

$$r=\begin{pmatrix}1/3 & 1/15 & 0\\ 0 & 13/30 & 1/6\end{pmatrix},$$

которое, как нетрудно заметить, представляет собой комонотонное распределение (16).

Минимальному значению $c=c_{\min}=-7/\sqrt{102}$ соответствует решение

$$r=\begin{pmatrix}0 & 7/30 & 1/6\\ 1/3 & 4/15 & 0\end{pmatrix},$$

представляющее собой антикомонотонное распределение (14).

Результаты работы позволяют эффективно строить дискретные вероятностные модели с заданной корреляционной структурой. Предложенная методика без труда обобщается на многомерный случай $d>2$.

Библиографические ссылки

1. Новоселов А. А. Воспроизведение дискретных распределений с заданной ковариационной структурой // Материалы II Межрегион. конф. / КГТЭИ. Красноярск, 2009. С. 229–234.

2. Новоселов А. А. Дискретные распределения с заданной корреляцией, наименее отклоняющиеся от независимого // Тр. XIII Междунар. конф. по эвентологической математике и смежным вопросам / КГТЭИ ; СФУ. Красноярск, 2009. С. 126–131.

3. Nelsen R. B. An Introduction To Copulas. Springer, 1998.

A. A. Novosyolov

CONSTRUCTION OF MULTIDIMENSIONAL DISCRETE DISTRIBUTIONS WITH PREASSIGNED CORRELATION STRUCTURE

The paper is devoted to methods of construction of multidimensional discrete distributions with preassigned correlation structure and marginal distributions. The methods are based on mixture of the basic distributions and on decision of some optimization problems.

Keywords: discrete distribution, correlation, copula, mixture.

© Новоселов А. А., 2010

УДК 62-50:519.224

А. А. Новоселов

ПАРАМЕТРИЗАЦИЯ МОДЕЛЕЙ УПРАВЛЯЕМЫХ СИСТЕМ

Описывается применение метода ортогональных рядов для построения моделей управляемых систем параметрического вида в условиях непараметрической неопределенности. Ключевым элементом метода является выбор длины ортогонального ряда по данным наблюдений, т. е. определение параметрической структуры модели. Продемонстрировано применение метода к оцениванию плотности распределения и функции регрессии. Предложены пути обобщения оценок на многомерный случай.

Ключевые слова: плотность распределения, функция регрессии, ортогональный ряд, непараметрическая оценка.

В классической теории управления [1] обычно используются параметрические модели, которые строятся в условиях параметрической неопределенности. Здесь предполагается известным факт принадлежности модели к некоторому заранее определенному конечномерному классу моделей. Процедура идентификации (оценивания) модели сводится при этом к оцениванию по данным наблюдений некоторого конечного набора параметров.

Столь полную априорную информацию можно считать доступной далеко не всегда, поэтому параллельно развивается и непараметрическая теория управления [2; 3], в которой априорный класс моделей предполагается бесконечномерным; при этом говорят о непараметрической неопределенности. Малое количество априорной информации приходится компенсировать при этом большим количеством статистических наблюдений.

Основными строительными блоками для создания математических моделей систем управления по наблюдениям являются оценки плотности распределения и функции регрессии [1–3]. Непараметрическая оценка плотности распределения ядерного типа была впервые рассмотрена в [4; 5], а соответствующая непараметрическая оценка функции регрессии – в работах [6; 7].

Классические параметрические оценки этих функций гораздо проще для человеческого восприятия и лучше приспособлены для анализа, чем непараметрические оценки. Поэтому постоянно предпринимаются попытки построения оценок параметрического вида в условиях непараметрической неопределенности. Это направление можно назвать параметризацией моделей.

Одной из ветвей этого направления является использование оценок в виде отрезков ортогональных рядов, называемых еще проекционными оценками ввиду прозрачной геометрической аналогии. Оценка такого типа для плотности распределения была предложена Н. Н. Ченцовым в [8], а оценка для функции регрессии – в [9].

Проекционные оценки были подвергнуты тщательному изучению. Так, в статьях [10–13] исследовались вопросы сходимости оценок в различных смыслах, в публикациях [14; 15] проводилось сравнение различных непараметрических оценок. В [16–19] исследовались проекционные оценки в конкретных базисах. Во всех этих исследованиях длина отрезка ортогонального ряда выбиралась произвольно. В работах [20–24] предложен некий принцип оптимального выбора длины отрезка ортогонального ряда в проекционных оценках, исследованы свойства получаемых оптимальных проекционных оценок. В настоящей работе приведен обзор результатов [20–24], и представлены соображения по поводу обобщения оптимальных проекционных оценок на многомерный случай.

Проекционное приближение. Пусть L_r^2 – гильбертово пространство функций f, заданных на подмножестве U вещественной оси и интегрируемых с квадратом с весом r в смысле

$$\int_U f^2(x) r(x) dx < \infty,$$

где r – неотрицательная весовая функция на U. Скалярное произведение и норма задаются в L_r^2 выражениями:

$$(f,g)=\int_U f(x)g(x)r(x)dx,\ \ \|f\|=\sqrt{(f,f)}.$$

Любая функция f из L_r^2 представима в виде ряда [25]

$$f(x)=\sum_{j=0}^{\infty}a_j g_j(x),$$

в смысле сходимости последовательности его частичных сумм

$$f_N(x)=\sum_{j=0}^{N}a_j g_j(x) \qquad (1)$$

по норме пространства L_r^2, что означает $\|f-f_N\|\to 0$ при $N\to\infty$. Здесь $G=\{g_j,\ j=0,\ 1,\ \ldots\}$ обозначена произвольная полная ортонормированная система в L_r^2, а $a_j=(f,g_j),\ j=0,\ 1,\ \ldots$ представляют собой коэффициенты Фурье функции f относительно системы G.

Выбор ортонормированной системы G зависит от конкретной задачи; в различных ситуациях могут быть использованы, например, классические ортогональные многочлены [26] или тригонометрическая система функций [27].

Оценка плотности распределения. Пусть X, $X_1,\ \ldots,\ X_n,\ \ldots$ – последовательность независимых, одинаково распределенных на $U\subseteq R$ случайных величин, и пусть распределение X имеет неизвестную плотность f. Рассмотрим задачу оценивания неизвестной плотности распределения f по выборке наблюдений $X_1,\ \ldots,\ X_n$ объема n в ситуации, когда априори неизвестен какой-либо конечномерный класс плотностей, которому заведомо принадлежит f, т. е. в ситуации непараметрической неопределенности.

В [4; 5] для этого была предложена ядерная оценка плотности

$$\hat f_{n,h}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)=\\ =\frac{1}{n}\sum_{i=1}^{n}K_h\left(x-X_i\right),\quad x\in U, \qquad (2)$$

где K – некоторая плотность распределения на R, называемая ядром оценки, $K_h(u)=K(u/h)/h$; h – параметр оценки, называемый параметром размытости или параметром ширины окна.

В [8] рассматривалась проекционная оценка плотности, основанная на отрезке ряда (1), в котором коэффициенты Фурье a_j заменены на оценки по выборке наблюдений вида

$$\hat a_j=\frac{1}{n}\sum_{i=1}^{n}g_j(X_i)r(X_i),\ j=0,1,\ldots$$

Таким образом, проекционная оценка плотности имеет вид

$$\hat f_{n,N}(x)=\sum_{j=0}^{N}\hat a_j g_j(x),\quad x\in U. \qquad (3)$$

Отметим, что в результате несложных преобразований эту оценку можно представить следующим образом:

$$\hat f_{n,N}(x)=\frac{1}{n}\sum_{i=1}^{n}K_N(x,X_i),$$

$$K_N(x,u)=\sum_{j=0}^{N}g_j(x)g_j(u)r(x),\quad x,u\in U.$$

Это представление по виду напоминает второе представление в оценке (2), как среднее выборочное значение некоторого ядра. Существенное отличие состоит в том, что ядро в последнем выражении не является плотностью распределения, в частности, оно может принимать отрицательные значения. Таким же недостатком обладает и проекционная оценка (3), она не является, вообще говоря, плотностью распределения, поскольку может принимать отрицательные значения. В некоторых случаях избавиться от этого недостатка позволяет следующий прием. В качестве оценки плотности рассматривается среднее арифметическое первых частичных сумм ряда

$$\tilde f_{n,N}(x)=\frac{1}{N+1}\sum_{j=0}^{N}\hat f_{n,j}(x),\quad x\in U.$$

Такой прием позволяет обеспечить неотрицательность оценки плотности, например, при использовании тригонометрической ортонормированной системы [27]. При этом оценка оказывается построенной посредством так называемых сумм Фейера [27].

Длина разложения N выбирается в [8] произвольно, там также приводятся некоторые асимптотические соображения по поводу выбора N. В работах [20–22; 24] предложен и исследован оптимальный принцип выбора длины отрезка ряда, основанный на следующем наблюдении. Обозначим $t_j=g_j(X)r(X),\ j=0,\ 1,\ \ldots$ и вычислим среднеквадратичное отклонение проекционной оценки плотности (3) от неизвестной плотности f. Получаем выражение

$$T_n(N)=E\left\|f-\hat f_{n,N}\right\|^2=\|f\|^2-\frac{n+1}{n}\sum_{j=0}^{N}a_j^2+\frac{1}{n}\sum_{j=0}^{N}Et_j^2. \qquad (4)$$

Оказывается [20; 24], что величина $T_n(N)$ при фиксированном n достигает минимума по N, это означает, что при любом объеме выборки имеется оптимальная длина отрезка ряда в проекционной оценке (3), причем оптимальность понимается в смысле среднеквадратического критерия (4). Для приближенной минимизации критерия (4) с использованием выборки наблюдений заметим [20; 24], что слагаемое $\|f\|$ в правой части (4) не влияет на положение точки минимума, и его можно отбросить, а остальные члены оцениваются следующим образом:

$$\hat T_n(N)=-\frac{n+1}{n}\sum_{j=0}^{N}\hat a_j^2+\frac{2}{n}\sum_{j=0}^{N}\tilde t_j^2,$$

$$\tilde t_j^2=\frac{1}{n}\sum_{i=1}^{n}g_j^2(X_i)r^2(X_i). \qquad (5)$$

Минимизация $\hat{T}_n(N)$ по N дает оптимальное значение $\hat{N}$ длины отрезка ряда, подстановка которого в выражение (4) дает оптимальную проекционную оценку

$$\hat{f}_n(x) = \sum_{j=0}^{\hat{N}} \hat{a}_j g_j(x), \quad x \in U. \qquad (6)$$

В качестве примера рассмотрим тригонометрическую ортонормированную систему на [–1; 1], которая состоит из функций

$$\left\{\frac{1}{2}, \sin(\pi j x), \cos(\pi j x), x \in [-1;1], j = 1, 2, \ldots\right\}.$$

При использовании этой ортонормированной системы удобно включать члены в ортогональное разложение парами, при этом оптимальная проекционная оценка приобретает вид

$$\hat{f}_n(x) = \frac{1}{2} + \sum_{j=1}^{\hat{N}} \left(\hat{c}_j \cos(\pi j x) + \hat{s}_j \sin(\pi j x)\right). \qquad (7)$$

Здесь

$$\hat{c}_j = \frac{1}{n}\sum_{i=1}^{n} \cos(\pi j X_i), \;\; \hat{s}_j = \frac{1}{n}\sum_{i=1}^{n} \sin(\pi j X_i), \;\; j = 1, 2, \ldots$$

Рассмотрим случайную величину X, имеющую распределение с треугольной плотностью на [–1; 1], т. е. $f(x) = 1 - |x|$, $x \in [-1; 1]$. Методами статистического моделирования нетрудно получить выборку наблюдений такой случайной величины, для целей иллюстрации мы возьмем объем выборки равным $n = 50$. График оценки критериальной функции (5), построенный по полученной выборке, приведен на рис. 1. Видно, что в оптимальное разложение входит только первая пара тригонометрических функций ($N = 1$). Графики модельной плотности распределения f и ее оптимальной проекционной оценки показаны на рис. 2. С ростом объема выборки качество оценивания довольно быстро возрастает. Более подробный анализ оптимальных проекционных оценок плотности можно найти в работах [20–24].

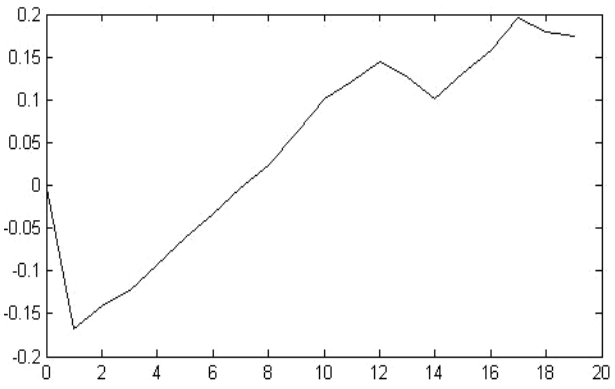


Рис. 1. График оценки критериальной функции (5), построенный по выборке наблюдений объема $n = 50$

Оценка функции регрессии. Пусть теперь (X, Y), (X_1, Y_1), …, (X_n, Y_n), … – последовательность независимых одинаково распределенных случайных векторов. Обозначим $f(x) = E(Y|X = x)$, $x \in U$ функцию регрессии Y на X. Для непараметрического оценивания этой функции по наблюдениям (X_i, Y_i), $i = 1, \ldots, n$ в [6; 7] была предложена ядерная оценка

$$\hat{f}_n(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)}, \quad x \in U.$$

Здесь объекты K, h имеют тот же смысл, что и для ядерной оценки плотности (2).

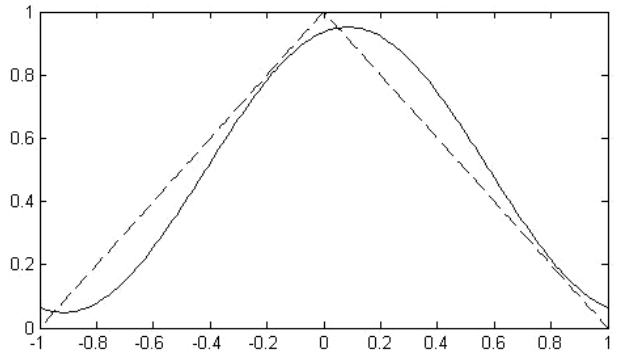


Рис. 2. График модельной треугольной плотности распределения и ее оптимальной проекционной оценки (7), объем выборки $n = 50$

Для простоты предположим, что $U = [0; 1]$ и случайная величина X имеет равномерное распределение на [0; 1] так, что ее плотность равна $P(x) = 1$, $x \in [0; 1]$. Положим также, что весовая функция постоянна: $r(x) = 1$, $x \in [0; 1]$. При этом коэффициенты Фурье имеют вид

$$a_j = \int_0^1 f(x) g_j(x) dx = E(Y g_j(X)), \; j = 0, 1, \ldots$$

Поэтому проекционная оценка функции регрессии имеет вид (3), где оценки коэффициентов Фурье вычисляются по формулам:

$$\hat{a}_j = \frac{1}{n}\sum_{i=1}^{n} Y_i g_j(X_i), \; j = 0, 1, \ldots$$

Все свойства критерия выбора длины ортогонального разложения сохраняются и в случае оценивания функции регрессии, оптимальная проекционная оценка вычисляется по той же формуле (6), а качество приближения получается даже лучше, чем для плотности распределения, что связано со спецификой данной задачи.

Случай известной, но не равномерной плотности распределения X отличается лишь повышенной громоздкостью формулировок. В случае неизвестной плотности X можно использовать фейеровские оценки [23; 24].

Многомерное обобщение. Представляет интерес обобщение описанных оценок на многомерный случай, когда область распределения случайной величи-

ны X имеет размерность, превышающую 1. Для осуществления такого обобщения по существу нужно лишь уметь строить многомерные ортонормированные системы. Это легко достигается на основе одномерных ортонормированных систем, что мы и продемонстрируем для двумерного случая.

Пусть $\{g_j(x), j = 0, 1, \ldots\}$, $x \in U$ и $\{h_k(y), k = 0, 1, \ldots\}$, $y \in V$ – одномерные полные ортонормированные системы на соответствующих областях определения. Тогда полная ортонормированная система на $U \times V$ образуется из всевозможных произведений одномерных функций, а именно:

$$G_{jk}(x, y) = g_j(x)h_k(y),$$
$$j = 0, 1, \ldots, k = 0, 1, \ldots, (x, y) \in U \times V.$$

Следует иметь в виду, что с ростом размерности очень быстро нарастают комбинаторные сложности в вычислении ортонормированных функций, поэтому применимость описанного способа оценивания в пространствах высокой размерности сомнительна. Кроме того, метод проекционного оценивания не обладает свойствами локальных оценок, вследствие чего не может использовать «трубчатое» строение выборочных данных. По-видимому, возможно построение специальных ортонормированных систем, адаптированных под имеющуюся структуру данных, что позволит использовать такое специальное строение, однако, нам неизвестны исследования в этом направлении.

В работе дан краткий обзор оптимальных проекционных оценок плотности распределения и функции регрессии, которые применяются при построении моделей управляемых систем по наблюдениям. Описаны основные алгоритмы, участвующие в построении оценок, приведены примеры и иллюстрации. Отмечается возможность многомерного обобщения процедуры оценивания и имеющиеся ограничения по размерности.

Библиографические ссылки

1. Цыпкин Я. З. Адаптация и обучение в автоматических системах. М. : Наука, 1968.

2. Живоглядов В. П., Медведев А. В. Непараметрические алгоритмы адаптации. Фрунзе : Илим, 1974.

3. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука, 1983.

4. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statist. 1962. № 33. P. 1065–1076.

5. Rosenblatt M. Remarks on some nonparametric estimates of a density function // Ann. Math. Statist. 1956. № 27. P. 832–837.

6. Надарая Э. А. Об оценке регрессии // Теория вероятностей и ее применения. 1964. Т. 10. С. 157–159.

7. Watson G. S. Smooth Regression Analysis // Sankhya. Ser. A. 1964. № 26. P. 359–378.

8. Ченцов Н. Н. Оценка неизвестной плотности распределения по наблюдениям // ДАН СССР. 1962. № 147. С. 45–48.

9. Greblicky W. Nonparametric system identification by orthogonal series // Probl. Control Inform. Theory. 1979. № 8. P. 67–73.

10. Anderson G. L., Figueiredo R. J. P. An adaptive orthogonal series estimator for probability density functions // Ann. Statist. 1980. № 8. P. 347–376.

11. Bosq M. D. Sur l'estimation d'une densite multivariee pur une serie de fonctions orthogonales // C. R. Acad. Sci. Ser. A. Paris, 1969. № 268. P. 555–557.

12. Crain B. R. A note on density estimation using orthogonal expansions // J. Amer. Statist. Assoc. 1973. № 68. P. 964–965.

13. Watson G. S. Density estimation by orthogonal series // Ann. Math. Statist. 1969. № 40. P. 1496–1498.

14. Fryer M. J. A review of some nonparametric methods of density estimation // J. Inst. Math. Appl. 1970. № 20. P. 335–354.

15. Wahba G. Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation // Ann. Statist. 1975. № 3. P. 15–29.

16. Hall P. Estimating a density on the positive half line by the method of orthogonal series // Ann. Inst. Statist. Math. 1980. № 32. Part A. P. 351–362.

17. Hall P. On trigonometric series estimates of densities // Ann. Statist. 1981. № 9. P. 683–685.

18. Kronmal R., Tarter M. The estimation of probability densities and cumulative by Fourier series methods // J. Amer. Statist. Assoc. 1968. № 63. P. 925–952.

19.Walter G. G. Properties of Hermite series estimation of probability density // Ann. Statist. 1977. № 5. P. 1258–1264.

20. Новоселов А. А. О выборе структуры моделей управляемых систем // Стохастические системы управления. Новосибирск : Наука, 1979. С. 72–77.

21. Новоселов А. А. Об оптимальных ортогональных оценках // Математическая статистика и ее приложения. Томск : Изд-во Том. ун-та, 1981. С. 151–155. Вып. 7.

22. Новоселов А. А. Оптимальная параметризация плотности вероятности и функции регрессии // Адаптация и обучение в системах управления и принятия решений. Новосибирск : Наука, 1982. С. 139–147.

23. Новоселов А. А. Оценивание кривой регрессии методом сумм Фейера // Математическая статистика и ее приложения. Томск : Изд-во Том. ун-та, 1986. С. 177–183. Вып. 10.

24. Новоселов А. А. Разработка и исследование методов параметризации структур адаптивных систем : автореф. дис. канд. физ.-мат. Наук. Томск : Изд-во Том. ун-та, 1987.

25. Канторович Л. В., Акилов Г. П. Функциональный анализ. М. : Наука, 1984.

26. Суетин П. К, Классические ортогональные многочлены. М. : Наука, 1979.

27. Бари Н. К. Тригонометрические ряды. М. : Физматгиз, 1961.

A. A. Novosyolov

PARAMETRIZATION OF MODELS OF CONTROLLED SYSTEMS

The paper describes application of orthogonal series method for construction of controlled systems models under non-parametric uncertainty. A key element of the method is draw of orthogonal expansion length based on observations, in other words, defining parametric structure of the model. The method is demonstrated for estimation of distribution density and regression function. Directions for generalizing onto multi-dimensional case are also presented.

Keywords: distribution density, regression function, orthogonal series, non-parametric estimate.

© Новоселов А. А., 2010

УДК 004.932.2

И. А. Пестунов, В. Б. Бериков, Ю. Н. Синявский

СЕГМЕНТАЦИЯ МНОГОСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ АНСАМБЛЯ НЕПАРАМЕТРИЧЕСКИХ АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ*

Предложен метод сегментации многоспектральных изображений на основе ансамбля непараметрических алгоритмов; дано теоретическое обоснование метода. Результаты статистического моделирования на модельных данных и реальных изображениях подтверждают эффективность предложенного метода.

Ключевые слова: сегментация многоспектральных изображений, непараметрические алгоритмы кластеризации, ансамблевый подход.

Сегментация является одним из важнейших этапов анализа цифровых изображений [1]. Она заключается в разбиении изображения на сегменты на основе подобия спектральных, текстурных и других характеристик пикселов. Методы сегментации нашли широкое применение во многих прикладных областях, в том числе в дистанционном зондировании Земли (ДЗЗ) [2; 3], интерес к которому в последние годы непрерывно возрастает.

Один из наиболее распространенных подходов к сегментации многоспектральных изображений основан на статистических методах кластеризации [4; 5]. В этом случае задачу кластеризации, как правило, приходится решать при отсутствии каких-либо априорных сведений о числе классов и их вероятностных характеристиках. Для этого наиболее подходящими являются непараметрические алгоритмы, позволяющие получить хорошие результаты при минимальной априорной информации. Их общим недостатком является высокая чувствительность к входным параметрам, что существенно усложняет процесс настройки алгоритма для решения конкретной задачи.

Известно [6–8], что устойчивость решений в задачах кластеризации может быть повышена благодаря формированию ансамбля алгоритмов и построению на его основе коллективного решения. При этом используются результаты, полученные различными алгоритмами либо одним алгоритмом с различными значениями параметров, по разным подсистемам переменных и т. д. В настоящее время ансамблевый подход является одним из наиболее перспективных направлений в кластерном анализе [9].

В данной работе предложен алгоритм сегментации многоспектральных изображений с использованием ансамбля непараметрических алгоритмов кластеризации, основанных на оценках плотности Розенблатта–Парзена [10; 11]. Для формирования ансамбля используются результаты выполнения непараметрического алгоритма MeanSC (представляющего собой оптимизацию предложенного ранее алгоритма [12]) с различными значениями параметра сглаживания h. Итоговое коллективное решение строится на основе попарной классификации объектов. Дано теоретическое обоснование предложенного алгоритма, приведены результаты статистического моделирования на модельных данных и реальных изображениях, подтверждающие его эффективность.

Непараметрический подход к задаче кластеризации данных ДЗЗ. Предположим, что произведена k-спектральная съемка участка местности, содержащего N элементов разрешения, тогда результат съемки можно представить в виде множества $X = \{x^{(i)} = (x_1^{(i)}, ..., x_k^{(i)}) \in R^k,\ i = \overline{1, N}\}$, где $x_j^{(i)}$ – значение яркости i-го элемента разрешения в j-м диапазоне спектра $(j = \overline{1, k})$.

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 09-07-12087-офи_м).

Пусть каждый вектор $x^{(i)}$ – реализация k-мерного случайного вектора x, плотность распределения которого $f(x)$, $x=(x_1,\ldots,x_k)\in R^k$ неизвестна и нет какой-либо априорной информации о ее параметрическом виде. В этих условиях для оценивания плотности $f(x)$ в точке $x\in R^k$ целесообразно воспользоваться непараметрической оценкой Розенблатта–Парзена $\hat{f}_N(x)$, определяемой выражением

$$\hat{f}_N(x)=\frac{1}{Nh^k}\sum_{i=1}^{N}\Phi\left(\frac{x-x^{(i)}}{h}\right),$$

где h – параметр сглаживания; $\Phi(x)$ – колоколообразная функция (ядро), удовлетворяющая определенным условиям сходимости [10; 13].

Среди ядер, удовлетворяющих этим условиям, наибольшей популярностью пользуются радиально-симметричные ядра, представимые в виде

$$\Phi(x)=c_k\phi\left(\|x\|^2\right),$$

где константа $c_k>0$; $\phi:[0;\infty)\to R$ – непрерывная функция, удовлетворяющая условиям:

$$\phi(t)\ge 0;\ \phi(t_1)\ge\phi(t_2),\ \text{если } t_1<t_2;\ \int_0^{\infty}\phi(t)dt<\infty.$$

К таким ядрам относятся, например, ядро Епанечникова [13]:

$$\Phi_{\mathrm{E}}=\begin{cases}\frac{1}{2}V_k^{-1}(k+2)(1-\|x\|^2), & \text{если } \|x\|\le 1,\\ 0, & \text{иначе,}\end{cases}$$

где V_k – объем k-мерного единичного шара и многомерное нормальное ядро;

$$\Phi_{\mathrm{N}}(x)=(2\pi)^{-k/2}e^{-\frac{\|x\|^2}{2}}.$$

При использовании радиально-симметричных ядер, оценки неизвестной плотности распределения $f(x)$ и ее градиента $\nabla f(x)$ могут быть записаны ([14]) в виде

$$\hat{f}_N(x)=\frac{c_k}{Nh^k}\sum_{i=1}^{N}\phi\left(\left\|\frac{x-x_i}{h}\right\|^2\right);$$

$$\hat{\nabla}f(x)=\nabla\hat{f}_N(x)=\frac{2c_k}{Nh^{k+2}}\sum_{i=1}^{N}(x-x_i)\,\phi'\left(\left\|\frac{x-x_i}{h}\right\|^2\right).$$

Обозначим $\psi(t)=-\phi'(t)=-d\phi(t)/dt$, предполагая, что функция $\phi(t)$ дифференцируема $\forall t\in[0;\infty)$ за исключением конечного множества точек.

Тогда, согласно [15; 16]

$$m_h(x)=\frac{\sum_{i=1}^{N}x_i\psi\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{N}\psi\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x-$$

вектор среднего сдвига.

Этот вектор интересен тем, что его направление совпадает с направлением градиента оценки плотности $f(x)$ в точке x.

Итерационную процедуру, заключающуюся в переходе от $x_0\in R^k$ к $x_0^1=x_0+m_h(x_0)$, затем от x_0^1 к $x_0^2=x_0^1+m_h(x_0^1)$ и т. д. до точки x_0^*, для которой $m_h(x_0^*)=0$, называют алгоритмом среднего сдвига. Доказано [14], что эта процедура сходится к локальным максимумам (модам) плотности распределения $f(x)$. Путь, пройденный от точки x_0 до моды x_0^*, будем называть траекторией среднего сдвига и обозначать $\left[x_0,\ldots,x_0^*\right]$.

Процедура среднего сдвига порождает естественное разбиение множества X на компоненты связности: точки x_i и x_j связны, если итеративные процессы среднего сдвига, начинающиеся с этих точек, сходятся к одной и той же моде. Эта процедура достаточно трудоемка, поэтому ее непосредственное применение ограничено выборками небольшого объема. В работах [17; 18] процедура среднего сдвига применяется не ко всей исходной выборке, а к некоторому ее подмножеству значительно меньшего объема.

В следующем разделе приведено описание быстрого алгоритма кластеризации многоспектральных данных MeanSC (являющегося оптимизацией предложенного ранее алгоритма [12]), в котором стартовое множество точек для запуска процедуры среднего сдвига порождается клеточной структурой данных, формируемой в пространстве спектральных признаков. В этом алгоритме вектор $m_h(x)$ и оценка плотности $\hat{f}_N(x)$ вычисляются с использованием финитных радиально-симметричных ядер.

Описание алгоритма MeanSC. Предлагаемый алгоритм опирается на использование двух характерных особенностей многоспектральных данных. Первая из них заключается в ограниченности диапазонов изменения значений спектральных признаков (значения лежат в диапазоне целых чисел от 0 до $K-1$, где K – число уровней квантования видеосигнала, обычно не превышающее 256), а вторая – в высокой частоте повторяемости векторов спектральных яркостей. Повторяемость обусловливается ограниченностью диапазона спектральных яркостей, наличием корреляции между спектральными диапазонами, а также относительной однородностью и достаточной протяженностью природных объектов. Для описания алгоритма MeanSC с параметрами h, ε, T введем следующие определения.

Определение 1. Пусть в точке $x^*\in R^k$ достигается локальный максимум оценки плотности $\hat{f}_N(x)$. Тогда точка $x\in R^k$ *связна* с x^*, если процедура среднего сдвига, стартовавшая из x, сходится к x^*. В дальнейшем через a^* будем обозначать моду плотности, к ко-

торой сойдется процедура среднего сдвига, стартовавшая в *a*.

Определение 2. *Компонентой связности*, определяемой локальным максимумом x^* ($\hat{f}_N(x^*) \ge \varepsilon$), назовем непустое подмножество точек $Q(x^*) \subseteq X$, связных с x^*. Точку $x \in X$ будем считать *«шумом»*, если она связна с локальным максимум x_0^*: $\hat{f}_N(x_0^*) < \varepsilon$. Здесь $\varepsilon > 0$ – порог «шума».

Определение 3. *Кластером* назовем подмножество $C \subseteq X$, которое либо задано одной компонентой связности, либо задано множеством компонент связности $\mathbb{C}$ и удовлетворяет условиям:

1) $\forall\, x \in C$ выполнено $Q(x^*) \in \mathbb{C}$;

2) $\forall\, x_1^*, x_2^*$ таких, что $Q(x_1^*), Q(x_2^*) \in \mathbb{C}$ существует непрерывная траектория $P \subset R^k$, соединяющая x_1^* и x_2^*, вдоль которой $\dfrac{\hat{f}_N(x)}{\min\left(\hat{f}_N(x_1^*), \hat{f}_N(x_2^*)\right)} > T$. Здесь $T \ge 1$ – параметр, задаваемый пользователем и отвечающий за уровень детализации результата.

В соответствии с введенными определениями, алгоритм MeanSC можно записать в виде следующей последовательности шагов.

1. Формируем клеточную структуру данных в пространстве спектральных признаков. Для этого разбиваем все пространство значений спектральных признаков $[0; K-1]_1 \times \ldots \times [0; K-1]_k$ на гиперкубические клетки со стороной $2h$ (h – параметр сглаживания). Вводим общую нумерацию клеток (последовательно от одного слоя клеток к другому) и с каждой клеткой связываем набор попавших в нее спектральных векторов из X.

2. Формируем таблицу «весов» векторов множества X. Здесь под «весом» вектора x понимаем число вхождений x в множество X. При обработке спутниковых изображений таблица «весов» позволяет значительно (иногда в несколько десятков раз) сократить объем вычислений при выполнении процедуры среднего сдвига и вычислении оценок плотности распределения.

3. Формируем множество начальных (стартовых) векторов S для запуска процедуры среднего сдвига. Для каждой клетки, которая содержит векторы из X, вычисляем вектор средних значений по всем точкам, попавшим в эту клетку. Совокупность полученных таким образом векторов образуют множество S.

4. Для каждого вектора $s \in S$ находим моду s^* оценки плотности распределения $\hat{f}_N(x)$, связную с s. Из найденных мод формируется множество $Z_0 = \left\{ s^* \mid s \in S, \hat{f}_N(s^*) \ge \varepsilon \right\}$. По мере нахождения мод заполняем множество $\overline{S} = \bigcup_{s \in S} \{\overline{s}_0 = s, \ldots, \overline{s}_i = \overline{s}_{i-1} + m_h(\overline{s}_{i-1}), \ldots, s^*\}$, которое содержит все точки, пройденные при выполнении процедуры среднего сдвига.

5. Связываем каждую точку $x \in X$ с ближайшей точкой из множества $\overline{S}$, используя для сокращения вычислительных расходов введенную клеточную структуру. В результате множество X разбивается на компоненты связности в соответствии с определением 2.

6. Формируем кластеры из выделенных компонент связности. Если искомая траектория P, соединяющая $Q(x_1^*)$ и $Q(x_2^*)$ (определение 3), существует, она, наиболее вероятно, проходит через общую границу $Q(x_1^*)$ и $Q(x_2^*)$. Поэтому для ее нахождения выбираем точки $x_1 \in Q(x_1^*)$ и $x_2 \in Q(x_2^*)$, расположенные ближе всего к общей границе, и проверяем выполнение условий из определения 3 для $P = \left[x_1, \ldots, x_1^*\right] \cup \left[x_1, x_2\right] \cup \left[x_2, \ldots, x_2^*\right]$. Здесь $\left[x_1, x_2\right]$ – отрезок прямой, соединяющей точки x_1 и x_2. Исходя из определения процедуры среднего сдвига, точка $x \in P$ с наименьшей плотностью находится на отрезке $\left[x_1; x_2\right]$. Точки отрезка проверяем последовательно с шагом h.

Заметим, что благодаря финитности используемого ядра, при вычислении оценки плотности и векторов среднего сдвига достаточно использовать только векторы из клеток, которые являются соседними к клетке, содержащей точку x. Это позволяет значительно уменьшить объем вычислений.

Метод построения коллективного решения. Результаты кластеризации, получаемые с помощью алгоритма MeanSC, являются неустойчивыми к изменению параметра сглаживания h. Известно [6–8], что устойчивость результатов кластеризации может быть повышена путем использования ансамблевого подхода. При этом для построения коллективного решения используются различные принципы. Так, в [7] предлагается принцип максимизации количества взаимной информации, которую разделяет итоговое группировочное решение с исходными кластеризациями. В ряде работ используется принцип, основанный на нахождении согласованной матрицы подобия (или различия) объектов. В данной работе будет использован именно этот принцип; ансамбль предлагается формировать из L частных решений, полученных в результате выполнения алгоритма MeanSC с различными значениями параметра сглаживания h.

Предлагаемый метод построения коллективного решения может быть описан следующим образом.

Пусть с помощью некоторого алгоритма кластеризации $\mu = \mu(\theta)$, зависящего от случайного вектора параметров $\Theta \in \mathbf{\Theta}$ (где $\mathbf{\Theta}$ – некоторое допустимое множество параметров), получен набор частных решений $\mathbb{G} = \{G^{(1)}, \ldots, G^{(l)}, \ldots, G^{(L)}\}$, где $G^{(l)}$ – l-й вариант кластеризации, содержащий $M^{(l)}$ кластеров.

Обозначим через $H(\theta_l)$ бинарную матрицу $H(\theta_l)=\{H_{i,j}(\theta_l)\}$ размерности $N\times N$, которая водится для l-й группировки, следующим образом:

$$H_{i,j}(\theta_l)=\begin{cases}0, & \text{если объекты отнесены в один кластер,}\\ 1, & \text{иначе,}\end{cases}$$

где $i,j=1,\ldots,N,\ i\neq j$.

После построения L частных решений можно сформировать согласованную матрицу различий

$$\mathbf{H}=\{\mathbf{H}_{i,j}\},\quad \mathbf{H}_{i,j}=\frac{1}{L}\sum_{l=1}^{L}H_{i,j}(\theta_l),$$

где $i,j=1,\ldots,N$. Величина $\mathbf{H}_{i,j}$ равна частоте классификации x_i и x_j в разные группы в наборе группировок $\mathbb{G}$. Близкое к нулю значение величины означает, что данные объекты имеют большой шанс попадания в одну и ту же группу. Близкое к единице значение этой величины говорит о том, что шанс оказаться в одной группе у объектов незначителен.

После вычисления согласованной матрицы различий, для нахождения коллективного решения будем применять стандартный агломеративный метод построения дендрограммы, который в качестве входной информации использует попарные расстояния между объектами [9]. При этом расстояния между группами будем определять по принципу «средней связи», т. е. как среднее арифметическое попарных расстояний между объектами, входящими в группы. Процесс объединения продолжается до тех пор, пока расстояние между ближайшими группами не превысит заданное пороговое значение T_d. Этот метод является очень привлекательным, потому что он позволяет выделять иерархическую структуру кластеров, которая упрощает процесс интерпретации результатов.

Для исследования свойств предложенного метода формирования коллективного решения рассмотрим его вероятностную модель.

Предположим, что имеется некоторая скрытая (непосредственно ненаблюдаемая) переменная U, которая задает принадлежность каждого объекта к некоторому из $M\geq 2$ классов. Каждый класс характеризуется определенным законом условного распределения $p(x\,|\,U=r)=f_r(x),\ r=1,\ldots,M$. Рассмотрим следующую вероятностную модель генерации данных. Пусть для каждого объекта определяется класс, к которому он относится, в соответствии с априорными вероятностями $P_r=\mathbb{P}(U=r),\ r=1,\ldots,M$, где $\sum_{r=1}^{M}P_r=1$. Затем в соответствии с распределением $f_r(x)$ определяется значение x. Указанная процедура проводится независимо для каждого объекта.

Пусть с помощью некоторого алгоритма кластерного анализа μ строится разбиение множества объектов X на M подмножеств. Поскольку нумерация кластеров не играет роли, удобнее рассматривать отношение эквивалентности, т. е. указывать, относит ли алгоритм μ каждую пару объектов в один и тот же класс либо в разные классы. Определим для каждой пары объектов a и b следующую величину:

$$\mathbf{H}_{a,b}(\mu)=\begin{cases}0, & \text{если объекты отнесены в один кластер,}\\ 1, & \text{иначе,}\end{cases}$$

где $a,b\in X,\ a\neq b$.

Выберем произвольную пару a и b различных объектов выборки.

Пусть $P_U=\mathbb{P}(U(a)\neq U(b))$ – вероятность отнесения объектов к различным классам. Например, при $M=2$ указанная вероятность равна

$$P_U=1-\mathbb{P}(U(a)=1\,|\,a)\mathbb{P}(U(b)=1\,|\,b)-\mathbb{P}(U(a)=$$
$$=2\,|\,a)\mathbb{P}(U(b)=2\,|\,b)=1-\sum_{r=1}^{2}\frac{f_r(a)f_r(b)P_r^2}{p(a)p(b)},$$

где $p(\omega)=\sum_{r=1}^{2}f_r(\omega)P_r,\ \omega=a,b$.

Обозначим вероятность ошибки, которую может совершить алгоритм μ при классификации a и b через $P_{\text{er}}(\mu)$, где

$$P_{\text{er}}(\mu)=\begin{cases}P_U, & \text{если } \mathbf{H}_{a,b}(\mu)=0,\\ 1-P_U, & \text{если } \mathbf{H}_{a,b}(\mu)=1.\end{cases}$$

Легко заметить, что

$$P_{\text{er}}(\mu)=(1-\mathbf{H}_{a,b}(\mu))P_U+\mathbf{H}_{a,b}(\mu)(1-P_U)=$$
$$=P_U+(1-2P_U)\mathbf{H}_{a,b}(\mu).$$

Алгоритм μ зависит от случайного вектора параметров $\Theta\in\boldsymbol{\Theta}$: $\mu=\mu(\Theta)$. Чтобы подчеркнуть зависимость результатов работы от параметра Θ, в дальнейшем будем обозначать $\mathbf{H}_{a,b}(\mu(\Theta))=\mathbf{H}_{a,b}(\Theta)$, $P_{\text{er}}(\mu(\Theta))=P_{\text{er}}(\Theta)$.

Пусть в результате L-кратного применения алгоритма μ со случайно и независимо отобранными параметрами $\theta_1,\ldots,\theta_L$ получен набор решений $H(\theta_1),\ldots,H(\theta_L)$. Для определенности, будем считать, что L – нечетно. Коллективным (ансамблевым) решением по большинству голосов будем называть функцию

$$\mathbf{H}(H(\theta_1),\ldots,H(\theta_L))=\begin{cases}0, \text{ если } \dfrac{1}{L}\displaystyle\sum_{l=1}^{L}H(\theta_l)<\frac{1}{2},\\ 1, \text{ иначе.}\end{cases}$$

В рамках описанной модели для предложенного коллективного решения справедливы следующие утверждения [6].

Утверждение 1. Математическое ожидание и дисперсия величины вероятности ошибки для алгоритма $\mu(\Theta)$ равны соответственно:

$$\mathbb{E}_{\Theta}P_{\text{er}}(\Theta)=P_U+(1-2P_U)P_H,$$
$$\text{Var}_{\Theta}P_{\text{er}}(\Theta)=(1-2P_U)^2P_H(1-P_H),$$

где $P_H=\mathbb{P}(H(\Theta)=1)$.

Обозначим через $P_{\text{er}}(\Theta_1, \ldots, \Theta_L)$ случайную функцию, принимающую при фиксированных аргументах значение, равное вероятности ошибки, которую может совершить ансамблевый алгоритм при классификации *a* и *b*. Здесь через $\Theta_1, \ldots, \Theta_L$ обозначены статистические копии случайного вектора Θ. Рассмотрим поведение вероятности ошибки для коллективного решения.

Утверждение 2. Математическое ожидание и дисперсия величины вероятности ошибки для коллективного решения равны соответственно:

$$\mathbb{E}_{\Theta_1,\ldots,\Theta_L} P_{\text{er}}(\Theta_1, \ldots, \Theta_L) = P_U + (1 - 2P_U) P_{\mathbf{H},L},$$

$$\text{Var}_{\Theta_1,\ldots,\Theta_L} P_{\text{er}}(\Theta_1, \ldots, \Theta_L) = (1 - 2P_U)^2 P_{\mathbf{H},L}(1 - P_{\mathbf{H},L}),$$

где $P_{\mathbf{H},L} = \mathbb{P}\left(\frac{1}{L}\sum_{l=1}^{L} H(\theta_l) \geq \frac{1}{2}\right) = \sum_{l=\left[\frac{L}{2}\right]+1}^{L} C_L^l P_H^l (1 - P_H)^{L-l}$,

$[\cdot]$ означает целую часть числа.

Воспользуемся следующей априорной информацией об алгоритме кластерного анализа. Будем считать, что ожидаемая вероятность ошибочной классификации $\mathbb{E}_\Theta P_{\text{er}}(\Theta) < 1/2$, т. е. ожидается, что алгоритм μ проводит классификацию с лучшим качеством, нежели алгоритм случайного равновероятного выбора. Из утверждения 1 следует, что выполняется один из двух вариантов: $P_H > 1/2$ и $P_U > 1/2$; $P_H < 1/2$ и $P_U < 1/2$. Рассмотрим, для определенности, первый случай.

Утверждение 3. Если $\mathbb{E}_\Theta P_{\text{er}}(\Theta) < 1/2$ и при этом $P_H > 1/2$ и $P_U > 1/2$, то при увеличении мощности ансамбля ожидаемая вероятность ошибочной классификации уменьшается, стремясь в пределе к $1 - P_U$, а дисперсия величины вероятности ошибки стремится к нулю.

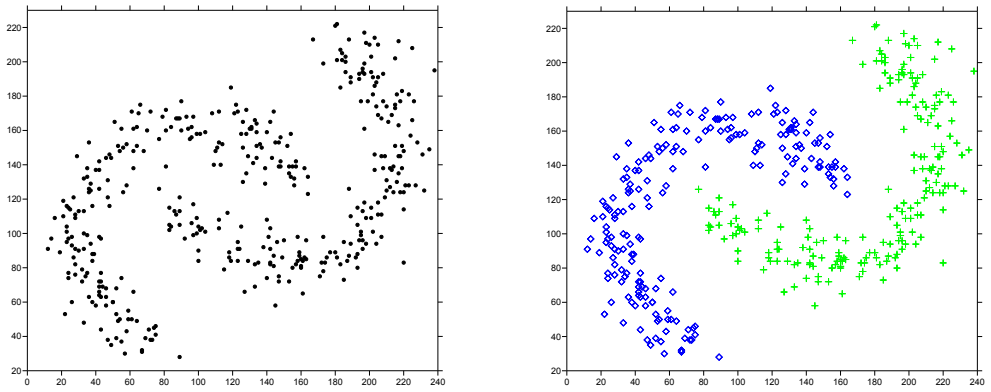
Последнее утверждение позволяет сделать вывод о том, что при выполнении вполне естественных условий использование ансамблевого подхода позволяет улучшить качество кластеризации.

Результаты экспериментальных исследований. В соответствии с описанной выше схемой, разработан и программно реализован на языке программирования C++ ансамблевый алгоритм $\text{EMeanSC}\left(\bar{h} = \{h_1, \ldots, h_L\}, \varepsilon, T, T_d\right)$. Здесь $h_1, \ldots, h_L$ – значения параметра сглаживания; ε – порог «шума»; T – порог объединения компонент связности, а T_d – параметр дендрограммы.

Ниже приведены результаты нескольких экспериментов на модельных данных и реальных изображениях. Эксперименты 1 и 2 подтверждают эффективность алгоритма для выделения классов сложной формы. Эксперимент 3 демонстрирует возможность разделения пересекающихся классов разной плотности. Эксперимент 4 демонстрирует применение алгоритма для обработки спутниковых изображений. В эксперименте 5 показано влияние параметра T_d на уровень детализации картосхемы.

Эксперимент 1. Использовались двумерные данные, состоящие из 400 точек, сгруппированных в два равновеликих линейно неразделимых класса (рис. 1), имеющих форму бананов (рис. 1, *а*). Модель построена с помощью инструментария PRTools (PRTools: the Matlab Toolbox for Pattern Recognition – http://www.prtools.org). Результаты кластеризации, построенные по ансамблю из шести элементов, представлены на рис. 1, *б*. Выделено 2 кластера, содержащих по 200 точек.

Эксперимент 2. Использовались двумерные данные (рис. 2), состоящие из 200 двумерных точек, сгруппированных в два спиралевидных класса по 100 точек (рис. 2, *а*). Сложность этой модели заключается в том, что плотность точек по мере удаления от центра спирали уменьшается. На рис. 2, *б* представлены результаты выполнения алгоритма MeanSC с двумя значениями параметра сглаживания. Несложно заметить, что искомые классы оказались раздробленными. Ансамбль, сформированный на основе указанных результатов, успешно выделил кластеры, совпадающие с искомыми классами (рис. 2, *в*).



а *б*

Рис. 1. Двумерные данные, состоящие из 400 точек:
а – исходные данные; *б* – результат выполнения EMeanSC с параметрами
$\bar{h} = \{10; 10{,}5; 11; 11{,}5; 12; 12{,}5\}$, $\varepsilon = 0$, $T = 4$, $T_d = 0{,}99$ (2 кластера)

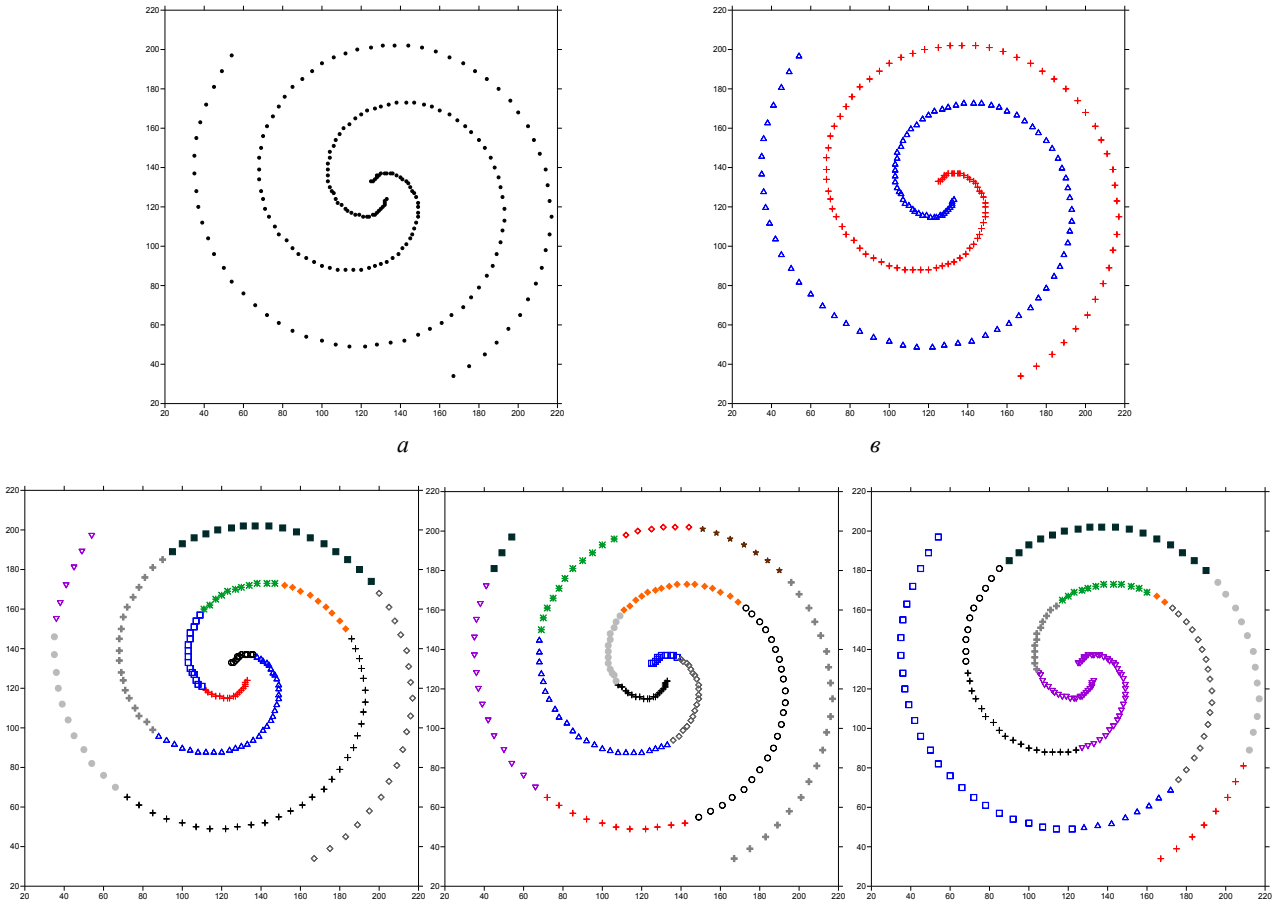


Рис. 2. Двумерные данные, состоящие из 200 точек:
а – исходные данные; *б* – результаты выполнения алгоритма MeanSC (12, 14 и 12 кластеров соответственно)
с параметрами $h \in \{8; 8{,}5; 9\}$, $\varepsilon = 0$, $T = 5{,}5$; *в* – результат выполнения алгоритма EMeanSC
с параметрами $\overline{h} = \{8; 8{,}5; 9\}$, $\varepsilon = 0$, $T = 5{,}5$, $T_d = 0{,}999$ (2 кластера)

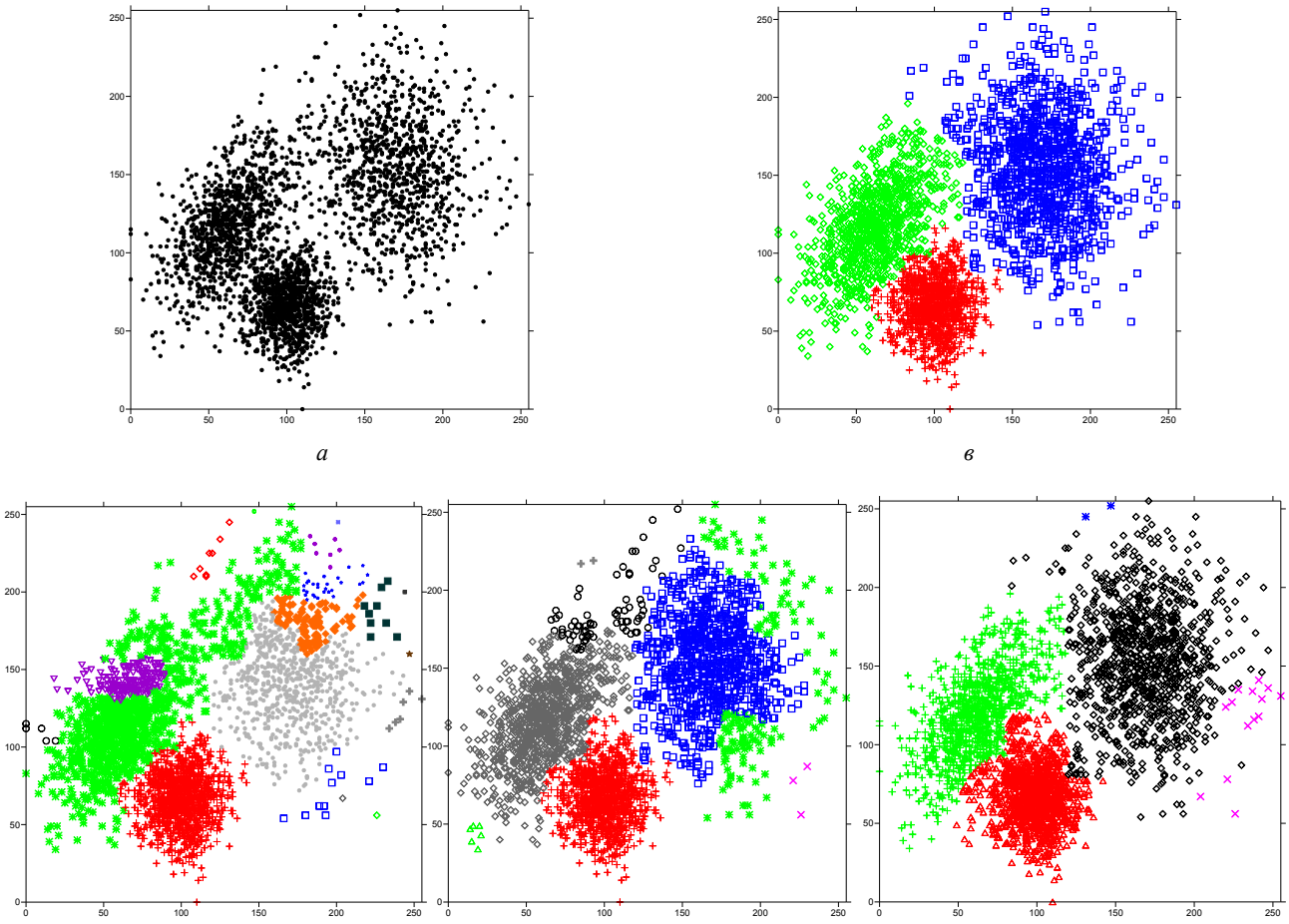
Эксперимент 3. Использовались двумерные данные, состоящие из 3 000 точек (рис. 3), сгруппированных в 3 нормально распределенных класса (рис. 3, *а*). Большинство кластеризаций, на основе которых формировался ансамбль, содержат достаточно грубые ошибки (дробление и пересечение классов) (рис. 3, *б*). При этом результаты выполнения ансамблевого алгоритма (рис. 3, *в*) не содержат грубых ошибок.

Эксперимент 4. Использовался фрагмент снимка Болотнинского района Новосибирской области (размером 500 × 450), полученного со спутника ALOS/ANVIR-2 17 июня 2007 г. (рис. 4). Исследуемый участок ограничен 55°52'14.8" и 55°55'2.13" северной широты и 83°50'45.41" и 83°54'51.6" восточной долготы. Исходный фрагмент представлен на рис. 4, *а*. Обработка выполнялась по трем спектральным каналам. Обработка производилась на ПК с тактовой частотой 2,4 ГГц (объем оперативной памяти 2 Гб). Время обработки с двумя значениями параметра сглаживания составляет 18 с. Алгоритм выделил пять классов (рис. 4, *б*).

Эксперимент 5. Использовалось цветное изображение (рис. 5, *а*) размером 510×604 пикселов. Кластеризация выполнялась в цветовом пространстве $R \times G \times B$. Каждый кластер соответствовал однородной области на изображении. Срезам дендрограммы, построенной в ходе выполнения алгоритма EMeanSC, на уровнях, соответствующих различным значениям параметра T_d, соответствуют рис. 5, *б–г*. Параметр дендрограммы управляет степенью раздробленности получаемых кластеров, что позволяет получить картосхему с необходимым пользователю уровнем детализации.

В работе представлен метод комбинирования ансамблевого и непараметрического подходов для кластеризации изображений, позволяющий повысить качество и устойчивость получаемых результатов. Дано его теоретическое обоснование. В соответствии с этим методом создан алгоритм EMeanSC, основанный на непараметрических оценках плотности Розенблатта–Парзена. Показано, что алгоритм способен разделять кластеры сложной структуры и может быть использован для сегментации многоспектральных изображений.

Заметим, что предложенный метод построения ансамблевых алгоритмов допускает распараллеливание наиболее трудоемких этапов обработки, позволяющее повысить быстродействие при реализации его на многопроцессорных вычислительных системах.



а *в* *б*

Рис. 3. Двумерные данные, состоящие из 3 000 точек:
а – исходные данные; *б* – 3 из 5 элементов ансамбля (19, 8 и 5 кластеров, соответствующих значениям $h \in \{10; 15; 20\}$);
в – результат выполнения EMeanSC с параметрами $\overline{h} = \{5; 10; 15; 20; 25\}$, $\varepsilon = 0$, $T = 2,5$, $T_d = 0,75$ (3 кластера)

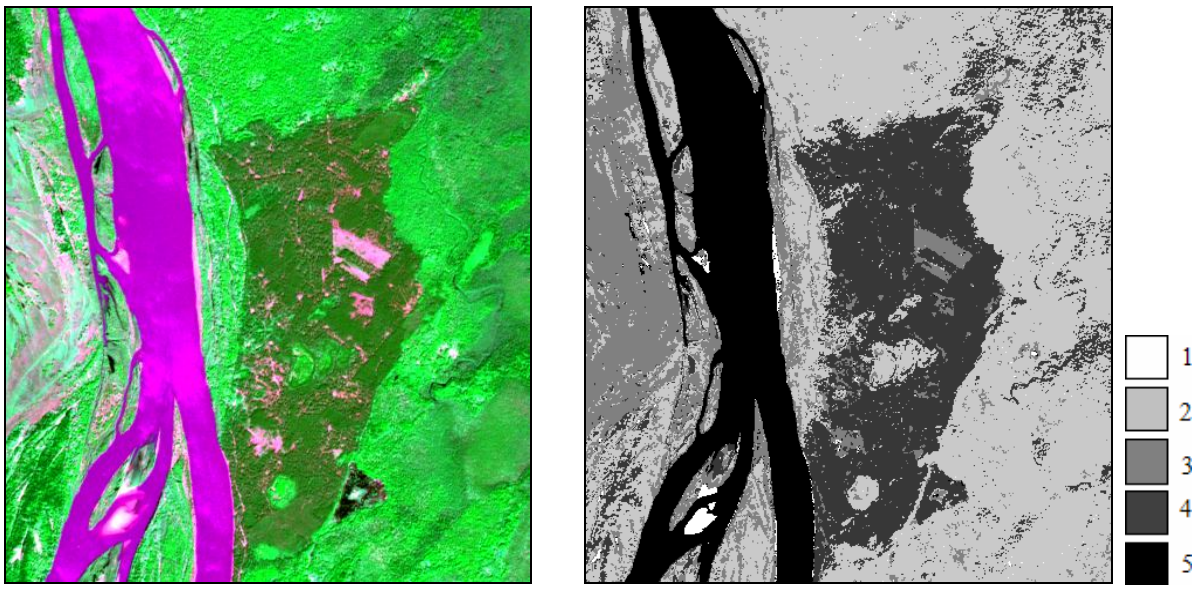
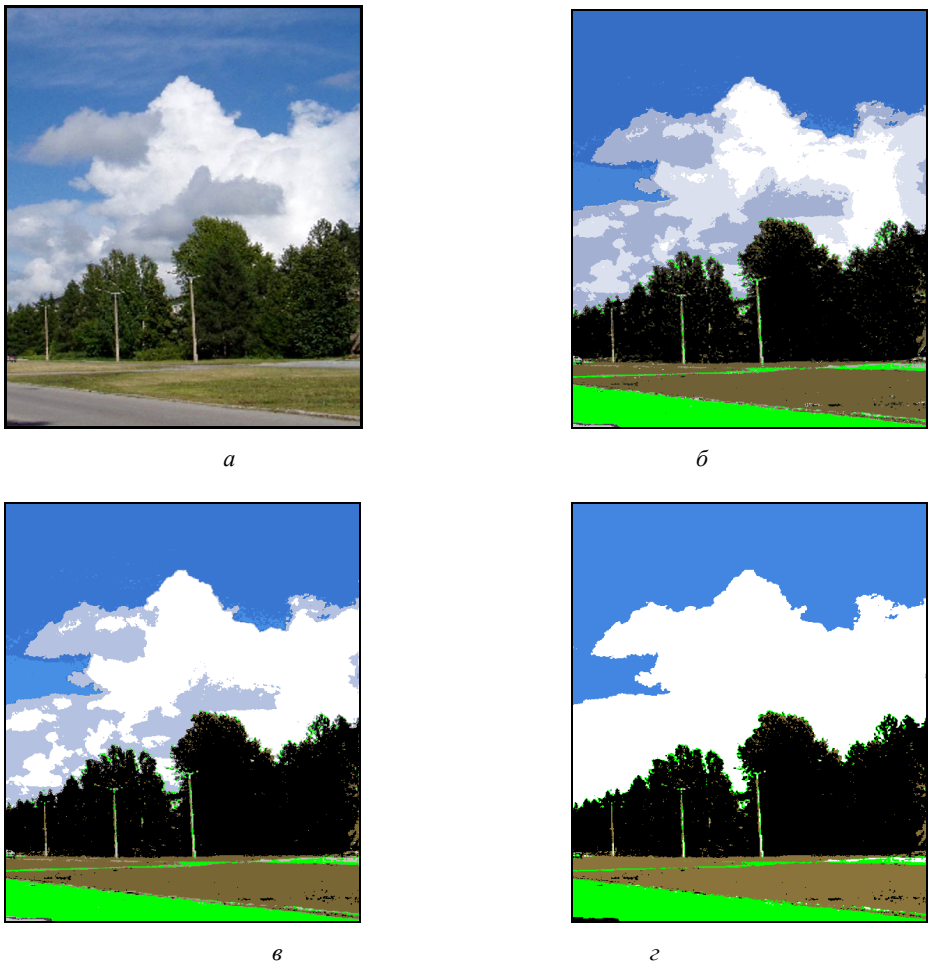


Рис. 4. Фрагмент снимка Болотнинского района Новосибирской области:
а – снимок ALOS; *б* – результаты выполнения EMeanSC с параметрами $\overline{h} = \{5; 10; 15\}$, $\varepsilon = 0$, $T = 1,3$, $T_d = 0,5$.
Выделено 6 кластеров: *1* – песчаные отложения; *2* – березовые и березово-осиновые травяные леса; *3* – луговая и кустарниковая растительность в пойме реки; *4* – сосновые травяные и травяно-кустарничковые леса; *5* – река Обь



а *б*

в *г*

Рис. 5. Изображение:
а – исходное изображение; *б*, *в*, *г* – результаты выполнения EMeanSC с параметрами $\overline{h} = \{5; 7; 9\}$, $\varepsilon = 0$, $T = 1,5$, $T_d = 0,4$ (18 кластеров), 0,6 (12 кластеров) и 0,95 (6 кластеров) соответственно

Библиографические ссылки

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М. : Техносфера, 2006. С. 812.

2. Dey V., Zhang Y., Zhong M. A review on image segmentation techniques with remote sensing perspective // ISPRS TC VII Symposium – 100 Years ISPRS, Vienna. Austria. July 5–7 2010. IAPRS. Vol. XXXVIII. Part 7A. P. 31–42.

3. Rekik A., Zribi M., Hamida A., Benjelloun1 M. Review of satellite image segmentation for an optimal fusion system based on the edge and region approaches // IJCSNS International Journal of Computer Science and Network 242 Security. 2007. Vol. 7. № 10. P. 242–250.

4. Jain A. K., Duin R. P. W., Mao J. Statistical Pattern Recognition: A Review // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000. Vol. 22. № 1. P. 4–37.

5. Clausi D. A. K-means Iterative Fisher (KIF) unsupervised clustering algorithm applied to image texture segmentation // Pattern Recognition. 2002. Vol. 35. № 9. P. 1959–1972.

6. Бериков В. Б. Построение ансамбля деревьев решений в кластерном анализе // Вычислительные технологии. 2010. Т. 15. № 1. С. 40–52.

7. Strehl A., Ghosh J. Clustering ensembles – a knowledge reuse framework for combining multiple partitions // The Journal of Machine Learning Research. 2002. Vol. 38. P. 583–617.

8. Hong Y., Kwong S. To combine steady-state genetic algorithm and ensemble learning for data clustering // Pattern Recognition Letters. 2008. Vol. 29(9). P. 1416–1423.

9. Jain A. K. Data clustering: 50 years beyond K-means // Pattern Recognition Letters. 2010. Vol. 31, Is. 8. P. 651–666.

10. Parzen E. On the estimation of a probability density function and the mode // The Annals of Mathematical Statistics. 1962. Vol. 33. P. 1065–1076.

11. Rosenblatt M. Remarks on some nonparametric estimates of a density function // The Annals of Mathematical Statistics. 1956. Vol. 27. P. 832–837.

12. Пестунов И. А., Синявский Ю. Н. Непараметрический алгоритм кластеризации данных дистанци-

онного зондирования на основе grid-подхода // Автометрия. 2006. Т. 42. № 2. С. 90–99.

13. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее применение. 1969. Т. 14. № 1. С. 156–160.

14. Comaniciu D., Meer P. Mean shift: A Robust Approach toward Feature Space Analysis // IEEE Transactions on Pattern Analysis Machine Intelligence. 2002. Vol. 24. № 5. P. 603–619.

15. Fukunaga K., Hosteeler L. D. The estimation of the gradient of a density function, with applications in patter recognition // IEEE Trasactions on Informational Theory. 1975. Vol. 21. P. 32–40.

16. Cheng Y. Mean shift, mode seeking, and clustering // IEEE Tans. Pattern Analysis and Machine Intelligence. 1995. Vol. 17. P. 790–799.

17. Comaniciu D., Meer P. Distribution Free Decomposition of Multivariate Data // Pattern Analysis and Applications. 1999. Vol. 2. P. 22–30.

18. Freedman D., Kisilev P. Fast Mean Shift by Compact Density Representation // IEEE Conference on Computer Vision and Pattern Recognition. 2009. P. 1818–1825.

I. A. Pestunov, V. B. Berikov, Yu. N. Sinyavskiy

ALGORITHM FOR MULTISPECTRAL IMAGE SEGMENTATION BASED ON ENSEMBLE OF NONPARAMETRIC CLUSTERING ALGORITHMS

The method for constructing an ensemble of nonparametric clustering algorithms is proposed. Its theoretical substantiation is resulted. Results of the model data and real images confirm the efficiency of the proposed method.

Keywords: multispectral image segmentation, nonparametric clustering algorithms, ensemble approach.

© Пестунов И. А., Бериков В. Б., Синявский Ю. Н., 2010

УДК 519.24

А. Н. Сергеев

О НЕПАРАМЕТРИЧЕСКИХ АЛГОРИТМАХ ПРИНЯТИЯ РЕШЕНИЙ

Рассматриваются особенности и параметры, влияющие на деятельность организаций, осуществляется постановка задачи моделирования и управления организационной системой. Приводятся математические непараметрические модели организационных систем.

Ключевые слова: организационные системы, управление, измерения, неопределенность, случайные факторы, непараметрическое моделирование, алгоритмы принятия решений.

В понятии «организационная система» используются одновременно два нетривиальных термина: «организация» и «система».

Организация может рассматриваться как процесс либо как сущность [1]. Как процесс организация – это совокупность действий, ведущих к образованию и совершенствованию взаимосвязей между частями целого. Как сущность организация – это целевое объединение под единым началом ресурсов для реализации определенной программы на основании определенных правил и процедур.

Здесь надо отметить, что некоторые организации могут сами являться ресурсами для более крупных структур, в которые они входят. Одним из наиболее важных ресурсов, несомненно, является человеческий.

Термин «система» имеет множество вариантов определений в разной литературе. Рассел Л. Акофф [1] попытался сформулировать «ядро» определения: «Система есть целое, состоящее из двух или более частей, которое удовлетворяет следующим 6 условиям:

– целое обладает одним или более определяющими свойствами или функциями;

– каждая часть в этом множестве может влиять на поведение или свойства целого;

– существует подмножество частей, которое достаточно в одном или нескольких внешних условиях для выполнения определяющей функции целого;

– способ, которым любая существенная часть воздействует на поведение или свойства системы, зависит от поведения или свойств по крайнем мере одной другой существенной части системы;

– воздействие любого подмножества существенных частей на систему в целом зависит от поведения по крайней мере еще одного другого такого подмножества;

– система есть целое, которое не может быть разделено на независимые части без потери ее существенных свойств или функций».

Акофф вводит для ресурса два различаемых свойства-термина: «целенаправленный» и «целеориентированный».

«Некая сущность является целеустремленной, если она может выбирать цели и средства в двух или более разных обстоятельствах» [1]. Если же сущность, имея

способность выбирать средства и способы, производит только один результат в любой среде, то она является целенаправленной. Уже оттолкнувшись от этих терминов, производится типология систем по свойствам своих частей и целого (см. таблицу).

Организационная система – социальная система, это чрезвычайно сложный «организм», который состоит из целеустремленных частей и заставляет искать принципиально новые методы управления взамен используемых в 1960–80-е и даже 1990-е гг. XX в.

И ключевым отличием в новых методах организационного управления должен стать акцент на управлении отношениями частей системы друг с другом, с внешним окружением и самой системы с другими системами. Внешняя среда включает многие условия и организации, в том числе политические, экономические и экологические условия, конкурирующие организации, поставщиков и потребителей, социальную инфраструктуру и т. д.

В широком смысле, отношения – взаимодействие или противодействие между частями системы внутри и вне ее, при создании, функционировании, развитии и разрушении самой системы. И когда мы начинаем говорить об организациях, отношения бывают трех уровней – здравого смысла, взаимного уничтожения, заранее спроектированного взаимодействия.

Основные свойства организационных систем следующие: стремление сохранить свою структуру (закон самосохранения); потребность в управлении; сложная зависимость от свойств входящих в нее элементов и подсистем (система может обладать свойствами, не присущими ее элементам, и может не иметь свойств этих элементов) [1].

Каждая система имеет входное воздействие (необходимо разделять на управляемые и неуправляемые), влияние случайных помех, технологию их обработки, конечные результаты и обратную связь. Изначальная постановка задач идентификации и моделирования организационных процессов будет опираться на достижения и результаты классической теории идентификации в узком и широком смыслах [2].

Общая схема исследуемого процесса, принятая в теории моделирования и идентификации, приведена на рис. 1 [3].

На рис. 1 приняты следующие обозначения: A – неизвестный оператор процесса; $x(t)$ – выходная переменная процесса; $u(t)$ – управляющее воздействие; $\mu(t)$ – входная переменная процесса; $\xi(t)$ – случайное воздействие; t – дискретное время; h^{μ}, h^{u}, h^{x}– ошибки измерения показателей, соответствующие различным переменным (индекс); μ_t, u_t, x_t – означают измерение $\mu(t)$, $u(t)$, $x(t)$ в дискретное время t. Используемые при этом каналы измерения включают в себя различные средства и методы контроля, учета и измерения.

Контроль переменных x, u, μ осуществляется через интервал времени Δt, т. е. (x_i, u_i, μ_i), $i = \overline{i,s}$ – выборка измерений переменных процесса (x_1, u_1, μ_1), (x_2, u_2, μ_2), …, (x_s, u_s, μ_s), …, s – объем выборки. В этом случае $x(t)$ определяется следующим образом:

$$x(t) = A\big(u(t), \mu(t), \xi(t), t\big). \qquad (1)$$

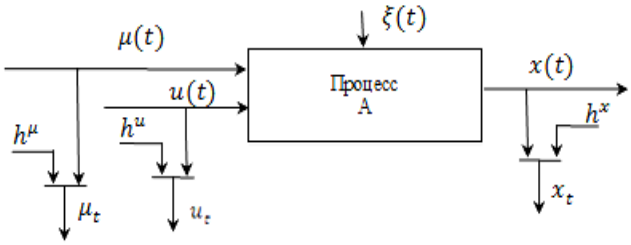


Рис. 1. Классическая схема стохастического процесса и контроля переменных

Заметим, что выход объекта $x(t)$ зависит от всех входных переменных $\mu(t)$, $u(t)$, $\xi(t)$ и оператор $A(\cdot)$ не зависит от t, т. е. неизменен во времени.

Искомая же модель процесса (1) будет определяться как

$$\hat{x}(t) = B_s\left(u(t), \mu(t), \vec{x}_s, \vec{u}_s, \vec{\mu}_s\right), \qquad (2)$$

где B – оператор, поиск и настройка которого будет являться основной задачей исследователя; $\vec{x}_s = (x_1, x_2, ..., x_s)$, $\vec{u}_s = (u_1, u_2, ..., u_s)$, $\vec{\mu}_s = (\mu_1, \mu_2, ..., \mu_s)$ – временные векторы значений переменных; s – объем выборки. Влияние переменной $\xi(t)$ учитывается через измерения фактических значений переменных x, u, μ: $\vec{x}_s$, $\vec{u}_s$, $\vec{\mu}_s$.

Решая задачи идентификации в широком смысле, исследователь сталкивается с недостатком или отсутствием информации о параметрической структуре исследуемого процесса. В таких условиях могут применяться методы непараметрической идентификации, используя которые модель (2) может быть записана в виде [3]

$$\hat{x} = \frac{\sum_{i=1}^{s} x_i \Phi\left(\frac{u-u_i}{c_s}\right)\Phi\left(\frac{\mu-\mu_i}{c_s}\right)}{\sum_{i=1}^{s} \Phi\left(\frac{u-u_i}{c_s}\right)\Phi\left(\frac{\mu-\mu_i}{c_s}\right)}. \qquad (3)$$

Используя некий обобщенный аргумент y, можем записать условия, которым должна удовлетворять ядерная функция («ядро») $\Phi(y)$: ограниченность $0 \le \Phi(y) < \text{const}$; четность $\Phi(y) = \Phi(-y)$; $\int_{-\infty}^{\infty} \Phi(y)dy = 1$.

Типы систем и моделей

Системы и модели	Части системы	Целое
Детерминированные	Нецелеустремленные	Нецелеустремленные
Живые	Нецелеустремленные	Целеустремленные
Социальные	Целеустремленные	Целеустремленные
Экологические	Целеустремленные	Нецелеустремленные

Параметр размытости c_s для модели (3) обладает свойствами: $c_s \to 0$ и $sc_s^2 \to \infty$ при $s \to \infty$ (степень соответствует числу переменных).

Если рассматривать управление организационными системами, то имеет смысл говорить, как минимум, о двух подсистемах (составляющих):

– техническая подсистема – включает оборудование, технологии и изделия, имеющие инструкции для пользователя; набор решений в технической подсистеме ограничен и последствия всех решений, как правило, предопределены; решения носят строго формализованный характер и выполняются в строго определенном порядке;

– социальная подсистема характеризуется наличием человека как субъекта и объекта управления в совокупности взаимосвязанных элементов.

Если процессы контроля и управления технической подсистемой уже достаточно изучены и здесь достигнуты определенные успехи, то управление социальной подсистемой до сих пор представляет определенную сложность. Одной из неприятных особенностей действия социальной подсистемы можно считать появление неизмеримых входных воздействий и специфику поведения самой системы (человеческий фактор).

Говоря об организационных системах, нельзя не вспомнить об их структурах, которые отражают специфику отношений внутри предприятия и стиль управления. По типу представления, разделения функций и ответственности среди организационных структур выделяют *линейные*, *функциональные*, *матричные*, *проектные*, *дивизиональные* [4].

Также стали рассматривать многомерные структуры, выделяя в разные «измерения» ориентации: функциональную, продуктовую (сервисную) и рыночную [1]. Сложнее представляется задача управления социальной подсистемой в многомерном пространстве.

Тип организационной структуры, используемой в отдельно взятой организации, предопределяет принципы и специфику процессов взаимодействия. Существенное влияние на характер организационной системы оказывает соотношение горизонтальных и вертикальных связей.

Горизонтальные связи – это взаимодействие относительно равнозначных частей организационной системы (сотрудников, подразделений, групп) друг с другом напрямую, минуя привлечение вышестоящих руководителей (не используя эскалацию).

Вертикальные связи – это эскалация проблемы (задачи, требования, просьбы и т. д.) по вертикали подчинения и получение решения (ответа, отказа, указания, инструкции …) от вышестоящего руководства.

От того, чего больше (какого вида связей) в организации, будет зависеть оперативность, гибкость, качество, своевременность, адекватность осуществляемых действий в конкретных обстоятельствах. Однако однозначно ответить, чего должно быть больше или меньше, каким должно быть оптимальное соотношение в той или иной организационной системе, весьма непросто.

Напрямую преобладание горизонтальных или вертикальных связей зависит от используемого в организационной системе стиля управления [5]. Среди всех стилей управления принято выделять: *директивный* (*авторитарный*), *демократический* (*коллегиальный*), *либеральный* (*попустительский*). Каждый из них хорош по-своему при конкретных обстоятельствах, в которых оказывается организация. Однако вне зависимости от используемого стиля управления, руководитель организации вынужден принимать решения для достижения заданных показателей (требуемого выхода) в условиях имеющихся ресурсов (входные параметры), информации и оказывающих влияние на процессы внутренних и внешних воздействий. Алгоритм, по которому действует конкретное лицо, принимающее решение (ЛПР), как минимум, сложно формализуем. Кроме того, алгоритм действий зачастую зависит от самого ЛПР как субъекта, что и предопределяет успешность или неуспешность принимаемого решения.

Само по себе наличие человека (коллектива) в организационной системе существенно усложняет задачу управления, так как способность человека принимать самостоятельные решения и видоизменять устоявшиеся процессы заставляет ЛПР оперативно реагировать на любые отклонения. В этих условиях решение задачи идентификации организационного процесса не может полностью исключить ЛПР из процесса управления. Можно говорить о построении модели, которая формализует процесс управления организационной системой, насколько это возможно без ущерба для показателей эффективности системы. Такая настроенная и «обученная» модель предназначается быть помощником ЛПР, снижая, тем самым, требования к самому ЛПР.

Развивая классическую постановку задач идентификации [2; 3], следует учитывать и влияние внешнего окружения организации (государство, конкуренты, партнеры, клиенты) и их обратной связи. Тогда в такой постановке задач неизбежно увеличение как количества переменных на входе, так и необходимый рост числа каналов измерения показателей. И в этих условиях схема исследуемого процесса в организации примет более общий расширенный (рис. 2) (стрелки внутри объекта символизируют наличие человеческого фактора, социальных связей).

На рис. 2, в дополнение к схеме на рис. 1, вводятся новые обозначения:

– входная переменная $\omega(t)$ существенно отличается по своему содержанию от переменных $\mu(t)$, $u(t)$ тем, что эта переменная известна, но не поддается регулярному измерению в соответствии, как правило, со сложной и длительной процедурой;

– переменная $\theta(t)$ представляет собой воздействие внешней среды на объект; θ_t – измерение переменной; h^θ – ошибки измерения переменной;

– $\lambda^j(t): j = 1, 2, \ldots, \Lambda$ – дополнительные измеряемые переменные «по длине» процесса, контролируемые на разных участках в разные кванты времени; λ_t^j означает измерение $\lambda^j(t)$ в дискретное время и представляет собой дополнительную информацию о процессе; h^{λ^j} – ошибки измерения этих переменных при использовании различных средств и методов контроля.

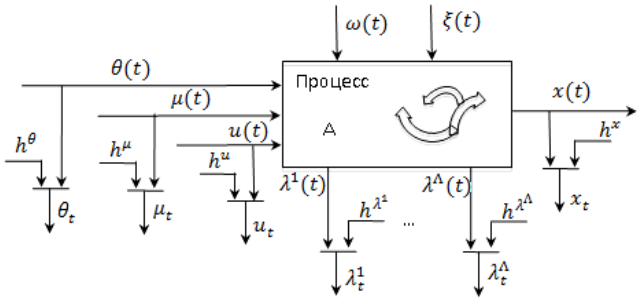


Рис. 2. Схема организационного процесса

Контроль переменных $(x, \lambda^1, ..., \lambda^\Lambda, u, \mu, \theta)$ осуществляется через интервал времени Δt, т. е. $x_i, \lambda_i^1, ..., \lambda_i^\Lambda, u_i, \mu_i, \theta_i$, $i = \overline{i, s}$ – выборка измерений переменных процесса $(x_1, \lambda_1^1, ..., \lambda_1^\Lambda, u_1, \mu_1, \theta_1)$, $(x_2, \lambda_2^1, ..., \lambda_2^\Lambda, u_2, \mu_2, \theta_2)$, …, $(x_s, \lambda_s^1, ..., \lambda_s^\Lambda, u_s, \mu_s, \theta_s)$, …, s – объем выборки. В этом случае $x(t)$ определяется не только значениями входных переменных, как в выражении (1), но и $\omega(t)$, $\theta(t)$ и может быть представлена в следующем виде:

$$x(t) = A\big(u(t), \mu(t), \theta(t), \omega(t), \xi(t), t\big). \qquad (4)$$

Моделирование процессов подобного типа существенно усложняется в зависимости от характера влияния $\omega(t)$ и $\theta(t)$ на $x(t)$.

С повышением уровня образованности человека, развития технологий, ростом объема доступной информации и стирания границ и барьеров для понимания происходит развитие атмосферы конфликтности. Учитывая такую тенденцию, эксперты призывают применять *социально-системные приемы* моделирования организаций [1].

Организация, основанная на социально-системной модели, обладает следующими свойствами:

– наличие демократических принципов управления, когда каждый голос влияет на принятие решения, а каждый руководитель подвластен мнению большинства (никакая власть не является окончательной);

– наличие в организации внутренней рыночной экономики, когда каждый вправе сделать выбор при покупке или продаже требуемой услуги внутри или вовне организации; в случае запрета вышестоящим начальством каждый вправе рассчитывать на компенсацию понесенных при этом затрат;

– многомерная организационная структура, объединяющая функциональные, продуктовые и рыночные (пользовательские) блоки в разных измерениях;

– интерактивное планирование, приспосабливающееся к доступным ресурсам;

– содержит систему поддержки решений, обеспечивающую обучение и адаптацию организации к внутренним и внешним изменениям.

Каждый из вышеприведенных пунктов способен улучшить работу организации, а использование их в комплексе может повлечь мультипликативный эффект [1].

Наиболее естественный, часто и достаточно эффективный, путь моделирования организационных процессов состоит в предварительной декомпозиции. Причем, выявление фрагментов исследуемого процесса осуществляется исходя из возможностей его математической формализации. Те же фрагменты, которые не поддаются последней, «отдаются на откуп» ЛПР. Следует иметь в виду, что переход от содержательной характеристики задачи к ее математической формализации часто чрезвычайно труден, а сам этот процесс не формализуем. Именно на этой стадии могут быть, из различных соображений, сделаны допущения (а это совершенно необходимо для математической постановки задачи), которые могут оказаться совершенно не адекватными существу исследуемого процесса. В последнем случае исследователь решает задачу, имеющую отдаленное отношение к действительности [6].

Таким образом, при математической формулировке задач крайне важно охватить реалистические черты исследуемого процесса моделирования и управления. Это означает, что математическая формализация должна действительно охватывать те черты исследуемого процесса, которые фактически имеют место. Иными словами, крайне важно ничего не упрощать при формулировке задачи. В этом контексте для достижения адекватности построенной модели реально действующим процессам (4) в организации важно учитывать при моделировании переменные $\omega(t)$, $\theta(t)$, $\lambda^j(t)$, $j = 1, 2, \ldots, \Lambda$. В силу специфики (неизмеряемости) переменных $\omega(t)$, $\xi(t)$ их влияние в модели может присутствовать через измерения других переменных, включая «выход» $x(t)$. Такая модель при постановке задачи идентификации в широком смысле будет выглядеть следующим образом:

$$\begin{aligned} \hat{x}(t) = B_s(u(t), \mu(t), \lambda^1(t), ..., \lambda^\Lambda(t), \\ \theta(t), t, \vec{x}_s, \vec{u}_s, \vec{\mu}_s, \vec{\lambda}_s^1, ..., \vec{\lambda}_s^\Lambda, \vec{\theta}_s), \end{aligned} \qquad (5)$$

где $\vec{x}_s = (x_1, x_2, ..., x_s)$, $\vec{u}_s = (u_1, u_2, ..., u_s)$, $\vec{\mu}_s = (\mu_1, \mu_2, ..., \mu_s)$, $\vec{\theta}_s = (\theta_1, \theta_2, ..., \theta_s)$, $\vec{\lambda}_s^j = (\lambda_1^j, \lambda_2^j, ..., \lambda_s^j)$, $j = \overline{1, \Lambda}$ – временные векторы значений переменных; s – объем выборки; B – оператор, поиск и настройка которого предопределяют уровень соответствия построенной модели моделируемому процессу.

С использованием методов непараметрической идентификации модель (5) принимает следующий вид (для упрощения восприятия приводится без учета переменных $\lambda^j(t),\ \ j=\overline{1,\Lambda}$):

$$\hat{x}=\frac{\sum\limits_{i=1}^{s}x_i\prod\limits_{j=1}^{J}\Phi\left(\frac{u^j-u_i^j}{c_s^j}\right)\prod\limits_{k=1}^{K}\Phi\left(\frac{\mu^k-\mu_i^k}{c_s^k}\right)\prod\limits_{l=1}^{L}\Phi\left(\frac{\theta^l-\theta_i^l}{c_s^l}\right)}{\sum\limits_{i=1}^{s}\prod\limits_{j=1}^{J}\Phi\left(\frac{u^j-u_i^j}{c_s^j}\right)\prod\limits_{k=1}^{K}\Phi\left(\frac{\mu^k-\mu_i^k}{c_s^k}\right)\prod\limits_{l=1}^{L}\Phi\left(\frac{\theta^l-\theta_i^l}{c_s^l}\right)}. \quad (6)$$

Здесь обозначения и требования те же, что и в формуле (3), и для обобщенной векторной переменной $y=(y_1,y_2,...,y_I)$: $\int\limits_{-\infty}^{\infty}\prod\limits_{i=1}^{I}\Phi(y_i)dy_i=\text{const}$. Требования же к параметру размытости c_s в случае моделирования организационных процессов могут не выполняться. Процесс подбора наилучших параметров размытости является отдельным предметом для исследований. Учет в модели переменных $\lambda^j(t)$, $j=\overline{1,\Lambda}$ осуществляется аналогичным образом, путем введения дополнительных произведений ядерных функций по этим переменным.

На практике при исследовании различных непрерывных и дискретно-непрерывных процессов, в зависимости от требований точности, в качестве ядерной функции могут использоваться различные варианты. Например, для управляющего воздействия *u* при величине выборки *s* это могут быть такие зависимости, как

$$\Phi\left(\frac{u-u_i}{c_s}\right)=\begin{cases}\frac{1}{2};\left|\frac{u-u_i}{c_s}\right|\le 1;\\ 0;\ \left|\frac{u-u_i}{c_s}\right|>1;\end{cases}$$

$$\Phi\left(\frac{u-u_i}{c_s}\right)=\begin{cases}1-\left|\frac{u-u_i}{c_s}\right|;\left|\frac{u-u_i}{c_s}\right|\le 1;\\ 0;\ \left|\frac{u-u_i}{c_s}\right|>1,\end{cases}$$

или другие более сглаженные «колоколообразные» функции.

При нахождении всех требуемых параметров и настройке модели вида (6) поиск необходимого управляющего воздействия для получения заданного векторного выхода *x* используется преобразованная зависимость следующего вида:

$$\hat{u}=\frac{\sum\limits_{i=1}^{s}u_i\prod\limits_{n=1}^{N}\Phi\left(\frac{x^n-x_i^n}{c_s^n}\right)\prod\limits_{k=1}^{K}\Phi\left(\frac{\mu^k-\mu_i^k}{c_s^k}\right)\prod\limits_{l=1}^{L}\Phi\left(\frac{\theta^l-\theta_i^l}{c_s^l}\right)}{\sum\limits_{i=1}^{s}\prod\limits_{n=1}^{N}\Phi\left(\frac{x^n-x_i^n}{c_s^n}\right)\prod\limits_{k=1}^{K}\Phi\left(\frac{\mu^k-\mu_i^k}{c_s^k}\right)\prod\limits_{l=1}^{L}\Phi\left(\frac{\theta^l-\theta_i^l}{c_s^l}\right)}. \quad (7)$$

Включение в модель измеряемых переменных $\lambda^j(t),\ \ j=\overline{1,K}$ осуществляется аналогичным образом – введением дополнительных произведений ядерных функций по этим переменным.

Попытаемся построить непараметрическую модель для конкретного организационного процесса. Возьмем, например, в качестве процесса *A* – управление конкретным проектом по оснащению некоего объекта инфраструктурным оснащением инженерными системами. В таком случае будем иметь дело с организационным процессом управления деятельностью проектной группы, состоящей из специалистов разного назначения и выполняющих различные специальные функции. Здесь можно выделить руководителя проекта, главного инженера проекта, отдельных инженерно-технических специалистов, производственный и монтажный персонал, специалистов логистики и снабжения, проектно-сметных специалистов, бухгалтеров и, возможно, других. В зависимости от подхода исследователя, всех указанных участников можно считать непосредственными членами проектной группы или, например, бухгалтеров выделять в так называемое «внешнее окружение», наряду с поставщиками, производителями, транспортными организациями и др.

Для данного примера параметры и переменные зависимости (4) могут приобретать следующие смысловые значения:

– *A* – неизвестный оператор процесса управления проектом;

– $x(t)$ – вектор, включающий $x^1(t)$ – параметры выпускаемого продукта, измеряемые уровнем соответствия решения рабочей проектной документации в %; $x^2(t)$ – соблюдение сроков реализации решений; $x^3(t)$ – исполняемость бюджета проекта; …; $x^N(t)$ – удовлетворенность клиента;

– $u(t)$ – векторное управляющее воздействие: $u^1(t)$ – степень укомплектованности проектной группы необходимыми специалистами; $u^2(t)$ – своевременность поставок материалов на объект; …; $u^J(t)$ – проводимые планерки и штабы для членов проектной группы;

– $\mu(t)$ – неуправляемые входные переменные процесса: $\mu^1(t)$ – квалификация членов группы; $\mu^2(t)$ – выделенные бюджеты проекта; … ; $\mu^K(t)$ – строительная готовность участков объекта;

– $\xi(t)$ – случайное воздействие, форс-мажорные обстоятельства: поломка инструмента, болезнь ключевого специалиста, взлом и ограбление оперативного склада и др.;

– h^μ, h^u, h^x – ошибки измерения показателей в различных каналах (аттестации, оценка износа оборудования и инструмента, учетные системы, медицинские освидетельствования персонала, рыночный мониторинг, отчетность); используемые при этом каналы включают в себя различные средства и методы контроля, учета и измерения (позволяют оценить достаточность квалификации и численности сотрудников, инструмента, оптимальность закупочных цен, сроков поставки, финансового соответствия поставленной задаче);

– $\lambda^j(t), \quad j=\overline{1,2,...,\Lambda}$ – измеряемые ключевые опережающие показатели деятельности проектной группы, контролируемые на разных участках деятельности (по количеству Λ в разные кванты времени (работоспособность инструмента, результаты экономии или издержек, исполнение нормативных показателей, фиксация скрытых работ, отклонения от графика и т. д.);

– $\omega(t)$ – вектор сложноизмеримых показателей: организационная культура в проектной группе, лояльность сотрудников к руководству и компании, соответствие норм оплаты труда среднерыночным показателям;

– $\theta(t)$ – измеряет воздействие на деятельность группы со стороны внешней среды: $\theta^1(t)$ – время, уходящее на планерки, проводимые высшим руководством; $\theta^2(t)$ – сдвиги сроков поставки транспортными компаниями, задержка на таможне; $\theta^3(t)$ – дополнительное время на устранение аварий из-за действий смежников на объекте; … ; $\theta^L(t)$ – лояльность заказчика к компании.

В данном случае N, J, K, L – размерности векторов соответствующих переменных $x(t)$, $u(t)$, $\mu(t)$, $\theta(t)$, измеряющихся и контролируемых при управлении процессом.

Обратим здесь внимание на важнейшее обстоятельство, возникающее при исследовании реальных процессов. Это средства контроля, измерения всех доступных переменных, характеризующих состояние процесса. Здесь следует отметить, что если переменные процесса доступны измерению техническими средствами, то это, чаще всего, позволяет установить желаемый интервал (дискретность) контроля. Некоторые же переменные, приобретая специфические семантические значения, не могут измеряться стандартными способами. И это накладывает дополнительные усложняющие ограничения на процессы моделирования подобных систем. На практике может использоваться введение специальных шкал измерения и приведение показателей к безразмерным величинам.

Выбранный для моделирования процесс совершенно отчетливо демонстрирует специфику некоторых переменных организационной системы как в части измерения и оценивания, так и в части использования весовых коэффициентов. Привычные требования теории непараметрической идентификации к ядерной функции, и в особенности к параметрам размытости, рушатся и трансформируются в несколько иную семантику и ограничения [3].

Так, например, нарушение своевременности поставок $u^2(t)$ для одного отрезка выполнения проекта недопустимо более чем на 4 часа, для другого может оказаться некритичным вплоть до нескольких дней. Таким образом, требования сходимости к параметру c_s оказываются неприменимыми. И допустимые отклонения в значениях некоторых переменных процесса зачастую могут быть определены только субъективно ЛПР.

Вообще говоря, в случае управления организационными процессами корректнее будет говорить об управляющем воздействии u^r на очередном r-м такте принятия решений. Предлагается ввести новое обозначение взамен c_s^{-1}. Пусть теперь это будет v^r, и этот коэффициент будет являться весовым для переменных процесса на r-м управленческом такте. Его значение определяется исследователем на основании опыта экспертов, накопленных знаний о процессе и соответствия конкретному такту управления процессом с учетом конкретных ситуаций и складывающихся условий.

Тогда и ядерная функция для обобщенной переменной y будет существенно отличаться от привычной $\Phi(y)$. Введем специально для моделирования организационных процессов обозначение ядерной функции $\Psi(y)$ с выполнением всех тех же условий, что и для $\Phi(y)$: ограниченность $0 \le \Psi(y) < \text{const}$; четность $\Psi(y) = \Psi(-y)$; $\int\limits_{-\infty}^{\infty} \Psi(y)dy = \text{const}$.

При таких условиях для организационного процесса модель (6) примет вид

$$x^r = \frac{\sum\limits_{i=1}^{s} x_i \prod\limits_{j=1}^{J} \Psi\left(v^r\left(u^{rj}-u_i^j\right)\right)\prod\limits_{k=1}^{K}\Psi\left(v^r\left(\mu^{rk}-\mu_i^k\right)\right)\times \prod\limits_{l=1}^{L}\Psi\left(v^r\left(\theta^{rl}-\theta_i^l\right)\right)}{\sum\limits_{i=1}^{s} \prod\limits_{j=1}^{J} \Psi\left(v^r\left(u^{rj}-u_i^j\right)\right)\prod\limits_{k=1}^{K}\Psi\left(v^r\left(\mu^{rk}-\mu_i^k\right)\right)\times \prod\limits_{l=1}^{L}\Psi\left(v^r\left(\theta^{rl}-\theta_i^l\right)\right)}, \quad (8)$$

где все переменные те же, что обозначены на рис. 2, *а* $\lambda^j(t), \quad j=\overline{1,2,...,\Lambda}$ не используется только ради упрощения восприятия формулы.

В рассматриваемом процессе можно выделить результаты (выходы), получаемые немедленно и доступные к измерению в короткие промежутки времени, например, $x^1(t)$, $x^3(t)$. Будем называть такие результаты внутренними. Другие выходные переменные, назовем их внешними, не всегда могут измеряться через кванты времени Δt, а только через ΔT $(\Delta T >> \Delta t)$, например, $x^N(t)$. Можно заметить и то, что $\omega(t)$, $\theta(t)$ влияют на внутренние выходные переменные в меньшей степени, чем на внешние. Аналогично можно разделить на внутренние и внешние управляющие воздействия. Тогда приходим к пониманию о неизбежности появления внешнего контура управления (хотя бы одного).

Изучаемый процесс A без внешних выходных переменных, влияющих и управляющих воздействий покажем на рис. 3.

На рис. 3 вектор $x(t)$ – внутренняя выходная переменная; $u(t)$ – внутреннее же управляющее воздейст-

вие на процесс A во времени t, остальные переменные те же; УУ – устройство управления, выдающее воздействие на процесс; $x^*(t)$ – желаемые параметры выхода.

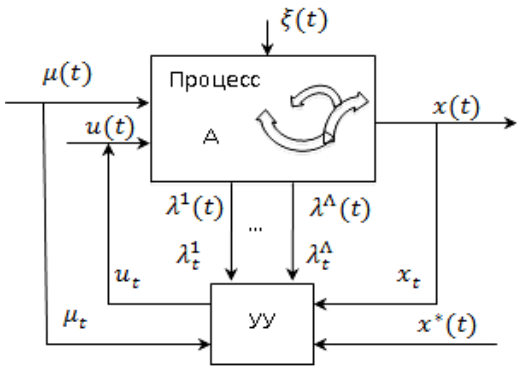


Рис. 3. Схема локального организационного процесса

Каналы и ошибки измерений не показаны на схеме только для облегчения иллюстраций и упрощения их восприятия. Однако будем помнить о них и учитывать их наличие в любых организационных процессах. Тогда в таких условиях и для таких процессов зависимость $x(t)$ будет выглядеть следующим образом:

$$x(t) = A(u(t), \mu(t), \xi(t), t). \tag{9}$$

Моделируя эту функцию выхода $x(t)$ локального организационного процесса, для более компактного представления модели будем использовать следующее написание:

$$x^r = \sum_{i=1}^{s} \alpha_i^r x_i \Big/ \sum_{i=1}^{s} \alpha_i^r, \tag{10}$$

где r – соответствие параметров модели r-му такту принятия решения; s – объем имеющейся выборки значений переменных; коэффициент α_i^r имеет следующий вид:

$$\alpha_i^r = \prod_{j=1}^{J} \Psi\left(v^r\left(u^{rj} - u_i^j\right)\right) \prod_{k=1}^{K} \Psi\left(v^r\left(\mu^{rk} - \mu_i^k\right)\right) \times \times \prod_{g=1}^{\Lambda} \prod_{l=1}^{L} \Psi\left(v^r\left(\lambda^{glr} - \lambda_i^{gl}\right)\right). \tag{11}$$

Далее в качестве объекта управления будем рассматривать весь процесс A вместе со всеми переменными, управляющим устройством, генерируемым им управляющим воздействием u_t и реакцией в виде выходного продукта $x(t)$. Назовем его макрообъектом, или макропроцессом. На него уже начинает оказывать существенное влияние внешнее окружение $\theta(t)$ и сложно измеримые переменные $\omega(t)$. Результатом функционирования такого макрообъекта дополнительно становятся внешние выходные переменные $z(t)$, а для управления ими создается дополнительный контур управления с внешними управляющими воздействиями $q(t)$ (рис. 4).

Напомним, что выходная векторная переменная $z(t)$ измеряется через существенно бóльшие промежутки времени и имеет отношение к обратному воздействию на внешнюю среду, желаемые параметры внешнего выхода обозначены $z^*(t)$. Переменная $q(t)$ оказывает воздействие на всю систему вместе с реакцией на управляющее воздействие $u(t)$, вообще говоря, с некоторым запаздыванием, т. е. во времени $t + \tau$, где τ – величина запаздывания.

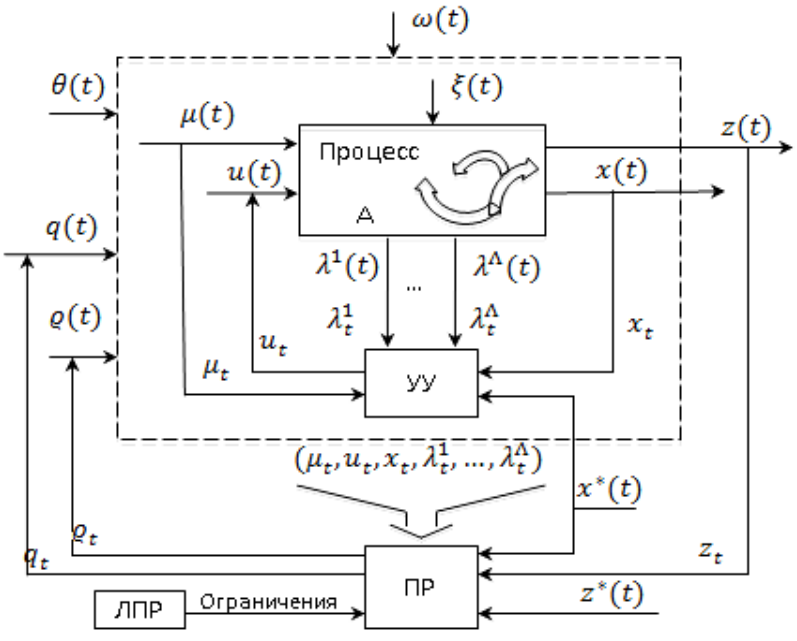


Рис. 4. Схема организационного процесса с двумя контурами управления

Отметим появление на рис. 4 еще одной переменной $\rho(t)$ («ро» малое). Это «корректирующее» управляющее воздействие на УУ с некоторым запаздыванием во времени. Тут же появляется блок принятия решений (ПР), который принимает в качестве входящих все параметры процесса A вместе с результатами воздействия управляющего устройства УУ.

Макропроцесс, представленный на рис. 4, может быть выражен следующей зависимостью:

$$z(t) = D(u(t), \mu(t), \xi(t), x(t), \theta(t), \omega(t), q(t), \rho(t), t), \quad (12)$$

где D – некий неизвестный оператор макропроцесса.

Тогда, используя для компактного написания коэффициенты α_i^r (см. модель (11)), модель функции выхода макропроцесса (12) примет вид

$$z^r = \frac{\sum_{i=1}^{s} z_i \alpha_i^r \prod_{j=1}^{Q} \Psi\left(v^r\left(q^{rj} - q_i^j\right)\right) \prod_{k=1}^{\mathrm{P}} \Psi\left(v^r\left(\rho^{rk} - \rho_i^k\right)\right) \times \prod_{n=1}^{N} \Psi\left(v^r\left(x^{rn} - x_i^n\right)\right)}{\sum_{i=1}^{s} \alpha_i^r \prod_{j=1}^{Q} \Psi\left(v^r\left(q^{rj} - q_i^j\right)\right) \prod_{k=1}^{\mathrm{P}} \Psi\left(v^r\left(\rho^{rk} - \rho_i^k\right)\right) \times \prod_{n=1}^{N} \Psi\left(v^r\left(x^{rn} - x_i^n\right)\right)}, \quad (13)$$

где все обозначения те же, Q, P («ро» большое) – размерности векторов $q(t)$, $\rho(t)$ («ро» малое), соответственно. Введя еще одно обозначение коэффициента β_i^r:

$$\beta_i^r = \prod_{j=1}^{Q} \Psi\left(v^r\left(q^{rj} - q_i^j\right)\right) \prod_{k=1}^{\mathrm{P}} \Psi\left(v^r\left(\rho^{rk} - \rho_i^k\right)\right) \times \prod_{n=1}^{N} \Psi\left(v^r\left(x^{rn} - x_i^n\right)\right), \quad (14)$$

можем записать модель макропроцесса, изображенного на рис. 4, в следующем виде:

$$z^r = \sum_{i=1}^{s} z_i \alpha_i^r \beta_i^r \Big/ \sum_{i=1}^{s} \alpha_i^r \beta_i^r . \quad (15)$$

При управлении организационным процессом на определенном такте принятия решения не все переменные могут влиять на процесс и использоваться не все управляющие воздействия, а лишь часть из них. Например, после завершения поставок всех необходимых материалов для выполнения работ по проекту учитывать переменную $u^1(t)$ своевременности поставок далее нелогично. Тогда имеет смысл говорить о составных векторах переменных, соответствующих этому такту.

Составным вектором $y^{(m)} = (y^1, y^2, ..., y^m)$ будем считать такой набор значений из вектора $y = (y^1, y^2, ..., y^k)$, при котором $m < k$. Чтобы избежать перегруженности и без того сложно читаемых формул, введем обозначение $y^{\langle\cdot\rangle}$, которое будет символизировать некоторый составной вектор переменной y. При таких условиях модель (15) перепишется в виде

$$z^{\langle\cdot\rangle r} = \sum_{i=1}^{s} z_i^{\langle\cdot\rangle} \alpha_i^{\langle\cdot\rangle r} \beta_i^{\langle\cdot\rangle r} \Big/ \sum_{i=1}^{s} \alpha_i^{\langle\cdot\rangle r} \beta_i^{\langle\cdot\rangle r} . \quad (16)$$

Последнее выражение (16) и имеет смысл использовать для настройки рабочей непараметрической модели под конкретный организационный процесс.

Библиографические ссылки

1. Акофф Рассел Л. Менеджмент в XXI в. (Преобразование корпорации) / пер. с англ. Ф. П. Тарасенко. Томск : Изд-во Том. ун-та, 2006.

2. Эйкхофф П. Основы идентификации систем управления. М. : Мир, 1975.

3. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука, 1983.

4. Румянцева З. П. Общее управление организацией. Теория и практика. М. : Инфра-М, 2007.

5. Адизес И. К. Стили управления. Эффективные и неэффективные. М. : Альпина Бизнес Букс : Альпина Паблишерз, 2009.

6. Медведев А. В. Теория непараметрических систем. Процессы // Вестник СибГАУ. Вып. 3. Красноярск, 2008. С. 4–9.

A. N. Sergeev

ON NONPARAMETRIC ALGORYTHMS OF DECISION-MAKING

Features and options affecting the activities of organizations, ongoing formulation of the problem of organizational system modeling and managing are considered. Some non-parametric mathematical models of organizational systems are given.

Keywords: organizational systems, management, measurements, uncertainty, chance factors, nonparametric modeling, algorithms of decision-making.

© Сергеев А. Н., 2010

УДК 519.234

В. А. Симахин

РОБАСТНЫЕ НЕПАРАМЕТРИЧЕСКИЕ ОЦЕНКИ ЛИНЕЙНЫХ ФУНКЦИОНАЛОВ

Рассматривается построение алгоритмов робастных непараметрических оценок линейных функционалов на основе взвешенного метода максимального правдоподобия.

Ключевые слова: робастный, непараметрический, оценка, линейный функционал.

Пусть $y_1, ..., y_M$ – выборка непараметрической оценки регрессии (НОР) с функцией распределения (ФР) $G(y)$ и $\theta = \int \varphi(\vec{t})dH(\vec{t}) < \infty$, где $\vec{t} = (t_1, ..., t_m)^T$; $H(\vec{t}) = G(t_1)\cdots G(t_m)$. Непараметрические оценки функционала θ при симметричных функциях $\varphi(\vec{t})$ получили название *U*-статистик [1; 2]. В классе робастных оценок θ применяется метод усечения выборки – усеченные *U*-статистики [3].

Обозначим через $f(x)$ и $F(x)$ плотность и ФР случайной величины $X = \varphi(Y_1, ..., Y_m)$, тогда $\theta = \int zdF(z)$. Выборку $Y_1, ..., Y_m$ преобразуем в выборку $x_1, ..., x_N$, где $x_j = \varphi(y_{i_1}, ..., y_{i_m})$; N – мощность множества $\{i_1 < i_2 < ... < i_m\}$. При таком преобразовании задача оценивания параметра θ сводится к задаче оценивания параметра сдвига распределения $F(x)$. В параметрической статистике такой прием широко используется для синтеза несмещенных оценок параметров как функций от достаточных статистик и в вычислительном отношении достаточно удобен, однако основная сложность здесь связана с переходом от распределения $G(y)$ к распределению $F(x)$ [4]. В связи с этим будем считать, что вид ФР $F(x)$ нам неизвестен и задача относится к классу непараметрических задач оценки параметра сдвига.

В настоящее время нет недостатка в робастных оценках параметра сдвига, что создает даже определенное неудобство для пользователей (см. например, [3; 5] и библиографические списки к ним). Отметим ряд особенностей таких оценок. Большинство из них робастны на классе и имеют низкую эффективность в отсутствии выбросов. Как выход предложены адаптивные оценки: в основном используется адаптация по параметру усечения, но не по виду $F(x)$ [3], или адаптация ведется по виду распределения $F(x)$, но функция и параметр усечения подбираются эвристически [6]. Эта работа Р. Берана интересна в двух аспектах: в ней, очевидно, впервые введены робастные непараметрические оценки плотности, а также использован метод подстановки на основе этих оценок для получения оценки параметра. Становится понятным, что робастные эффективные оценки должны быть адаптивными как по виду основного распределения, так и по отбраковке выбросов.

В данной статье на основе взвешенного метода максимального правдоподобия (ВММП) [7; 8] синтезированы адаптивные робастные непараметрические оценки и показано их использование для оценки линейных функционалов.

Взвешенный метод максимального правдоподобия. Пусть $F(x, \theta)$ – унимодальное непрерывное распределение с плотностью $f(x, \theta)$ и неизвестным параметром θ – принадлежит к классу унимодальных распределений и $x_1, ..., x_N$ – выборка НОР из распределения $F(x, \theta)$. Обозначим через $F_N(x)$ эмпирическую функцию распределения (ЭФР), а через $g(x, \theta)$ – априорную плотность распределения.

M-оценки неизвестного параметра θ можно определить на основе решения эмпирического уравнения вида

$$\int \varphi(x, \theta_N)dF_N(x) = 0, \qquad (1)$$

где $\varphi(x, \theta)$ – оценочная функция.

Анализ критерия радикальности и алгоритмов устойчивых оценок [5] позволяет сделать вывод, что все эти оценки можно получить на основе ВММП с оценочной функцией $\varphi(x, \theta)$ вида

$$\varphi(x, \theta) = \left[\frac{\partial}{\partial\theta}\ln g(x, \theta) + \beta\right] g^l(x, \theta), \qquad (2)$$

где l – параметр радикальности оценки; β – параметр, который определяется по условию несмещенности оценки, в нашем случае $\beta = 0$ [7].

Нетрудно заметить, что (2) определяет ВММП с весами $g^l(x, \theta)$. При $l = 0$ мы получаем оценки максимального правдоподобия (ОМП), при $l = 0,5$ – радикальные оценки, при $l = 1$ – оценки максимальной устойчивости (ОМУ) [5]. Физически роль параметра l вполне понятна и сводится к определению степени мягкого усечения как для удаленных выбросов, так и по форме априорного распределения. Таким образом, варьируя параметром l, можно получать эффективные оценки при локальных отклонениях распределения $F(x, \theta)$ от априорного в классе устойчивых оценок.

В непараметрическом случае, когда вид $g(x, \theta)$ неизвестен, заменим $g(x, \theta)$ в (2) непараметрической симметризованной оценкой Розенблатта–Парзена

$$g_N(x,\theta)=\frac{1}{h_N}\int K\left(\frac{2\theta-x-t}{h_N}\right)dF_N(t). \qquad (3)$$

Например, для нормального ядра уравнения для оценки параметров сдвига θ и масштаба λ принимают следующий вид [7; 8]:

$$\begin{cases}\dfrac{1}{N(N-1)}\sum\limits_{i=1\neq}^{N}\sum\limits_{j=1}^{N}(\theta_N-z_{ij})\cdot W_1(z_{ij})=0,\\ \dfrac{1}{N(N-1)}\sum\limits_{i=1\neq}^{N}\sum\limits_{j=1}^{N}\left[\left(\dfrac{\theta_N-z_{ij}}{\lambda_N}\right)^2-\dfrac{1}{l+1}\right]\cdot W_1(z_{ij})=0,\end{cases} \qquad (4)$$

где

$$W_1(z_{ij})=\exp\left\{-\frac{(\theta_N-z_{ij})^2}{\lambda_N^2}\right\}\times$$

$$\times\left[\frac{1}{N-1}\sum_{i\neq m=1}^{N}\exp\left\{-\frac{(\theta_N-z_{im})^2}{\lambda_N^2}\right\}\right]^{l-1};$$

$z_{ij}=\dfrac{x_i+x_j}{2}$ – полусуммы Уолша.

Рассмотрим обобщенную *M*-оценку θ_N параметра θ, которая определяется на основе решения эмпирического уравнения вида

$$\int\varphi(x,\theta_N,\vec{T}_N(x,\theta_N))dF_N(x)=0,$$

где $\vec{T}=(T_1,...,T_k)^T$; $T_i=\int S_i(x,t,\theta)dF(t)$;

$T_{iN}=\int S_i(x,t,\theta)dF_N(t)$.

В связи с ограниченностью объема статьи приведем без доказательства ряд результатов в окончательном виде.

Имеет место следующее представление:

$$\theta_N-\theta=\left[\int\frac{\partial}{\partial\theta}\varphi(x,\theta,\vec{T})dF(x)\right]^{-1}\cdot\int\psi(t,\theta)dF(t),$$

$$\psi(t,\theta)=\varphi(t,\theta,\vec{T}(t,\theta))+$$

$$+\sum_{i=1}^{k}\int S_i(x,t,\theta)\frac{\partial}{\partial T_i}\varphi(t,\theta,\vec{T}(t,\theta))dF(x).$$

При выполнении ряда ограничений $\sqrt{N}(\theta_N-\theta)$ имеет асимптотически нормальное распределение с дисперсией

$$\sigma^2=\left[\int\frac{\partial}{\partial\theta}\varphi(x,\theta,\vec{T})dF(x)\right]^{-2}\cdot\int\psi^2(t,\theta)dF(t). \qquad (5)$$

Техника доказательства основана на работах Г. М. Кошкина ([9]) и результаты имеют место для стационарных процессов со слабой зависимостью.

В параметрическом случае ($S_i=0$)

$$\varphi(x,\theta)=\left[\frac{\partial}{\partial\theta}g(x,\theta)\right]g^{l-1}(x,\theta). \qquad (6)$$

Выражение (5) определяет дисперсию параметрического ВММП (классические *M*-оценки) и при $l=0$ (5) совпадает с выражением для дисперсии ОМП, а при $l=1$ – с выражением для дисперсии ОМУ [7].

Для непараметрического ВММП

$$\varphi(x,\theta,T_1,T_2)=T_1(x,\theta)\cdot T_2^{l-1}(x,\theta),$$

$$S_1(x,t,\theta)=\frac{1}{h_N}K\left(\frac{2\theta-x-t}{h_N}\right),$$

$$S_2(x,t,\theta)=\frac{\partial}{\partial\theta}S_1(x,t,\theta).$$

Выражение (5) определяет дисперсию непараметрического ВММП в зависимости от *l*.

Зависимости дисперсии параметрической (рис. 1) и вариации непараметрической (типа «складного ножа» jackknife) (рис. 2) оценок ВММП для модели Тьюки с асимметричным засорением от параметра радикальности *l* ($0\le l\le 1$) приведены ниже (кривая *1* на рис. 1 – без выбросов, кривая *2* – 3 % выбросов, среднее – 4, кривая *3* – 10 % выбросов, среднее – 4; кривая 1 – на рис. 2 – без выбросов, кривая *2* – выброс – 5, кривая *3* – выброс – 11, $N=39+1$ выброс).

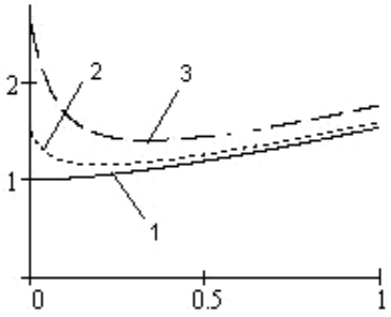


Рис. 1

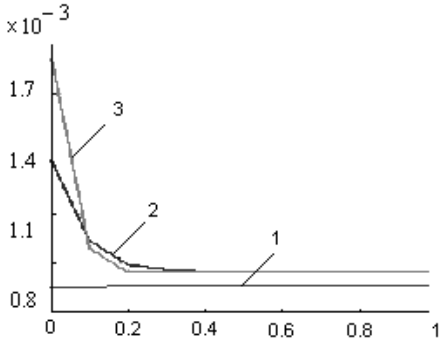


Рис. 2

Анализ дисперсии и вариации в зависимости от *l* (рис. 1, 2) показывает, что существует оптимальное *l*, доставляющее минимум дисперсии и вариации оценки.

Адаптивные оценки взвешенного метода максимального правдоподобного. Непараметрический подход на основе оценок Розенблатта–Парзена вида (3) позволяет осуществить адаптацию оценок ВММП по виду распределения. Адаптации по параметру радикальности l ($0 \le l \le 1$) производится с помощью бутстреп-метода. Для этого достаточно использовать простые бутстреп-процедуры типа «складного ножа» (jackknife) и алгоритмы поиска минимума вариации непараметрического ВММП. Моделирование также показывает, что при оптимальном l наблюдается и минимальное смещение оценки.

Примеры. Как отмечалось выше, значительный интерес представляет нахождение робастных непараметрических оценок для *U*-статистик. Применим для этого адаптивные оценки ВММП.

В первую очередь нас интересуют робастные непараметрические оценки функции распределения $G(t) = \int C(t-y)dG(y)$ и плотности в виде $g(t) = \int K((t-y)\cdot h_N^{-1})dG(y)$, где $C(y)$ – функция Хевисайда; $K(y)$ – ядерная функция. Зафиксируем значение $t = t_0$. От выборки $y_1, ..., y_M$ перейдем к выборкам $x_i = C(t_0 - y_i)$ для ФР и $x_i = K((t_0 - y_i)\cdot h_N^{-1})$ для плотности соответственно.

Представим результаты моделирования в зависимости от l для асимметричной модели выбросов Тьюки (N = 100, 10 % выбросов из нормального распределения со средним, равным пяти, рис. 3, 4). Хорошие результаты показывают радикальные оценки (l = 0,5), l оптимально при l = 0,35, при l = 1 происходит достаточно сильное подрезание.

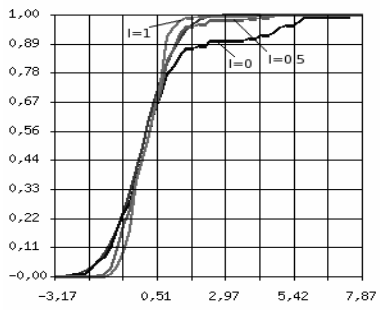


Рис. 3

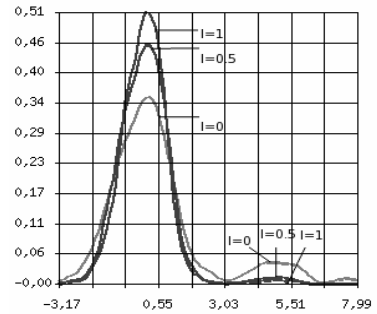


Рис. 4

Результаты моделирования для вариаций оценок дисперсии ($x_k = 0,5\cdot(y_i - y_j)$) и средней разницы Джини ($x_k = |y_i - y_j|$) приведены на рис. 5, 6 (N = 30 + + 1 выброс).

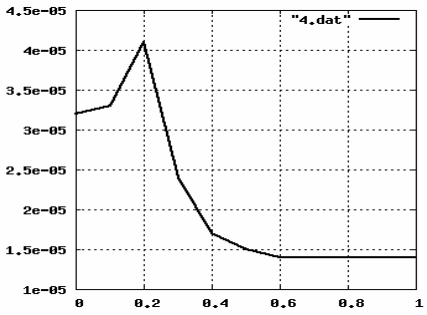


Рис. 5

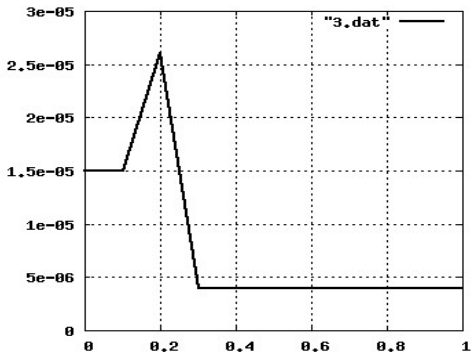


Рис. 6

Таким образом, предложен адаптивный робастный непараметрический алгоритм нахождения линейных функционалов, который позволяет адаптивно (путем мягкого усечения) настраивать оценку в зависимости от исходного распределения и выбросов. Рассмотрено робастное оценивание функции распределения, плотности распределения типа Розенблатта–Парзена, дисперсии, средней разницы Джини. Проведено моделирование оценок для асимметричной модели засорений Тьюки. На модели эксперимента Берана [7] проведено сравнение оценки Берана и вышеприведенной оценки. Они показывают одинаковые результаты, но в оценке Берана функция усечения и окно для нее (адаптация) подбирались эвристически [6]. Необходимо отметить, что представленный в данной статье подход позволяет применять робастные оценки ФР и плотности методом подстановки для получения адаптивных оценок неявных параметров от нелинейных функционалов.

Библиографические ссылки

1. Королюк В. С., Боровских Ю. В. Теория *U*-статистик. Киев : Наук. думка, 1989.

2. Непараметрическое оценивание функционалов по стационарным выборкам / Ю. Г. Дмитриев,

Г. М. Кошкин, В. А. Симахин и др. ; Тос. гос. ун-т. Томск, 1974.

3. Шуленин В. П. Введение в робастную статистику / Тос. гос. ун-т. Томск, 1993.

4. Воинов В. Г., Никулин М. С. Несмещенные оценки и их применения. М. : Наука, 1989.

5. Шурыгин А. М. Прикладная статистика. Робастность. Оценивание. Прогноз. М. : Финансы и статистика, 2000.

6. Beran R. An efficient and robust adaptive estimator of location // Ann. Stat. 1978. Vol. 6, № 2. P. 292–313.

7. Симахин В. А. Непараметрическая статистика. Ч. II. Теория оценок / Курган. гос. ун-т. Курган, 2004.

8. Симахин В. А. Взвешенный метод максимального правдоподобия // Высокие технологии XXI века : материалы IX Междунар. науч.-техн. конф. : в 2 т. Т. 2. Воронеж, 2008. С. 661–672.

9. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М. : Наука, 2004.

V. A. Simakhin

ROBUST NONPARAMETRIC ESTIMATION OF LINEAR FUNCTIONALS

Robust nonparametric algorithms for estimation of linear functionals on the basis of weighted maximum likelihood method is considered in the article.

Keywords: robust, nonparametric, linear functional.

© Симахин В. А., 2010

УДК 62-506.1

Н. А. Сергеева, Е. С. Терентьева

О НЕПАРАМЕТРИЧЕСКИХ ОЦЕНКАХ ФУНКЦИИ РЕГРЕССИИ И ЕЕ ПРОИЗВОДНЫХ ПРИ НАЛИЧИИ ПРОПУСКОВ ДАННЫХ

Рассмотрены непараметрические методы оценивания регрессии и ее производных по выборкам случайных величин с некоторыми особенностями при их измерении. Представлен бутстреп-метод, применяемый для решения задачи заполнения пропусков в неполных данных или устранения пустот в пространстве наблюдений.

Ключевые слова: непараметрическая оценка регрессии, Н-аппроксимация, бутстреп-метод, непараметрическая оценка производной функции регрессии, сходимость оценок.

Проблема моделирования дискретно-непрерывных процессов является одной из центральных в кибернетике. Определяющее значение при постановке задачи идентификации имеет математическая постановка, соответствующая различным априорным предпосылкам. Априорные сведения о процессе, по существу, определяют подход к задаче идентификации.

Ниже мы остановимся на задаче идентификации и связанной с ней задаче оценивания соответствующих вероятностных характеристик в условиях непараметрической неопределенности. В отличие от ставшего традиционным параметрического подхода к решению задачи идентификации в дальнейшем нам понадобятся некоторые качественные свойства поведения исследуемого процесса. Одним из главных этапов на пути решения этой задачи является оценивание регрессионных характеристик входных-выходных переменных процесса.

Непараметрический уровень априорной информации не предполагает наличия этапа выбора параметрической структуры модели, но требует некоторых сведений качественного характера о процессе, например от однозначности или неоднозначности его характеристик, линейности для динамических процессов или характере нелинейности. При идентификации линейных динамических объектов мы сталкиваемся с необходимостью оценивания производной функции регрессии. Это связано с оценкой весовой функции линейной системы по измерениям функции переходной характеристики последней. Непараметрическая модель в этом случае представляет собой оценку интеграла Дюамеля.

Существенная особенность данного исследования состоит в предположении, что исходные выборки содержат пропуски данных при контроле входных-выходных переменных объекта. Это приводит к необходимости построения модифицированных непараметрических оценок функции регрессии и ее производных.

Пусть имеется неравномерная выборка статистически независимых наблюдений (u_i, x_i), $i = \overline{1,s}$, входных и выходных переменных системы объемом s. Здесь u_i – значение вектора наблюдений входных воздействий размерности m в i-й точке выборки, а x_i – значение выходного воздействия в этой точке. Требу-

ется построить непараметрическую модель объекта и восстановить производную первого порядка стохастической зависимости $x(u)$ по имеющейся выборке наблюдений.

На начальном этапе восстановления $x(u)\ \forall x \in \Omega(x)$ принимается статистика [1]:

$$\tilde{x}(u) = \sum_{j=1}^{s} x_j \prod_{k=1}^{m} \Phi\left(\frac{u^k - u_j^k}{C_s^{\,k}}\right) \Bigg/ \sum_{j=1}^{s} \prod_{k=1}^{m} \Phi\left(\frac{u^k - u_j^k}{C_s^{\,k}}\right), \quad (1)$$

где $\Phi(z)$ – колоколообразная функция, удовлетворяющая следующим условиям [1]:

$$\Phi(z) < \infty, \forall z \in \Omega(z), \ \int_{\Omega(z)} \Phi(z)dz = 1,$$
$$\int_{\Omega(z)} \Phi^2(z)dz < \infty, \ \lim_{s\to\infty} \frac{1}{C_s} \cdot \Phi\left(\frac{u^j - t^j}{C_s}\right) = \delta(u^j - t^j), \quad (2)$$

здесь $\delta(u^j - t^j)$ – дельта-функция Дирака, z – аргументы колоколообразной функции.

Параметр размытости C_s должен удовлетворять условиям [1]:

$$C_s > 0, \ \lim_{s\to\infty} C_s = 0, \ \lim_{s\to\infty} s \cdot C_s^{\,m} = \infty. \quad (3)$$

На следующем этапе строится H-аппроксимация [2]:

$$x_k(u) = \frac{1}{kc_k^m} \sum_{i=1}^{k} x_i^* \prod_{j=1}^{m} H\left(\frac{u^j - u_i^{*j}}{c_k}\right), \quad (4)$$

где $\left\{(u_i^*, x_i^*)\right\}$ – рабочая равномерная выборка объемом $k < s$; $c_k^m \to 0$; $kc_k^m \to \infty$; $\forall j = 1,...,m$. Тогда $\frac{1}{c_k} H\left(\frac{u^j - t^j}{c_k}\right) \to \delta(u^j - t^j)$ с ростом k, где $H(z)$ – колоколообразная функция, отличающаяся от $\Phi(z)$ на множитель, равный константе

$$\frac{1}{kc_k^m} \sum_{i=1}^{k} \prod_{j=1}^{m} \Phi\left(\frac{u^j - u_i^{*j}}{c_k}\right) = \text{const} \quad \forall u \in \Omega(u).$$

Можно показать, что

$$\frac{1}{kc_k^m} \sum_{i=1}^{k} \prod_{j=1}^{m} H\left(\frac{u^j - u_i^{*j}}{c_k}\right) = 1 \quad \forall u \in \Omega(u). \quad (5)$$

К классу функций $H(z)$ и $\Phi(z)$ относятся, например, гауссова кривая и функция Соболева.

В качестве оценки производной регрессии предлагается взять аппроксимацию, построенную на рабочей равномерной выборке, в виде [3]:

$$D^{\alpha} x_k(u) = \frac{1}{kc_k^m} \sum_{i=1}^{k} x_i^* \prod_{j=1}^{m} H^{(/)}\left(\frac{u^j - u_i^{*j}}{c_k}\right), \quad (6)$$

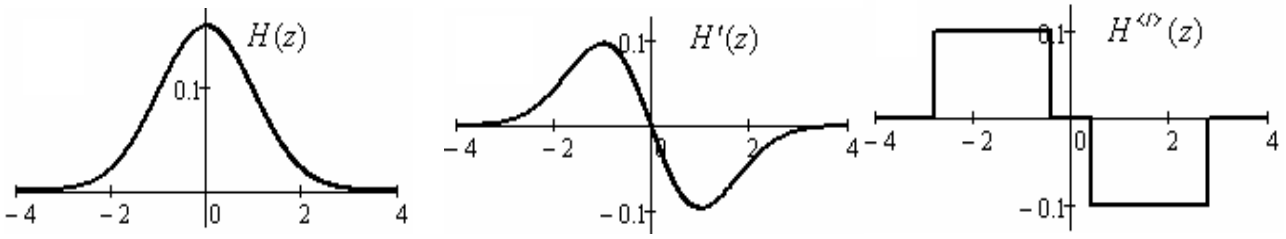
где $\alpha = (\alpha_1,...,\alpha_m)$, $\alpha_j = \{0,1\}$, $j = \overline{1,m}$, если $\alpha_j = 1$, то по переменной u^j берется производная первого порядка, если $\alpha_j = 0$, то производная нулевого порядка; $H^{(/)}(z)$ – функция, удовлетворяющая интегральным условиям, сформулированным ниже.

Если u – скалярная величина, т. е. $m = 1$, то оценка первой производной регрессии (6) при $\alpha = 1$ является асимптотически несмещенной и состоятельной при выполнении следующих условий:

$$\int_{R^1} H^{(/)}(z)dz = 0, \ c_k \int_{R^1} zH^{(/)}(z)dz = -1,$$
$$\int_{R^1} H^{(//)}(z)dz < \infty. \quad (7)$$

Техника доказательства основывается на результатах [3].

Заметим, что обозначенные выше условия сужают класс функций, которые могли бы применяться в случае восстановления кривой регрессии. В частности, гауссова кривая $H(z) = \frac{e^{-\frac{z^2}{2}}}{2\pi}$ (рис. 1, *а*) имеет производную $H'(z)$ (рис. 1, *б*), которая сложна с вычислительной точки зрения, однако ее можно заменить кусочно-постоянным аналогом $H^{</>}(z)$ (рис. 1, *в*). Отметим, что время вычисления $H^{</>}(z)$ в 1,3 раза меньше, чем $H'(z)$.



а *б* *в*

Рис. 1

Для выбора оптимального параметра размытости введем меру отклонения оценки (6) от истинной производной:

$$W(c_k) = \int_{G(u)} (Dx(u) - D^{\alpha} x_k(u))^2 du, \qquad (8)$$

где $G(u)$ – область сравнения. Тогда критерий оптимальности будет иметь вид

$$W(c_k) \to \min_{\Omega(c_k)}, \qquad (9)$$

где $\Omega(c_k) = (0, +\infty)$. Но при практическом применении данный критерий невозможно использовать, так как $D^{\alpha} x(u)$ неизвестна.

Преобразуем (9), используя формулу интегрирования по частям и тем самым понижая порядок производной. Тогда критерий оптимальности приобретет следующий вид:

$$\begin{aligned} I(c_k) = & \int_{G(u)} (D^{\alpha} x_k(u))^2 du - \\ & - 2 \int_{G(u^-)} \breve{x}_k(u) D^{\alpha} x_k(u)\big|_{G(u_j)} du^- + \\ & + 2 \int_{G(u)} \breve{x}_k(u) D^{\alpha^+} x_k(u) du \to \min_{c_k}, \end{aligned} \qquad (10)$$

где $u^- = (u_1, ..., u_{j-1}, u_{j+1}, ..., u_m)$, здесь j такое, что $\alpha_j \neq 0$, $\alpha^+ = (\alpha_1, ..., \alpha_{j-1}, \alpha_j + 1, \alpha_{j+1}, ..., \alpha_m)$; $\breve{x}_k(u)$ – оценка любого типа, которая в смысле некоторого критерия достаточно хорошо аппроксимирует нужную нам величину; $G(u) \subset \Omega(u)$ – область хорошего качества оценки $\breve{x}_k(u)$. Если это непараметрическая оценка, то вычисление оптимального параметра размытости становится рекуррентным. На первом шаге находится оптимальный параметр размытости c_k^0 для оценки кривой регрессии. На следующем шаге определяется оптимальный параметр размытости c_k^1 для оценки производной кривой регрессии первого порядка по какой-либо переменной. Критерий оптимальности на этом шаге имеет вид (10) при $\breve{x}(u) = \breve{x}(u, c_k^0)$ и $\alpha = (0, ..., 0, 1, 0, ..., 0)$.

При построении модели реального процесса иногда используются выборки случайной величины, результаты измерений которой распределены неравномерно. Это приводит к тому, что в некоторых подобластях пространства наблюдений образуются пустоты. В таких условиях приходится отказываться от применения стандартной непараметрической оценки регрессии.

При наличии пустот в пространстве наблюдения $\Omega(u)$ непараметрическая оценка регрессии основывается на использовании не конкретного значения выходной переменной в *j*-й точке выборки, а ее оценки. Проводить это оценивание будем по представленному ниже алгоритму.

Пусть мы находимся в *j*-й точке выборки. Определим множество Λ_{sj} соседних точек выборки, в которых колоколообразная функция не равна нулю:

$$\Lambda_{sj} = \left\{ u_i : \prod_{k=1}^{m} \Phi\left(\frac{u_j^k - u_i^k}{C_s^i(c_s, \rho_s)} \right) > 0 \right\}, \quad i = \overline{1,s},\ j = \overline{1,s}, \qquad (11)$$

где ρ_s – параметр оптимизации, а колоколообразная функция для каждой входной переменной u^k, $k = \overline{1,m}$, расширяется в направлении разрежений в выборке, т. е. ее ветви имеют разные константы Липшица. Приведем пример такой асимметричной функции:

$$\Phi\left(\frac{u^k - u^k{}_i}{C_s}\right) = \begin{cases} \left.\begin{array}{l} \frac{1}{2}\left(\cos\left(\frac{1}{l}\cdot\frac{u^k - u^k{}_i}{C_s}\right)\right) + \frac{1}{2}, \ if \ \ -l\cdot\pi \le \frac{u^k - u^k{}_i}{C_s} \le 0 \\ \frac{1}{2}\left(\cos\left(\frac{u^k - u^k{}_i}{C_s}\right)\right) + \frac{1}{2}, \ \text{if } 0 < \frac{u^k - u^k{}_i}{C_s} \le \pi \\ 0, \ \text{if} \ \ \left(\frac{u^k - u^k{}_i}{C_s} > \pi\right) \vee \left(\frac{u^k - u^k{}_i}{C_s} < -l\cdot\pi\right) \end{array}\right\}, \text{if } \ f_S(u^k + h, c_s) > f_S(u^k - h, c_s), \\ \left.\begin{array}{l} \frac{1}{2}\left(\cos\left(\frac{1}{l}\cdot\frac{u^k - u^k{}_i}{C_s}\right)\right) + \frac{1}{2}, \ \text{if} \ \ 0 \le \frac{u^k - u^k{}_i}{C_s} \le l\cdot\pi \\ \frac{1}{2}\left(\cos\left(\frac{u^k - u^k{}_i}{C_s}\right)\right) + \frac{1}{2}, \ \text{if} \ \ -\pi < \frac{u^k - u^k{}_i}{C_s} \le 0 \\ 0, \ \text{if} \ \ \left(\frac{u^k - u^k{}_i}{C_s} > l\cdot\pi\right) \vee \left(\frac{u^k - u^k{}_i}{C_s} < \pi\right) \end{array}\right\}, \text{if } \ f_S(u^k + h, c_s) < f_S(u^k - h, c_s), \end{cases}$$

где $h>0$ – радиус окрестности текущей точки u; $l\ge 1$ – коэффициент расширения колоколообразной функции, при l = 1 колоколообразная функция принимает симметричный вид. Функция множества в виде непараметрической оценки плотности [4] с малым параметром размытости

$$f_s(u^k,c_s)=\frac{1}{sc_s}\sum_{i=1}^{S}\Phi\left(\frac{u^k-u^k_i}{c_s}\right),\quad k=\overline{1,m}, \qquad (12)$$

применяется для определения сгущений и разрежений точек в выборке по каждой входной переменной u^k, $k=\overline{1,m}$. Форма и вид колоколообразной функции зависят от плотности точек выборки в h-окрестности текущей точки.

Рассмотрим выборку наблюдений входной переменной $u_i, i=\overline{1,20}$ (рис. 2) с функцией множества (12) (рис. 3). Тогда график колоколообразной функции в точке $u_i=0,26$ при разных значениях коэффициента расширения l будет иметь вид, представленный на рис. 4 колоколообразная функция расширилась по направлению уменьшения значения функции множества в окрестности текущей точки $u_i=0,26$.

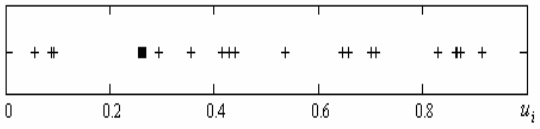


Рис. 2

+++ – точки, в которых производились измерения входной переменной объекта; ■■■ – ui = 0,26

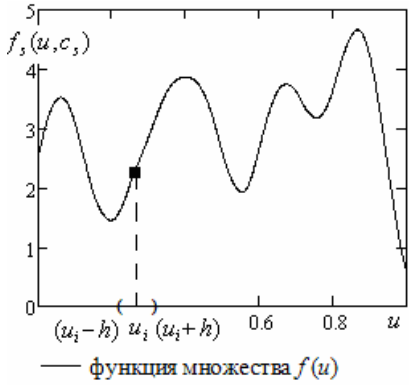


Рис. 3

На подобласти $\Omega(x,u)$, определяемой Λ_{sj}, сформированной в соответствии с правилом (11), строится поверхность $\phi_j(u)=\beta_j\Psi(u), j=\overline{1,s}$, где $\Psi(u)$ – вектор базисных функций; β_j^T – вектор параметров, определяемых по методу наименьших квадратов.

Для примера рассмотрим выборку наблюдений со скалярным входным и выходным воздействиями $(u_i,x_i), i=\overline{1,15}$ (рис. 5), где функции $\phi_j(u), j=\overline{1,15}$, имеют линейный вид. На рис. 5 отмечена текущая десятая точка выборки и точки вокруг нее, колоколообразная функция $\Phi\left((u-u_i)/C_s\right)$ в которых не равна 0. В данном случае их 4. Таким образом, вектор коэффициентов β_j^T при наборе базисных функций $\Psi(u)=(1,u)^T$ определяется по методу наименьших квадратов на основании пяти выделенных точек.

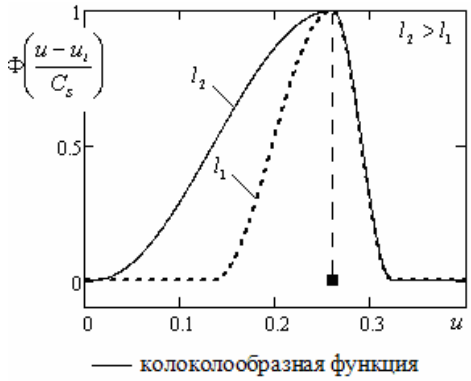


Рис. 4

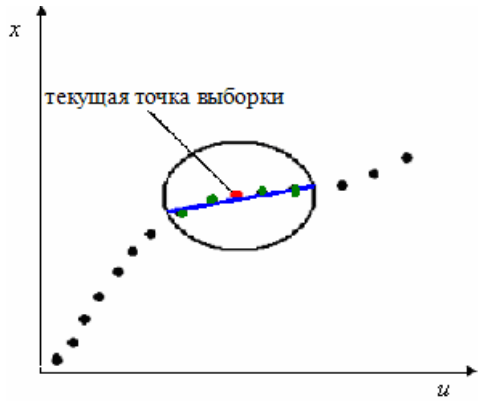


Рис. 5

Произведем модификацию непараметрической оценки регрессии следующим образом [5]:

$$\hat{x}(u,c_s,\rho_s)=\frac{\sum_{j=1}^{S}\left(\phi_j(u)\prod_{k=1}^{m}\Phi\left(\frac{u^k-u^k_j}{C_S{}^j(c_s,\rho_s)}\right)\right)}{\sum_{j=1}^{S}\prod_{k=1}^{m}\Phi\left(\frac{u^k-u^k_j}{C_S{}^j(c_s,\rho_s)}\right)}. \qquad (13)$$

Параметр размытости в формуле (13) оценивается для каждой j-й точки из выборки наблюдений [6]:

$$C_s{}^j(c_s,\rho_s)=\rho_s\cdot f_s^{-1}(u_j,c_s),\quad j=\overline{1,s}. \qquad (14)$$

Выбор оптимального параметра размытости c_s и коэффициента пропорциональности ρ_s осуществляется путем минимизации среднеквадратичного критерия:

$$W(c_s,\rho_s)=\frac{1}{s}\sum_{i=1}^{s}(x_i-\hat{x}(U_i,c_s,\rho_s))^2\to\min_{Cs,Ps}. \qquad (15)$$

Другим методом непараметрического оценивания регрессии между входными и выходными воздействиями объекта является предварительное заполнение пустот в пространстве наблюдений случайной величины. Для этого используется следующая бутстреп-процедура [7].

1. Точка u_l равномерной сетки, натянутой на область наблюдения $\Omega_s(u)$ входных переменных u^k, $k=\overline{1,m}$, будет считаться пропуском, если

$$\prod_{j=1}^{m} f_s(u_l^j, c_s) < a,\ a>0. \qquad (16)$$

Пусть таких точек будет K, т. е. $l=\overline{1,K}$. Параметр $a>0$ настраивается исходя из оптимизационной процедуры по среднеквадратичному критерию.

2. По присутствующим наблюдениям $(u_i, x_i), i=\overline{1,s}$ входных и выходных переменных системы строится регрессионная модель (1), однако параметр размытости оценивается по формуле (14) для каждой точки выборки наблюдений, что позволяет избежать неадекватного поведения оценки (1) в областях разрежений (пустот) пространства наблюдений $\Omega_s(u)$.

3. По построенной регрессионной модели находятся оценки $x_{si}=\tilde{x}(u_i),\ i=\overline{1,s}$.

4. Определяются ошибки $\varepsilon_i = x_{si} - x_i,\ i=\overline{1,s}$, для всех точек выборки.

5. Для каждого пропуска после подстановки значения сопутствующей входной переменной u_l в полученное регрессионное уравнение находится оценка $\tilde{x}(u_l),\ l=\overline{1,K}$.

6. Значения выходной переменной, которыми замещают пропуски, получается по формуле

$$x_{sl} = \tilde{x}(u_l) + \varepsilon_l,\ l=\overline{1,K}, \qquad (17)$$

где ε_l выбирается случайно из ошибок, рассчитанных в п. 4. Это можно реализовать следующим образом: с помощью генератора случайных чисел выбирается целое число q на интервале $[1; s]$, $\varepsilon_l = \varepsilon_q$, и операция повторяется для каждого $\varepsilon_l, l=\overline{1,K}$.

7. Данные, полученные после заполнения пропусков, $(u_l, x_{sl}), l=\overline{1,K}$, объединяются с исходной выборкой наблюдений $(u_i, x_i), i=\overline{1,s}$. По итоговой выборке объема $s+K$ строится регрессионная модель (1).

Приведем численные результаты моделирования при использовании модифицированной непараметрической оценки регрессии (13) в сравнении со стандартной непараметрической оценкой (1).

Пусть размерность входной переменной равна двум и имеется неравномерная выборка наблюдений $(u_i, x_i), i=\overline{1,s}$, объемом s = 200, $u=(u^1,u^2)$, на области $\Omega_s(u): \{u^1 \in [0;0,5], u^2 \in [0;0,5]\}$; $x=\varphi(u)+\xi$, где $\varphi(u)=\varphi(u^1,u^2)=\sin(7\cdot u^1+u^2)$; ξ – аддитивная центрированная помеха, имеющая нормальный закон распределения с математическим ожиданием $M(\xi)=0$ и ограниченной дисперсией.

Помеха накладывается следующим образом:

– измеряется интервал изменения сигнальной части $[a_1; a_2]$;

– задается уровень помех h на интервале $[0;1]$;

– с помощью генератора случайных чисел формируется вектор ξ значений случайной величины, распределенной по нормальному закону с $M(\xi)=0$, $\sigma(\xi)=\dfrac{h(a_2-a_1)}{6}$. Вектор ξ складывается с вектором значений сигнальной части.

Создадим пробел в выборке наблюдений, расположение точек $u_i, i=\overline{1,200}$, для которой приведено на рис. 6. С помощью генератора случайных чисел формируем вектор u^2, $|u^2|=200$, значений случайной величины, распределенной по равномерному закону на интервале $[0;0,5]$. Затем генерируем векторы u_1^1 и u_2^1, $|u_1^1|=|u_2^1|=100$, на интервалах $[0;0,2]$ и $[0,3;0,5]$ соответственно. Выборка наблюдений входных воздействий объемом 200 имеет вид $(u_1^1 \cup u_2^1; u^2)$.

Рис. 6

Пусть в каналах измерения выходного сигнала присутствует 5%-я помеха, т. е. $h=0,05$, а $\phi_j(u)=\beta_{2j}u^2+\beta_{1j}u^1+\beta_{0j},\ j=\overline{1,s}$, имеет линейный вид, где u^2, u^1 – входные переменные. Результат моделирования в виде среза ($u^2=0,25, u^1=[0;0,5]$) представлен ниже (рис. 7).

Среднеквадратичная оценка ошибки моделирования при использовании стандартной непараметрической оценки регрессии (1) равна 3,3 %. При использовании модифицированной оценки регрессии (13) с несимметричной колоколообразной функции оценка ошибки равна 1,5 %, что в 2 раза меньше, чем при ис-

пользовании стандартной оценки. Оценка ошибки моделирования здесь и в дальнейшем рассчитывалась следующим образом:

$$E = \frac{\sqrt{\sum_{i=1}^{100} (\varphi(u_i) - \hat{x}(u_i))^2}}{100} 100\ \%,$$

где $u_i, i = \overline{1,100}$ – вектор значений случайной величины, распределенной по равномерному закону на области $\Omega_s(u)$.

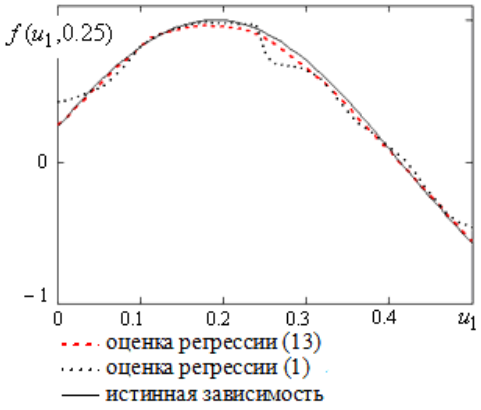


Рис. 7

Пусть размерность входной переменной равна четырем и имеется неравномерная выборка наблюдений $(u_i, x_i), i = \overline{1,s}$, объемом $s = 400$, $u = (u^1, u^2, u^3, u^4)$, на области $\Omega_s(u)$: $\{u^1 \in [0;0,5], u^2 \in [0;0,5], u^3 \in [0;0,5], u^4 \in [0;0,5]\}$. И пусть в этой области имеются искусственно созданные пустоты по двум входным переменным: $u_i^3 \in [0;0,2] \cup [0,25;0,5]$, $u_i^4 \in [0;0,3] \cup [0,4;0,5]$, $i = \overline{1,400}$; $x = \varphi(u) + \xi$, где ξ – аддитивная помеха, имеющая нормальный закон распределения с $M(\xi) = 0$ и $D(\xi) < \infty$, а истинная зависимость имеет вид

$$\varphi(u) = \varphi(u^1, u^2, u^3, u^4) = \sin(0,45 \cdot u^1) - \\ - \sin(0,5 \cdot u^2) + \sin(0,45 \cdot u^3) - \sin(0,5 \cdot u^4).$$

Пусть в каналах измерения выходного сигнала присутствует 5%-я помеха. Тогда результат моделирования при помощи модифицированной оценки регрессии (13) в виде среза ($u^1 = u^2 = 0,2$, $u^3 = 0,4$, $u^4 = [0;0,5]$) будет представлен в виде линейного аппроксимирующего полиноми (рис. 8) и квадратичного аппроксимирующего полинома (рис. 9).

Оценка ошибки моделирования при использовании модифицированной оценки регрессии (13) с квадратичным аппроксимирующим полиномом $\phi_j(u), j = \overline{1,s}$, в 8,2 раза меньше, чем с линейным полиномом, и в 15,6 раз меньше, чем при использовании стандартной непараметрической оценки (1).

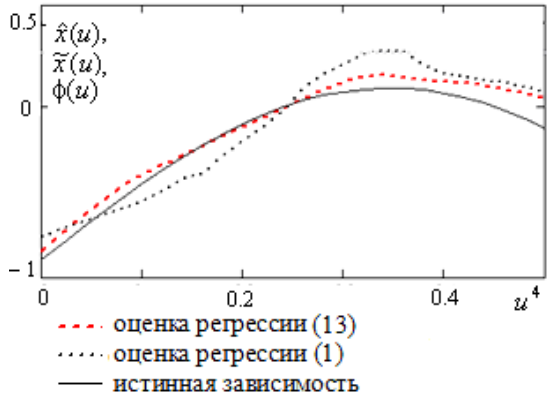


Рис. 8

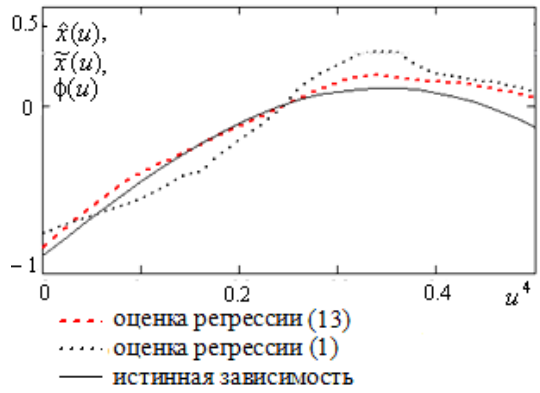


Рис. 9

Таким образом, результаты экспериментов, представленные на рис. 6–9, подтверждают, что использование модифицированной непараметрической оценки регрессии (13) с квадратичной аппроксимирующей функцией дает в несколько раз меньшую ошибку идентификации при наличии пустот в пространстве наблюдений, чем применение стандартной непараметрической оценки (1).

Выберем объект с одномерным входным и выходным воздействиями. Для этого примем $\varphi(u) = 5(u - 0,5)^2$. Пусть объем выборки равен $s = 30$: $(u_i, x_i), i = \overline{1,30}$, а сама выборка наблюдений отно- сительно равномерна и не имеет больших пробелов.

Результаты моделирования, приведенные на рис. 10, показывают, что среднеквадратичная оценка ошибки моделирования при использовании стандартной непараметрической оценки регрессии (1) в 1,5 раза больше, чем при применении оценки регрессии (13) с несимметричной колоколообразной функцией и линейным аппроксимирующим полиномом.

Таким образом, при разреженной выборке, без явных пробелов в пространстве наблюдений, ошибка моделирования с использованием модифицированной оценки регрессии (13) меньше, чем с применением стандартной оценки.

Проведем сравнение результатов численного моделирования, полученных с помощью модифициро-

ванной оценки регрессии (13) и бутстреп-метода со стандартной непараметрической оценкой (1).

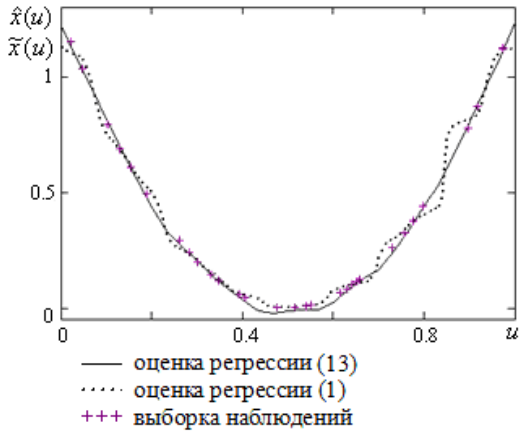


Рис. 10

Пусть имеется неравномерная выборка наблюдений одномерных входного и выходного воздействий объекта (u_i, x_i), $i = \overline{1,s}$, объемом $s = 20$, $u \in [0;1]$, $x = \varphi(u) + \xi$, где $\varphi(u) = \sin(5 \cdot u)$; ξ – 5 %-я помеха в каналах измерения выходного сигнала. Пусть $\phi_j(u) = \beta_{2j} u^2 + \beta_{1j} u + \beta_{0j}$, $j = \overline{1,s}$. Тогда результаты моделирования будут следующими (рис. 11, 12).

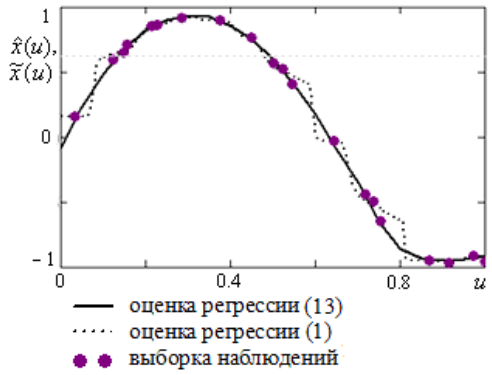


Рис. 11

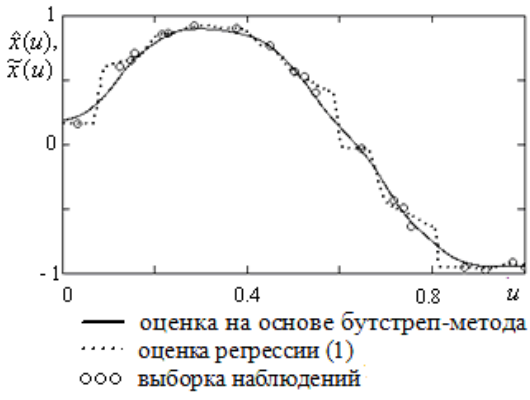


Рис. 12

Исходная выборка наблюдений входных воздействий может быть дополнена значениями наблюдений по бутстреп-методу (рис. 13).

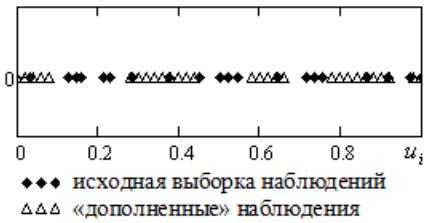


Рис. 13

Среднеквадратичная оценка ошибки моделирования при использовании стандартной непараметрической оценки регрессии (1) равна 3 %, при применении модифицированной оценки регрессии (13)–1,45 %, а при построении непараметрической оценки (1) на основе выборки, сформированной по бутстреп-методу, – 1,5 % (рис. 14).

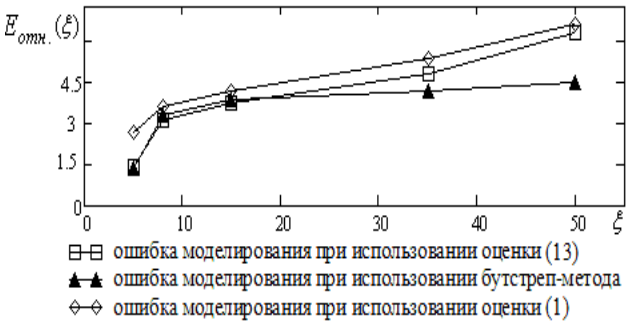


Рис. 14

Анализ графиков на рис. 14 позволяет сделать вывод, что при сравнительно малом уровне помехи в каналах измерения выходного сигнала (< 15 %) наименьшую ошибку моделирования дает модифицированная оценка регрессии (13). При увеличении уровня шума (> 15 %) ошибка моделирования на основе бутстреп-метода становится меньше, чем при использовании оценки (13). Следует отметить, что при высоком уровне помехи в каналах измерения (> 40 %) качество построения модели при помощи модифицированной оценки (13) становится практически эквивалентным качеству моделирования на основе стандартной непараметрической оценки (1).

Сравним результаты численного моделирования при использовании модифицированной оценки регрессии (13) в случае, когда функция локальной аппроксимации $\phi_j(u) = \beta_{2j} u^2 + \beta_{1j} u + \beta_{0j}$, $j = \overline{1,s}$, имеет квадратичный вид, и в случае, когда она линейна: $\phi_j(u) = \beta_{1j} u + \beta_{0j}$, $j = \overline{1,s}$.

Пусть имеется неравномерная выборка наблюдений одномерных входного и выходного воздействий объекта (u_i, x_i), $i = \overline{1,s}$, объемом $s = 20$, $u \in (0;1)$, $x = \varphi(u) + \xi$, где $\varphi(u) = \sin(5 \cdot u)$; ξ – 30%-я центрированная помеха в каналах измерения выходного сигнала.

Оценка ошибки моделирования при использовании непараметрической оценки регрессии (13) с квадратичным аппроксимирующим полиномом $\phi_j(u) = \beta_{2j}u^2 + \beta_{1j}u + \beta_{0j}$, $j = \overline{1,s}$, в 1,1 раза меньше, чем с линейным полиномом $\phi_j(u) = \beta_{1j}u + \beta_{0j}$, $j = \overline{1,s}$. При дальнейшем увеличении порядка полинома ошибка моделирования практически не изменяется, а при порядке полинома больше четырех ошибка моделирования увеличивается. Таким образом, предлагается использовать второй порядок аппроксимирующего полинома $\phi_j(u)$, $j = \overline{1,s}$.

Представим зависимости оценок ошибки моделирования при использовании стандартной непараметрической оценки (1) на основе исходной выборки наблюдений и выборки, дополненной по бутстреп-методу, а также модифицированной оценки регрессии с квадратичным аппроксимирующим полиномом от уровня шума, % (рис. 15).

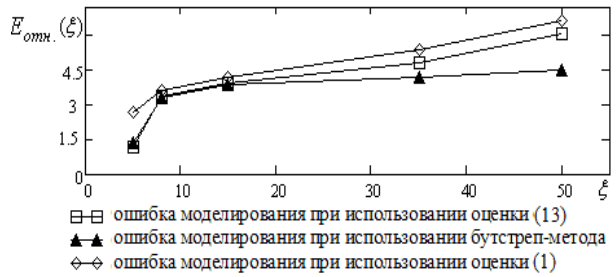


Рис. 15

Таким образом, разработанные методы построения непараметрических моделей позволяют довольно эффективно моделировать объекты в случае неравномерно распределенной выборки наблюдений входных и выходных воздействий. Применение несимметричной колоколообразной функции способствует большей согласованности модели и объекта по сравнению с симметричной функцией. Однако при увеличении размерности входной переменной на единицу время расчета модели при использовании модифицированной оценки регрессии (13) с несимметричной колоколообразной функцией увеличивается в среднем в 1,5 раза при фиксированном объеме выборки наблюдений.

В заключение дадим несколько рекомендаций по применению представленных оценок в различных условиях. В случае идентификации объекта на основе представительной выборки наблюдений можно использовать любую из предложенных оценок. При наличии пустот в пространстве наблюдений и малом уровне помехи в каналах измерения для моделирования больше подходит модифицированная оценка регрессии (13) с несимметричной колоколообразной функцией и квадратичным аппроксимирующим полиномом. Если в каналах измерения высокий уровень помех, то предлагается использовать непараметрическую оценку, построенную на основании дополненной по бутстреп-методу выборки наблюдений.

Библиографические ссылки

1. Надарая Э. А. Замечания о непараметрических оценках плотности вероятности и кривой регрессии // Теория вероятности и ее применение. 1970. Т. 15. Вып. 1. С. 139–142.

2. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука. Сиб. отд-ние, 1983.

3. Медведева Н. А. Непараметрические оценки производной кривой регрессии и модели динамики // Информатика и процессы управления. Красноярск, 1995. С. 74–81.

4. Parzen E. On estimation of a probability density function and mode // Ann. Math. Stat. 1962. Vol. 33. P. 1065–1076.

5. Катковник В. Я. Непараметрическая идентификация и сглаживание данных. М. : Наука, 1985.

6. Непараметрическое моделирование стохастических систем / В. А. Гутшмидт, Я. И. Демченко, М. В. Кураченко и др. // Теория вероятностей, случайные процессы, математическая статистика и приложения. Минск, 2008. С. 66–73.

7. Злоба Е., Яцкив И. Статистические методы восстановления пропущенных данных // Computer Modeling & New Technologies. 2002. Т. 6, № 1. С. 51–61.

N. A. Sergeeva, E. S. Terentyeva

NONPARAMETRIC ESTIMATIONS OF REGRESSION FUNCTION AND ITS DERIVATIVES IN THE PRESENCE OF DATA ADMISSIONS

In the article we consider nonparametric methods of estimation of a regression and its derivatives on samplings of random variables with some singularities at their measurement. A bootstrap-method applied to the decision of the passes filling task in incomplete data or elimination of emptiness in space of observations is presented.

Keywords: nonparametric estimation of a regression, H-approximation, a butstrep-method, nonparametric estimation of a derivative of a regression, convergence of estimations.

© Сергеева Н. А., Терентьева Е. С., 2010

УДК 519.233.22

П. О. Смирнов, Г. Л. Шевляков

ПРИБЛИЖЕНИЕ ОЦЕНКИ Q_N ПАРАМЕТРА МАСШТАБА С ПОМОЩЬЮ БЫСТРЫХ *M*-ОЦЕНОК

Рассмотрены популярные робастные оценки параметра масштаба, обладающие высокой эффективностью, в частности оценка Q_n. Предложено параметрическое семейство М-оценок, обеспечивающих более быстрые вычисления с допустимым снижением пороговой точки. Приведены результаты статистического моделирования.

Ключевые слова: робастность, М-оценки, параметр масштаба.

Один из подходов к оптимизации и принятию решений в условиях неопределенности вероятностных моделей сигналов и помех связан с использованием робастных статистических методов, обеспечивающих устойчивость и надежность результатов статистического анализа к возможным отклонениям от принятых гипотез о распределениях [1]. В свою очередь задачи робастного оценивания параметра масштаба распределений занимают второе по значению место после задач робастного оценивания параметра положения распределений [2]. В данной статье предлагаются быстрые высокоэффективные и робастные оценки параметра масштаба симметричных распределений.

Оценкой масштаба называется любая положительная статистика S_n, которая удовлетворяет равенству $S_n(ax_1, \ldots, ax_n) = aS_n(x_1, \ldots, x_n)$ для $a > 0$ [2–4]. Такая оценка определяет степень разброса значений в выборке и может использоваться как в непараметрических задачах, так и для оценивания значений неизвестного параметра масштаба семейства распределений вероятности. Примером такого параметра может служить параметр σ нормального распределения с плотностью вероятности $N(0, \sigma^2) = \exp(-x^2/2\sigma) / (2\pi\sigma^2)^{1/2}$, а возможной оценкой – стандартное отклонение *SD*.

Эта часто используемая оценка обладает серьезным недостатком: при небольших отклонениях от предполагаемой модели результат может оказаться далеким от истинного. Такими отклонениями могут быть загрязненные данные или ошибочные предположения о законе распределения генеральной совокупности. Классическая статистика рассматривает идеализированные условия, но, согласно [3], в статистических данных, как правило, встречается от 1 до 10 % больших ошибок, а некоторые ряды измерения после тщательной проверки не подтверждают принадлежность нормальному распределению, имея более тяжелые «хвосты».

Подобные отклонения от идеальной модели рассматриваются в робастной статистике. В зависимости от задачи в качестве робастных альтернатив для оценивания масштаба чаще всего используют межквартильный размах $IQR = F^{-1}(3/4) - F^{-1}(1/4)$ или абсолютное медианное отклонение от медианы $MAD = \mathrm{med}_i \mid x_i - \mathrm{med}\, x \mid$. Одним из критериев робастности в данном случае служит пороговая точка оценки, т. е. наибольший процент наблюдений, который можно заменить произвольно большими значениями, оказав лишь ограниченное влияние на значение самой оценки [2]. По этому критерию абсолютное медианное отклонение является более предпочтительным, поскольку оно имеет максимально возможную пороговую точку 50 %. Пороговые точки стандартного отклонения и межквартильного размаха равны 0 и 25 % соответственно.

Стандартное отклонение *SD* имеет минимальную возможную дисперсию для нормального распределения. Робастные же оценки обладают более высокой пороговой точкой, как правило за счет большей дисперсии, т. е. меньшей точности. В этом случае эффективность оценки *MAD* равна всего 36,7 %, поэтому возникает необходимость построения более эффективных оценок с максимальной пороговой точкой.

В работе [5] были предложены такие оценки, одна из которых, а именно Q_n, впоследствии стала часто использоваться на практике. Оценка Q определяется как первый квартиль расстояний между наблюдениями: $Q = \{| x_i - x_j |\}_{(k)}$, $k = C(h, 2)$, $h = [n/2] + 1$, и имеет асимптотическую эффективность 82,3 % и пороговую точку 50 %. Серьезным недостатком является большая вычислительная сложность такой оценки, так как задействовано n^2 разностей между парами значений. Отметим, что даже более эффективный алгоритм [6] требует в 3–5 раз больше времени, чем *MAD*.

В данной статье используется параметрическое семейство оценок масштаба, имеющих такую же или большую эффективность за счет снижения пороговой точки, которая, тем не менее, остается в разумных пределах. Введенный параметр позволяет сохранить баланс между пороговой точкой и эффективностью в зависимости от решаемой задачи.

Постановка задачи. В теории робастности важным инструментом для анализа оценок является подход на основе функции влияния. Функция влияния $IF(x; S, F)$ оценки S на модельном распределении F показывает устойчивость оценки к большим ошибкам в точке x и определяется как ее производная по Гато [3]. Построим оценку масштаба с функцией влияния, совпадающей с $IF(x; Q, F)$, тем самым обеспечив совпадение выражающихся через нее характеристик, в частности асимптотической дисперсии и эффективности.

Рассмотрим класс *M*-оценок масштаба *S*, задаваемых неявным уравнением

$$\sum \chi(x_i / S) = 0,$$

где χ – некоторая оценочная функция, обычно четная и неубывающая при $x > 0$. Выбирая вид функции χ, можно получать различные как робастные, так и неробастные оценки масштаба. Этот класс был введен Хьюбером и подробно рассмотрен в [2; 3].

Известно, что функция влияния таких оценок с точностью до нормирующего множителя совпадает с выбранной оценочной функцией $IF(x; S, F) \propto \chi(x)$. Воспользуемся этим фактом и построим M-оценку масштаба M_α, приближающую Q, положив

$$\chi_\alpha(x) = c_\alpha - \alpha^{-1}(\Phi(x+\alpha) - \Phi(x-\alpha)), \quad \alpha > 0,$$

где $\Phi(x)$ – стандартное нормальное распределение; c_α выбирается из условия состоятельности оценки $\int \chi(x)\, d(x) = 0$. При $\alpha = \Phi^{-1}(5/8)\ \sqrt{2}\ = 0{,}450\ 6$ функция влияния соответствующей оценки M_α будет совпадать с $IF(x; Q, \Phi)$ [5].

Основной результат. Для удобства будем рассматривать другое параметрическое семейство, взяв первые несколько членов разложения $\Phi(x)$ в ряд Тейлора:

$$\chi_\alpha(x) = c_\alpha - \frac{1}{3}\ (6 - \alpha^2\,(x^2 - 1))\ \varphi(x),\ c_\alpha = \\ = (12 - \alpha^2) / (12\sqrt{\pi}).$$

Такое представление позволит получить результат в явном виде через элементарные функции.

Из-за своей простоты также будет интересен частный случай при $\alpha = 0$:

$$\chi_0(x) = 1 / \sqrt{\pi} - 2\varphi(x).$$

Исследования показали, что не имеет смысла рассматривать разложение с большей точностью: выигрыш в характеристиках минимален, а объем вычислений возрастает. В качестве альтернативы можно взять функцию $\chi_{\alpha,\beta}(x)$ с произвольным полиномом второй степени от x^2, но в свете полученных результатов это также представляется избыточным.

Получим характеристики предложенной оценки. Формула для асимптотической дисперсии M-оценок масштаба выглядит следующим образом [3]:

$$V(M_\alpha, \Phi) = \int IF^2(x; M_\alpha, \Phi) d\Phi(x) = \frac{\int \chi_\alpha^2(x)\varphi(x)dx}{(\int x\chi'_\alpha(x)\varphi(x)dx)^2},$$

что дает выражение для эффективности

$$e_\alpha = \frac{1}{2V(M_\alpha, \Phi)} = \\ = \frac{81(\alpha^2-4)^2}{8(432(2\sqrt{3}-3) - 24(8\sqrt{3}-9)\alpha^2 + (16\sqrt{3}-9)\alpha^4}.$$

Максимальная достижимая эффективность составляет 95,9 %, но даже при $\alpha = 0$ она не опускается ниже уровня 80,8 % (рис. 1).

Пороговая точка оценки определяется соотношением

$$\varepsilon^*_\alpha = \frac{-\chi_\alpha(0)}{\chi_\alpha(\infty) - \chi_\alpha(0)} = \frac{12(\sqrt{2}-2) - (\sqrt{2}-4)\alpha^2}{4(\alpha^2-6)}$$

в случае четной монотонно возрастающей при $x > 0$ функции $\chi(x)$. Эти условия будут выполнены при $[0;\ \sqrt{2}]$, причем при нуле достигается максимум $\varepsilon^*_\alpha = 29{,}3$ % (рис. 2).

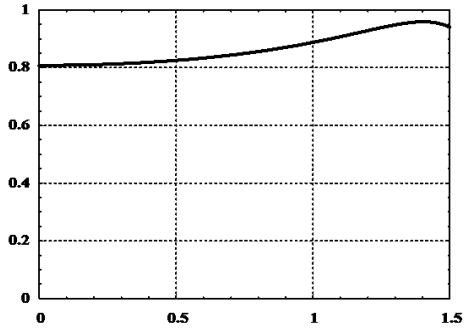


Рис. 1. Эффективность оценки в зависимости от параметра α

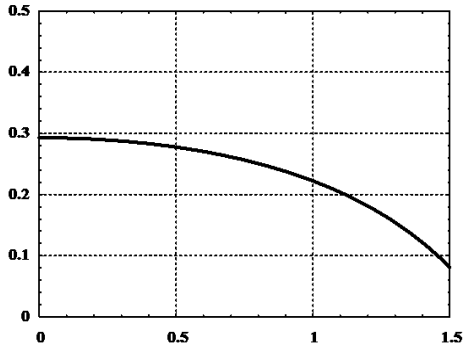


Рис. 2. Пороговая точка оценки в зависимости от параметра α

Вычисление оценки как решения неявного уравнения в большинстве случаев затруднительно, но при этом возможно применение итеративных схем. В частности, можно ограничиться первой итерацией, получив так называемую M-оценку [2]:

$$S_n^{(1)} = S_n^{(0)} - \frac{\sum \chi(x_i / S)}{\left.\frac{\partial}{\partial S}\sum \chi(x_i / S)\right|_{S=S_n^{(0)}}}.$$

Начальное приближение должно быть само по себе в высшей степени робастно. Подставляя предложенную оценочную функцию, получаем, что одношаговая оценка при $\alpha = 0$ задается как уточнение медианы абсолютных отклонений:

$$M_0 = 1{,}483 MAD \cdot \left(1 - \frac{Z_0 - n/\sqrt{2}}{Z_2}\right),$$

$$Z_k \sum u_i^k e^{-u_i^2/2}, \quad u_i = \frac{x_i - \operatorname{med} x}{1{,}483 MAD}.$$

Множитель перед MAD необходим для того, чтобы обеспечить состоятельность оценки на нормальном распределении.

Следует отметить, что функция влияния, а значит асимптотическая дисперсия и эффективность одношаговой оценки, будут несколько отличаться от первоначальных теоретических результатов. Но важным свойством таких оценок является то, что они наследуют пороговую точку начального приближения [6]. При выборе в качестве основы медианы абсолютных отклонений пороговая точка одношаговой оценки M_α повышается до 50 %.

Сравнение оценок. Приведем численные результаты статистического моделирования оценок при 50 000 испытаний на стандартном нормальном распределении (табл. 1). Параметр α имеет значение 0,450 6, что обеспечивает совпадение основных асимптотических характеристик оценок M_α и Q. Время вычисления оценок в миллисекундах соответствует конфигурации Intel Core i7 с частотой 2,8 ГГц. Влияние отклонений от идеальных условий проверялось на модели больших ошибок Тьюки вида $(1 - \varepsilon)\,\Phi(x) + \varepsilon\,\Phi(x/\sigma)$ при $\varepsilon = 0{,}1$, $\sigma = 3$. Результаты моделирования представлены в табл. 2. Таким образом, предложена оценка масштаба, имеющая такую же функцию влияния, асимптотическую дисперсию и эффективность, что и часто используемая оценка Q_n. Результаты моделирования показывают, что одношаговый алгоритм вычисления оценки не только требует значительно меньшего времени, но и обеспечивает меньшее смещение относительно оцениваемой величины, особенно на малых выборках. При подстановке медианы абсолютных отклонений в качестве начального приближения пороговая точка имеет максимально возможное значение 50 %, а свободный параметр позволяет повысить эффективность оценки до 95 %.

Библиографические ссылки

1. Цыпкин Я. З. Информационная теория идентификации. М. : Наука, 1995.

2. Хьюбер Дж. П. Робастность в статистике : пер. с англ. М. : Мир, 1984.

3. Робастность в статистике. Подход на основе функций влияния : пер. с англ. / Ф. Хампель, Э. Рончетти, П. Рауссеу, В. Штаэль. М. : Мир, 1989.

4. Шуленин В. П. Введение в робастную статистику. Томск : Изд-во Том. ун-та, 1993.

5. Rousseeuw P. J., Croux C. Alternatives to the median absolute deviation // J. of the American Statistical Association. 1993. Vol. 88, № 424. P. 1273–1283.

6. Rousseeuw P. J., Croux C. The bias of k-step M-estimators // Statistics & Probability Letters. 1994. Vol. 20, № 5. P. 411–420.

Таблица 1

Математическое ожидание и стандартизованная дисперсия оценок на нормальном распределении в зависимости от размера выборки

n	Среднее				Дисперсия				Время, мс			
	SD	*MAD*	*Q*	M_α	*SD*	*MAD*	*Q*	M_α	*SD*	*MAD*	*Q*	M_α
20	0,986	0,958	1,190	0,951	0,532	1,365	0,789	0,656	0,003	0,004	0,007	0,005
60	0,996	0,986	1,062	0,985	0,511	1,350	0,676	0,632	0,010	0,011	0,036	0,014
200	0,998	0,996	1,019	0,995	0,499	1,345	0,638	0,616	0,034	0,034	0,164	0,045
1 000	1,000	0,999	1,004	0,999	0,493	1,364	0,605	0,609	0,168	0,169	1,022	0,228
∞	1,000	1,000	1,000	1,000	0,500	1,361	0,608	0,608	–	–	–	–

Таблица 2

Математическое ожидание и стандартизованная дисперсия оценок на нормальном распределении с загрязнением ($\varepsilon = 10$ %, $\sigma = 3$)

n	Среднее				Дисперсия			
	SD	*MAD*	*Q*	M_α	*SD*	*MAD*	*Q*	M_α
20	1,290	1,047	1,345	1,083	1,601	1,447	0,993	0,886
60	1,324	1,070	1,195	1,110	1,750	1,410	0,875	0,829
200	1,336	1,078	1,144	1,120	1,815	1,426	0,824	0,807
1 000	1,340	1,081	1,126	1,124	1,823	1,415	0,794	0,808

P. O. Smirnov, G. L. Shevlyakov

APPROXIMATION OF THE Q_N-ESTIMATE OF SCALE WITH THE HELP OF FAST *M*-ESTIMATES

Popular highly efficient robust estimates of scale including the Q_n estimate are considered. A parametric family of M-estimates which allows to faster computations with acceptable decrease in breakdown points is offered. The results of the Monte-Carlo simulation are given.

Keywords: robustness, M-estimates, scale parameter.

© Смирнов П. О., Шевляков Г. Л., 2010

УДК 519.24

В. Ф. Первушин, Н. А. Сергеева, А. В. Стрельников

ПРЕЦИЗИОННЫЙ ГЕНЕРАТОР СЛУЧАЙНЫХ ЧИСЕЛ

Рассматривается алгоритм генерации выборки случайной величины, имеющей заданный закон распределения и численные характеристики (математическое ожидание, дисперсию и т. д.). Показывается высокая точность работы алгоритма на численных примерах моделирования выборки. Общая концепция генерации выборки позволяет выделить данный подход в отдельную категорию алгоритмов генерации случайных чисел, что повлекло объединение таких алгоритмов под названием «прецизионный генератор случайных чисел».

Ключевые слова: случайная величина, закон распределения, числовые характеристики, вероятность события, частота события, объем выборки, гистограмма.

Тему разработок, о которых пойдет речь в данной статье, подсказала задача стохастического моделирования. Если модель построена адекватно физическому процессу, то на ней можно проводить численные эксперименты, связанные с изучением поведения объекта или процесса при изменении входных параметров и влияния внешних воздействии. Также немаловажно более глубоко изучить физические и математические связи внутри самого процесса, найти закономерности и установить меру взаимного влияния параметров и переменных объекта, получить новые качественные знания о процессе, восполнить потерянные или неточные данные. Все эти задачи бывает затруднительно, а порой и невозможно изучать на реальном объекте. Понятно, что некоторые ситуации, если их реализовать на практике, потребуют временных и материальных затрат, изменения хода всего процесса или даже режима работы объекта, чего зачастую допустить нельзя. Именно в таких случаях важность задачи имитационного моделирования становится наиболее значимой. Ведь на модели можно реализовать различные ситуации, в том числе и нетипичные для данного процесса или объекта, и провести наблюдение за его поведением при измененных параметрах входа или зависимостях интересующих исследователей показателей.

Ниже предлагается общая схема процесса, принятая в теории идентификации [1] (рис. 1). Здесь приняты следующие обозначения: $x(t)$ – векторная выходная переменная процесса; $u(t)$ – векторное управляющее воздействие; $\mu(t)$ – векторная входная переменная процесса; $\omega^i(t): i=1,2,...,k$ – переменные процесса, контролируемые по длине объекта; $\xi(t)$ – векторное случайное воздействие; t, заключенное в круглые скобки, – непрерывное время; H со значком вверху – каналы связи, соответствующие различным переменным и включающие в себя средства контроля, приборы для измерения наблюдаемых переменных; значок t внизу переменных x,ω,u,μ означает дискретное время; $h(t)$ – со значком вверху – случайные помехи измерений соответствующих переменных процесса. Контроль переменных x,ω,u,μ осуществляется через интервал времени Δt, т. е. $x_i,\omega_i^1,...,\omega_i^k,u_i,\mu_i$, где $i=\overline{1,s}$ – выборка измерений переменных процесса $\left(x_1,\omega_1^1,...,\omega_1^k,u_1,\mu_1\right)$, $\left(x_2,\omega_2^1,...,\omega_2^k,u_2,\mu_2\right),...,\left(x_s,\omega_s^1,...,\omega_s^k,u_s,\mu_s\right),...$, здесь s – объем выборки.

Материалы, представленные в данной статье, имеют непосредственное отношение к величинам $h(t)$. Речь пойдет о моделировании случайных помех, имеющих определенный закон распределения и числовые характеристики.

Приведем далее некоторые размышления о методе статистического моделирования и его месте в научных исследованиях. Метод статистического моделирования широко используется в разнообразных задачах кибернетики для исследования алгоритмов идентификации, распознавания образов, управления и т. д. В последние годы появилось много эвристических алгоритмов, прежде всего в связи с тем, что новые формулировки задач не поддаются строгой математической постановке. А это, как правило, означает отсутствие процедуры аналитического синтеза тех или иных алгоритмов, доказательство соответствующих теорем сходимости, наличие которых ранее считалось мерой истинности, правильности и основанием того, что дальнейшие действия являются правомерными. В этой связи метод статистического моделирования следует считать доказательным этапом, а не иллюстрацией работы тех или иных алгоритмов. Последнее существенно повышает требования к проведению подобных исследований. Определяющей здесь является возможность повторения определенного цикла численных экспериментов другими исследователями.

Во многих случаях на этом пути безусловно важной является необходимость работы со случайными помехами, распределенными по конкретному закону, а не просто с датчиками случайных чисел, о которых говорят, что они распределены по равномерному, нормальному либо какому-то другому закону. Как оказалось на практике, существующие генераторы случайных чисел весьма условно можно назвать соответствующими заявленным законам распределения. Особенно заметно отклонение на выборках небольших объемов. Все классические алгоритмы генерации [2] верны прежде всего на бесконечно большом объеме выборки, ведь все теоремы сходимости содержат формулировку «Пусть объем выборки стремится в бесконечность» или тому подобное допущение. Однако если объем выборки и достаточен для того, чтобы

служить основой для построения модели, то он все равно далек от того, чтобы применить к нему термин «бесконечный». Это говорит о том, что следует задуматься о создании алгоритмов генерации конечной выборки случайной величины и способ создания такой величины будет существенно отличаться от классического подхода. Мы ни в коем случае не спорим с классикой, однако нам хотелось бы больше внимания уделить задачам практики и увязать классический подход с прикладными задачами.

При исследовании многих стохастических процессов предполагается, что плотность вероятности случайных факторов (помех) подчиняется какому-либо закону. В частности, что ошибка измерения электрических и неэлектрических величин электрическими способами подчиняется нормальному закону, в задачах теории надежности используется распределение Вейбулла (интенсивность отказа), логнормальное распределение характерно для многих физических и социально-экономических ситуаций (размер и вес частиц, образующихся при дроблении; заработная плата работника; размеры космических образований; долговечность изделия, работающего в режиме износа и старения, и др.) [3].

Для того чтобы сгенерировать множество случайных точек, свойства которых подчиняются одному из известных законов распределения, необходимо задать закон распределения и его параметры, количественно характеризующие случайную величину, такие как математическое ожидание, среднеквадратическое отклонение, мода и др., и определить количество случайных точек.

Рассмотрим генерацию множества точек для одномерной случайной величины (рис. 2).

Пусть необходимо сгенерировать статистически независимую выборку случайной величины X с плотностью вероятности распределения $f(x, m_x, \sigma_x, ...)$. Обозначим через n количество генерируемых точек, т. е. объем будущей выборки. Далее зададим интервал $[a, b]$, на котором будет генерироваться выборка.

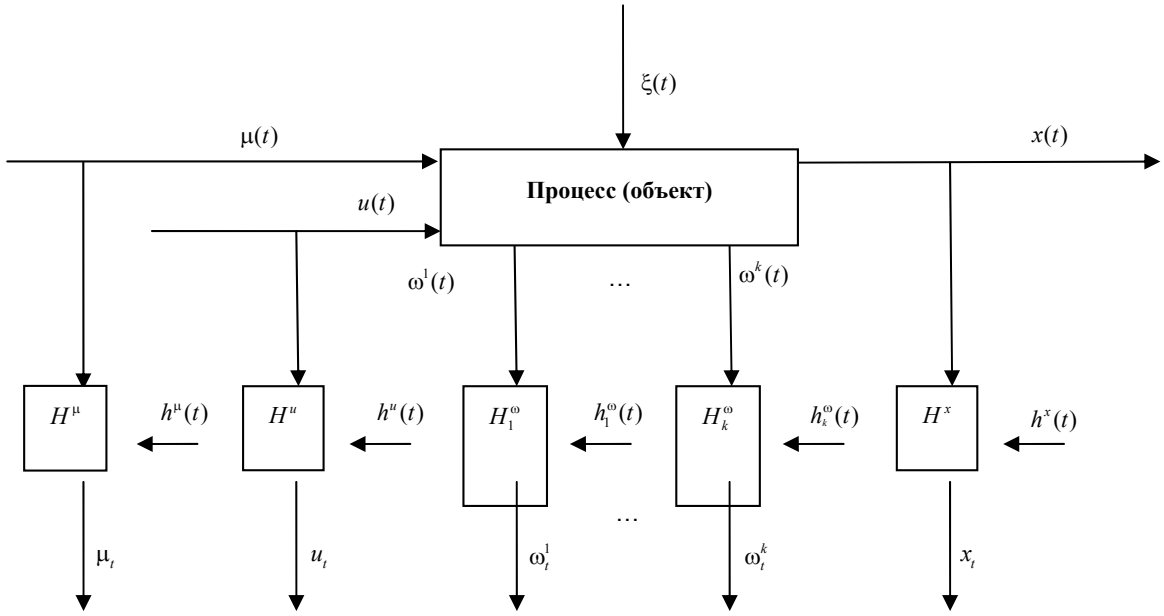


Рис. 1. Схема процесса

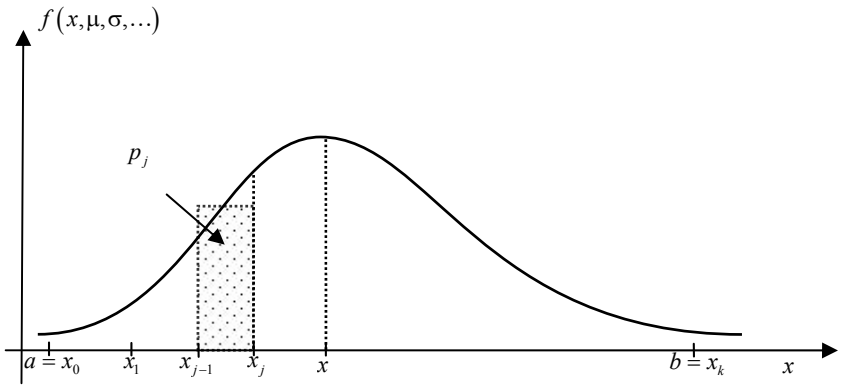


Рис. 2. Графическая интерпретация алгоритма

Выборка генерируемых точек должна наиболее полно охватывать всю область возможных значений, но по понятным причинам следует отрезать «хвосты» распределения, где вероятность выпадения значений X становится малой. В этом случае нужно определить этот порядок малости и, например, принять к получению выборки тот интервал, внутри которого $f(x, m_x, \sigma_x, ...)$ не меньше некоторого значения. Определим это значение пропорционально $f_{\max}(x, m_x, \sigma_x, ...)$:

$$f(x = a, m_x, \sigma_x, ...) = f(x = b, m_x, \sigma_x, ...) = \\ = 0,01 \cdot f_{\max}(x, m_x, \sigma_x, ...). \quad (1)$$

Для распределений, которые начинаются от $x = 0$, соответственно примем $a = 0$, а b получим из вышеприведенных рассуждений. Если границы интервалов не удается определить аналитически, то можно применить любые численные процедуры решения нелинейного уравнения. Для нахождения экстремума распределения также могут потребоваться численные методы, если это невозможно сделать аналитически.

Весь интервал разобьем на множество равных подынтервалов $[a = x_0, x_1, ..., x_k = b]$. Количество интервалов, равное k, будет задаваться пользователем. Внутри каждого подынтервала среднее значение функции плотности определяется по формуле

$$\frac{f(x_{j-1}) + f(x_j)}{2} = f(\tilde{x}_j), \quad (2)$$

где $\tilde{x}_j \in [x_{j-1}, x_j]$.

Вероятность попадания случайной величины X в *j*-й интервал есть площадь, ограниченная функцией плотности на отрезке $[x_{j-1}, x_j]$. Обозначим ее через p_j:

$$P(x_{j-1} \le X \le x_j) = p_j = \frac{f(x_{j-1}) + f(x_j)}{2} \cdot (x_j - x_{j-1}). \quad (3)$$

На основании закона больших чисел с ростом объема выборки значений случайной величины *n* частота наступления случайного события стремится к вероятности этого события. Тогда можно принять, что

$$p_j = \frac{n_j}{n}, \quad (4)$$

где n_j – количество попаданий значений случайной величины X в интервал $[x_{j-1}, x_j]$ из общего объема значений.

Определим количество точек n_j, которые нужно поместить в интервал $[x_{j-1}, x_j]$ для того, чтобы сгенерированная выборка соответствовала заданному закону распределения (например, равномерному закону или любой встроенной функции генерации конкретного языка программирования):

$$n_j = \left[p_j \cdot n \right]. \quad (5)$$

В данном случае качество встроенного генератора случайных чисел становится некритичным.

Совокупность точек, полученных описанным выше способом на всех интервалах, образует выборку случайной величины X, распределенной по заданному закону $f(x, m_x, \sigma_x, ...)$, и обладает всеми требуемыми значениями параметров. Выбор точек из представленного набора значений может производиться также любой встроенной функцией извлечения данных из массива значений.

Необходимо проверить, что полученная выборка чисел действительно имеет требуемый закон распределения и обладает нужными характеристиками, для чего по этой выборке строится гистограмма распределения, которая сравнивается с заказанной функцией плотности.

Приведем результаты численного моделирования предложенного алгоритма и оценки характеристик случайной величины по сгенерированной выборке.

Рассмотрим случайную величину, распределенную по логнормальному закону $X \sim \log N(\mu, \sigma^2)$, $\mu = 0, \sigma^2 = 0,5$ [4; 5]. Плотность и закон распределения логнормального закона имеют вид

$$f(X, \mu, \sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi} X \sigma} \exp\left(\dfrac{-(\ln X - \mu)^2}{2\sigma^2} \right), & X > 0, \\ 0, & X \le 0, \end{cases}$$

$$F(X, \mu, \sigma) = \int_{-\infty}^{X} f(t, \mu, \sigma) dt. \quad (6)$$

Оценка плотности распределения, приведенная на рис. 3, построена на основе выборки объемом $s = 100$, количество подынтервалов $k = 20$. Выборочные оценки параметров закона распределения, усредненные по 1 000 экспериментам, составили $\hat{\mu} = -0,002$, $\hat{\sigma}^2 = 0,49$.

График, приведенный на рис. 4, построен на основе выборки объемом $s = 500$, количество подынтервалов $k = 30$. Выборочные оценки параметров закона распределения, также усредненные по 1 000 экспериментам, составили $\hat{\mu} = -0,011$, $\hat{\sigma}^2 = 0,497$.

Далее рассмотрим случайную величиу, распределенную по двухпараметрическому закону Вейбулла [4; 6]. Этот закон и плотность его распределения имеют вид

$$F(X, a, b) = \begin{cases} 1 - e^{-\left(\frac{X}{b}\right)^a}, & X \ge 0, \\ 0, & X < 0, \end{cases}$$

$$f(X, a, b) = \begin{cases} \dfrac{a}{b} \left(\dfrac{X}{b} \right)^{a-1} e^{-\left(\frac{X}{b}\right)^a}, & X \ge 0, \\ 0, & X < 0, \end{cases} \quad (7)$$

где *a* – параметр формы; *b* – параметр масштаба.

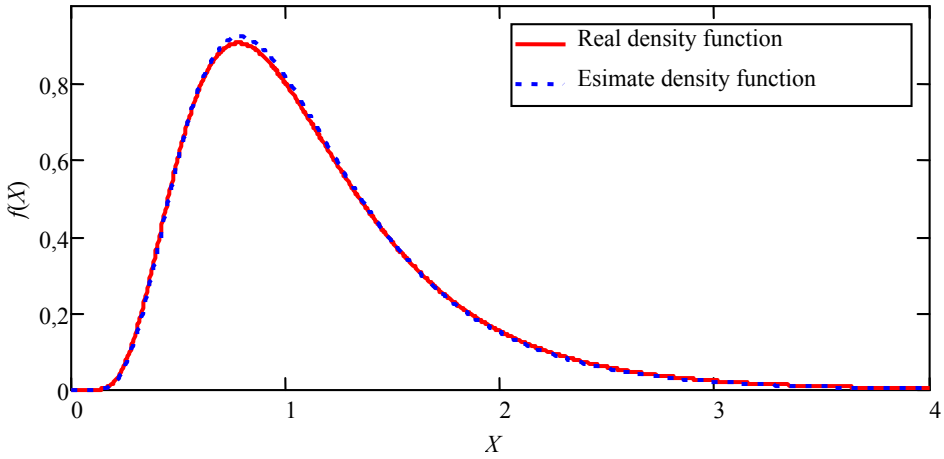


Рис. 3. Плотность распределения логнормального закона и ее оценка:
$S = 100$; $k = 20$

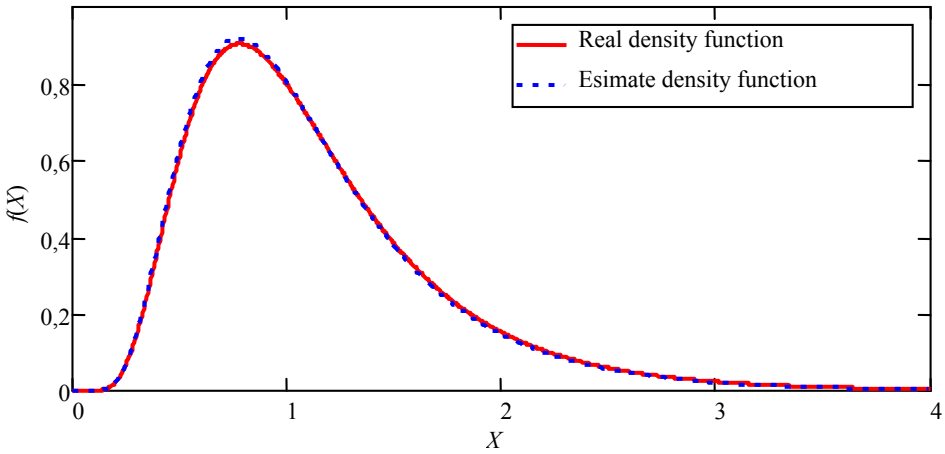


Рис. 4. Плотность распределения логнормального закона и ее оценка:
$S = 500$; $k = 30$

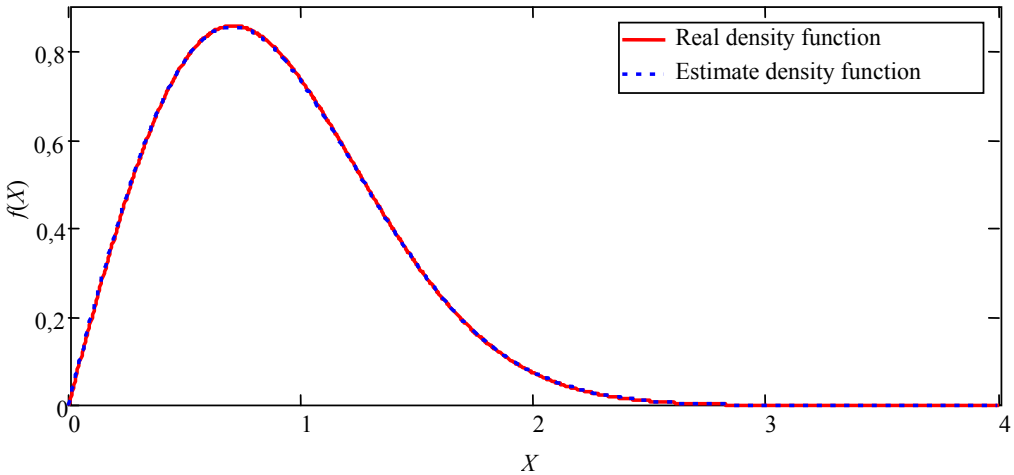


Рис. 5. Функция плотности распределения Вейбулла и ее оценка:
объем генерируемой выборки $s = 100$; количество подынтервалов $k = 12$; истинные значения параметров $a = 2$, $b = 1$; выборочные значения параметров $\hat{a} = 1,991$, $\hat{b} = 1$

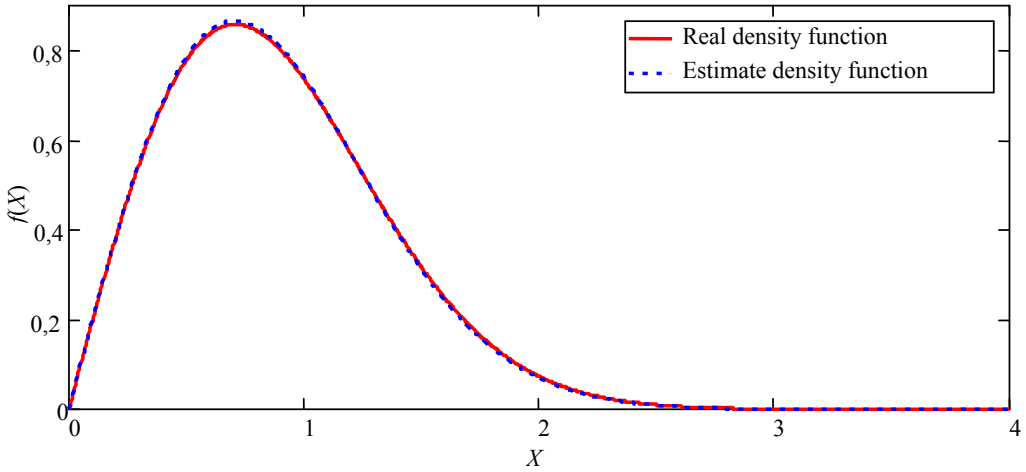


Рис. 6. Функция плотности распределения Вейбулла и ее оценка:
объем генерируемой выборки $s = 500$; количество подынтервалов $k = 30$; истинные значения параметров $a = 2$, $b = 1$; выборочные значения параметров $\hat{a} = 2{,}012$, $\hat{b} = 0{,}994$

В результате проведения экспериментов получены следующие графики (рис. 5, 6).

Построим гистограммы для логнормального распределения (рис. 7, 8) и распределения Вейбулла (рис. 9, 10).

Рис. 7. Гистограмма для логнормального закона:
параметры $\mu = 0$, $\sigma^2 = 0{,}5$; объем выборки $n = 100$,
количество подынтервалов $k = 20$

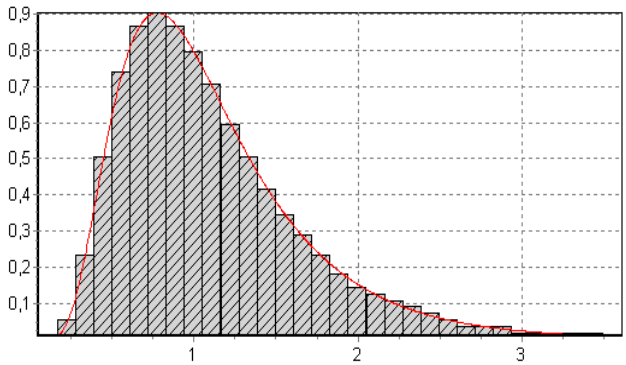


Рис. 8. Гистограмма для логнормального закона:
параметры $\mu = 0$, $\sigma^2 = 0{,}5$; объем выборки $n = 500$,
количество подынтервалов $k = 30$

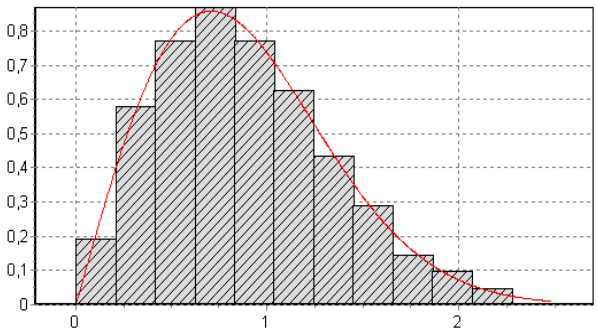


Рис. 9. Гистограмма закона Вейбулла
параметры $a = 2$, $b = 1$; объем выборки $n = 100$,
количество подынтервалов $k = 12$

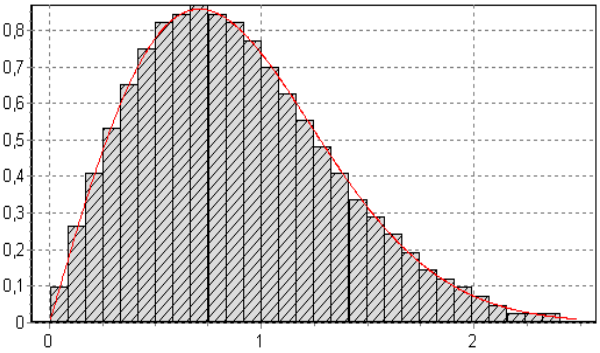


Рис. 10. Гистограмма закона Вейбулла:
параметры $a = 2$, $b = 1$; объем выборки $n = 500$,
количество подынтервалов $k = 30$

Дальнейшее развитие исследований касается построения генераторов для как можно большего количества известных функций распределения, изучения свойств сходимости каждого конкретного распределения, точности оценок числовых параметров распределения и случайной величины.

Библиографические ссылки

1. Медведев А. В. Теория непараметрических систем. Моделирование // Вестник СибГАУ. 2010. Вып. 4. С. 4–9.

2. Ермаков С. М., Михайлов Г. А. Курс статистического моделирования. М. : Наука, 1976.

3. Фадеева Л. Н. Математика для экономистов: Теория вероятностей и математическая статистика. М. : Эксмо, 2006.

4. Вероятность и математическая статистика : энциклопедия / под ред. Ю. В. Прохорова М. : Большая Рос. энцикл., 2003.

5. Первушин В. Ф., Сергеева Н. А. Генератор случайных чисел, распределенных по логнормальному закону // Решетневские чтения : материалы XIII Междунар. науч. конф. : в 2 ч. Ч. 2 / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2009. С. 448–449.

6. Сергеева Н. А., Стрельников А. В. Генератор случайных чисел, распределенных по закону Вейбулла // Решетневские чтения : материалы XIII Междунар. науч. конф. : в 2 ч. Ч. 2 / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2009. С. 451–453.

V. F. Pervushin, N. A. Sergeeva, A. V. Strelnikov

THE RANDOM SELECTION PRECISION GENERATOR

The algorithm of random value selection with adjusted distributions generation and numerical characteristics is considered (mathematical estimation, dispersion etc.).The algorithm proceedings of high precisions is showed on calculated examples of selection modeling. The selection generation common concept allows to extract this approach to the special category of random value generation algorithms. It was joined to the term "The Random selection precision generator".

Keywords: random value, retribution law, numeric characteristics, event probability, event frequency, selection size, bar chart.

© Первушин В. Ф., Сергеева Н. А., Стрельников А. В., 2010

УДК 519.6

В. Б. Бериков

АЛГОРИТМ АДАПТИВНОГО ПЛАНИРОВАНИЯ АНСАМБЛЯ ТАКСОНОМИЧЕСКИХ ДЕРЕВЬЕВ РЕШЕНИЙ*

Рассматривается подход к решению задач кластерного анализа, основанный на применении ансамбля таксономических деревьев решений. Предлагается алгоритм адаптивного планирования ансамбля, использующий расстояния между логическими высказываниями, описывающими кластеры. Приводятся результаты статистического моделирования, подтверждающие эффективность алгоритма.

Ключевые слова: кластерный анализ, ансамбль, дерево решений.

Одной из актуальных проблем кластерного анализа (таксономии, автоматической классификации «без учителя») является группировка объектов, описываемых разнотипными (количественными или качественными) переменными. Другая актуальная проблема связана с повышением устойчивости группировочных решений, так как в большинстве алгоритмов кластерного анализа результаты могут сильно меняться в зависимости от выбора начальных условий, порядка объектов, параметров работы алгоритма и т. п.

Наиболее перспективный подход к кластерному анализу при наличии разнотипных переменных основан на применении деревьев решений, которые позволяют получать легко интерпретируемую логическую модель группировки, выделять наиболее информативные факторы и не требуют задания метрики в разнотипном пространстве. Особенностью логико-вероятностного подхода, основанного на деревьях решений, является возможность не только разбивать заданное множество объектов на кластеры, но и строить иерархическое дерево, описывающее структуру разбиения.

Повысить устойчивость кластеризации можно с помощью ансамблей алгоритмов. При этом используются результаты группировки, полученные различными алгоритмами или одним алгоритмом, но с различными параметрами настройки, по различным подсистемам переменных и т. д. После построения ансамбля проводится нахождение итогового коллективного решения. Такой способ описан, например, в работе [1].

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты 08-07-00136а, 10-01-00113а, 09-07-12087-ofi_m) и Междисциплинарного интеграционного проекта Сибирского отделения Российской академии наук № 83.

В работе [2] предложен алгоритм построения коллектива таксономических деревьев решений. При построении ансамбля использовалось введенное в [3] понятие расстояния между логическими высказываниями, описывающими кластеры. В данной статье описано усовершенствование этого алгоритма путем применения методики адаптивного планирования, при которой ведется целенаправленная селекция наиболее перспективных элементов ансамбля.

Основные понятия. Пусть имеется множество $s = \{o^{(1)}, ..., o^{(N)}\}$ некоторых объектов, выбранных из генеральной совокупности. Каждый объект описывается с помощью набора переменных $X_1, \ldots, X_n$. Этот набор может включать в себя переменные разных типов (количественные и качественные, под которыми будем понимать номинальные и булевы, а также порядковые переменные). Пусть D_j обозначает множество значений переменной X_j – некоторый интервал числовой оси в случае количественной переменной или конечный набор значений (имен) в случае качественной переменной. Пусть $D = D_1 \times ... \times D_n$. Обозначим через $x = x(o) = (x_1(o), ..., x_n(o))$ набор наблюдений переменных для объекта o, где $x_j(o)$ – значение переменной X_j для данного объекта.

В задаче кластерного анализа требуется разбить объекты на некоторое число кластеров K ($K << N$) так, чтобы заданный критерий качества группировки принял бы оптимальное значение. Под критерием качества обычно понимается некоторый функционал, зависящий от разброса объектов внутри группы и расстояний между группами. Число классов может быть выбрано заранее или не задано (в последнем случае оптимальное количество кластеров должно быть определено автоматически).

Введем понятие таксономического дерева решений. Рассмотрим дерево, в котором каждой внутренней вершине (узлу) соответствует некоторая переменная X_j, а ветвям, выходящим из данной вершины, – определенное высказывание вида $X_j(o) \in E_j^{(i)}$, где o – некоторый объект; $i = 1, ..., v$, $v \ge 2$ – число ветвей, выходящих из данной вершины; $E_j^{(1)}, ..., E_j^{(v)}$ – попарно непересекающиеся подмножества множества D_j (интервалы значений в случае количественной переменной). Каждому k-му листу (концевой вершине) дерева соответствует группа объектов выборки $C^{(k)} = \{o^{(i_1)}, ..., o^{(i_{n^{(k)}})}\}$, где $n^{(k)}$ – число объектов, входящих в группу, $k = 1, ..., K$. Объекты данной группы удовлетворяют цепочке высказываний, проверяемых по пути из корневой вершины в этот лист, т. е. логическому утверждению вида «Если $X_{j_1}(o) \in E_{j_1}^{(i_1)}$ И $X_{j_2}(o) \in E_{j_2}^{(i_2)}$ И ... И $X_{j_{q_k}}(o) \in E_{j_{q_k}}^{(i_{q_k})}$, то объект o относится к k-му классу», где q_k – длина данной цепочки, $o \in C^{(k)}$. Под таксоном $T^{(k)}$, соответствующим k-му листу дерева, будем понимать прямоугольную подобласть, заключающую объекты из $C^{(k)}$ в многомерном пространстве:

$$T^{(k)} = T(C^{(k)}) = T_1^{(k)} \times ... \times T_j^{(k)} \times ... \times T_n^{(k)},$$

где $T_j^{(k)} = \{X_j(o) \mid o \in C^{(k)}\}$ для качественной переменной X_j; $T_j^{(k)} = [\min\limits_{o \in C^{(k)}} X_j(o); \max\limits_{o \in C^{(k)}} X_j(o)]$ для количественной переменной X_j.

В работе [2] описан регуляризующий критерий качества группировки, основанный на минимизации функционала, зависящего от суммарного относительного объема таксонов и от их числа:

$$Q = \sum_{k=1}^{K} \prod_{j=1}^{n} \frac{|T_j^{(k)}|}{|D_j|} + \alpha \frac{K}{N},$$

где $\alpha > 0$ – регуляризующий параметр, а также предложен алгоритм направленного поиска оптимального дерева, в котором применяется рекурсивная схема перебора различных вариантов.

Построение ансамбля деревьев решений. Пусть с помощью алгоритма построения таксономического дерева решений строится разбиение множества объектов s на K подмножеств. Для построения каждого варианта случайным образом формируется своя подсистема переменных (либо случайным образом выбираются параметры алгоритма). Под группировочным решением будем понимать набор $G = \{C^{(1)}, ..., C^{(k)}, ..., C^{(K)}\}$, где $C^{(k)} = \{o^{(i_1)}, ..., o^{(i_{N_k})}\}$; N_k – число объектов, входящих в k-й кластер, $k = 1, ..., K$. Пусть построен набор группировочных решений $\mathbf{G} = \{G^{(1)}, G^{(2)}, ..., G^{(L)}\}$, где $G^{(l)}$ – l-й вариант группировки. Обозначим через $S^{(l)}$ бинарную матрицу размерности $N \times N$, которая водится для l-й группировки следующим образом: $S^{(l)}(q, m) = 0$, если объекты $o^{(q)}$ и $o^{(m)}$ принадлежат одному кластеру, $S^{(l)}(q, m) = 1$, если иначе, где q, m = 1, 2, …, N; l = 1, 2, …, L. После построения L группировочных решений формируется согласованная матрица различий

$$S(q, m) = \frac{1}{L} \sum_{l=1}^{L} S^{(l)}(q, m),$$

где q, m = 1, 2, …, N. Величина $S(q, m)$ равна частоте классификации объектов $o^{(q)}$ и $o^{(m)}$ на разные группы в наборе группировок $\mathbf{G}$. Близкое к единице значение этой величины означает, что данные объекты имеют большой шанс попадания в одну и ту же группу, а близкое к нулю говорит о том, что шанс оказаться в одной группе у этих объектов незначителен.

После вычисления согласованной матрицы подобия для нахождения итогового варианта группировки можно применять алгоритм построения дендрограммы, который в качестве входной информации использует расстояния между объектами. В предлагаемом нами алгоритме используется аналогичный принцип,

однако вместо бинарной матрицы $S^{(l)}$, в которой отражаются события типа вхождения (либо невхождения) пары объектов в одну и ту же группу, берется более информативная матрица расстояний между кластерами, к которым отнесены объекты. Для вычисления расстояний между кластерами в разнотипном пространстве переменных будем использовать введенное в работе [3] расстояние между логическими высказываниями экспертов.

Пусть в одном и том же варианте группировки имеются два таксона $T^{(s)}$ и $T^{(r)}$, $s, r = 1, \ldots, K$, описываемые некоторыми логическими высказываниями. Введем расстояние между данными таксонами как $\rho(T^{(s)}, T^{(r)}) = f(\rho_1(T_1^{(s)}, T_1^{(r)}), ..., \rho_j(T_j^{(s)}, T_j^{(r)}), ..., \rho_n(T_n^{(s)}, T_n^{(r)}))$, где f – некоторая функция; $\rho_j(T_j^{(s)}, T_j^{(r)})$ – расстояние между j-ми компонентами множеств $T^{(s)}$ и $T^{(r)}$.

Зададим способ вычисления величины ρ_j в зависимости от типа переменной. Если X_j – номинальная переменная, то расстояние определяется как взвешенная мера симметрической разности:

$$\rho_j(T_j^{(s)}, T_j^{(r)}) = \frac{|T_j^{(s)} \Delta T_j^{(r)}|}{|D_j|}.$$

Если X_j – количественная переменная, а интервалы

$$T_j^{(s)} = [a^{(s)}, b^{(s)}], \quad T_j^{(r)} = [a^{(r)}, b^{(r)}],$$

то

$$\rho_j(T_j^{(s)}, T_j^{(r)}) = \frac{|a^{(s)} - a^{(r)}| + |b^{(s)} - b^{(r)}|}{2|D_j|}.$$

В работе [3] было доказано, что для введенного расстояния выполняются все свойства меры, в том числе неравенство треугольника.

В качестве функции f предлагается рассматривать функцию $f_1 = \left(\sum_{j=1}^{n} w_j \rho_j \right)^{1/2}$, где $0 \le w_j \le 1$ – некоторые веса ($\sum_{j=1}^{n} w_j = 1$), либо функцию $f_2 = \max_j \rho_j$, удобную тем, что она не требует задания весов. Для данных функций также выполняются свойства меры.

Таким образом, согласованная матрица различий с учетом расстояний между кластерами определяется как

$$\tilde{S}(q,m) = \frac{1}{L} \sum_{l=1}^{L} \tilde{S}^{(l)}(q,m),$$

где $\tilde{S}^{(l)}(j,m) = \rho(T^{(s)}, T^{(r)})$, $T^{(s)}, T^{(r)}$ – таксоны, к которым принадлежат объекты $o^{(q)}$ и $o^{(m)}$ соответственно.

В алгоритме построения ансамбля [2] выбор подсистемы переменных осуществляется случайно. Однако этот выбор (испытание) можно организовать так, чтобы лучшие по критерию качества группировки варианты имели бы большую вероятность отбора. В основе предлагаемой модификации лежит идея адаптивного случайного поиска наиболее информативной подсистемы переменных, предложенная Г. С. Лбовым.

Первоначально все переменные имеют одинаковый вес, т. е. вероятность отбора в подсистему, используемую при построении дерева. Предполагается, что число переменных достаточно велико. После проведения очередного испытания вес отобранной в корне дерева переменной уменьшается, за счет чего достигается более высокая степень разнообразия ансамбля. Если при этом критерий качества ухудшается, то эта переменная наказывается, т. е. вероятность ее отбора еще более уменьшается пропорционально критерию. В случае когда критерий качества улучшается, соответствующая переменная поощряется, т. е. ее вес уменьшается не столь сильно. То, что выбирается переменная, соответствующая именно корню дерева, объясняется тем, что она является наиболее информативной и оказывает определяющее влияние на дальнейшее ветвление.

Алгоритм планирования состоит из следующих основных шагов.

Шаг 1. Присвоить всем переменным одинаковые веса $P_{X_j} = 1/n, \quad j = 1, ..., n.$

Шаг 2. Выбрать случайным образом d переменных в соответствии с назначенными весами, где d – заданный параметр.

Шаг 3. Построить дерево решений в случайном подпространстве размерности d, в котором X^* – переменная, соответствующая корню дерева. Определить для построенного дерева критерий качества группировки Q.

Шаг 4. Присвоить переменной X^* новый вес:

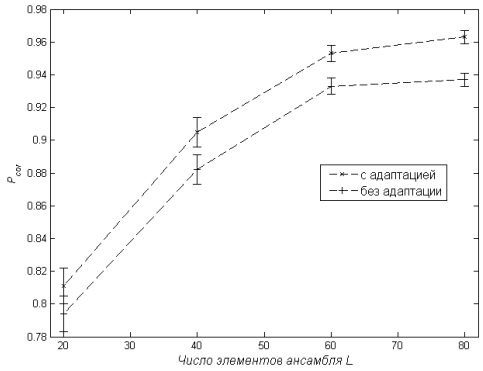
$$P_{X^*} = \max\{P_{X^*} - \tau Q, 0\},$$

где τ – параметр адаптации.

5. Повторить шаги 2…4 до тех пор, пока не будет построено заданное число элементов ансамбля.

Экспериментальное исследование алгоритма. Для определения качества алгоритма была разработана процедура статистического моделирования. Эта процедура состоит в многократном генерировании случайных выборок в соответствии с заданным распределением для каждого класса; построении с помощью исследуемого алгоритма согласованного группировочного решения для каждой выборки; нахождении усредненного по всем выборкам показателя качества. Распределение для каждого из $K = 2$ классов является многомерным нормальным с одной и той же ковариационной матрицей Σ. Число переменных $n = 100$, объем выборки для каждого класса – 25. Вектор математических ожиданий для каждого класса выбирается случайно из множества, соответствующего вершинам единичного гиперкуба. Ковариационная матрица является диагональной: $\Sigma = \sigma \mathbf{I}$, где $\sigma = 1$. Из общего числа переменных 20 являются шумовыми (для них $\sigma = 10$). Номера шумовых переменных выбираются случайно. Каждое дерево строится в слу-

чайно выбранном подпространстве переменных размерности 3. Параметры $\alpha = 2,\ \tau = 0,05$, в качестве функции *f*, задающей вид расстояния между высказываниями, выбрана f_1.



Алгоритм построения коллектива таксономических деревьев решений с адаптацией и без адаптации (указаны 95%-е доверительные интервалы для вероятности правильной классификации)

Качество группировки определяется как частота правильной классификации P_{cor}. Усреднение проводится по 100 случайным выборкам, являющихся реализациями смеси указанных распределений. Оценивается 95%-й доверительный интервал для вероятности правильной классификации. Полученные результаты моделирования позволяют сделать вывод о том, что при достаточно большом числе элементов ансамбля адаптивный алгоритм дает значимо более высокое качество классификации по сравнению с разработанным ранее вариантом, в котором адаптация не используется (см. рисунок).

Таким образом, предложен алгоритм адаптивного планирования ансамбля таксономических деревьев решений, использующий расстояния между логическими высказываниями, описывающими кластеры. Результаты статистического моделирования подтвердили эффективность разработанной процедуры. В дальнейшем планируется применить данный алгоритм в задаче анализа биомедицинской информации, относящейся к устойчивости паразитарной системы клещевого энцефалита, а также для обработки спутниковых и натурных данных.

Библиографические ссылки

1. Strehl A., Ghosh J. Clustering ensembles – a knowledge reuse framework for combining multiple partitions // J. of Machine Learning Research. 2002. Vol. 3. P. 583–617.

2. Бериков В. Б. Кластерный анализ с использованием коллектива деревьев решений // Науч. вестн. Новосиб. гос. техн. ун-та. 2009. № 3 (36). С. 67–76.

3. Лбов Г. С., Бериков В. Б. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. Новосибирск : Изд-во Ин-та математики, 2005.

V. B. Berikov

ALGORITHM FOR ADAPTIVE PLANNING OF AN ENSEMBLE OF TAXONOMIC DECISIONS TREES

We suggest an approach to cluster analysis based on the ensemble of taxonomic decisions trees. The adaptive algorithm for the ensemble planning that uses distances between logic statements describing clusters is offered. The results of statistical modeling confirm the efficiency of the algorithm.

Keywords: cluster analysis, ensemble, decisions tree.

© Бериков В. Б., 2010

УДК 681.3

А. В. Бобров, Е. А. Перепелкин

ВОССТАНОВЛЕНИЕ ИЗОБРАЖЕНИЙ НА ОСНОВЕ СПЕКТРАЛЬНОЙ МАТРИЧНОЙ НОРМЫ

Рассматривается проблема восстановления изображения. Описывается нелинейный гауссовский фильтр, построенный на основе спектральной матричной нормы. Приводятся результаты численных экспериментов, подтверждающие преимущество данного фильтра по сравнению с линейным гауссовским фильтром.

Ключевые слова: восстановление изображения, гауссовский фильтр, спектральная норма матрицы.

Пусть изображение задано в виде неотрицательной вещественной матрицы *A* размером $n \times m$ с элементами $a_{ij} \in [0 \quad a_{\max}]$. Будем считать, что изображение содержит искаженные пиксели. Координаты искаженных пикселей известны, соответствующие элементы матрицы изображения равны нулю. Необходимо восстановить искаженные пиксели.

Для восстановления изображения применяют пространственные и частотные фильтры [1; 2].

Например, оценка значения пикселя a_{ij} в пространственном линейном сглаживающем фильтре определяется по формуле

$$\hat{a}_{ij} = \sum_{p=-r}^{r} \sum_{q=-r}^{r} w_{p+r+1,q+r+1} a_{i+p,j+q}, \qquad (1)$$

где r – радиус фильтра; w_{kl} – весовые коэффициенты фильтра

$$w_{kl} \geq 0, \ \sum_{k=1}^{2r+1} \sum_{l=1}^{2r+1} w_{kl} = 1.$$

В случае равенства весов

$$\hat{a}_{ij} = \frac{1}{(2r+1)^2} \sum_{p=-r}^{r} \sum_{q=-r}^{r} a_{i+p,j+q}.$$

Матрица W размером $(2r+1)\times(2r+1)$, составленная из весов w_{kl}, называется ядром фильтра. Часто применяется гауссовский фильтр, ядро которого определяется функцией

$$w_{kl} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-r-1)^2 + (l-r-1)^2}{2\sigma^2} \right).$$

Радиус гауссовского фильтра r можно задать равным 3σ.

В данной статье описывается нелинейный гауссовский сглаживающий фильтр, построенный на основе спектральной матричной нормы.

Рассмотрим линейный фильтр (1). Заметим, что оценка значения a_{ij} может быть получена в результате минимизации нормы Фробениуса

$$\hat{a}_{ij} = \arg\min_{s} \left\| B_{ij}(s) \right\|_F \qquad (2)$$

матрицы $B_{ij}(s) = sE - (2r+1)^2 W.A_{ij}$, где E – матрица с единичными элементами;

$$A_{ij} = \begin{bmatrix} a_{i-rj-r} \cdots a_{i-rj+r} \\ . \ . \ . \ . \ . \ . \\ a_{i+rj-r} \cdots a_{i+rj+r} \end{bmatrix};$$

точка в выражении $W.A_{ij}$ обозначает операцию поэлементного произведения матриц.

В вычислительной математике и методах анализа данных наряду с нормой Фробениуса применяется спектральная матричная норма [3], обладающая свойством грубости (робастности) по отношению к изменениям в элементах матрицы [4]. Заменим норму Фробениуса в выражении (1) на спектральную норму. Получим оценку

$$\hat{a}_{ij} = \arg\min_{s} \left\| B_{ij}(s) \right\|_2, \qquad (3)$$

где $\left\| B_{ij} \right\|_2$ – максимальное сингулярное число матрицы B_{ij}.

Фильтр (3) является нелинейным гауссовским фильтром. Вычислить значение $\hat{a}_{ij}$ несложно, поскольку s – это скалярная переменная, принимающая значения в заданном интервале $s \in [0 \quad a_{\max}]$.

Качество восстановления изображения можно оценить по нескольким критериям:

– критерию отношения сигнала к шуму [2]:

$$d(A, \hat{A}) = 10 \log_{10} \frac{255^2 nm}{\sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ij} - \hat{a}_{ij})^2}; \qquad (4)$$

– критерию максимального отклонения:

$$d_{\max}(A, \hat{A}) = \max_i \max_j \left| a_{ij} - \hat{a}_{ij} \right|; \qquad (5)$$

– критерию на основе нормы Фробениуса:

$$d_F(A, \hat{A}) = \frac{\left\| A - \hat{A} \right\|_F}{\left\| A \right\|_F}; \qquad (6)$$

– критерию на основе спектральной нормы:

$$d_2(A, \hat{A}) = \frac{\left\| A - \hat{A} \right\|_2}{\left\| A \right\|_2}. \qquad (7)$$

Два последних критерия позволяют оценить качество восстановления изображения в процентном отношении.

В качестве примера рассмотрим черно-белое изображение, представленное в виде матрицы размером 400×600 (рис. 1). Элементы матрицы принимают значения от 0 до 255, что соответствует восьмибитному кодированию изображения. Файл с изображением взят с сайта http://freefoto.com.

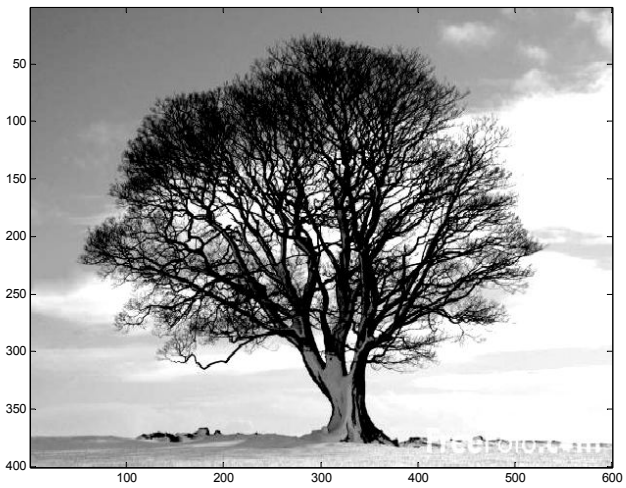


Рис. 1. Исходное изображение

Введем следующие обозначения: A – исходное изображение; $\overline{A}$ – изображение с искаженными пикселями; $\hat{A}$ – восстановленное изображение. Искаженные пиксели имеют нулевое значение, что соответствует черному цвету (рис. 2).

Восстановление изображения выполнялось на основе линейного (ЛГФ) (рис. 3) и нелинейного гауссовских фильтров (НГФ) (рис. 4) с параметрами $r = 2, \ \sigma = 2/3$.

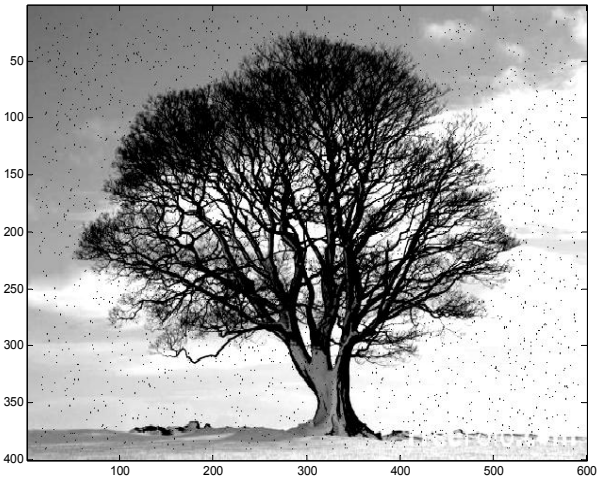


Рис. 2. Изображение с искаженными пикселями

Рис. 3. Изображение, восстановленное на основе линейного гауссовского фильтра

Расчеты выполнялись в системе MATHLAB с использованием базовых функций чтения, записи и визуализации изображений [5].

Результаты расчетов, приведенные в таблице, показывают, что нелинейный гауссовский фильтр дает лучший результат по критериям (4)…(7), чем линейный гауссовский фильтр.

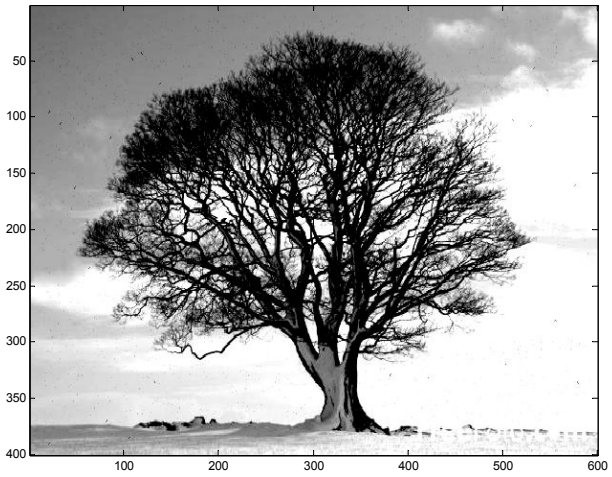


Рис. 4. Изображение, восстановленное на основе нелинейного гауссовского фильтра

Таким образом, нелинейный гауссовский фильтр, построенный на основе спектральной матричной нормы, может быть использован для восстановления искаженных пикселей изображений. Численные эксперименты подтверждают работоспособность данного фильтра и возможность его применения для решения практических задач цифровой обработки изображений.

Библиографические ссылки

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М. : Техносфера, 2005.

2. Алгоритмические основы растровой машинной графики / Д. В. Иванов, А. С. Карпов, Е. П. Кузьмин и др. М. : БИНОМ. Лаборатория знаний, 2007.

3. Воеводин В. В., Кузнецов Ю. А. Матрицы и вычисления. М. : Наука, 1984.

4. Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. М. : Наука, 1986.

5. Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATHLAB. М. : Техносфера, 2006.

Качество восстановления изображения

Фильтр	$d(A,\overline{A})$	$d(A,\hat{A})$	$d_{\max}(A,\overline{A})$	$d_{\max}(A,\hat{A})$	$d_F(A,\overline{A})$	$d_F(A,\hat{A})$	$d_2(A,\overline{A})$	$d_2(A,\hat{A})$
ЛГФ	24,3	32,6	255	147	0,096	0,037	0,013	0,0049
НГФ	24,3	38,2	255	141	0,096	0,019	0,013	0,0029

A. V. Bobrov, E. A. Perepelkin

IMAGE RECONSTRUCTION BASED ON SPECTRAL MATRIX NORM

A problem of image reconstruction is considered. Nonlinear Gaussian filter, built on the basis of the spectral matrix norm, is described. Results of numerical experiments confirming advantage of the given filter in comparison with linear Gaussian filter are discussed.

Keywords: images reconstruction, Gaussian filter, spectral matrix norm.

© Бобров А. В., Перепелкин Е. А., 2010

УДК 519.214, 519.673

А. Н. Бондаренко, В. А. Дедок

МОДЕЛИРОВАНИЕ НЕАДАМАРОВСКОГО КВАНТОВОГО БЛУЖДАНИЯ ДЛЯ РЕШЕТОК ВЫСШИХ РАЗМЕРНОСТЕЙ*

Методом компьютерного моделирования исследованы асимптотические свойства вероятности возвращения в модели квантового случайного блуждания на решетках разных размерностей. Результаты численных экспериментов позволяют сформулировать гипотезы локализации на двумерной решетке для модели неадамаровского блуждания с различными эволюционными матрицами.

Ключевые слова: квантовое случайное блуждание, вероятность возвращения.

Модель квантового случайного блуждания активно изучается в последнее десятилетие в связи с возможным применением ее результатов в теории квантовых вычислений и ускорении алгоритмов, основанных на случайном блуждании. Более того, указанная модель имеет неожиданные приложения в теории прямых и обратных задач рассеяния [1].

Одномерное квантовое случайное блуждание. Прежде чем определить дискретное квантовое блуждание, опишем классический случай [2]. Симметричное классическое блуждание может быть реализовано следующим образом. Частица начинает движение, например, из начала координат, на каждом шагу подбрасывая монетку, она равновероятно выбирает направление движения. Затем она делает один шаг в выбранном направлении и т. д.

В более общем случае, когда частица выбирает направление не равновероятно, а с вероятностями p и $q = 1 - p,$ вероятность нахождения классической блуждающей частицы в момент времени $t+1$ в ячейке с номером n выглядит следующим образом:

$$P_n(t+1) = pP_{n+1}(t) + qP_{n-1}(t). \qquad (1)$$

Тем самым, задав распределение вероятностей нахождения классической блуждающей частицы в начальный момент времени, можно вычислить распределение вероятностей для любого наперед заданного момента времени t.

Квантовое обобщение случайного блуждания [3] рассматривает квантовую частицу, обладающую дополнительной степенью свободы, и характеризуется двукомпонентной волновой функцией

$$\Psi(n,t) = \begin{bmatrix} \psi_L(n,t) \\ \psi_R(n,t) \end{bmatrix}$$

амплитуд в ячейке с номером n во время t. Поведение во времени функции $\Psi(n,t)$ задается унитарным преобразованием.

Определение 1. *Адамаровское случайное блуждание* – это случайное блуждание, задаваемое правилом

$$\Psi(n,t+1) = \begin{pmatrix} 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \Psi(n-1,t) + \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \Psi(n+1,t) =$$
$$= M_{-}\Psi(n-1,t) + M_{+}\Psi(n+1,t).$$

Вероятность нахождения квантовой частицы в узле с номером n в момент времени t выражается следующей формулой:

$$P(n,t) = \left|\psi_L(n,t)\right|^2 + \left|\psi_R(n,t)\right|^2.$$

Если частица начинает свое движение из начала координат с исходным состоянием «влево», то начальные условия будут выглядеть следующим образом:

$$\Psi(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \Psi(n,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, n \neq 0,$$

и адамаровское квантовое случайное блуждание станет сводиться к решению двумерной системы уравнений в конечных разностях.

В общем случае характеристики квантового случайного блуждания (унитарного преобразования над волновой функцией) зависят от четырех параметров. Однако, как было отмечено в [3], достаточно рассмотреть зависимость от одного параметра, характеризующего семейства блужданий. Таким образом, общее случайное блуждание может быть описано следующим преобразованием:

$$M(\theta) = T \circ (I \otimes U_\theta),$$

где $U_\theta = e^{i\frac{\theta}{2}\sigma_y}$.

Оператор $M(\theta)$ описывает эволюцию блуждания: $\Psi(t+1) = M(\theta)\Psi(t)$. Адамаровское случайное блуждание соответствует значению параметра $\theta = \pi/2$:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -i\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = -ie^{i\frac{\pi}{2}\sigma_z}U_{\frac{\pi}{2}},$$

где дополнительное вращение может быть нивелировано подходящим переопределением фазы состояния.

*Работа поддержана Российским фондом фундаментальных исследований (грант № 08-01-00312).

В более удобном для использования виде эволюция квантового блуждания может быть записана как

$$\Psi(n,t+1)=\begin{pmatrix}0 & 0\\ \sin\frac{\theta}{2} & -\cos\frac{\theta}{2}\end{pmatrix}\Psi(n-1,t)+ \\ +\begin{pmatrix}\cos\frac{\theta}{2} & \sin\frac{\theta}{2}\\ 0 & 0\end{pmatrix}\Psi(n+1,t). \qquad (2)$$

Используя выражение для волновой функции (2), можно вычислить распределение вероятностей нахождения квантовой частицы для различных начальных состояний частицы и различных значений параметра θ.

Результаты расчетов для вероятности возвращения квантовой частицы даже для не очень большого числа шагов, приведенные на рис. 1, позволяют сделать вывод, что распределение вероятностей нахождения квантовой частицы в заданной точке достаточно сильно зависит как от начального состояния частицы, так и от параметра θ.

Здесь и далее нас будет интересовать такая характеристика, как вероятность возвращения квантовой блуждающей частицы в исходную точку за t шагов по времени, а также асимптотические свойства этой характеристики. Интерес к данному показателю обусловлен связью вероятности возвращения и свойства локализации [1].

Возвратные марковские цепи. Локализация в модели квантового блуждания. В теории классических цепей Маркова большое внимание уделяется возвратным состояниям. Рассмотрим цепь Маркова $\{X_n\}_{n=0}^{\infty}$. Если в n-м испытании реализовалось событие E_j, то будем считать, что $X_n = j$. Введем следующие обозначения:

$$f_j(n) = P\left(X_n = j, X_{n-1} \neq j, ..., X_1 \neq j \mid X_0 = j\right),$$

$$F_j = \sum_{n=0}^{\infty} f_j(n).$$

Определение 2. Состояние E_j называется *возвратным*, если $F_j = 1$, и *невозвратным*, если $F_j < 1$.

Пусть $p_{ij} = P\left(X_k = j \mid X_0 = i\right)$. Важным способом проверки возвратности является следующая теорема [2].

Теорема 1. Состояние E_j возвратно тогда и только тогда, когда $\sum_{j=1}^{\infty} p_{jj} = \infty$.

Численно исследуя асимптотические свойства убывания членов ряда p_{jj}, можно сделать выводы о его сходимости или расходимости, а следовательно, и выводы о возвратности цепи Маркова.

В случае классического случайного блуждания имеет место следующая теорема [2].

Теорема 2. Описанное случайное блуждание образует возвратную цепь Маркова тогда и только тогда, когда $p = q = \frac{1}{2}$.

По аналогии с вышеприведенными рассуждениями и определением локализации в моделях, рассматриваемых в [4], можно определить следующие понятия.

Определение 3. Квантовое случайное блуждание называется *локализованным*, если $\lim_{n\to\infty} p_r(n) > 0$, где $p_r(n) = P(0,n)$ – вероятность возвращения квантовой частицы в начальную точку за n шагов.

Определение 4. Квантовое случайное блуждание называется *слабо локализованным*, если $\sum_{n=0}^{\infty} p_r(n) = \infty$.

Свойство возвратности одномерного квантового блуждания достаточно подробно описано в работе [1] и представлено следующей теоремой.

Теорема 3 (одномерная квантовая теорема Пойа). Пусть $\theta = 0$, тогда квантовое блуждание на прямой с любым начальным состоянием не возвратно. Пусть $0 < \theta < \pi$, тогда квантовое блуждание на прямой с любым начальным состоянием возвратно (слабо локализовано). При этом главный член асимптотики не зависит от начального состояния. Пусть $\theta = \pi$, тогда квантовое блуждание на прямой с любым начальным состоянием возвратно (слабо локализовано). При этом $P_0(n,t) = 0$ при $|n| \geq 2$.

Квантовое случайное блуждание высших размерностей. Одномерное классическое блуждание на прямой описывается эволюционным уравнением для вероятности нахождения в точках с целыми координатами (1). Аналогичное уравнение, но уже зависящее от четырех параметров, имеет место для классического блуждания по точкам с целыми координатами на плоскости (а в общем случае – и для блуждания по точкам с целыми координатами в пространстве размерности n):

$$P_{x,y}(t+1) = p_1 P_{x+1,y}(t) + p_2 P_{x-1,y}(t) + \\ + p_3 P_{x,y+1}(t) + p_4 P_{x,y-1}(t).$$

Вероятности p_i соответствуют вероятности классической частицы совершить шаг вправо, влево, вверх и вниз. Для квантового случайного блуждания роль вероятностей играют матрицы $P = M_+$ и $Q = M_-$.

Соответствующее уравнение для амплитуд вероятностей выглядит следующим образом:

$$\Psi(n,t+1) = P\Psi(n+1,t) + Q\Psi(n-1,t).$$

Множество начальных состояний квантовой частицы имеет вид

$$\Phi = \left\{\varphi = \begin{bmatrix}\alpha\\ \beta\end{bmatrix} \in C^2 : |\alpha|^2 + |\beta|^2 = 1\right\}.$$

Матрица преобразования для одномерного адамаровского блуждания будет такой:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}.$$

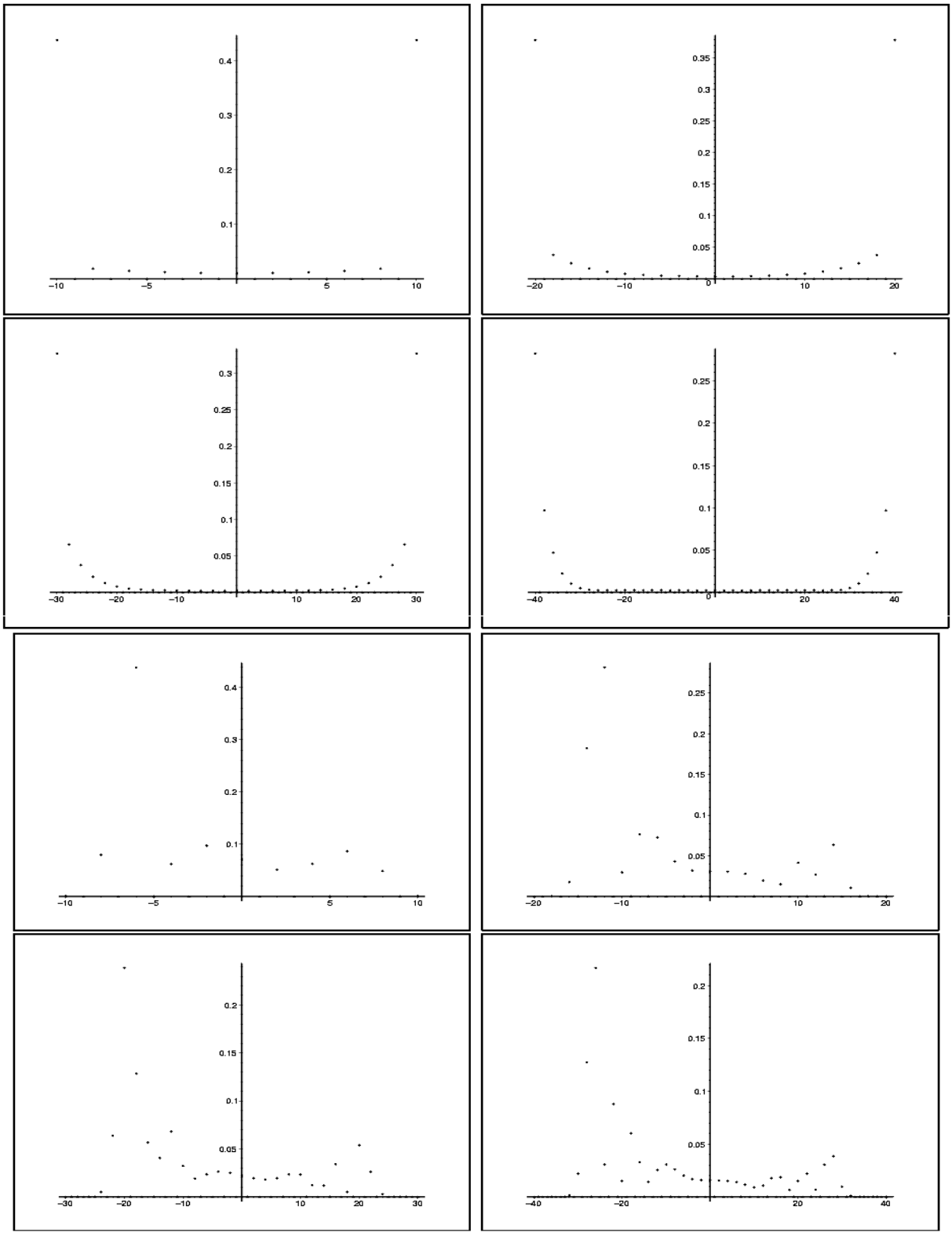


Рис. 1. Распределение вероятностей возвращения для блуждающей квантовой частицы:

верхние четыре рисунка – $\theta = \frac{\pi}{13}$, $\Psi(0,0) = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix}$, число шагов по времени 50, 100, 150, 200;

нижние четыре рисунка – $\theta = \frac{\pi}{2}$, $\Psi(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, число шагов по времени 50, 100, 150, 200

Обобщение квантового случайного блуждания на многомерный случай определяется матрицами большей размерности [5]:

$$H_d = H \otimes H \underset{d}{\otimes} \ldots \otimes H,$$

являющимися тензорным произведением матриц одномерного случайного блуждания. В этом случае множество начальных состояний квантовой частицы выглядит как

$$\Phi^{(d)} = \left\{ \varphi_1 \otimes \varphi_2 \otimes \ldots \otimes \varphi_d : \varphi_i \in \Phi, i = 1, 2, \ldots, d \right\}.$$

Определение 5. *Двумерным адамаровским случайным блужданием* называется блуждание, описываемое матрицей

$$H_{\text{Н}} = H_1 = H \otimes H = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Вероятность нахождения квантовой частицы в соответствующем узле (n,m) определяется аналогично одномерному случаю:

$$P((n,m),t) = |\Psi_1((n,m),t)|^2 + |\Psi_2((n,m),t)|^2 + \\ + |\Psi_3((n,m),t)|^2 + |\Psi_4((n,m),t)|^2 .$$

Другим примером двумерного квантового блуждания является блуждание, описываемое матрицами H_2 (Гровера) и H_3:

$$H_2 = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}, \quad H_3 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}.$$

Как и раньше, нас будет интересовать такая характеристика блуждания, как вероятность возвращения в исходную точку за определенное число шагов, например за 20 (рис. 2).

Сложность изучения квантового случайного блуждания обусловливается еще большим числом все возможных конфигураций и степеней свободы по сравнению с блужданием на одномерной решетке (см. рис. 2).

Проследить асимптотические свойства вероятности возвращения в исходную точку можно на рис. 3, где изображены результаты компьютерного моделирования. Эти результаты показывают, что вероятность возвращения для блуждания с матрицами H_1 и H_3 убывает на бесконечности как $1/t^2$ и, следовательно, невозвратно.

Особенно интересным в качестве объекта численного исследования является случайное блуждание Гровера, которое имеет принципиально иной асимптотический характер вероятности возвращения в исходную точку (пропорционально t^0).

Таким образом, результаты моделирования позволяют сформулировать следующие гипотезы.

Гипотеза 1. Асимптотика вероятности возвращения квантовой частицы в начало координат не зависит от начальных условий.

Гипотеза 2. Подмножество параметров квантового блуждания, при котором оно обладает свойством возвратности при любых начальных данных, имеет меру нуль.

Библиографические ссылки

1. Бондаренко А. Н., Дедок В. А. Квантовая теорема Пойа // Сиб. электрон. мат. изв. 2009. Т. 6. С. 199–210.

2. Боровков А. А. Курс теории вероятностей. М. : Наука, 1972.

3. Nayak A., Vishwanath A. Quantum Walk on the Line (Extended Abstract) [Electronic resource] // Cornell University Library : site. URL: http://arxiv.org/pdf/quant-ph/0005106 (date of visit: 20.11.2010).

4. Ishii K. Localization of Eigenstates and Transport Phenomena in the One-Dimensional Disordered System // Prog. Theor. Phys. Suppl. 1973. № 53. P. 77–138.

5. Sanders Quantum walks in higher dimensions / T. D. Mackay, S. D. Bartlett, L. T. Stephanson, B. C. Sanders // J. Phys. A: Math. Gen. 2002. Vol. 35. P. 2745–2753.

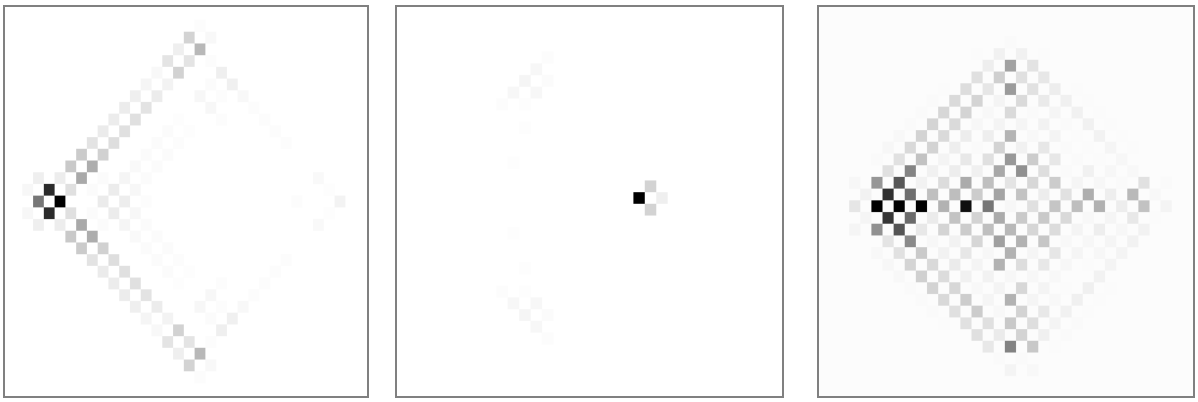


Рис. 2. Распределение вероятностей для блуждания с матрицами преобразования H_1, H_2 и H_3 за 20 шагов по времени

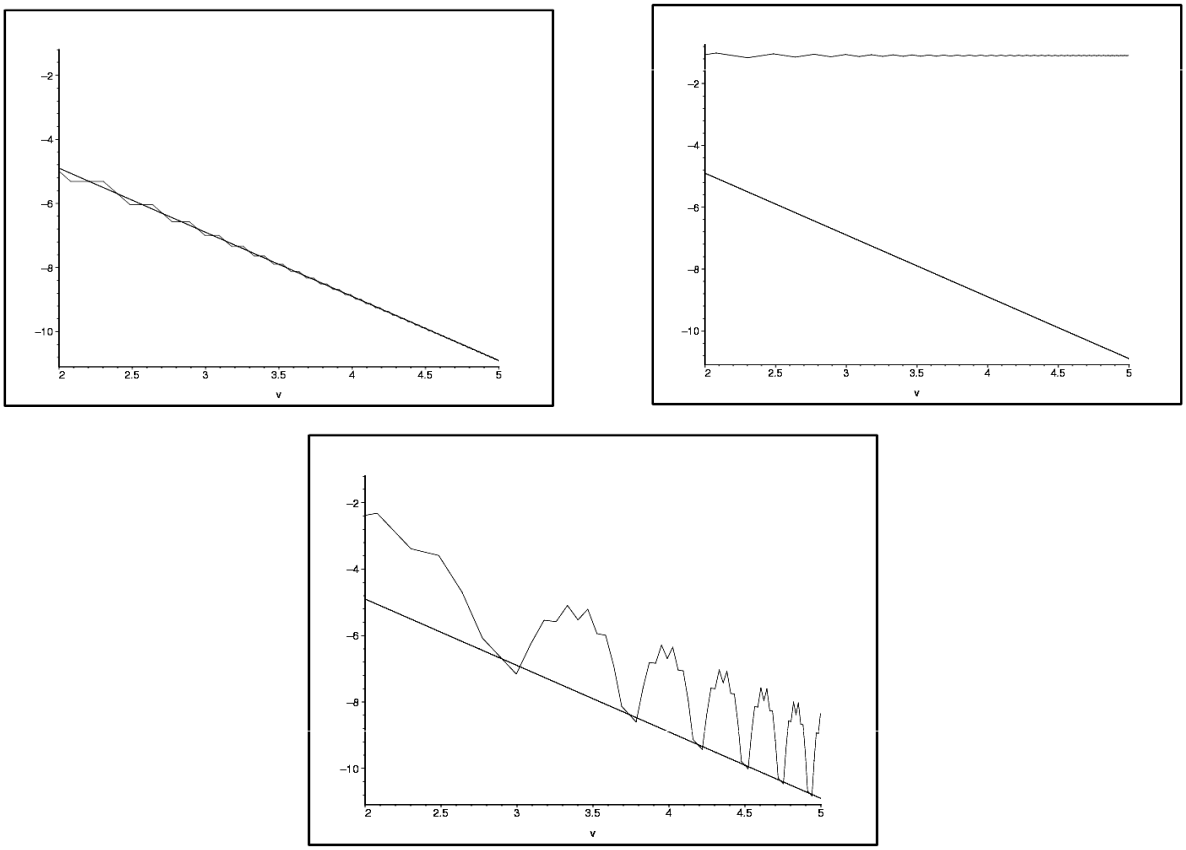


Рис. 3. Вероятность возвращения в исходную точку в логарифмическом масштабе для блужданий с матрицами H_1, H_2 и H_3 и асимптотическое приближение на бесконечности вида $1/t^2$

A. N. Bondarenko, V. A. Dedok

MODELLING OF NEADAMAROV QUANTUM WANDERING FOR GRILL GRIDS OF ADVANCED DIMENSIONS

The work is devoted to studies of asymptotic properties of probability of returning in model of quantum casual wandering on grill grids of different dimensions by a method of computer modeling. Results of numerical experiments allow to formulate localization hypotheses on a two-dimensional grill grid for model of Neadamarov wanderings with various evolutionary matrixes.

Keywords: quantum casual wandering, probability of returning.

© Бондаренко А. Н., Дедок В. А., 2010

УДК 517.9

М. А. Бондаренко

МЕТОДЫ ХАОТИЧЕСКОЙ ДИНАМИКИ В ЗАДАЧЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ ВОЛЬФА*

Представлены результаты компьютерного исследования временного ряда, составленного из чисел Вольфа. В соответствии с подходом, основанным на теореме Такенса, решалась задача восстановления размерности аттрактора динамической системы, описывающей динамику Солнца. Показано, что за последние 3 000 лет принципиальных изменений в эволюции Солнца не произошло.

Ключевые слова: теорема Такенса, реконструкция аттрактора, числа Вольфа.

Математические модели, содержащие три и более обыкновенных дифференциальных уравнений (ОДУ), способны демонстрировать хаотические режимы колебаний, которые на первый взгляд имеют вид случайных процессов. Переход в фазовое пространство позволяет получать наглядную информацию об особенностях сложной динамики соответствующих систем, прежде всего о геометрии предельных множеств фазовых траекторий, которые соответствуют установившимся режимам [1].

В последнее время это свойство хаотичности нелинейных диссипативных динамических систем активно используется для анализа временных рядов. Основным моментом здесь является предположение о том, что данный временной ряд порожден некоторым функционалом от решения неизвестной нам динамической системы с аттрактором. Алгоритм Грассберга–Прокаччиа [2], используемый в данной статье, позволяет получить оценку для корреляционной размерности этого аттрактора. Она и является тем параметром, который характеризует качественное поведение траекторий такой динамической системы. В работе в качестве временного ряда автором были выбраны числа Вольфа, называемые также международными числами солнечных пятен, относительными числами солнечных пятен или цюрихскими числами.

Солнечные пятна – это темные области на Солнце, температура которых понижена примерно на 1 500 К по сравнению с окружающими участками фотосферы. Количество пятен на Солнце и связанное с ним число Вольфа – одни из главных показателей солнечной магнитной активности. Прямые наблюдения солнечных пятен доступны в течение последних четырех столетий (рис. 1), но более длительный временной ряд необходим, например, для выявления возможного воздействия Солнца на климат и для проверки моделей солнечного динамо.

В работе [3] приведены результаты реконструкции чисел Вольфа на период, охватывающий последние 11 400 лет. Эти результаты были получены на основании дендрохронологического анализа радиоуглеродной концентрации ископаемых останков растений. На основании объединенной физико-биологической модели была построена связь процессов набора определенной радиоуглеродной концентрации с числами Вольфа.

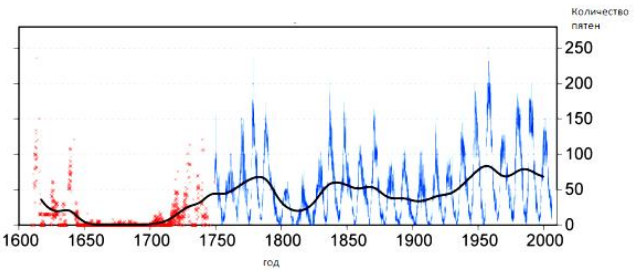


Рис. 1. Количество солнечных пятен в зависимости от года

Согласно полученной реконструкции, уровень солнечной активности за последние 70 лет является исключительным случаем. Предыдущий период столь же высокой активности произошел более 8 000 лет назад. Было подсчитано, что в течение последних 11 400 лет Солнце провело лишь порядка 10 % времени на столь же высоком уровне магнитной активности, и почти все предыдущие периоды такой активности были гораздо короче, чем нынешний эпизод.

Редкость текущего эпизода высокого среднего числа солнечных пятен может свидетельствовать о том, что Солнце внесло свой вклад в необычные климатические изменения в течение XX в. Но насколько значимы эти отклонения? Являются ли они некоторым сигналом того, что на Солнце происходят какие-то изменения в механизме термоядерного синтеза или же это простые временные флуктуации? Будем считать, что числа Вольфа порождены некоторой динамической системой, описывающей временную эволюцию Солнца.

Была поставлена следующая задача: сравнить размерность аттрактора Солнца за последние 130 лет с его глобальным поведением за 11 400 лет. Разные величины N с аттрактором размерности N будут говорить о том, что в последние 130 лет в механизме Солнца происходят качественные изменения.

Для решения этой задачи выбраны два временных ряда для чисел Вольфа:

– соответствующий 11 400 годам через промежуток 100 лет;

– соответствующий 130 годам через промежуток 1 год.

*Работа поддержана Российским фондом фундаментальных исследований (грант № 08-01-00312).

Для каждого из этих случаев определялась временная задержка τ (рис. 2–4). Необходимо, чтобы система достаточно долго находилась на аттракторе, но при этом слишком детальное описание ее движения приведет к определенным трудностям.

Числовым показателем количества пятен на Солнце является число Вольфа. Для данного дня это число определяется

$$W = k(f + 10g), \qquad (1)$$

где k – нормировочный коэффициент; f – количество наблюдаемых пятен; g – количество наблюдаемых групп пятен.

Для нахождения величины временной задержки τ на интервале 9400 г. до н. э. – 2000 г. н. э. с шагом 10 лет были использованы метод взаимной информации $I(\tau)$:

$$I(\tau) = -\sum_{i=1}^{L}\sum_{j=1}^{L} P(A_i B_j)\cdot \log_2 \frac{P(A_i B_i)}{P(A_i)P(B_j)}, \qquad (2)$$

где $P(\cdot)$ – вероятность соответствующего события, и метод автокорреляционной функции $R(\tau)$:

$$R(\tau) = \frac{1}{N_1}\sum_{t=1}^{N_1} y(t)y(t+\tau). \qquad (3)$$

Анализ графика зависимости $I(\tau)$ и $R(\tau)$ от выбора временной задержки τ (рис. 2) показывает, что первый минимум встречается при τ = 5.

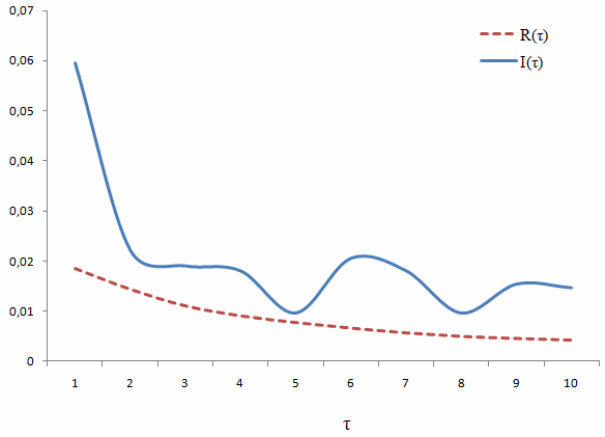


Рис. 2. График зависимости функции взаимной информации $I(\tau)$ и автокорреляционной функции $R(\tau)$ от временной задержки τ для интервала 9400 г. до н. э. – 2000 г. н. э. с шагом 10 лет

Для определения размерности пространства вложения был использован метод Грассберга–Прокаччиа [2]. На графике $d(m)$, представленном на рис. 3, насыщение наступает при m = 3, при этом d = 1,81. Исходя из этого можно сделать вывод, что размерность пространства вложения составит $d \geq 2\ [d(3)] + 1 = 3$.

Для нахождения временной задержки τ на интервале 1870–2000 гг. с шагом 1 год были использованы методы взаимной информации $I(\tau)$ (2) и автокорреляционной функции $R(\tau)$ (3) (рис. 4). Очевидно, что в качестве временной задержки в этом случае можно использовать τ = 3 или τ = 4.

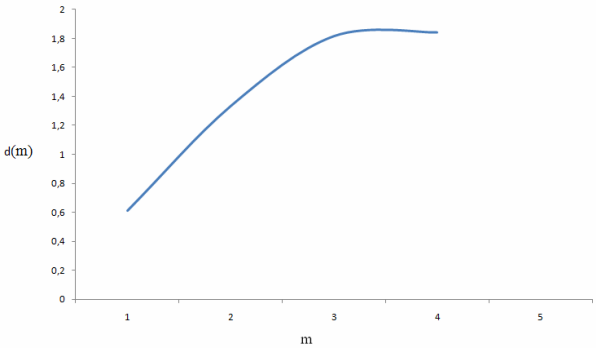


Рис. 3. График зависимости коэффициента корреляционной размерности d от размерности m для интервала 9400 г. до н. э. – 2000 г. н. э. с шагом 10 лет

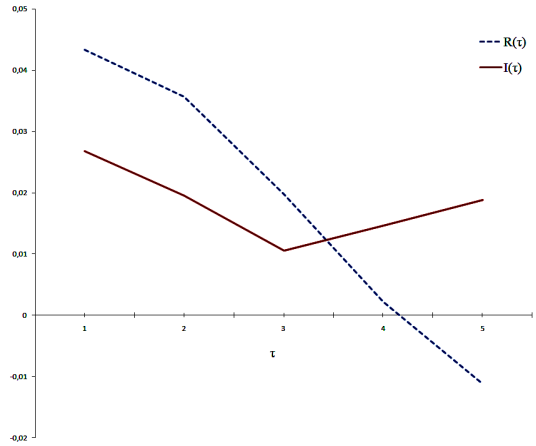


Рис. 4. График зависимости автокорреляционной функции $R(\tau)$ и функции взаимной информации $I(\tau)$ от временной задержки τ для интервала 1870–2000 гг. с шагом 1 год

Вычислим размерность пространства вложения методом Грассберга–Прокаччиа [2] для временного интервала 1870–2000 гг. (рис. 5). Можно отметить, что $d(m)$ также достигает своего насыщения при размерности 1,81. В этом случае имеем следующую оценку размерности пространства вложения: $d \geq 2\ [1{,}81] + 1 = 3$.

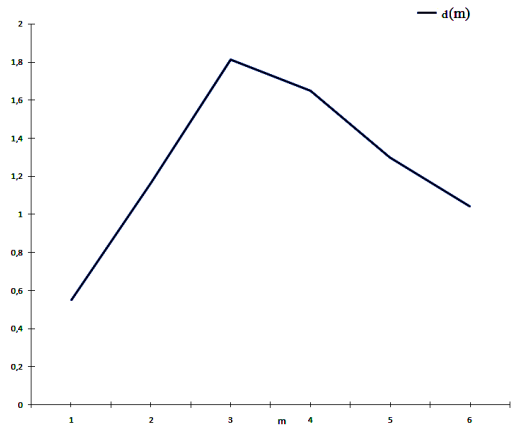


Рис. 5. График зависимости коэффициента корреляционной размерности d от размерности m для временного ряда, составленного из чисел Вольфа, для интервала 1870–2000 гг.

Библиографические ссылки

1. Kennel M. B., Brown R., Abarbanel H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction // Phys. Rev. A. 1992. Vol. 45. № 6. P. 3403–3411.

2. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors // Physica D. 1983. Vol. 9. P. 189.

3. An unusually active sun during recent decades compared to the previous 11 000 years / S. Solanki, I. Usoskin, B. Kromer et al. // Nature. 2004. Vol. 431. № 7012. P. 1084–1087.

M. A. Bondarenko

METHODS FOR CHAOTIC DYNAMICS IN THE PROBLEM OF ANALYSIS OF WOLFER'S TIME SERIES

The results of computer studies of time series composed of the Wolfer's numbers are presented in the article. In accordance with the approach based on Takens's theorem, the author solves the problem of reconstructing dimension of attractor of a dynamical system describing the dynamics of the Sun. It is shown that for the last 3 000 years, fundamental changes in the evolution of the Sun have not happened.

Keywords: Taken's theorem, reconstruction of the attractor, Wolfer's numbers.

© Бондаренко М. А., 2010

УДК 519.67-519.24, 519.6

А. А. Викентьев

О ВВЕДЕНИИ МЕТРИК НА ВЫСКАЗЫВАНИЯХ ЭКСПЕРТОВ С ВЕРОЯТНОСТЯМИ*

Предлагаются способы введения метрик на высказываниях экспертов, заданных формулами исчисления высказываний с вероятностями. Результаты переносятся на формулы над бесконечными носителями.

Ключевые слова: высказывания экспертов, кластеризация, расстояния, метрики.

В настоящее время проявляется все больший интерес к анализу экспертной информации, заданной в виде вероятностных логических высказываний экспертов. Интересны исследования о высказываниях экспертов, представленных формулами исчисления высказываний (ИВ) с вероятностями. Возникают задачи об алгоритмах распознавания закономерностей, согласования логических (экспертных) знаний и их кластеризации [1–6]. Для решения этих задач необходимы метрики на знаниях.

В данной статье рассматриваются логические высказывания экспертов, представленные формулами исчисления высказываний с вероятностями, предлагаются способы задания расстояний на формулах-высказываниях с вероятностями, а также устанавливаются свойства введенных расстояний, для чего используются вероятностный и теоретико-модельный подходы [2; 5; 7].

Основные определения. Будем рассматривать знания экспертов, представленные формулами ИВ с вероятностями (вероятностные высказывания), т. е. высказывания вида «ϕ с вероятностью p_ϕ», где ϕ – формула ИВ. Для таких высказываний будем использовать следующую запись:

$$B_i = \langle \phi, p_\phi \rangle, \quad B_j = \langle \psi, p_\psi \rangle.$$

Пусть Σ – база знаний, состоящая из формул ИВ, т. е. в Σ содержатся все формулы, с которыми будут работать эксперты; $S(\phi)$ – носитель формулы ϕ, т. е. множество элементарных высказываний, используемых при написании формулы ϕ, $S(\Sigma) = \bigcup_{\phi \in \Sigma} S(\phi)$ – носитель совокупности знаний.

Рассмотрим совокупность $P(S(\Sigma)) = 2^{S(\Sigma)}$ – множество всевозможных подмножеств множества $S(\Sigma)$. Для простоты обозначения примем, что $\left|P(S(\Sigma))\right| = 2^{|S(\Sigma)|} = n$. Элементы множества $P(S(\Sigma))$ называются *моделями*. Более подробно о теории моделей см. в [2–4].

Пусть эксперты говорят о вероятностях (частости) формул на множестве всех n моделей, и каждое высказывание присутствует только с одной вероятностью. Тогда будем интерпретировать вероятность, данную экспертом, следующим образом: $B = \langle \phi, p_\phi \rangle$. Это означает, что высказывание φ истинно на $n_\phi = \left[n \cdot p_\phi \right]$ моделях, где $n = 2^{|S(\Sigma)|}$ – число всех моделей.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты 07-01-00331а, 08-07-00136а).

Пусть имеется два вероятностных логических высказывания $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$. Дадим способ вычисления расстояния $\rho(B_i, B_j)$ между такими высказываниями.

Интерпретируя данные экспертами вероятности описанным выше способом, получаем, что высказывание ϕ истинно на $n_\phi = \lfloor n \cdot p_\phi \rfloor$ моделях, а высказывание ψ – на $n_\psi = \lfloor n \cdot p_\psi \rfloor$ моделях. Отметим, однако, что при таком подходе мы не знаем, на каких именно моделях каждое высказывание истинно, а также число моделей, на которых эти высказывания истинны одновременно.

Решим следующую задачу: пусть высказывание ϕ истинно на n_ϕ моделях, высказывание ψ истинно на n_ψ моделях и k – число моделей, на которых эти высказывания истинны одновременно. Вычислим расстояние между высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Обозначим рассматриваемые далее расстояния через $\rho_k(B_i, B_j)$, где $k = 0, 1, 2, \ldots, \min(n_\phi, n_\psi)$.

Как и в [2–4], расстояние $\rho_k(B_i, B_j)$ для каждого $k = 0, 1, 2, \ldots, \min(n_\phi, n_\psi)$ определим через разность:

$$\rho_k(B_i, B_j) = \frac{n_\phi - k + n_\psi - k}{n} = \frac{n_\phi + n_\psi - 2k}{n}.$$

Теорема 1. Для расстояний $\rho_k(B_i, B_j)$ справедливы свойства:

1) $0 \le \rho_k(B_i, B_j) \le 1$;

2) $\rho_k(B_i, B_j) = \rho_k(B_j, B_i)$;

3) $\rho_k(B_i, B_j) \le \rho_k(B_i, B_s) + \rho_k(B_s, B_j)$;

4) $B_i \equiv B_j \Leftrightarrow \rho_k(B_i, B_j) = 0$ ($B_i \equiv B_j \Leftrightarrow \varphi \equiv \psi$ и $p_\phi = p_\psi$, т. е. формулы φ и ψ истинны на одних и тех же моделях);

5) $B_i \equiv \neg B_j \Leftrightarrow \rho_k(B_i, B_j) = 1$;

6) $\rho_k(B_i, B_j) = 1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j)$;

7) $\rho_k(B_i, B_j) = \rho_k(B_i \wedge B_j, B_i \vee B_j)$.

Докажем неочевидное свойство 3. Определение $\rho_k(B_i, B_j)$ можно переписать следующим образом:

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = \frac{n_{\phi \Delta \psi}}{n} = \frac{n_{(\neg\phi \wedge \psi) \vee (\phi \wedge \neg\psi)}}{n}$$

(по определению симметрической разности). Тогда для произвольного высказывания $B_s = \langle \chi, p_\chi \rangle$ нетрудно доказать, что $n_{\phi \Delta \psi} \le n_{\phi \Delta \chi} + n_{\chi \Delta \psi}$, откуда

$$\rho_k(B_i, B_j) = \frac{n_{\phi \Delta \psi}}{n} \le \frac{n_{\phi \Delta \chi}}{n} + \frac{n_{\chi \Delta \psi}}{n} = \rho_k(B_i, B_s) + \rho_k(B_s, B_j).$$

Докажем свойство 4 сначала слева направо ($\Rightarrow$). Если $B_i \equiv B_j$, то $\phi \equiv \psi$ и, значит, $n_\phi = n_\psi = k$. Следовательно,

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n} = 0.$$

Докажем свойство 4 в обратную сторону ($\Leftarrow$). Если $\rho_k(B_i, B_j) = 0$, то $n_\phi + n_\psi - 2k = 0$. Так как k может принимать значения $0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = 0 \Leftrightarrow n_\phi = n_\psi = k)$. Следовательно, $\phi \equiv \psi$ и, значит, $B_i \equiv B_j$.

Докажем свойство 5 слева направо ($\Rightarrow$). Если $B_i \equiv \neg B_j$, то $\phi \equiv \neg\psi$. Тогда $n_\phi = n - n_\psi$ и $k = 0$, следовательно

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 0}{n} = \frac{n}{n} = 1.$$

Докажем свойство 5 в обратную сторону ($\Leftarrow$). Если $\rho_k(B_i, B_j) = 1$, то $n_\phi + n_\psi - 2k = n$. Так как k может принимать значения $0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$, то $(n_\phi + n_\psi - 2k = n \Leftrightarrow n_\phi + n_\psi = n$ и $k = 0)$. Следовательно, $\phi \equiv \neg\psi$ и $B_i \equiv \neg B_j$.

Докажем свойство 6:

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n},$$

$$\rho_k(B_i, \neg B_j) = \frac{n_\phi + n_{\neg\psi} - 2(n_\phi - k)}{n} = \frac{n_\phi + (n - n_\psi) - 2(n_\phi - k)}{n} =$$

$$= \frac{n - n_\phi - n_\psi - 2k}{n} = 1 - \frac{n_\phi + n_\psi - 2k}{n} = 1 - \rho_k(B_i, B_j),$$

Откуда нетрудно доказать, что

$$1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j).$$

Докажем свойство 7.

Легко доказать, что $n_{(\phi \wedge \psi) \Delta (\phi \vee \psi)} = n_{\phi \Delta \psi}$. Тогда

$$\rho_k(B_i \wedge B_j, B_i \vee B_j) = \frac{n_{(\phi \wedge \psi) \Delta (\phi \vee \psi)}}{n} = \frac{n_{\phi \Delta \psi)}}{n} = \rho_k(B_i, B_j).$$

Далее предложим несколько способов вычисления расстояния $\rho(B_i, B_j)$ между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$.

Так как нам неизвестно число k, т. е. число моделей, на которых высказывания ϕ и ψ истинны одновременно, и если нет никаких предпочтений для значения k (хотя они и могут быть высказаны экспертами), то мы можем поступить следующим образом.

Предположим, что все значения числа k равновероятны. Тогда расстояние между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ определим как усреднение расстояний $\rho_k(B_i, B_j)$ по всем значениям k:

$$\rho(B_i, B_j) = \frac{\sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j)}{\min(n_\phi, n_\psi)+1}.$$

Для этого расстояния также справедлива теорема 1, и слагаемые под знаком суммы можно взять с весами, в которых учтены мнения экспертов.

Если экспертами указано, какое значение k предпочтительнее, то в качестве $\rho(B_i, B_j)$ берется расстояние $\rho_k(B_i, B_j)$. Это возможно в случае, когда мы знаем, что пересечение состоит из k моделей.

Можно подойти к вопросу определения расстояния $\rho_k(B_i, B_j)$ и с вероятностно-статистической точки зрения и для каждого k вычислить частоту того, что высказывания ϕ и ψ одновременно истинны на k моделях.

Найдем частоту (вероятность) p_k того, что в выбранных n_ϕ и n_ψ моделях (они выбираются из n моделей) будет k моделей, на которых высказывания ϕ и ψ истинны одновременно, где $k = 0, 1, 2, \ldots, K$, $\min(n_\phi, n_\psi)$.

Сначала определим вероятностное пространство $\langle \Omega, A, p \rangle$, где Ω – пространство элементарных исходов – моделей, $\Omega = \{0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)\}$ – число возможных совпадений моделей в наборах из n_ϕ и n_ψ моделей; A – система пар подмножеств множества моделей Ω, образующая σ-алгебру событий; p – вероятность на $\langle \Omega, A \rangle$.

Определим для $k = 0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$ на декартовом произведении $\Omega \otimes A$ случайную величину ξ:

$$\xi(k) = \rho_k(B_i, B_j),$$

т. е. функция ξ каждому k из Ω будет ставить в соответствие число – расстояние $\rho_k(B_i, B_j)$.

Вероятность этого события (с пересечением k) на классе из n моделей можно вычислить так:

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}},$$

где $C_n^{n_\phi}$ – число способов выбрать n_ϕ моделей из n моделей.

Действительно, любой набор, состоящий из n_ϕ моделей, может сочетаться с любым набором, состоящим из n_ψ моделей, значит число $(C_n^{n_\phi} \cdot C_n^{n_\psi})$ определяет количество всех способов выбрать два набора, один из которых состоит из n_ϕ моделей, а другой – из n_ψ моделей. Выбрать k моделей, которые будут общими в этих наборах, из n моделей можно C_n^k способами. Тогда остальные $(n_\phi - k)$ и $(n_\psi - k)$ моделей должны быть дизъюнктными. Следовательно, остальные $(n_\phi - k)$ моделей для пополнения набора, состоящего из k, до n_ϕ моделей можно выбрать $C_{n-k}^{n_\phi - k}$ способами, а $(n_\psi - k)$ моделей для получения набора, состоящего из n_ψ моделей, с учетом наших предположений, – $C_{n-n_\phi}^{n_\psi - k}$ способами. Итак, имеется всего $\left(C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}\right)$ способов выбрать два набора, один из которых состоит из n_ϕ моделей, а другой – из n_ψ моделей, и k моделей в этих наборах моделей совпадут. Поэтому вероятность того, что k моделей совпадет в наборах из n_ϕ и n_ψ элементных моделей, составит

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}}.$$

В результате получим, что расстояния

$$\rho_k(B_i, B_j) = \frac{n_\phi + n_\psi - 2k}{n}$$

будут появляться с вероятностями

$$p_k = \frac{C_n^k C_{n-k}^{n_\phi - k} C_{n-n_\phi}^{n_\psi - k}}{C_n^{n_\phi} C_n^{n_\psi}},$$

где $k = 0, 1, 2, \ldots, K, \min(n_\phi, n_\psi)$. Эти вероятности (и близкие к ним числа) можно использовать в качестве весов-коэффициентов расстояний для получения результирующего расстояния для данных формул с вероятностями (подробнее об этом см. дальше).

Заметим, что при таком подходе главную роль играют не сами формулы, а числа, определяющие количество моделей и их пересечения. Не имея другой информации, мы рассмотрели все возможные подмножества для подсчета частости (вероятности) появления расстояния для конкретного k. Используя свойство инвариантности расстояний между формулами и вероятностей высказываний (формул) [2–4], можно получить аналогичный результат, но с меньшим носителем знаний, включающим только те, которые встречаются в двух формулах, для которых и ищется расстояние. Будем считать, что мы так и сделали с самого начала. И тогда наш расчет будет оптимальным.

Зная вероятности p_k для каждого расстояния $\rho_k(B_i, B_j)$, в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять, например, наиболее вероятное расстояние $\rho(B_i, B_j) = \rho_m(B_i, B_j)$, где $p_m = \max_k p_k$, для которого справедлива теорема 1.

Для получения других расстояний можно использовать расстояние, усредненное для некоторых выбранных подмножеств из всех полученных расстояний.

Если мы возьмем произвольные веса p_k (исходя из экспертных оценок или дополнительных сведений экспертов, которые могут и не совпадать!) для расстояний $\rho_k(B_i, B_j)$ таким образом, чтобы они подчинялись закону распределения, то получим общий случай для адаптивного поиска нужного расстояния между формулами с вероятностями. Тогда в качестве расстояния между вероятностными высказываниями $B_i = \langle \phi, p_\phi \rangle$ и $B_j = \langle \psi, p_\psi \rangle$ можно взять величину, равную математическому ожиданию (центру тяжести) или среднему значению случайной величины ξ:

$$\rho(B_i, B_j) = M\xi = \sum_{k=0}^{\min(n_\phi, n_\psi)} \rho_k(B_i, B_j) \cdot p_k.$$

Теорема 2. Для расстояния $\rho(B_i, B_j)$ справедливы свойства:

1) $0 \le \rho(B_i, B_j) \le 1$;

2) $\rho(B_i, B_j) = \rho(B_j, B_i)$;

3) $\rho(B_i, B_j) \le \rho(B_i, B_s) + \rho(B_s, B_j)$;

4) если $\rho(B_i, B_j) = 0$, то $B_i \equiv B_j$;

5) $\rho(B_i, B_j) = 1 - \rho(B_i, \neg B_j) = \rho(\neg B_i, \neg B_j)$;

6) $\rho(B_i, B_j) = \rho(B_i \wedge B_j, B_i \vee B_j)$.

Докажем неочевидное свойство 3. Для расстояния $B_i = \langle \phi, p_\phi \rangle$ по свойству 3 теоремы 1 имеем $\rho_k(B_i, B_j) \le \rho_k(B_i, B_s) + \rho_k(B_s, B_j)$. Тогда по свойствам математического ожидания:

а) $\xi \le \eta \Rightarrow M\xi \le M\eta$;

б) $M(\xi + \eta) = M\xi + M\eta$, откуда получаем требуемое свойство для расстояния.

Докажем свойство 4. Пусть $\xi(k) = \rho_k(B_i, B_j) \ge 0$ и $M\xi = 0$. Тогда по свойству математического ожидания $\xi(k) = \rho_k(B_i, B_j) = 0$ с вероятностью, равной 1, а по свойству 4 теоремы 1 $B_i \equiv B_j$.

Докажем свойство 5. Так как для расстояния $\rho_k(B_i, B_j)$ по свойству 6 теоремы 1 справедливо равенство

$$\rho_k(B_i, B_j) = 1 - \rho_k(B_i, \neg B_j) = \rho_k(\neg B_i, \neg B_j),$$

тогда по свойствам математического ожидания:

а) если $p(\xi = \eta) = 1$ и $\exists M\xi$ (математическое ожидание существует), то $M\xi = M\eta$;

б) $M(a + b\xi) = a + bM\xi$, откуда получаем требуемое свойство для расстояния $\rho(B_i, B_j)$.

Таким образом, в данной статье исследованы способы введения метрик на классах эквивалентных высказываний экспертов, заданных формулами ИВ с вероятностями. Такое исследование необходимо для решения задач согласования вероятностных высказываний экспертов, кластеризации и для построения баз знаний и экспертных систем. Результаты переносятся на формулы над бесконечными носителям и формулы с переменными языка первого порядка с использованием измеримых (для фиксированной теории) подклассов измеримых (в том числе и метрических) моделей.

Библиографические ссылки

1. Блощицын В. Я., Лбов Г. С. О мерах информативности логических высказываний // Технология разработки экспертных систем : докл. Респ. школы-семинара. Кишинев, 1987. С. 12–14.

2. Лбов Г. С., Старцева Н. Г. Логические решающие функции и вопросы статистической устойчивости решений. Новосибирск : Изд-во Ин-та математики, 1999.

3. Vikent'ev A. A., Lbov G. S. Setting the metric and informativeness on statements of experts // Pattern Recognition and Image Analysis. 1997. Vol. 7 (2). P. 175–189.

4. Викентьев А. А., Лбов Г. С. О метризациях булевой алгебры предложений и информативности высказываний экспертов // Докл. Рос. акад. наук. 1998. Т. 361 (2). С. 174–176.

5. Бериков В. Б. Кластерный анализ с использованием коллектива деревьев решений // Науч. вестн. Новосиб. гос. техн. ун-та. 2009. № 3 (36). С. 67–76.

6. Лбов Г. С., Бериков В. Б. Устойчивость решающих функций в задачах распознавания образов и анализа разнотипной информации. Новосибирск : Изд-во Ин-та математики, 2005.

7. Ершов Ю. Л., Палютин Е. А. Математическая логика. М. : Наука,1991.

A. A. Vikentiev

INLETTING METRICS FOR EXPERT STATEMENTS WITH PROBABILITIES

In the work the author discusses ways of inletting metrics for expert statements represented as the formulas of Sentence Logic. Methods for inletting metrics on such formulas are offered and properties of the entered metrics are investigated. The research can be applied for solving problems of the best reconciliation of expert statements, for constructing decision functions in pattern recognition and for building expert systems.

Keywords: cluster analysis, expert statements, distance, metrics.

© Викентьев А. А., 2010

УДК 519.6

В. Б. Звонков, Е. С. Семенкин

О РЕШЕНИИ СЛОЖНЫХ ЗАДАЧ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ МОДИФИЦИРОВАННЫМИ ЭВРИСТИЧЕСКИМИ СТОХАСТИЧЕСКИМИ ПРОЦЕДУРАМИ ПРЯМОГО ПОИСКА*

Проведен анализ эффективности самонастраивающегося генетического алгоритма и адаптивной гибридной стохастической процедуры на репрезентативном множестве тестовых и реальных прикладных экономических задач. Сравнение со стандартным генетическим алгоритмом показало целесообразность применения предложенных модификаций.

Ключевые слова: генетический алгоритм, автоматическая настройка, адаптивная гибридная стохастическая процедура, многокритериальная оптимизация.

На современном этапе развития человеческого общества все большее значение в повседневной жизни преобретают сложные технические, технологические, экономические и социальные процессы.

Эти процессы предъявляют высокие требования к системам управления, которые должны обрабатывать значительные объемы информации, мгновенно принимать решения и т. д. А поскольку человеческий потенциал является ограниченным, т. е. в силу своих биологических особенностей человек не способен обрабатывать гигабайты «сырой» информации, мгновенно принимать решения, работать в опасной для здоровья среде, то необходимо исключить участие человека из таких процессов. Однако существующие системы автоматического управления во многих случаях не могут обеспечить достаточную интеллектуальность при принятии решений.

В настоящее время актуальной научной задачей является разработка управленческих систем не с жесткой, а с гибкой архитектурой, способных изменять свои свойства в процессе решения задач на основании накопленной информации и имеющихся знаний с целью приспособления к условиям окружающей среды. В качестве интеллектуальных агентов при решении многих задач широкое применение нашли эволюционные алгоритмы.

Модификации стандартного эволюционного алгоритма. Для эффективного решения задач безусловной и условной однокритериальной оптимизации, а также безусловной многокритериальной оптимизации (в сочетании с методом учета многокритериальности SPEA [1]) был разработан, протестирован и апробирован при решении ряда экономических задач, основанных на реальных данных, самонастраивающийся генетический алгоритм [2], обеспечивающий автоматическую настройку основных параметров эволюционного алгоритма (типа селекции, уровня мутации и типа скрещивания) в ходе решения оптимизационной задачи, без использования априорной информации об эффективности алгоритмических схем эволюционного алгоритма, а также о свойствах задачи и поискового пространства.

Принцип работы предложенного алгоритма состоит в вероятностном выборе основных генетических операторов на основании информации о средней пригодности порожденных потомков с помощью каждой рассматриваемой и автоматически настраиваемой алгоритмической схемы. Изначально вероятности выбора любой настройки или комбинации настроек равны. Адаптация алгоритма обеспечивается при помощи перераспределения вероятностей, когда для лучших комбинаций они увеличиваются, а для худших уменьшаются.

Для эффективного решения задач безусловной и условной многокритериальной оптимизации была разработана, протестирована и апробирована на реальных экономических задачах адаптивная гибридная стохастическая процедура [3].

Адаптивная гибридная стохастическая процедура в качестве алгоритмического ядра использует самонастраивающийся генетический алгоритм [2], а в качестве метода учета многокритериальности задачи – метод SPEA [1]. Новизна данной процедуры состоит в оригинальной самонастройке основных параметров генетического алгоритма, а также в использовании локального поиска для улучшения парето-оптимальных решений, найденных эволюционным алгоритмом [3]. В отличие от классического подхода, локальный поиск здесь используется на каждом поколении, а не только по окончании работы генетического алгоритма. С его помощью осуществляется учет ограничений задачи путем перевода недопустимых индивидов в область допустимых решений.

Рассматриваемые задачи и методика численных экспериментов. В качестве тестовых задач безусловной и условной однокритериальной оптимизации использовался стандартный набор задач [4]. Это задачи нелинейной оптимизации многоэкстремальных функций многих вещественных переменных (после бинаризации длина хромосом составляет 46…50 бит). Целевые функции не имеют свойства дифференцируемости, но обладают большим количеством локальных оптимумов, слабо отличающихся от глобального и нерегулярно расположенных. Разворот функций относительно начала координат, протяженные овраги в области оптимума и наличие системы нелинейных ограничений-неравенств также затрудняют процесс оптимизации.

*Работа выполнена при финансовой поддержке АВЦП «Развитие научного потенциала высшей школы» (НИР 2.1.1/2710) и ФЦП «Научные и научно-педагогические кадры инновационной России» (НИР НК-136П/3).

В качестве тестовых задач безусловной и условной многокритериальной оптимизации рассматривались двух-, трех-, четырех- и пятикритериальные задачи с квадратичными целевыми функциями и нелинейными ограничениями-неравенствами.

Двухкритериальная безусловная задача:

$$\begin{cases} f_1(x,y) = (x-3)^2 + (y-3)^2 \to \min, \\ f_2(x,y) = (x+4)^2 + (y+4)^2 \to \min. \end{cases}$$

Трехкритериальная безусловная задача:

$$\begin{cases} f_1(x,y) = (x-3)^2 + (y-3)^2 \to \min, \\ f_2(x,y) = (x+4)^2 + (y+4)^2 \to \min, \\ f_3(x,y) = (x+4)^2 + (y-3)^2 \to \min. \end{cases}$$

Четырехкритериальная безусловная задача:

$$\begin{cases} f_1(x,y) = (x-3)^2 + (y-3)^2 \to \min, \\ f_2(x,y) = (x+4)^2 + (y+4)^2 \to \min, \\ f_3(x,y) = (x+4)^2 + (y-3)^2 \to \min, \\ f_4(x,y) = (x-3)^2 + (y+4)^2 \to \min. \end{cases}$$

Пятикритериальная безусловная задача:

$$\begin{cases} f_1(x,y) = (x-3)^2 + (y-3)^2 \to \min, \\ f_2(x,y) = (x+4)^2 + (y+4)^2 \to \min, \\ f_3(x,y) = (x+4)^2 + (y-3)^2 \to \min, \\ f_4(x,y) = (x-3)^2 + (y+4)^2 \to \min, \\ f_5(x,y) = (x-5)^2 + y^2 \to \min. \end{cases}$$

В качестве прикладных задач были использованы следующие экономические задачи:

– формирование оптимального инвестиционного портфеля предприятия (Химзавода – филиала ОАО «Красмаш»);

– формирование оптимального кредитного портфеля банка (Красноярского филиала Банка Москвы);

– оптимизация управления кредитным финансированием проектов с длительным инвестиционным циклом (ОАО «Информационные спутниковые системы» имени академика М. Ф. Решетнева»).

Постановка задач и исходные данные заимствованы из источников [5; 6].

Все прикладные задачи являются задачами условной одно- и многокритериальной (кроме задачи ОАО «Информационные спутниковые системы») нелинейной булевой оптимизации алгоритмически заданных функций в бинарном пространстве высокой размерности (длина хромосом составляет 50…66 бит) с системой ограничений-неравенств.

Для решения тестовых и реальных задач в однокритериальной постановке использовались следующие настройки эволюционных алгоритмов: размер популяции – 100, количество поколений – 100, вероятность скрещивания – 1, количество статистически независимых прогонов алгоритма – 500 или 1 000. Для решения тестовых и реальных задач в многокритериальной постановке применялись следующие настройки: размер популяции – 100, количество поколений – 100, вероятность скрещивания – 1, количество статистически независимых прогонов алгоритма – 100. Размер недоминируемого внешнего множества варьировался на каждой задаче с целью определения наиболее эффективного. В данной статье приведен априорный размер внешнего множества в каждом прогоне для каждой из задач.

Для решения каждой задачи проводился полный набор численных экспериментов с помощью стандартного генетического алгоритма и предложенных в данной статье модификаций, включающий в себя все рассматриваемые комбинации параметров генетических алгоритмов (табл. 1, 2), с последующим усреднением показателей эффективности алгоритмов (надежности, времени работы, скорости сходимости к оптимуму, процента паретовских точек, процента допустимых точек, разброса точек в пространстве независимых переменных и разброса точек в пространстве целевых функций) (в тех задачах, где возможно введение таких показателей эффективности) по независимым случайным реализациям алгоритмов и применением методов статистической обработки экспериментальных данных [7–9].

В качестве методов учета ограничений для задач условной однокритериальной оптимизации использовались статические штрафы, динамические штрафы, адаптивные штрафы, смертельные штрафы, поведенческая память, лечение, лечение + смертельные штрафы [10].

Таблица 1

Количество комбинаций параметров эволюционных алгоритмов на одной задаче безусловной оптимизации

Тип генетического алгоритма	Общее число комбинаций параметров генетического алгоритма
Стандартный	360
Автонастройка мутации	120
Автонастройка скрещивания	120
Автонастройка селекции	72
Автонастройка мутации и скрещивания	40
Автонастройка селекции и мутации	24
Автонастройка селекции и скрещивания	24
Автонастройка селекции, мутации и скрещивания	8

Таблица 2

Количество комбинаций параметров эволюционных алгоритмов на одной задаче условной оптимизации

Тип генетического алгоритма	Общее число комбинаций параметров генетического алгоритма
Стандартный	1 800
Автонастройка мутации	600
Автонастройка скрещивания	600
Автонастройка селекции	360
Автонастройка мутации и скрещивания	200
Автонастройка селекции и мутации	120
Автонастройка селекции и скрещивания	120
Автонастройка селекции, мутации и скрещивания	40

Для получения требуемого количества допустимых индивидов применялась процедура минимизации штрафной функции, функция пригодности для которой имеет вид

$$\text{fitness}(x) = \frac{1}{\text{sum} + 1},$$

где $\text{sum} = \sum_{j=1}^{m} f_j^2(x)$ – штрафная функция, здесь $f_j(x)$ – штраф за нарушение j-го ограничения.

При этой процедуре самонастраивающийся генетический алгоритм работает со всей популяцией до тех пор, пока не будет получено требуемое количество допустимых индивидов.

Рассматривались два способа определения метрики в методе лечения:

– при первом способе расстояние между индивидами задается значением целевой функции (лечение-1);

– при втором способе расстояние между индивидами определяется их координатами (лечение-2).

При решении реальных задач использовался способ «лечение-2», а метод статических штрафов не применялся в силу результатов проведенных ранее исследований, показавших более высокую эффективность способа «лечение-2» по сравнению со способом «лечение-1» и нецелесообразность реализации автоматической настройки коэффициентов в методе статических штрафов. Кроме того, были установлены оптимальные параметры методов учета ограничений для тестовых и реальных задач однокритериальной условной оптимизации [11] (табл. 1, 2).

Экспериментальные результаты. Приведем результаты решения некоторых из описанных выше задач многокритериальной оптимизации (рис. 1, 2) (результаты решения тестовых и прикладных задач однокритериальной оптимизации представлены в работе [11]). Эти результаты отражают наиболее эффективные настройки каждого алгоритма на каждой задаче, полученные с помощью многократных прогонов алгоритмов и последующего статистического анализа в пакете Statistica. На рис. 1 сплошными линиями показано истинное множество и фронт Парето, точками обозначены наилучшие результаты работы генетических алгоритмов; на рис. 2 черными точками отмечены точки, полученные эволюционными алгоритмами, белыми – точки, определенные методом полного перебора.

Для эффективной реализации методики решения сложных задач оптимизации и проведения научных исследований была создана программная система Intelligence Technologies – Self-Adapting Genetic Algorithm (IT-SAGA). С ее помощью были получены следующие результаты:

– доказано выполнение условия статистического ансамбля по непараметрическим критериям в пакете Statistica;

– наилучшей модификацией на классе задач однокритериальной оптимизации по критериям надежности работы, скорости сходимости к оптимуму и времени работы является самонастраивающийся генетический алгоритм с автоматическим выбором селекции, мутации и скрещивания;

– в тестовых задачах многокритериальной оптимизации множество и фронт Парето, полученные адаптивной гибридной стохастической процедурой, более точно описывают истинные парето-оптимальные множество и фронт в сравнении с самонастраивающимся генетическим алгоритмом в сочетании с методом SPEA;

– в прикладных задачах при размерности поискового пространства, равной 2^{25}, фронт Парето, полученный самонастраивающимся генетическим алгоритмом в сочетании с методом SPEA, содержит точки, не принадлежащие истинному парето-оптимальному фронту, найденному методом полного перебора. Фронт Парето, полученный адаптивной гибридной стохастической процедурой, совпадает с точностью до одной или двух точек с парето-оптимальным фронтом, определенным методом полного перебора;

– наилучшей модификацией на классе задач многокритериальной оптимизации является адаптивная гибридная стохастическая процедура по следующим критериям эффективности: процент паретовских точек, процент допустимых точек, разброс точек в пространстве независимых переменных, разброс точек в пространстве целевых функций;

– при размерности поискового пространства 2^{50} и выше, в условиях отсутствия информации о множестве и фронте Парето и неприменимости полного перебора (из-за невозможности получения решения за приемлемое время), предложенные в данной статье модификации генетического алгоритма обеспечивают наилучшее известное на настоящий момент решение, найденное другими генетическими алгоритмами [5; 6] за время, сравнимое со временем работы данных мо-

дификаций при решении прикладных задач в однокритериальной постановке [11].

Таким образом, самонастраивающийся генетический алгоритм, адаптивная гибридная стохастическая процедура, методы учета ограничений для задач условной оптимизации, метод SPEA, метод полного перебора и оригинальный алгоритм локального поиска, являясь алгоритмическим ядром разработанной программной системы Intelligence Technologies – Self-Adapting Genetic Algorithm, прошли успешную апробацию при решении репрезентативного множества тестовых и прикладных задач однокритериальной и многокритериальной оптимизации (всего решены 22 тестовых и 3 практических задачи).

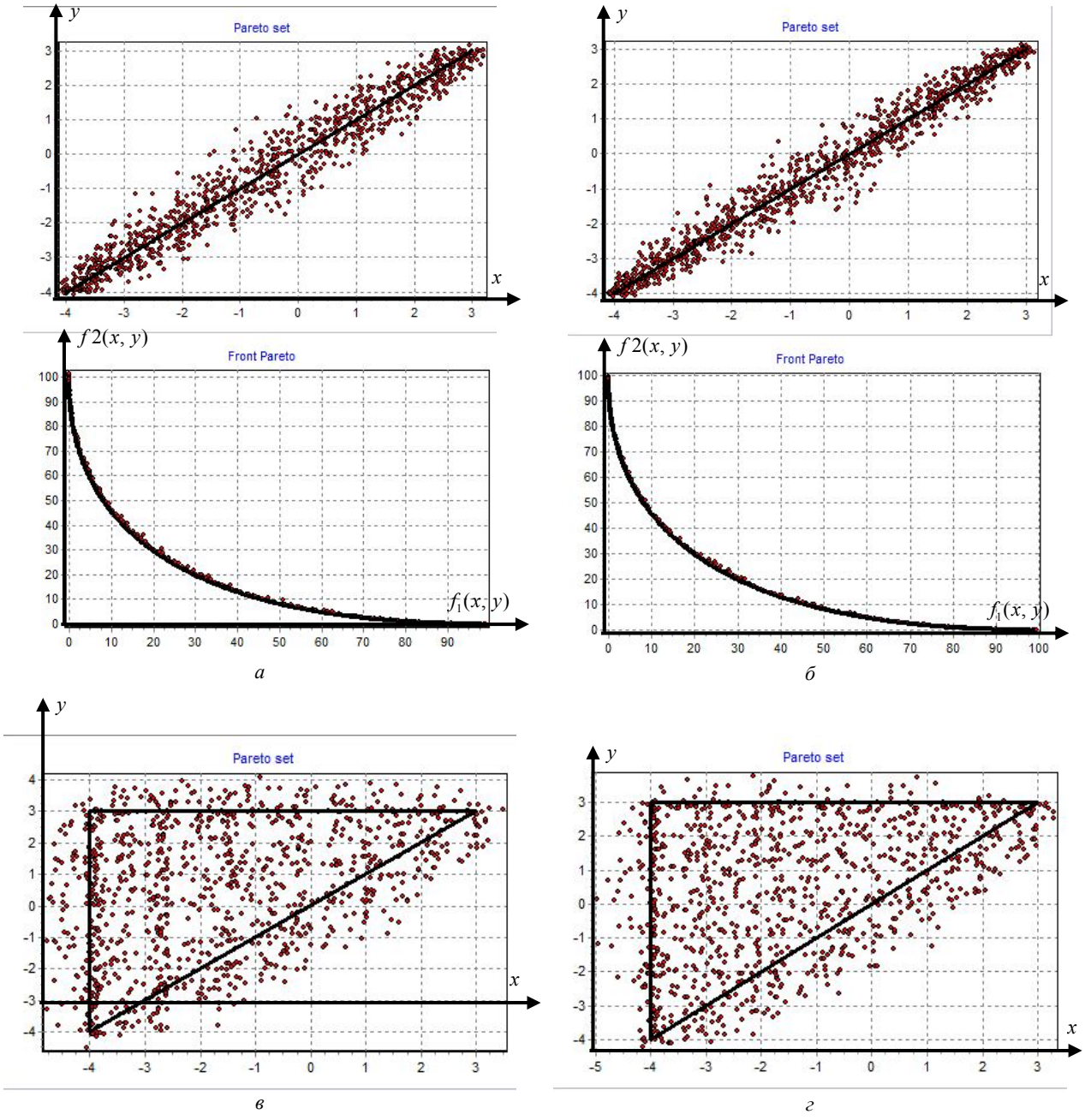
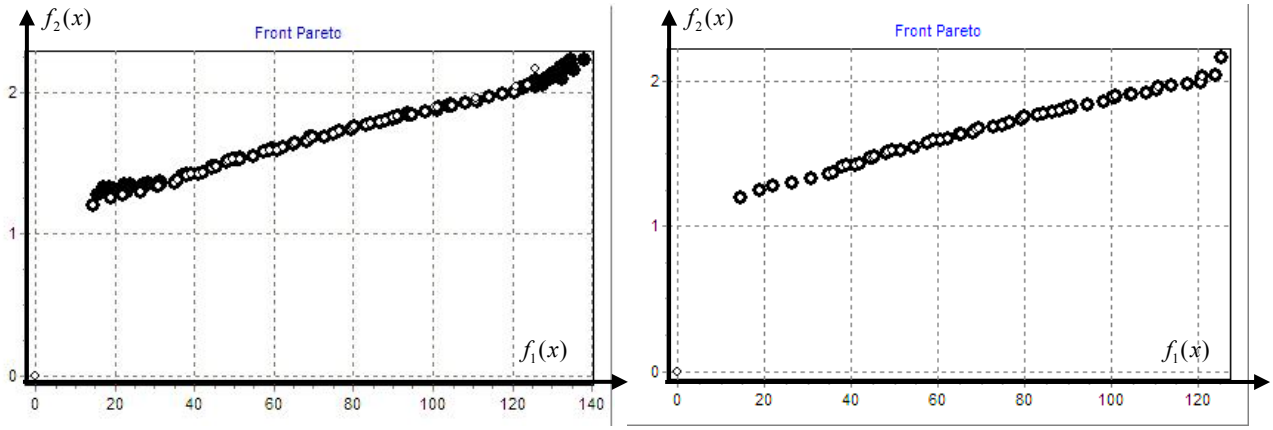


Рис. 1. Результаты решения некоторых тестовых задач многокритериальной оптимизации стандартным (*а*, *в*) и самонастраивающимся (автоматическая настройка мутации и скрещивания) (*б*, *г*) генетическими алгоритмами в сочетании с методом SPEA:
а, *б* – двухкритериальная задача, размер внешнего множества – 50 точек (верхний рисунок соответствует множеству Парето, нижний – фронту Парето); *в*, *г* – трехкритериальная задача, размер внешнего множества – 50 точек (множество Парето)



а *б*

Рис. 2. Результаты решения многокритериальной задачи формирования оптимального инвестиционного портфеля предприятия самонастраивающимся генетическим алгоритмом (автоматическая настройка мутации и скрещивания) в сочетании с методом SPEA (*а*) и адаптивной гибридной стохастической процедурой (*б*) (фронт Парето): размер внешнего множества – 50 точек; $f_1(x)$ – ожидаемая прибыль; $f_2(x)$ – рискованность программы

В ходе проведения исследований было доказано выполнение условий статистического ансамбля, что является подтверждением неслучайности превосходства предложенных в данной статье модификаций над стандартным генетическим алгоритмом по всем критериям эффективности (надежность, время работы, скорость сходимости к оптимуму, процент паретовских точек, процент допустимых точек, разброс точек в пространстве независимых переменных и разброс точек в пространстве целевых функций). Поэтому можно сделать обоснованный вывод о практической ценности самонастраивающегося генетического алгоритма и адаптивной гибридной стохастической процедуры.

Библиографические ссылки

1. Zitzler E., Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach // IEEE Transactions on Evolutionary Computation. 1999. Vol. 3, № 4. P. 257–271.

2. Звонков В. Б. Самонастраивающийся генетический алгоритм решения сложных задач оптимизации // Всероссийский конкурс компьютерных программ : тр. III тура Всерос. студ. олимпиады. Вологда, 2010.

3. Звонков В. Б., Семенкин Е. С. Адаптивная гибридная стохастическая процедура // Решетневские чтения : сб. тез. докл. XIV Междунар. науч. конф. : в 2 ч. Ч. 2 / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2010. С. 389–390.

4. Real-Parameter Black-Box Optimization Benchmarking 2009: Experimental Setup : INRIA research report RR-6828 / N. Hansen [et al.]. Saclay, France, 2009.

5. Клешков В. М., Семенкин Е. С. Модели и алгоритмы распределения общих ресурсов при управлении инновациями реструктурированного машиностроительного предприятия // Пробл. машиностроения и автоматизации. 2006. № 3. С. 24–31.

6. Пуртиков, В. А. Оптимизация управления формированием кредитного портфеля банка: дис. … канд. техн. наук. Красноярск, 2001.

7. Вуколов Э. А. Основы статистического анализа. М. : Форум : Инфра-М, 2004.

8. Тюрин Ю. Н., Макаров А. А. Анализ данных на компьютере. М. : Инфра-М, 2003.

9. Электронный учебник по статистике [Электронный ресурс] // StatSoft : общедоступ. сервис. URL: http://www.statsoft.ru/home/textbook/default.htm (дата обращения: 07.09.2010).

10. Michalewicz Z., Schoenauer M. Evolutionary algorithms for constrained parameter optimization problems // Evolutionary Computation. 1996. Vol. 4, № 1. P. 1–32.

11. Звонков В. Б. Исследование эффективности методов учета ограничений адаптивного эволюционного алгоритма на задачах условной однокритериальной оптимизации // Научные исследования и их практическое применение. Современное состояние и пути развития : сб. науч. тр. по материалам Междунар. науч.-практ. конф. : в 2 т. Т. 2. Технические науки. Одесса : Черноморье, 2010. С. 70–83.

V. B. Zvonkov, E. S. Semenkin

SOLVING COMPLEX MULTIOBJECTIVE OPTIMIZATION PROBLEMS WITH THE MODIFIED HEURISTIC STOCHASTIC ALGORITHMS OF DIRECT SEARCH

In this paper the analysis of efficiency of automatically adjusting genetic algorithm and adaptive hybrid stochastic procedure has been fulfilled on representative set of test and real world economical problems. It has been proved that the developed modifications outperform the standard genetic algorithm.

Keywords: genetic algorithm, automatical adjustment, adaptive hybrid stochastic procedure, multi-criteria optimization.

© Звонков В. Б., Семенкин Е. С., 2010

УДК 519.673, 519.245

Л. А. Козинкин

МЕТОД ПРЯМОГО СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ В ЗАДАЧЕ ТОМСОНА*

Представлены минимальные и равновесные конфигурации систем зарядов на сфере, найденные с помощью метода Монте-Карло. Приведен анализ впервые полученных в ходе работы результатов.

Ключевые слова: задача Томсона, метод прямого статистического моделирования.

На рубеже XIX–XX вв. при изучении планетарной модели атома английским физиком Дж. Томсоном была сформулирована следующая задача: необходимо определить, какие конфигурации N одинаковых зарядов на сфере и в каких количествах дают минимум потенциальной энергии системы.

Математически эта задача сводится к поиску величины

$$W_N = \inf_{y^{(1)}, ..., y^{(N)} \in S^2} W(y^{(1)}, ..., y^{(N)}),$$

$$W(y^{(1)}, ..., y^{(N)}) = \sum_{\substack{i,\ j=1 \\ i \neq j}}^{N} \frac{1}{| y^{(i)} - y^{(j)}}.$$

Еще сам Томсон проводил физические эксперименты по нахождению таких конфигураций для небольшого количества зарядов. Решающее значение для решения этой задачи уже в середине XX в. имела теория приближения функций П. Л. Чебышева, на основе которой частные случаи (при N = 2, 3, 4, 6, 12) были исследованы аналитически. Соответствующие конфигурации приведены в таблице на светло-сером фоне.

Дальнейшие исследования привели лишь к частичным результатам для случая N = 120 в 4-мерном пространстве (заряды составляют правильный многогранник, имеющий соответствующее число вершин) и N = 196 560 в 24-мерном пространстве (заряды расположены на концах минимальных векторов решетки Лича).

Также существуют конфигурации, полученные экспериментальным путем, но не доказанные математически (они выделены в таблице темным фоном).

В данной статье для решения задачи Томсона использовался метод прямого статистического моделирования. Была разработана компьютерная модель, основанная на кинематике движения зарядов по поверхности сферы и их взаимодействия между собой.

Первая реализация модели выполнена в среде Borland C++ Builder. Она проводила итеративный поиск решения задачи Томсона для заданного количества зарядов. За каждую итерацию рассчитывались силы взаимодействия зарядов друг с другом, и на основе их результирующей выполнялся одновременный пропорциональный сдвиг зарядов. Благодаря классам визуализации и системе отображения результатов математически доказанные случаи были подтверждены экспериментально.

Следующая реализация модели разработана в среде Visual Studio с применением библиотек MPI в связи с необходимостью статистического исследования полученных конфигураций. Это существенно увеличило производительность моделирования за счет использования распределенных по локальной сети вычислений.

Таким образом, для каждого числа N генерируется 500 000 случайных начальных конфигураций на сфере и для каждой такой конфигурации рассчитывается около 100 000 итераций по времени. Такой подход позволяет с высокой вероятностью сделать заключение о минимальности потенциальной энергии той или иной системы зарядов, а также найти равновесные конфигурации (рис. 1–8).

Исследования разработанной модели позволили найти ряд неизвестных конфигураций в задаче Томсона. Результаты компьютерного моделирования некоторых систем зарядов получены впервые и приведены в таблице на белом фоне.

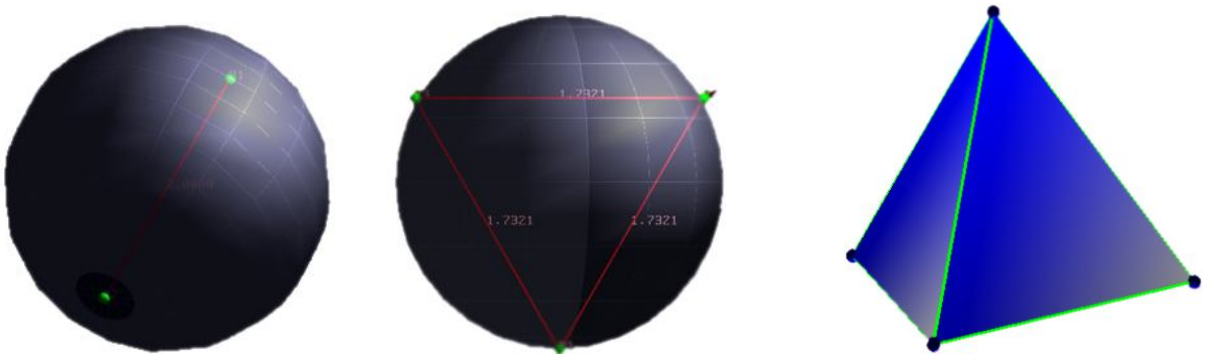
Исследованные конфигурации систем зарядов

Число зарядов	Потенциальная энергия	Характеристика конфигурации
2	1	Полюса сферы
3	$2\sqrt{3}$	Правильный треугольник на экваторе
4	$3\sqrt{6}$	Правильный тетраэдр
5	12,949 381 8	Два заряда на полюсах, три заряда образуют правильный треугольник на экваторе
5	12,967	Правильная пирамида
6	$3 + 12\sqrt{2}$	Правильный октаэдр
7	28,905 956 3	Два заряда на полюсах, пять зарядов образуют правильный пятиугольник на экваторе
8	39,350 574 5	Антипризма
9	51,520	Заряды образуют три пирамиды с прямоугольниками в основаниях, которые в свою очередь образуют друг с другом правильную треугольную призму

*Работа поддержана Российским фондом фундаментальных исследований (грант № 08-01-00312).

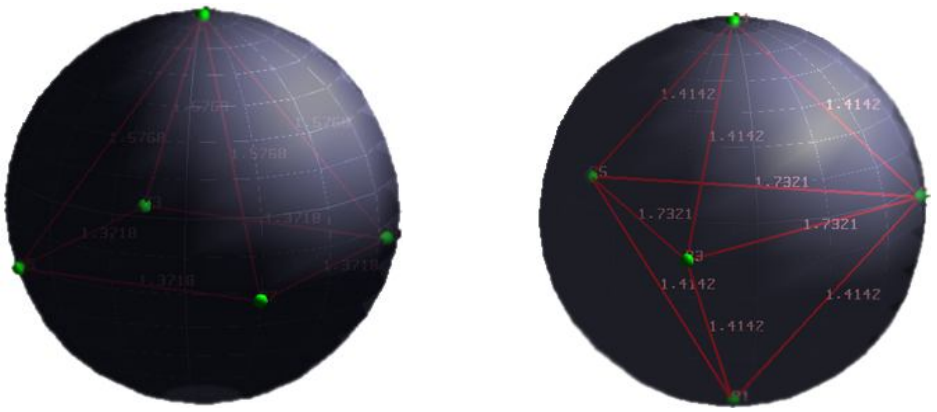
Окончание таблицы

Число зарядов	Потенциальная энергия	Характеристика конфигурации
10	65,433 899 0	Два заряда на полюсах, остальные заряды образуют антипризму, равноудаленную основаниями от полюсов
11	81,192 901 6	Конфигурация, как в предыдущем случае, отличающаяся только асимметрией за счет добавления одиннадцатого заряда между основанием антипризмы и зарядом на одном из полюсов
12	$6 + 15\sqrt{(10 - 2\sqrt{5})} + 15\sqrt{(10 + 2\sqrt{5})}$	Правильный икосаэдр
13	117,706	Предположительно деформированный икосаэдр
14	138,612 686 2	Два заряда на полюсах, остальные образуют шестиугольную антипризму с основаниями, равноудаленными от полюсов
15	161,340 484 6	Конфигурация, как в предыдущем случае, но деформированная за счет пятнадцатого заряда, разместившегося рядом с одним из полюсов
16	185,823	Предположительно деформированная двумя зарядами конфигурация $N = 14$
16	185,841	Равновесная конфигурация с антипризматическими основаниями северного и южного полюсов
17	212,101	Конфигурация, образованная двумя правильными пирамидами, с пятиугольниками в основании, вершинами на полюсах и квадратами, повернутыми на 45° и соединенными друг с другом, на экваторе
18	240,169	Антипризматически расположенные правильные пирамиды с квадратом в основании и одинаково ориентированные пирамиды с неправильными одинаковыми прямоугольниками в основании на экваторе
19	270,179	Правильная пирамида с правильным шестиугольником в основании, расположенная на одном полюсе параллельно квадрату на другом
20	301,763	Более устойчивая, чем додекаэдр, конфигурация



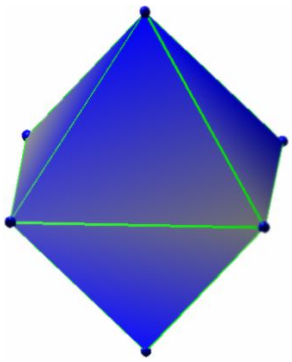
$N = 2$ $N = 3$ $N = 4$

Рис. 1

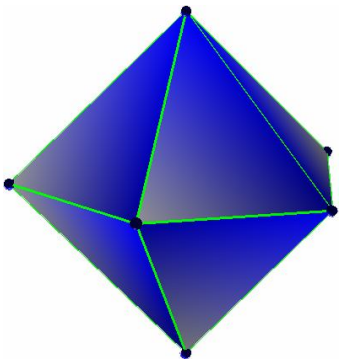


$N = 5$ (равновесная конфигурация) $N = 5$ (конфигурация с минимальной энергией)

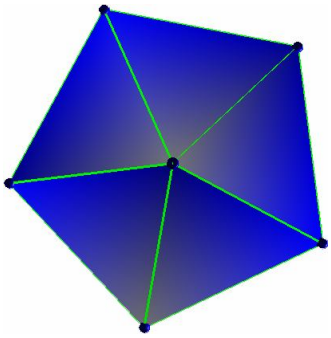
Рис. 2



$N = 6$

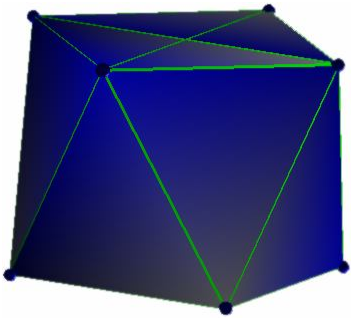


$N = 7$

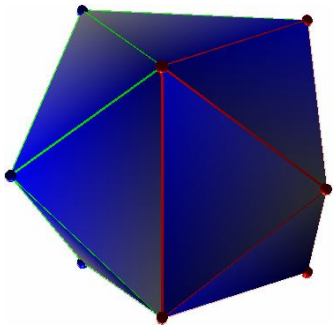


$N = 7$ (вид сверху)

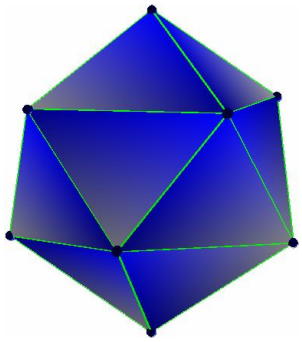
Рис. 3



$N = 8$

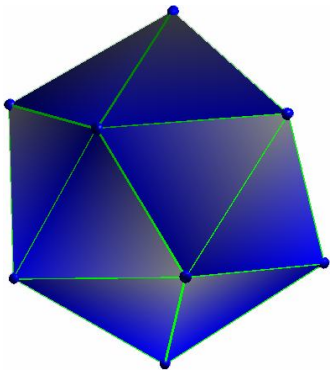


$N = 9$

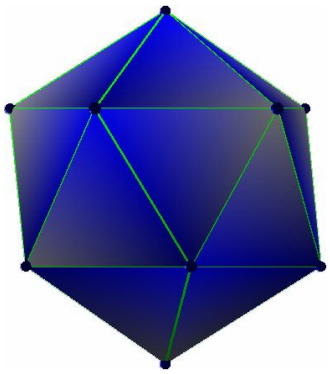


$N = 10$

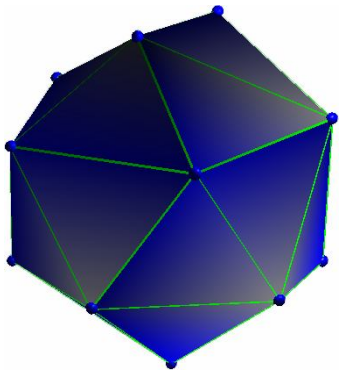
Рис. 4



$N = 11$

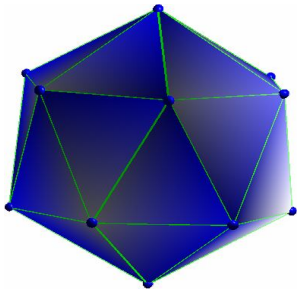


$N = 12$

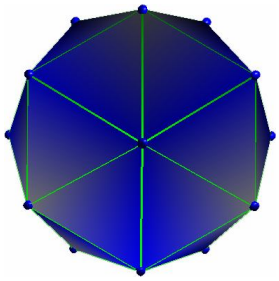


$N = 13$

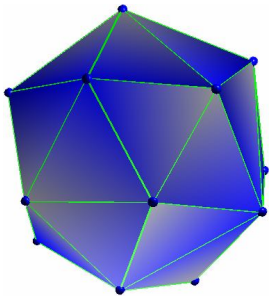
Рис. 5



$N = 14$

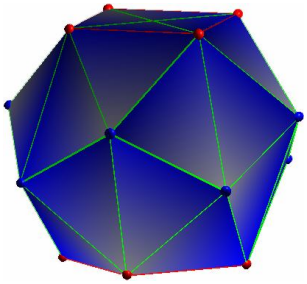


$N = 14$ (вид сверху)

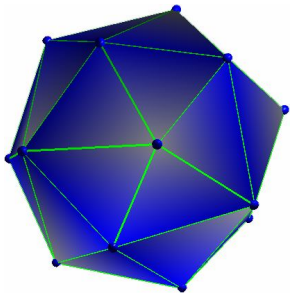


$N = 15$

Рис. 6

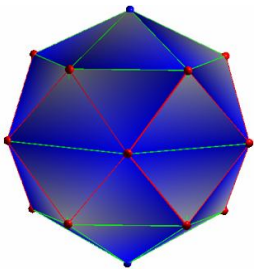


$N = 16$ (равновесная конфигурация)

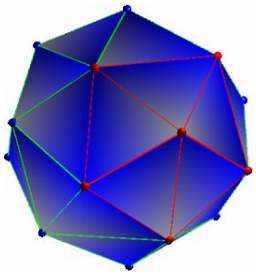


$N = 16$ (минимальная конфигурация)

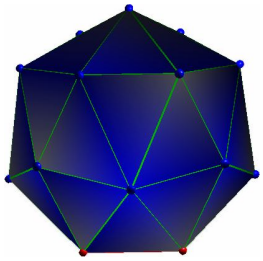
Рис. 7



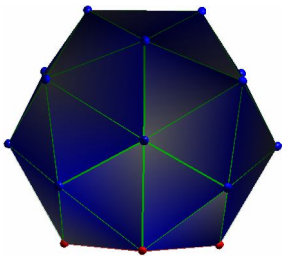
$N = 17$



$N = 18$



$N = 19$



$N = 20$

Рис. 8

Таким образом, были исследованы системы зарядов, обладающие минимальной потенциальной энергией. Большой интерес также представляют равновесные конфигурации, являющиеся локальными минимумами, а также вероятности распределения зарядов в той или иной конфигурации.

L. A. Kozinkin

DIRECT STATISTICAL SIMULATION METHOD IN THOMSON'S PROBLEM

In the article the author presents results of search of minimal and stable arrangements of points on a sphere with the help of Monte Carlo method. Analysis of new arrangements is carried out.

Keywords: Thomson's problem, direct statistical simulation method.

© Козинкин Л. А., 2010

УДК 621.396.6/0751

Е. И. Кротова

МЕТОД ОЦЕНКИ ВЛИЯНИЯ АДДИТИВНЫХ ПОМЕХ НА ВХОДЕ ПРИЕМНИКА СИГНАЛОВ С ПОМОЩЬЮ ИДЕНТИФИКАЦИИ ВИДОВ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

Рассмотрен метод оценки влияния помех с различными законами распределения на входе приемника телекоммуникационной системы по комбинированному параметру идентификации на сигнал с модуляцией минимальным частотным сдвигом.

Ключевые слова: информация, закон распределения, идентификация, помехи.

Система связи должна быть спроектирована так, чтобы она могла наилучшим образом противостоять действию помех. Для этого необходимо знать характеристики этих помех и анализировать их влияние на информационный сигнал.

В качестве основных параметров, характеризующих качество связи, используются отношение сигнал/шум, коэффициент вариации и т. п.

В современных системах связи применяются сигналы, обеспечивающие высокую помехоустойчивость, к которым, в частности, относятся сигналы с манипуляцией минимальным частотным сдвигом (ЧМн). При индексе модуляции, равном 0,5, ЧМн-сигналы имеют большее сосредоточение энергии в центральной области спектра, что обеспечивает высокую помехоустойчивость к несосредоточенным помехам.

В случае же сосредоточенных помех помехоустойчивость оказывается прямо пропорциональной индексу модуляции.

Обычно влияние помех оценивается пороговыми устройствами, в основе работы которых лежат различные методы принятия решений, использующие различные критерии; например критерий Неймана-Пирсона [1].

Для того чтобы получить аналитические выражения для оптимального алгоритма (правила) обработки смеси при принятии решений (гипотез) о действии сигнала или его отсутствии, нужно вероятностно описать смесь на входе системы при наличии в ней сигнала и помехи или действии только одной помехи. Принятие решения о наличии сигнала сопровождается минимальным средним риском:

$$l(x) = \frac{p\left(x_1 x_2 ... x_{k_n} / sn \right)}{p\left(x_1 x_2 ... x_{k_n} / n \right)} > \Pi , \qquad (1)$$

где $\Pi = r_{пр}P(s)/r_{л.о}P(0)$ – порог; $l(x)$ – отношение правдоподобия; $p(x_1x_2... x_{k_n}/sn)$, $p(x_1x_2... x_{k_n})$ – плотности вероятности смеси сигнала и помехи и одной помехи соответственно [1].

Проводя анализ дезинформационного действия случайных помех с различными законами распределения вероятностей, К. Шеннон пришел к выводу, что вносимая помехой дезинформация определяется не только мощностью этой помехи, т. е. ее среднеквадратичным отклонением (СКО) σ, но и видом закона распределения помехи [2]. Анализируя математические операции, предусмотренные выражением (1), можно синтезировать оптимальную схему обнаружителя дискретного радиосигнала [3].

В данной статье представлен метод, позволяющий оценивать и контролировать сигнал и помехи на входе приемного устройства системы связи ЧМн-сигналов в реальном времени по комбинированному параметру идентификации плотности вероятности их выборочных значений.

Частотная манипуляция с минимальным сдвигом – это двоичная цифровая частотная манипуляция. Такой сигнал может быть представлен в виде [4]:

$$y(t) = \cos(\omega_0 t + \frac{C_k \pi}{2T} t + \varphi_0) , \qquad (2)$$

где $\varphi(t)$ – изменение фазы:

$$\varphi(t) = \frac{C_k \pi [t - (k-1)T]}{2T} + \sum_{j=1}^{k-1} \frac{C_j \pi}{2} \varphi_0 ; \qquad (3)$$

$kT \le t \le (k+1)T$, $k = 1, 2, 3, ...$; $\omega_0 = 2\pi f_0$ – средняя частота сигнала; T – длительность единичного символа; φ_0 – начальная фаза.

Для того чтобы использовать тот или иной метод приема, необходимо знать вид помехи, воздействующей на сигнал, т. е. закон ее распределения. При наблюдении проводится анализ законов распределения исходного сигнала и смеси. Существуют несколько методов идентификации законов распределения, однако интерес представляют только те из них, которые обеспечивают однозначность идентификации.

Анализ дезинформационного действия случайных помех с различными законами распределения вероятностей приводит к выводу, что вносимая помехой дезинформация зависит не только от мощности этой помехи, т. е. ее СКО, но и от вида закона распределения этой помехи [4].

Возможность идентификации формы распределения экспериментальных данных ограничена прежде всего малостью объема выборки. При большом объеме выборки, например в несколько тысяч наблюдений, построение гистограммы позволяет получить достаточно плавную кривую, отражающую все характерные особенности наблюдаемого закона, но потребует значительного времени обработки данных.

Однозначность идентификации при малых объемах выборки можно обеспечить, если в качестве параметра идентификации использовать коэффициент Z, получаемый при суммировании отношения коэффициента энтропии k к контрэксцессу χ с коэффициентом асимметрии A, который умножается на масштабирующий коэффициент 4 [5]:

$$Z = \frac{k}{\chi} + 4A . \qquad (4)$$

При введении данного параметра точность идентификации резко повышается благодаря однозначности идентификации при различных объемах выборки.

Алгоритм идентификации вида распределения случайного процесса по параметру Z можно реализовать, используя выборочные значения наблюдаемого процесса. Этот алгоритм состоит из пяти шагов.

Шаг 1. Определяется значение контрэксцесса χ и асимметрии A:

– находится математическое ожидание:

$$m_1^* = \frac{1}{n} \sum_{i=1}^{n} x_i , \qquad (5)$$

где n – число измерений; x_i – значение случайной величины;

– определяется оценка момента третьего порядка:

$$\mu_3^* = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - m_1^* \right)^3 ; \qquad (6)$$

– находится оценка момента четвертого порядка:

$$\mu_4^* = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - m_1^* \right)^4 ; \qquad (7)$$

– вычисляется оценка дисперсии:

$$\sigma^{2*} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - m_i^* \right)^2 ; \qquad (8)$$

– находятся эксцесс:

$$\varepsilon = \frac{\mu_4^*}{\sigma^{*4}} , \qquad (9)$$

и контрэксцесс:

$$\chi = \frac{1}{\sqrt{\varepsilon}} ; \qquad (10)$$

– определяется асимметрия:

$$A = \frac{\mu_3^*}{\sigma^{*3}}. \qquad (11)$$

Шаг 2. Определяется энтропийный коэффициент $k_э$ – числовая характеристика формы распределения по той же выборке, для чего рассматривается массив исследуемого процесса и гистограмма и используется оценка среднеквадратического значения:

$$k_э = \frac{d \cdot n}{2 \cdot \sigma^*} 10^{-\frac{1}{n}\sum_{j=1}^{m} n_j \lg n_j}, \qquad (12)$$

где d – ширина столбца гистограммы; n – объем выборки; σ – среднеквадратическое отклонение; m – число столбцов гистограммы $m = 4\ln(n)$; n_j – число наблюдений в столбце.

Шаг 3. Определяется параметр Z по формуле (4).

Шаг 4. Вычисляется величина абсолютного значения отклонения отношений $\left| d_i(Z) \right|$, которая характеризует отличие исследуемого распределения от теоретического для данного закона распределения и заданного объема выборки:

$$\left| Z_{эт} - Z \right| = d_i(Z), \qquad (13)$$

где $Z_{эт}$ – параметр отношения для теоретического закона распределения; Z – параметр исследуемого распределения.

Шаг 5. Величина $d_i(Z)$ сравнивается с допустимым значением отклонения $d_{эi}(Z)$, результат сравнения S_i является определяющим параметром, характеризующим вид распределения.

Экспериментальные кривые, полученные при исследовании зависимости предложенного параметра идентификации вида закона распределения Z, определяемого по формуле (4), от объема выборки (рис. 1), при объемах выборки $n > 20$ не пересекаются и не лежат близко друг к другу, что позволяет однозначно идентифицировать вид распределения при объемах выборки $n > 20$. При этом относительная погрешность определения параметра Z в зависимости от объема выборки не превышает 8 % (рис. 2).

Рассматриваемый метод был применен для исследования влияния аддитивных помех с различными законами распределения: нормальным, лапласовским и арксинусоидальным – на ЧМн-сигнал.

Параметры эксперимента: несущая частота ЧМн-сигнала $f_н$ = 10 МГц; объем выборки экспериментальных данных n = 1 000; полоса частот 100 КГц; индекс модуляции d = 0,5; распределение сигнала в отсутствии помех идентифицировано как арксинусоидальное. Исследовалась аддитивная смесь сигнала и помехи.

В результате проведения экспериментов было установлено, что помехоустойчивость ЧМн-сигнала при аддитивном воздействии лапласовской помехи при отношении «сигнал/помеха» (ОСП) более 1 по мощности хуже, чем при воздействии нормальной помехи (рис. 3). При ОСП < 1 помехоустойчивость ЧМн-сигнала при воздействии лапласовской и нормальной помех примерно одинакова. Арксинусная помеха при всех отношениях «сигнал/помеха» оказывает минимальное воздействие на вид распределения ЧМн-сигнала (минимум параметра Z на графиках соответствует близости распределения смеси сигнала и помехи к арксинусоидальному распределению ЧМн-сигнала).

Полученные результаты также свидетельствуют об уменьшении коэффициента вариации γ при ОСП < 0,5, что соответствует уменьшению величины разброса данных смеси сигнала и помехи от среднего значения (рис. 4). При ОСП > 0,5 коэффициент вариации не изменяется. Таким образом, ОСП = 0,5 является пороговым для ЧМн-сигнала при воздействии аддитивной помехи.

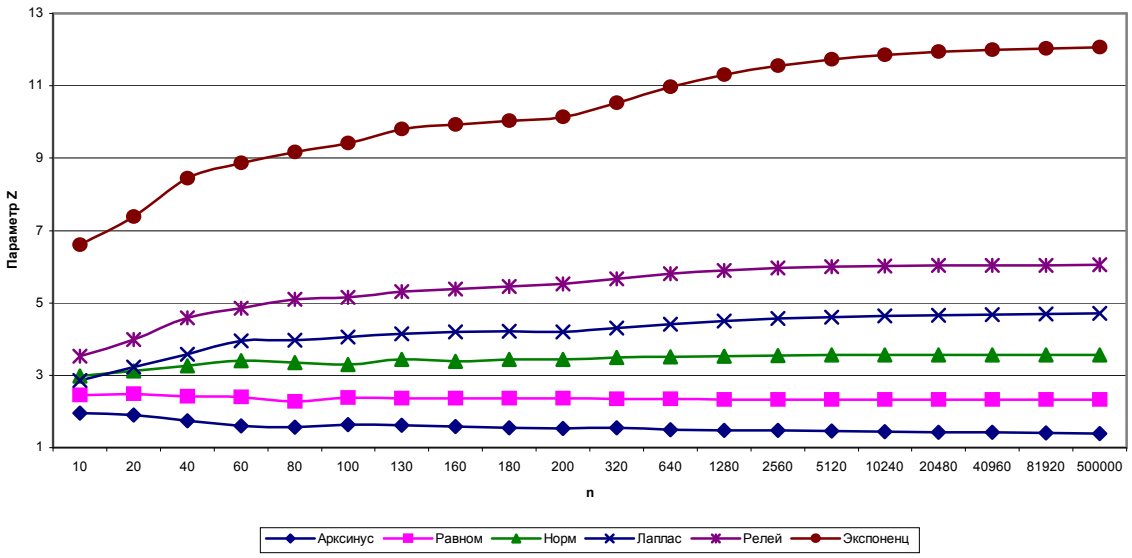


Рис. 1. Зависимость параметра Z от объема выборки при различных законах распределения

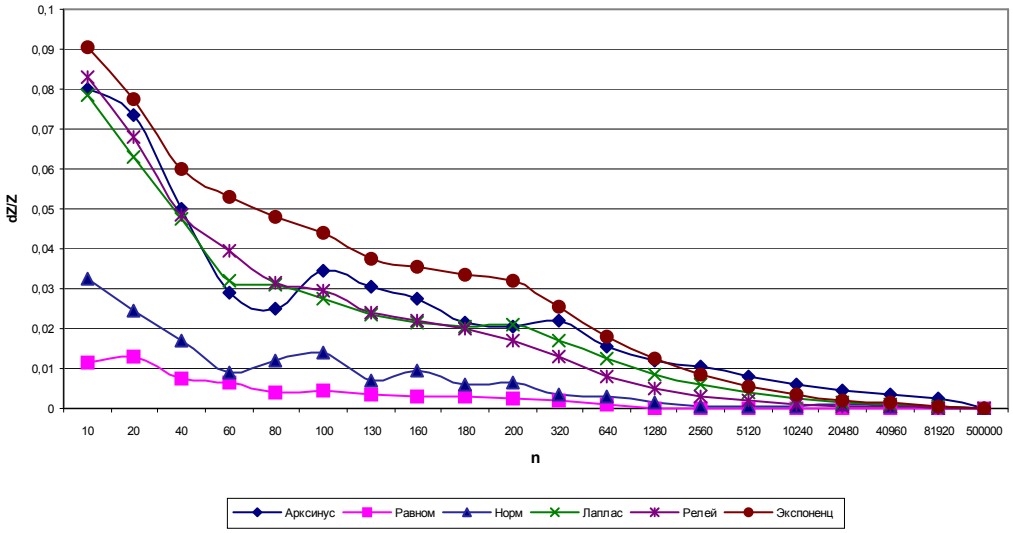


Рис. 2. Зависимость относительной ошибки определения параметра *Z* от объема выборки

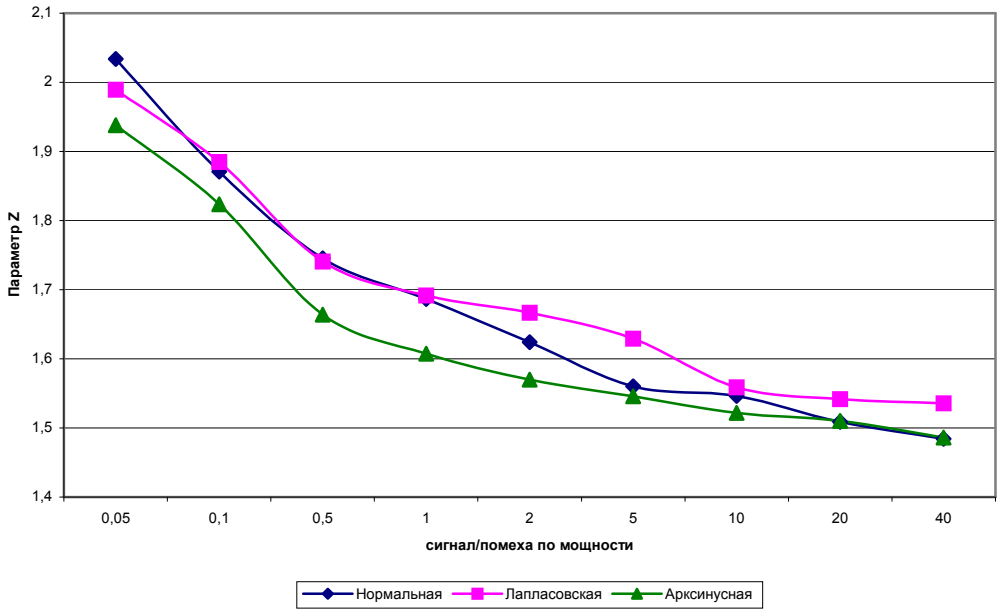


Рис. 3. Зависимость параметра *Z* от отношения «сигнал/помеха» по мощности при воздействии на ЧМн-сигнал помех с различными законами распределения для полосы частот 100 КГц

Если рассматривать зависимость относительной ошибки определения параметра *Z* от отношения «сигнал/помеха» по мощности для полос частот 100 КГц и помех с нормальным, лапласовским и арксинусоидальным распределениями (рис. 5), то можно отметить увеличение относительной ошибки определения параметра *Z* при уменьшении ОСП до значения 0,1 и дальнейшее уменьшение ошибки после достижения этого значения, что объясняется изменением вида закона распределения аддитивной смеси сигнала и помехи при данных значениях соотношений «сигнал/помеха»: распределение меняется с арксинусоидального на равномерное. При этом значении происходит подавление ЧМн-сигнала помехой. Относительная ошибка определения параметра *Z* во всех случаях не превышает значения $dZ/Z = 0{,}22$.

Анализ предложенного метода оценки влияния помех на ЧМн-сигнал с помощью параметра идентификации вида закона распределения *Z* показал, что данный параметр может быть использован для качественной и количественной оценки наличия помехи в аддитивной смеси ЧМн-сигнала и помех с различными законами распределения. В отличие от коэффициента вариации γ параметр *Z* изменяет численное значение не только при ОСП < 1, но и при ОСП >1. Параметр *Z* также обладает большей чувствительностью к наличию помехи по сравнению с коэффициентом вариации γ.

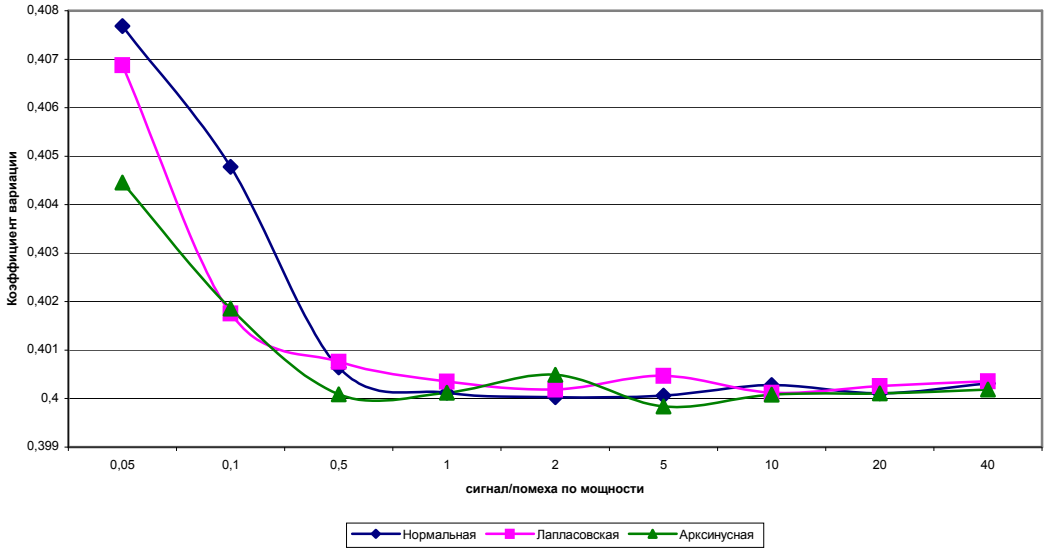


Рис. 4. Зависимость коэффициента вариации γ от ОСП по мощности при воздействии на ЧМн-сигнал помех с различными законами распределения для полосы частот 100 КГц

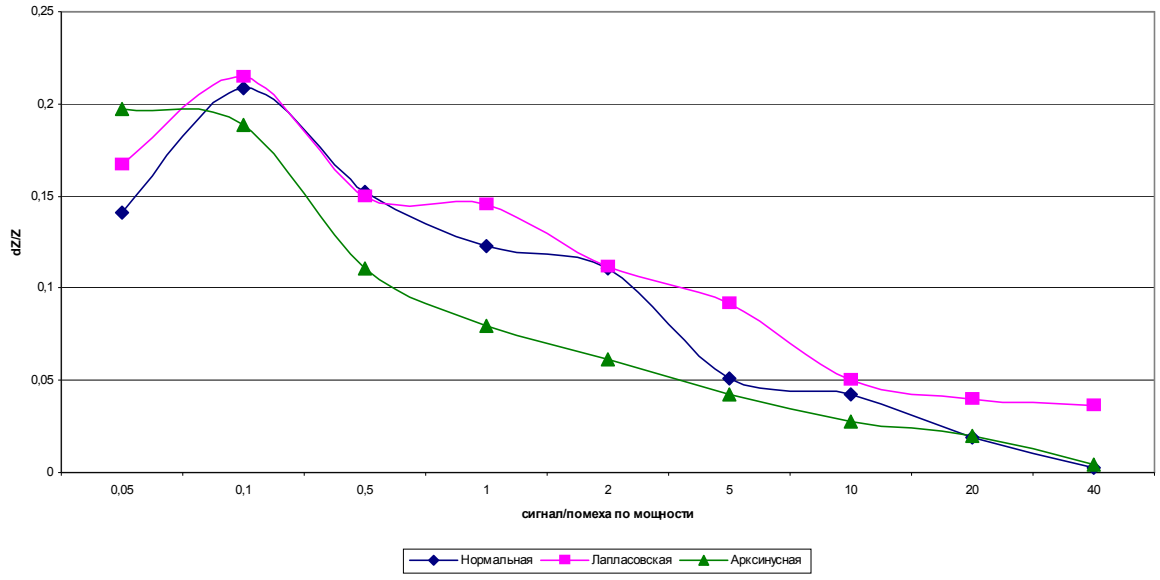


Рис. 5. Зависимость относительной ошибки определения параметра *Z* от ОСП по мощности для полосы частот 100 КГц при воздействии на ЧМн-сигнал помех с различными законами распределения

Моделирование проводилось с погрешностью идентификации вида распределения по параметру *Z* менее 8 % при объемах выборки $n > 100$ для всех рассматриваемых распределений. По степени уменьшения негативного воздействия на ЧМн-сигнал исследуемые помехи располагались в следующем порядке: лапласовская, нормальная и арксинусная.

Таким образом, использование данного метода дает следующие преимущества:

– однозначность идентификации при малых объемах выборки;

– легкую реализацию с помощью ЭВМ, при этом исключаются аналоговые методы обработки, что ведет к удешевлению системы (для сложных систем);

– универсальность алгоритма (для новых условий достаточно модифицировать программу).

Кроме того, информация о виде распределения помехи, которая присутствует в канале связи, и параметр идентификации *Z* могут использоваться для повышения эффективности работы автоматических систем.

Библиографические ссылки

1. Тихонов В. И. Статистическая радиотехника. М. : Сов. радио, 1966.

2. Феер К. Беспроводная цифровая связь. М. : Радио и связь, 2000.

3. Фомин А. Ф. Помехоустойчивость систем передачи непрерывных сообщений. М. : Сов. радио, 1975.

4. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. Л. : Энергоатомиздат. Ленингр. отд-ние, 1985.

5. Кротова Е. И. Идентификация типа распределений результатов экспериментальных исследований // Изв. вузов. Химия и хим. технология. 1998. № 1. С. 57–59.

E. I. Krotova

METHOD OF ESTIMATION OF INFLUENCE OF ADDITIVE DISTURBANCE ON INPUT OF A RECEIVER OF SIGNALS WITH THE HELP OF IDENTIFICATION OF KINDS OF DISTRIBUTION LOWS

In the article a method of estimation of influence of additive disturbance with various laws of distribution on an input of the receiver of telecommunication system according to combined parameter of identification for signals with modulation by the minimal frequency shift is considered.

Keywords: information, *law of distribution*, *identification*, *disturbance.*

© Кротова Е. И., 2010

УДК 519.233.5

Н. Н. Щелканов

РОБАСТНЫЙ МЕТОД ПОСТРОЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ МЕЖДУ ДВУМЯ ФИЗИЧЕСКИМИ ВЕЛИЧИНАМИ С УЧЕТОМ ИХ СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ

Представлена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1 X$ для общего случая, когда разброс точек в корреляционной связи величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами. Все известные выражения для коэффициентов регрессии оказались частными случаями полученной формулы.

Ключевые слова: робастный метод, линейная регрессия, случайные погрешности.

При работе с разными массивами данных часто возникает необходимость нахождения коэффициентов линейной регрессии между двумя случайными физическими величинами. В большинстве случаев коэффициенты регрессии имеют конкретный физический смысл и для корректной интерпретации полученных результатов очень важно найти их значения наилучшим образом.

Существует несколько формул для определения коэффициентов регрессии [1–3], но не для всех формул есть общее понимание, в каких случаях их следует использовать. В настоящее время отсутствует единый подход к нахождению коэффициентов линейной регрессии для общего случая, когда разброс точек в корреляционной связи между двумя величинами обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами.

Постановка задачи. Рассмотрим две случайные физические величины X_0 и Y_0, между которыми существует статистическая корреляционная связь. Предположим, что эта связь может быть описана линейной зависимостью

$$Y_0 = K_0 + K_1 X_0. \quad (1)$$

Требуется найти коэффициенты регрессии K_0 и K_1, которые наилучшим образом отражают физическую взаимосвязь между ними.

Так как X_0 и Y_0 измеряются со случайными погрешностями, то на практике мы имеем дело с величинами X и Y, для которых уравнение регрессии запишется в виде

$$Y = K_0 + K_1 X. \quad (2)$$

Запись уравнений (1) и (2) с одинаковыми коэффициентами регрессии показывает, что последние не должны зависеть от случайных погрешностей измеренных величин X и Y. В дальнейшем будем говорить о нахождении только коэффициента регрессии K_1, так как K_0 вычисляется после нахождения K_1 по известной формуле

$$K_0 = \overline{Y} - K_1 \cdot \overline{X}, \quad (3)$$

где $\overline{X}$ и $\overline{Y}$ – средние значения X и Y.

Новый подход к нахождению коэффициента регрессии K_1. Этот подход заключается в следующих моментах:

– случайные величины X и Y нормируются на значения $\sqrt{\delta_X^2 + \delta_{X_0}^2}$ и $\sqrt{\delta_Y^2 + \delta_{Y_0}^2}$ соответственно, где δ_X и δ_Y – случайные среднеквадратические погрешности измерения X и Y для рассматриваемого массива данных; δ_{X_0} и δ_{Y_0} – некоторые величины, характеризующие разброс точек в корреляционной связи физических величин X_0 и Y_0 за счет неконтролируемых физических параметров;

– при нахождении коэффициента регрессии K_1 используется ортогональная среднеквадратическая регрессия, т. е. минимизируется сумма квадратов отклонений, перпендикулярных искомой прямой.

Тогда уравнение линейной регрессии запишется в виде

$$\frac{Y}{\sqrt{\delta_Y^2+\delta_{Y_0}^2}} = K_0{}' + K_1{}' \cdot \frac{X}{\sqrt{\delta_X^2+\delta_{X_0}^2}}. \qquad (4)$$

Здесь величины δ_{X_0} и δ_{Y_0} находятся из решения системы двух уравнений:

– первое уравнение имеет вид

$$\left|\rho_{X_0Y_0}\right| \cdot \sigma_{X_0} \cdot \sigma_{Y_0} = \sqrt{\sigma_{X_0}^2-\delta_{X_0}^2} \cdot \sqrt{\sigma_{Y_0}^2-\delta_{Y_0}^2}, \qquad (5)$$

где $\sigma_{X_0} = \sqrt{\sigma_X^2-\delta_X^2}$ и $\sigma_{Y_0} = \sqrt{\sigma_Y^2-\delta_Y^2}$ – среднеквадратические отклонения величин X_0 и Y_0; σ_X и σ_Y – среднеквадратические отклонения величин X и Y; $\rho_{X_0Y_0}$ – коэффициент корреляции между X_0 и Y_0. Коэффициент корреляции $\rho_{X_0Y_0}$ находится из известного уравнения [1]:

$$\rho_{XY}\sigma_X\sigma_Y = \rho_{X_0Y_0}\ \sigma_{X_0}\ \sigma_{Y_0}, \qquad (6)$$

где ρ_{XY} – коэффициент корреляции между X и Y. Заметим, что из уравнения (6) следует уравнение (5);

– второе уравнение запишем в виде

$$\frac{\delta_{X_0}}{\sigma_{X_0}} = \frac{\delta_{Y_0}}{\sigma_{Y_0}} \qquad (7)$$

и назовем условием пропорциональности величин δ_{X_0}, δ_{Y_0} и σ_{X_0}, σ_{Y_0}. Введение величин σ_{X_0}, σ_{Y_0} и запись условия (7) являются ключевыми моментами, так как это позволяет получить обобщенное решение для коэффициентов линейной регрессии уравнения (2).

Результаты. После решения системы уравнений (5) и (7) имеем:

$$\delta_{X_0} = \sigma_X \cdot \sqrt{\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)\cdot\left(1-\frac{|\rho_{XY}|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\cdot\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}, \qquad (8)$$

$$\delta_{Y_0} = \sigma_Y \cdot \sqrt{\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)\cdot\left(1-\frac{|\rho_{XY}|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\cdot\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}. \qquad (9)$$

С учетом (8) и (9) найдем значения $\sqrt{\delta_X^2+\delta_{X_0}^2}$ и $\sqrt{\delta_Y^2+\delta_{Y_0}^2}$:

$$\sqrt{\delta_X^2+\delta_{X_0}^2} = \sigma_X \cdot A, \qquad (10)$$

$$\sqrt{\delta_Y^2+\delta_{Y_0}^2} = \sigma_Y \cdot B, \qquad (11)$$

где

$$A = \sqrt{1-\left|\rho_{X_0Y_0}\right|\cdot\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)} = \sqrt{1-|\rho_{XY}|}\cdot\sqrt{\frac{1-\delta_X^2/\sigma_X^2}{1-\delta_Y^2/\sigma_Y^2}}; \qquad (12)$$

$$B = \sqrt{1-\left|\rho_{X_0Y_0}\right|\cdot\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)} = \sqrt{1-|\rho_{XY}|}\cdot\sqrt{\frac{1-\delta_Y^2/\sigma_Y^2}{1-\delta_X^2/\sigma_X^2}}. \qquad (13)$$

Используя (10) и (11), уравнение линейной регрессии (4) запишем в виде

$$\frac{Y}{\sigma_Y \cdot B} = K_0{}' + K_1{}' \cdot \frac{X}{\sigma_X \cdot A}. \qquad (14)$$

Уравнение (14) легко привести к виду (2):

$$Y = K_0{}' \cdot \sigma_Y \cdot B + K_1{}' \cdot \frac{\sigma_Y \cdot B}{\sigma_X \cdot A} \cdot X = K_0 + K_1 \cdot X, \qquad (15)$$

где

$$K_0 = K_0{}' \cdot A \cdot \sigma_Y \cdot B; \qquad (16)$$

$$K_1 = K_1{}' \cdot \frac{\sigma_Y \cdot B}{\sigma_X \cdot A}. \qquad (17)$$

Применяя ортогональную среднеквадратическую регрессию к уравнению (14) и используя соотношение (17), получим выражение для искомого коэффициента регрессии:

$$K_1 = \frac{\sigma_Y}{\sigma_X}\cdot\frac{B}{A}\cdot\frac{1}{2\cdot\rho_{XY}}\times \times\left\{\left(\frac{A}{B}-\frac{B}{A}\right)+\sqrt{\left(\frac{A}{B}-\frac{B}{A}\right)^2+4\cdot\rho_{XY}^2}\right\}, \qquad (18)$$

где A и B определяются выражениями (12) и (13).

Впервые формула (18) была представлена в [4], а подробно описана в [5].

Анализ полученных результатов. Выражение (18) позволяет устанавливать однозначную связь между величинами X и Y и определять условия использования известных типов линейной регрессии.

Покажем, что все известные аналитические выражения для коэффициента регрессии K_1 уравнения (2) являются частными случаями формулы (18).

Так, для случая когда разброс точек в корреляционной связи X и Y обусловлен только их случайными погрешностями, т. е. $\rho_{X_0Y_0} = 1$, получим выражение для коэффициента регрессии K_1, приведенное в [1]:

$$K_1 = \frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+ +\sqrt{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)^2+4\cdot\rho_{XY}^2}\right\}. \qquad (19)$$

При $\rho_{X0Y0} = 1$, $\delta_X = 0$ и $\delta_Y \neq 0$ имеем

$$K_1 = \lim_{\delta x\to 0}\frac{\delta_Y}{\delta_X}\cdot\frac{1}{2\cdot\rho_{XY}}\cdot\left\{\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)+ +\left(-\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}+\frac{\sigma_X}{\sigma_Y}\cdot\frac{\delta_Y}{\delta_X}\right)\sqrt{1+4\cdot\rho_{XY}^2\cdot\left(\frac{\sigma_Y}{\sigma_X}\cdot\frac{\delta_X}{\delta_Y}\right)^2}\right\}.$$

Разлагая выражение под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1 = \lim_{\delta x \to 0} \frac{\delta_Y}{\delta_X} \cdot \frac{1}{2 \cdot \rho_{XY}} \cdot \left\{ \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} - \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) + \right.$$

$$\left. + \left(-\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} + \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) \left[1 + 2 \cdot \rho_{XY}^2 \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} \right)^2 \right] \right\} = \quad (20)$$

$$= \frac{\sigma_Y}{\sigma_X} \cdot \rho_{XY}.$$

Это известная формула для коэффициента K_1 уравнения прямой регрессии $Y = K_0 + K_1 X$, которая находится путем минимизации суммы квадратов отклонений вдоль оси Y от искомой прямой [2].

При $\rho_{X_0Y_0} = 1$, $\delta_Y = 0$ и $\delta_X \neq 0$

$$K_1 = \lim_{\delta Y \to 0} \frac{\delta_Y}{\delta_X} \cdot \frac{1}{2 \cdot \rho_{XY}} \cdot \left\{ \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} - \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) + \right.$$

$$\left. + \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} - \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) \sqrt{1 + 4 \cdot \rho_{XY}^2 \cdot \left(\frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right)^2} \right\}.$$

Проведя процедуру разложения выражения под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1 = \lim_{\delta Y \to 0} \frac{\delta_Y}{\delta_X} \cdot \frac{1}{2 \cdot \rho_{XY}} \cdot \left\{ \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} - \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) + \right.$$

$$\left. + \left(\frac{\sigma_Y}{\sigma_X} \cdot \frac{\delta_X}{\delta_Y} - \frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right) \left[1 + 2 \cdot \rho_{XY}^2 \left(\frac{\sigma_X}{\sigma_Y} \cdot \frac{\delta_Y}{\delta_X} \right)^2 \right] \right\}. \quad (21)$$

$$= \frac{\sigma_Y}{\sigma_X} \cdot \frac{1}{\rho_{XY}}.$$

Формула (21) – также известная формула для коэффициента $1/K_1^*$ уравнения обратной регрессии $X = K_0^* + K_1^* \cdot Y$, которая получается путем минимизации суммы квадратов отклонений вдоль оси X от искомой прямой [2].

При $\rho_{X_0Y_0} = 1$ и $\delta_X = \delta_Y \neq 0$ для коэффициента K_1 уравнения ортогональной регрессии $Y = K_0 + K_1 X$ получим формулу

$$K_1 = \frac{1}{2 \cdot \rho_{XY}} \cdot \left\{ \left(\frac{\sigma_Y}{\sigma_X} - \frac{\sigma_X}{\sigma_Y} \right) + \sqrt{\left(\frac{\sigma_Y}{\sigma_X} - \frac{\sigma_X}{\sigma_Y} \right)^2 + 4 \cdot \rho_{XY}^2} \right\}, \quad (22)$$

которая определяется путем минимизации суммы квадратов отклонений, перпендикулярных искомой прямой [3].

Если для массива данных выполняется соотношение $\frac{\delta_X}{\sigma_X} = \frac{\delta_Y}{\sigma_Y}$, то из выражения (18) вытекает простая формула для коэффициента регрессии:

$$K_1 = \frac{\sigma_Y}{\sigma_X}. \quad (23)$$

Так как соотношение $\frac{\delta_X}{\sigma_X} = \frac{\delta_Y}{\sigma_Y}$ выполняется для большинства экспериментальных данных, то формулу (23) можно рекомендовать к использованию при отсутствии информации о величинах случайных погрешностей X и Y. Заметим, что эта формула представляет собой среднее геометрическое формул (20) и (21).

Диапазон изменчивости коэффициента регрессии. Для случая когда разброс точек в корреляционной связи величин X и Y обусловлен только их случайными погрешностями, т. е. $\rho_{X_0Y_0} = 1$, коэффициент регрессии будет изменяться в следующих пределах:

$$\frac{\sigma_Y}{\sigma_X} \cdot |\rho_{XY}| \le |K_1| \le \frac{\sigma_Y}{\sigma_X} \cdot \frac{1}{|\rho_{XY}|}, \quad (24)$$

а при $\rho_{X_0Y_0} < 1$ – в пределах

$$\frac{\sigma_Y}{\sigma_X} \cdot |\rho_{XY}| < |K_1| < \frac{\sigma_Y}{\sigma_X} \cdot \frac{1}{|\rho_{XY}|}. \quad (25)$$

Из выражений (24), (25) следует, что коэффициенты для прямой и обратной регрессий принимают соответственно минимальное и максимальное значения.

Кратко сформулируем основные выводы:

– получена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1 X$ при условии, что разброс точек в корреляционной связи случайных величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами;

– все известные выражения для коэффициентов регрессии являются частными случаями полученной формулы. Определены условия использования данных выражений;

– обобщенная формула позволяет получать робастные, достоверные и физически корректные коэффициенты регрессии. Эта формула представляет интерес для специалистов, занимающихся обработкой разных массивов данных и может быть использована для их корректной физической интерпретации, независимо от области знания.

Библиографические ссылки

1. Кендалл М., Стьюарт А. Статистические выводы и связи. М. : Наука, 1973.

2. Зайдель А. Н. Погрешности измерений физических величин. Л. : Наука. Ленингр. отд-ние, 1985.

3. Крамер Г. Математические методы статистики. М. : Мир, 1975.

4. Щелканов Н. Н. Построение регрессионной зависимости между аэрозольными оптическими толщами атмосферы с учетом их случайных погрешностей // Аэрозоли Сибири : тез. докл. II заседания раб. группы проекта / Ин-т оптики атмосферы Сиб. отд-ния Рос. акад. наук. Томск, 1995. С. 16.

5. Щелканов Н. Н. Обобщенный метод построения линейной регрессии и его применение для построения однопараметрических моделей аэрозольного ослабления // Оптика атмосферы и океана. 2005. Т. 18. № 1–2. С. 86–90.

6. Кудрявцев В. А., Демидович Б. П. Краткий курс высшей математики. М. : Наука, 1975.

N. N. Shchelkanov

ROBUST METHOD FOR CONSTRUCTION OF LINEAR REGRESSION BETWEEN TWO PHYSICAL PARAMETERS IN VIEW OF THEIR RANDOM ERRORS

In the article the author presents a generalized equation for determination of the regression coefficients of the linear equation $Y = K_0 + K_1 X$ for a general case, when spread of points in correlation between X and Y is caused both by random measurement errors and by uncontrollable physical factors. All the known equations for the regression coefficients appeared to be particular cases of the equation obtained.

Keywords: robust method, linear regression, random errors.

Щелканов Н. Н., 2010

УДК 519.6

В. М. Белолипецкий, С. Н. Генова

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ГИДРОЛЕДОТЕРМИЧЕСКОГО РЕЖИМА РЕКИ АНГАРЫ

Рассматривается компьютерное моделирование гидроледотермического режима реки Ангары выше и ниже плотины строящейся Богучанской ГЭС.

Ключевые слова: гидроледотермика, ГЭС, длина полыньи.

Зарегулирование плотинами гидроузлов вносит существенные изменения в природные условия прилегающих районов. Изменяются температурные и скоростные режимы реки как выше, так и ниже гидроузла. Смена температурного режима оказывает влияние на развитие речной флоры и фауны, а понижение температуры в летний период приводит к уменьшению самоочищающей способности реки. Для полной оценки экологического воздействия строящейся Богучанской ГЭС необходимо знать прогноз гидротермического и ледового режима реки выше и ниже плотины и выработать рекомендации по уменьшению негативного влияния строительства ГЭС.

Для выполнения прогнозных расчетов используются разработанные в Институте вычислительного моделирования Сибирского отделения Российской академии наук математические модели и вычислительные программы, которые применялись для исследования гидроледотермических процессов в бьефах Красноярской, Туруханской и Средне-Енисейской ГЭС [1].

Математические модели гидротермических процессов в водоемах строятся на основе уравнений механики жидкости и теплопереноса с учетом специфики рассматриваемых задач. Для описания гидротермического режима водохранилища используются уравнения гидротермодинамики в приближении пограничного слоя с учетом коэффициента вертикального турбулентного обмена, который определяется по формуле Обухова. В модели учитываются тепловые потоки через свободную поверхность и напряжение трения ветра. Для численного решения уравнения теплопроводности применяются неявные разностные схемы. Полученные системы линейных алгебраических уравнений решаются методом прогонки. Для медленных течений в водохранилищах большой протяженности задача сводится к двумерной модели течения путем применения оператора осреднения по ширине. Скорости течения находятся из решения задачи для функции тока.

Для исследования неустановившихся течений в нижнем бьефе ГЭС используется одномерная постановка соответствующих гидравлических задач, основанных на использовании классических уравнений Сен-Венана. Математическая модель температурного режима водного потока основана на упрощенном одномерном уравнении переноса для средней по сечению температуры воды, учитывающем теплообмен с атмосферой. При наличии притоков температура воды определяется как средневзвешенная.

При изучении ледотермических режимов открытых водотоков основной интерес представляет определение изменений во времени положения створа кромки ледяного покрова (ЛП), толщины ледяного покрова по длине водотока, расхода и общего количества образующейся шуги. При этом различают три основных ледотермических режима в каналах и нижних бьефах ГЭС:

– перемещение кромки льда вверх по течению (наступление) имеет место, когда кромка ЛП расположена ниже створа нулевой изотермы;

– отступление кромки ЛП (температура воды у кромки положительна);

– кромка ЛП не перемещается.

Математическое моделирование ледотермического режима реки основывается на методе расщепления по физическим процессам: исходная задача разделяется на фрагменты, каждый из которых описывает отдель-

ные физические явления: гидротермический режим реки, режим ледообразования, динамика кромки ЛП. Для определения динамики толщины ледяного покрова применяется упрощенная модель, основанная на квазистационарном температурном режиме в затвердевшей области. На границе раздела «вода–лед» выполняется равенство тепловых потоков с учетом скрытой теплоты фазового перехода (условие Стефана).

Верификация компьютерной модели на данных для Усть-Илимской ГЭС. Для верификации модели расчета полыньи в нижнем бьефе ГЭС были использованы метеоданные и расходы Усть-Илимской ГЭС, предоставленые отделом разработки и внедрения гидрологических прогнозов Красноярского центра по гидрометеорологии и мониторингу окружающей среды с региональными функциями Росгидромета и значения средних декадных температур в нижнем бьефе Усть-Илимской ГЭС за 1975–1980 гг. [2] (табл. 1).

Результаты расчетов длины полыньи в нижнем бьефе Усть-Илимской ГЭС хорошо согласуются с натурными (рис. 1).

Прогноз ледотермического режима реки Ангары в нижнем бьефе Богучанской ГЭС. В период, когда температура воды на поверхности понижается до 0 °С (ноябрь), начинается ледообразование. Для зимнего периода после образования ледяного покрова температура воды в водохранилище рассчитывается по уравнению теплопроводности с постоянным коэффициентом турбулентного обмена. Значения средней температуры воды на приплотинном участке и средней температуры поверхностного слоя (глубина поверхностного слоя 30 м) приведены в табл. 2. Селективный отбор воды из верхнего слоя возможен для $Q \le 2\,500$ м^3/с.

Расчеты динамики полыньи в нижнем бьефе Богучанской ГЭС выполнены для реальных метеоданных и для различных значений температуры воды, поступающей из водохранилища (в зависимости от глубины водохранилища на приплотинном участке и от положения водозаборных отверстий). Расчеты показали, что для поверхностного водозабора длина полыньи сокращается на 10…20 км. В зимний период влияние водохранилища на ледотермический режим реки Ангары в нижнем бьефе Богучанской ГЭС распространяется на 20…70 км в зависимости от метеоусловий и условий водозабора (рис. 2, 3).

Таблица 1

Температура воды, поступающей в нижний бьеф Усть-Илимской ГЭС

Месяц	Октябрь			Ноябрь			Декабрь			Январь		
Декада	1	2	3	1	2	3	1	2	3	1	2	3
Температура, °С	7,19	6,72	5,7	4,52	3,46	2,64	2,23	2,16	2,13	2,08	2,03	1,96
Месяц	Февраль			Март			Апрель			Май		
Декада	1	2	3	1	2	3	1	2	3	1	2	3
Температура, °С	1,89	1,74	1,66	1,68	1,56	1,44	1,36	1,33	1,21	1,19	1,32	1,54

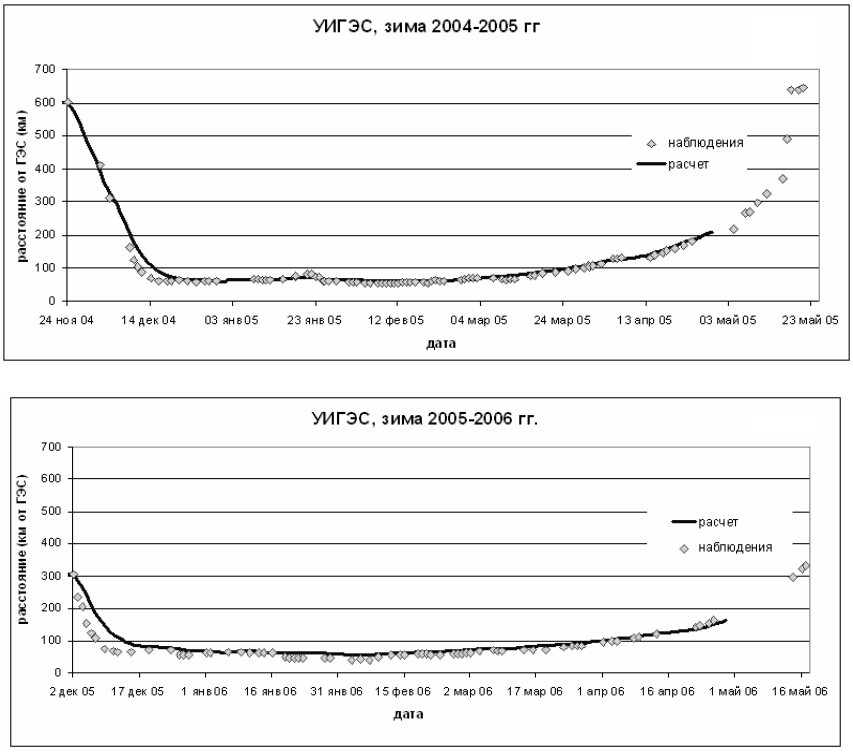


Рис. 1. Динамика кромки льда в нижнем бьефе Усть-Илимской ГЭС (расчетные и натурные данные)

Таблица 2

Значения средней температуры воды по глубине водохранилища (Т_ср) и средней температуре поверхностного слоя (Т_ср_пов) на приплотинном участке (НПУ 208 м)

Дата		Т_ср	Т_ср_пов	Дата		Т_ср	Т_ср_пов
Ноябрь	15	3,12	1,64	Февраль	15	1,9	1
	30	2,92	1,54		30	1,72	0,9
Декабрь	15	2,7	1,43	Март	15	1,56	0,82
	30	2,5	1,32		30	1,41	0,74
Январь	15	2,29	1,21	Апрель	15	1,27	0,67
	30	2,09	1,1		30	1,14	0,6

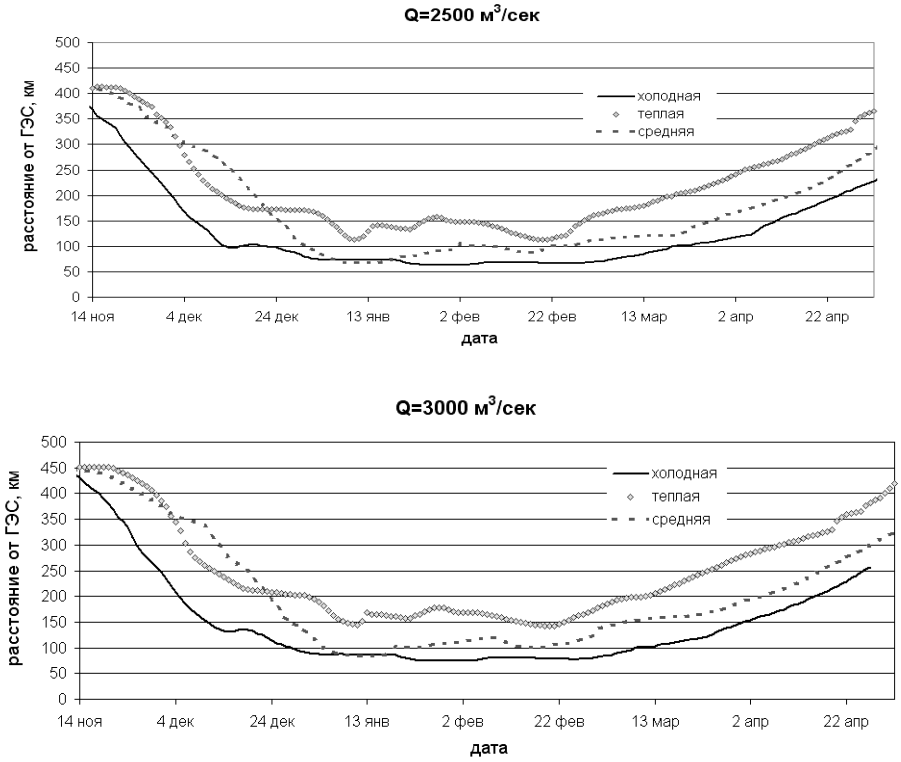


Рис. 2. Длина полыньи для холодной, средней и теплой зим при постоянном расходе

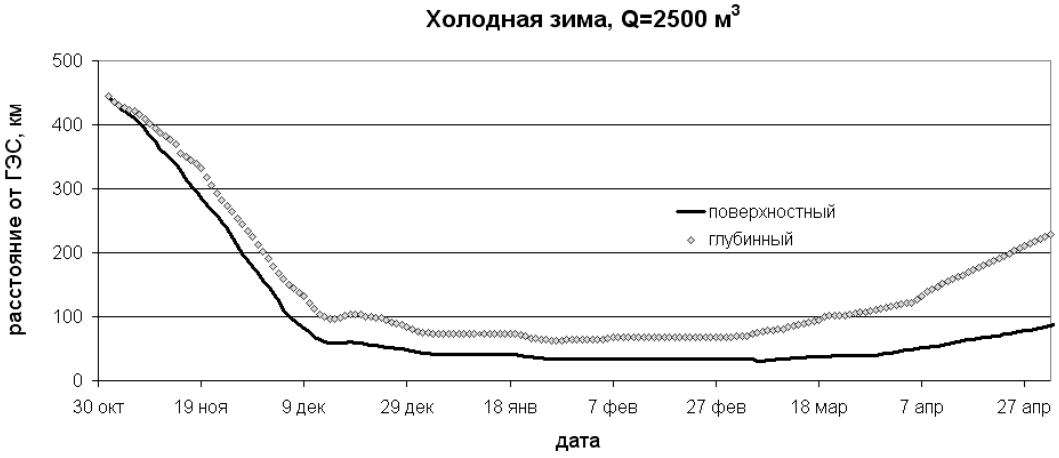


Рис. 3. Длина полыньи по метеоусловиям для холодной зимы при глубинном и поверхностном водозаборе

Таким образом, в данной статье представлены результаты прогноза ледотермического режима Богучанской ГЭС и температурного режима водохранилища для летнего и зимнего периодов при нормальном подпорном уровне (НПУ) воды 185,0 и 208,0 м. Следует отметить, что летом в водохранилище будет формироваться существенная температурная стратификация.

Исследование влияния плотины Богучанской ГЭС на ледотермический режим реки Ангары на 30…40 км при поверхностном водозаборе и на 60…70 км для проектных условий водозабора показало, что в летний период для проектных условий водозабора температура воды, сбрасываемой из водохранилища в нижний бьеф, будет на 3…5 °С ниже бытовой для НПУ 185,0 м и на 4…6 °С для НПУ 208,0 м. Для поверхностного расположения водозаборных отверстий температура воды, поступающей в нижний бьеф, близка к бытовой (14…17 °С в створе плотины ГЭС и 19…20 °С в устье Ангары). Изменения температурного режима в верхнем и нижнем бьефах Богучанской ГЭС в летний период могут оказать влияние на водные экосистемы.

Библиографические ссылки

1. Моделирование задач гидроледотермики водотоков / В. М. Белолипецкий, С. Н. Генова, В. Б. Туговиков, Ю. И. Шокин ; Ин-т вычисл. технологий Сиб. отд-ния Рос. акад. наук. Красноярск, 1993.

2. Лед в водохранилищах и нижних бьефах ГЭС / Я. Л. Готлиб, Р. В. Донченко, А. И. Пехович, И. Н. Соколов. Л. : Гидрометеоиздат, 1983.

V. M. Belolipetskii, S. N. Genova

NUMERICAL INVESTIGATION OF HYDROTHERMAL AND ICE REGIMES OF THE ANGARA RIVER

We consider a computer simulation of hydrothermal and ice regimes of the Angara river up and downsream from the dam of the Boguchanskaya HPP construction.

Keywords: hydroicethermics, HPP, length of water opening.

© Белолипецкий В. М., Генова С. Н., 2010

УДК 681.5.01

Н. Д. Демиденко, Ю. А. Терещенко

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ПРОЦЕССАМИ С РЕЦИРКУЛЯЦИЕЙ ВЗАИМОДЕЙСТВУЮЩИХ ПОТОКОВ

Представлена математическая модель для процессов с рециркуляцией взаимодействующих потоков. Сформулирована задача оптимального управления и получены необходимые условия оптимальности в форме Вейерштрасса. Разработан численный метод решения задачи оптимального управления.

Ключевые слова: оптимальное управление, системы с распределенными параметрами, процессы разделения многокомпонентных слоев.

Одной из характерных особенностей процессов ректификации является рециркуляция взаимодействующих потоков, что приводит к специфичным граничным условиям в краевых задачах. В этом случае технологический процесс описывается уравнениями в частных производных, а граничные условия – дифференциальными уравнениями с обыкновенными производными [1]. С учетом этих особенностей авторами сформулирована и решена задача оптимального управления.

Математическая модель управляемого процесса представлена следующими дифференциальными уравнениями в нормальной форме:

$$x_t' = \frac{1}{H_x}\left[\left(L+L^*\right)\zeta^{(1)} + \frac{\partial L^*}{\partial l}x + K\left(y - y^*\left(x\right)\right) + F\Phi_x x_f\right] \equiv X,$$
$$x_l' = \zeta^{(1)},$$

$$y_t' = \frac{1}{H_y}\left[\left(W-L-F\right)\zeta^{(2)} + K\left(y^*\left(x\right)-y\right)\right] \equiv Y, \qquad (1)$$
$$y_l' = \zeta^{(2)}, \quad 0<t<T, \quad 0<l<1,$$

при краевых условиях:

$$x_{kt}' = \frac{1}{H_{x_k}}\left[\left(L+F\right)x + \left(W-L-F\right)y - Wx_k\right] \equiv X_k, \qquad (2)$$
$$y - a\left[y_k^*\left(x_k\right) - x_k\right] - x_k = 0, \quad 0<t<T, \quad l=0,$$

$$x_{dt}' = \frac{1}{H_{x_d}}\left[L+F-W\right]\left(y_d - x_d\right) \equiv X_d,$$
$$\left(L+F-W\right)\left(y_d - y\right) - L\left(x_d - x\right) = 0, \qquad (3)$$
$$y_d - y - E_d\left(y_d^*\left(x_d\right) - y\right) = 0, \quad 0<t<T, \quad l=1,$$

при начальных условиях:

$$x(l,0)=x_0(l),\quad y(l,0)=y_0(l),\qquad 0<l<1,$$
$$x_d(0)=x_{d_0},\qquad x_k(0)=x_{k_0},\quad 0\le a,\quad E_d\le 1. \tag{4}$$

и ограничениях на управления:

$$(L-L_{\min})(L_{\max}-L)-u^2=0,$$
$$(W-W_{\min})(W_{\max}-W)-z^2=0. \tag{5}$$

Здесь u, z – вспомогательные управления; x, y – концентрации целевого продукта в жидкой и паровой фазах, мольные доли; x_f – концентрация целевого продукта в разделяемом сырье, мольные доли; F – поток сырья, кмоль/ч; L, V – потоки жидкости и пара в аппарате, кмоль/ч; $\Phi_x(l)$ – функция распределения потока по длине объекта; H_x, H_y – удерживающие способности аппарата, кмоль/м; H_{x_k}, H_{x_d} – удерживающие способности в кубе и дефлегматоре, кмоль/м; W – отбор продукта внизу объекта, кмоль/ч; D – отбор целевого продукта вверху объекта, кмоль/ч.

Сформулируем следующую задачу оптимального управления: во множестве кусочно-непрерывных функций L, W, удовлетворяющих условиям (5), найти такие, что соответствующее им решение задачи (1)…(4) дает минимум функционалу

$$F=\int_0^T\int_0^1\left(y(l,t)-\Theta^*(l,t)\right)^2 dldt,$$

где $\Theta^*(l,t)$ – заданное значение концентрации целевого продукта.

Для получения необходимых условий оптимальности в форме Лагранжа–Эйлера рассмотрим вспомогательный функционал [1; 2]:

$$I=I_1+I_2=\iint_\Omega \tilde{L}\,dldt+\int_{\partial\Omega}\tilde{l}\,dt,$$

где

$$\tilde{L}=\left(y-\Theta^*\right)^2+\xi^{(1)}(x_t'-X)+\xi^{(2)}\left(x_l'-\zeta^{(1)}\right)+$$
$$+\eta^{(1)}\left(y_t'-Y\right)+\eta^{(2)}\left(y_l'-\zeta^{(2)}\right),$$
$$\tilde{l}=\lambda_K^{(1)}\left(x_{kt}'-X_k\right)+\lambda_k^{(2)}\left(y-x_k-a\left(y_k^*(x_k)-x_k\right)\right)+$$
$$+\lambda_d^{(1)}\left(x_{dt}'-x_d\right)+\lambda_d^{(2)}\left((L+F-W)(y_d-y)-L(x_d-x)\right)+$$
$$+\lambda_d^{(3)}\left(y_d-y-E_d\left(y_d^*(x_d)-y\right)\right)+\gamma\left((L-L_{\min})(L_{\max}-L)-u^2\right)+$$
$$+\varepsilon\left((W-W_{\min})(W_{\max}-W)-z^2\right),$$

здесь $\xi^{(1)}$, $\xi^{(2)}$, $\eta^{(1)}$, $\eta^{(2)}$ – функции, определенные на $\Omega\ (\Omega=\{(l,t)\,|\,0<l<1,0<t<T\})$; $\lambda_k^{(i)}$, $\lambda_d^{(i)}$, γ, ε – функции, определенные на $[0,T]$ (их можно считать определенными на $\partial\Omega$), причем $\varepsilon=\lambda_k^{(i)}=0$ на $\partial\Omega/\{(l,t)\,|\,l=0\}$, $\gamma=\lambda_d^{(i)}=0$ на $\partial\Omega/\{(l,t)\,|\,l=1\}$.

Пусть L, W – оптимальные управления; u, z – соответствующие им (согласно (5)) вспомогательные управления; x, y, x_k, x_d, y_d – оптимальное решение задачи (1)…(4), соответствующее этим уравнениям; δL, δW – вариации управлений L, W; δu, δz – соответствующие вариации фиктивных управлений u, z; δx, δy, δx_k, δx_d, δy_d – соответствующие вариации решений. Тогда, используя аргументацию вариационного исчисления, получим сопряженную задачу относительно функций Лагранжа ξ, η, $\lambda_k^{(i)}$, $\lambda_d^{(i)}$, γ, ε:

– при $0<t<T$, $0<l<1$

$$\xi_t'-\frac{L+L^*}{H_x}\xi_l'=k_y\left(y^*\right)'\left(\frac{\xi}{H_x}-\frac{\eta}{H_y}\right),$$
$$\eta_t'-\frac{L+F-W}{H_y}\eta_l'=-k_y\left(\frac{\xi}{H_x}-\frac{\eta}{H_y}\right)+2\left(y-\Theta^*\right); \tag{6}$$

– при $l=0$, $0<t<T$

$$\frac{\xi}{H_x}-\frac{\lambda_k^{(1)}}{H_{x_k}}=0,$$
$$\left(\frac{\eta}{H_y}-\frac{\lambda_k^{(1)}}{H_{x_k}}\right)(L+F-W)-\lambda_k^{(2)}=0,$$
$$\frac{d\lambda_k^{(1)}}{dt}=\frac{W}{H_{x_k}}\lambda_k^{(1)}+\lambda_k^{(2)}\left(-a\left(y_k^*\right)'+a-1\right), \tag{7}$$
$$\lambda_k^{(1)}(T)=0;$$

– при $t=T$, $0<l<1$

$$\xi=0,\quad \eta=0, \tag{8}$$

– при $l=1$, $0<t<T$

$$\left(\frac{\xi}{H_x}+\lambda_d^{(2)}\right)L-\lambda_d^{(3)}E_d\left(y_d^*\right)'=0,$$
$$\left(\frac{\eta}{H_y}+\lambda_d^{(2)}\right)(L+F-W)+(1-E_d)\lambda_d^{(3)}=0,$$
$$\frac{d\lambda_d^{(1)}}{dt}=\frac{L+F-W}{H_{x_d}}\lambda_d^{(1)}-\lambda_d^{(2)}L,\quad \lambda_d^{(1)}(T)=0, \tag{9}$$
$$\left(\frac{\lambda_d^{(1)}}{H_{x_d}}-\lambda_d^{(2)}\right)(L+F-W)-\lambda_d^{(3)}=0;$$

– при $0<t<T$

$$\int_0^1\left(-\frac{\xi}{H_x}x_l'+\frac{\eta}{H_y}y_l'\right)dl-\frac{\lambda_k^{(1)}}{H_{x_k}}\left(x(0,t)-y(0,t)\right)+\frac{\lambda_d^{(1)}}{H_{x_d}}+$$
$$+(y_d+x_d)+\lambda_d^{(2)}\left(y(1,t)-y_d+x_d-x(1,t)\right)-$$
$$-\gamma\left(L_{\min}+L_{\max}-2L\right)=0, \tag{10}$$

$$\int_0^1 -\frac{\eta}{H_y} y_l'\, dl - \frac{\lambda_d^{(1)}}{H_{x_d}}\left(y_d - x_d\right) + \frac{\lambda_k^{(1)}}{H_{x_k}}\left(x_k - y(0,t)\right) - $$
$$- \lambda_d^{(2)}\left(y(1,t) - y_d\right) + \varepsilon\left(W_{\min} + W_{\max} - 2W\right) = 0,$$
$$\gamma u = 0, \quad \varepsilon z = 0.$$

Эти необходимые условия отличаются от известных условий стационарности соотношениями (10). Это отличие связано с тем, что управления L, W являются граничными управлениями, но одновременно входят и в уравнения процесса, так что их вариации в Ω и на $\partial\Omega$ не являются независимыми.

Теперь получим необходимые условия оптимальности в форме Вейерштрасса. Пусть L, W – оптимальные управления. Приращение функционала, вызванное переходом от оптимальных управлений L, W и соответствующих им решений x, y, x_k системы (1)…(4) к произвольным допустимым управлениям $\hat{L} = L + \Delta L$, $\hat{W} = W + \Delta W$ и соответствующим им решениям $x + \Delta x$, $y + \Delta y$, $x_k + \Delta x_k$ представим в виде

$$\Delta I = \int_0^T \left[A\Delta L + B\Delta W\right]dt + \iint_\Omega \varepsilon_1 dl dt + \int_{\partial\Omega} \varepsilon_2 dt,$$

где

$$A = \frac{1}{H_y}\int_0^1 \eta y_l' dl - \frac{1}{H_x}\int_0^1 \xi x_l' dl + \frac{1}{H_{x_k}}\lambda_k\left(y(0,t) - x(0,t)\right) -$$
$$- \frac{\lambda_d}{H_{x_d}}\left(x(1,t) - y(1,t)\right);$$

$$B = \frac{1}{H_y}\int_0^1 \eta y_l' dl - \frac{\lambda_k}{H_{x_k}}\left(x_k(t) - y(0,t)\right) - \frac{\lambda_d}{H_{x_d}}\left(y(1,t) - x(1,t)\right);$$

$$\varepsilon_1 = \Delta y^2 + \frac{\Delta L}{H_x}\Delta x \xi_l' - \frac{\Delta W}{H_y}\Delta y \eta_l' + 0(\Delta x);$$

$$\varepsilon_2 = -\frac{\Delta L}{H_x}\left(\Delta x(1,t)\xi(1,t) - \Delta x(0,t)\xi(0,t)\right) +$$
$$+ \frac{\Delta W}{H_y}\left(\Delta y(1,t)\eta(1,t) - \Delta y(0,t)\eta(0,t)\right) +$$
$$+ \frac{\lambda_k}{H_{x_k}}\left(\Delta W \Delta x_k - \Delta L \Delta x(0,t) + \Delta W \Delta y(0,t)\right) +$$
$$+ \frac{\lambda_d}{H_{x_d}}\Delta W\left(\Delta y(1,t) - \Delta x(1,t)\right) + 0(\Delta x).$$

Здесь Δx, Δy, Δx_k определяются из уравнений

$$H_x \Delta x_t' - \left(L + \Delta L + L^*\right)\Delta x_l' =$$
$$= \left[\frac{\partial L^*}{\partial l} - k_y\left(y^*\right)'\right]\Delta x + k_y \Delta y + x_l' \Delta L + 0(\Delta x), \tag{11}$$

$$H_y \Delta y_t' + (V + \Delta V)\Delta y_l' = k_y\left(y^*\right)'\Delta x - k_y \Delta y - y_l' \Delta V + 0(\Delta x),$$
$$0 < l < 1,\ 0 < t < T,$$
$$H_{x_k}\Delta x_k' = \left(L + \Delta L + F\right)\Delta x(0,t) -$$
$$- (V + \Delta V)\Delta y(0,t) + (W + \Delta W)\Delta x_k + \tag{12}$$
$$+ x(0,t)\Delta L - y(0,t)\Delta V + x_k \Delta W,$$

$$\Delta y(0,t) = \left(a\left(y_k^*\right)' - a + 1\right)\Delta x_k + 0(\Delta x_k),\ l = 0,$$
$$0 < t < T,$$
$$H_{x_d}\Delta x_t'(1,t) = (V + \Delta V)\left(\Delta y(1,t) - \Delta x(1,t)\right) + \tag{13}$$
$$+ \left(y(1,t) - x(1,t)\right)\Delta V,$$
$$l = 1,\ 0 < t < T,$$
$$\Delta x(l,0) = \Delta y(l,0) = \Delta x_k(0) = \Delta x(1,0) = 0 \tag{14}$$
$$0 < l < 1,$$
$$\Delta V = \Delta L - \Delta W. \tag{15}$$

Производные $\Delta x_l'$, $\Delta y_l'$ считаем ограниченными.

Для оценки приращений используем следующий результат, легко получаемый из известного неравенства Гронуолла: если

$$|\varphi(t)| \le M\left(P_1 + \int_{t_0}^t \left(|\varphi(\tau)| + \psi(\tau)\right)d\tau\right),\ P_1, M = \text{const},$$
$$\psi(\tau) \ge 0,\ t_0 \le \tau, t \le T, \tag{16}$$

тогда

$$|\varphi(t)| \le M_1\left(P_1 + \int_{t_0}^t \psi(\tau)\, d\tau\right),$$
$$M_1 = \text{const}, \qquad t_0 \le t \le T. \tag{17}$$

Из первого уравнения (12) получим неравенство

$$|\Delta x_k(t)| \le N_1 \int_0^t \left(|\Delta x_k(\tau)| + |\Delta x(0,\tau)| + |\Delta y(0,\tau)|\right)d\tau +$$
$$+ N_2 \int_0^T \left(|\Delta L(t)| + |\Delta W(t)|\right)dt,$$

а в силу (17) будем иметь

$$|\Delta x_k(t)| \le N_3 \int_0^t \left(|\Delta x(0,\tau)| + |\Delta y(0,\tau)|\right)d\tau +$$
$$+ N_4 \int_0^T \left(|\Delta L(t)| + |\Delta W(t)|\right)dt.$$

Из второго уравнения (12) и (17) следует оценка

$$|\Delta y(0,t)| \le N_5 \int_0^t |\Delta x(0,\tau)|\, d\tau +$$
$$+ N_6 \int_0^T \left(|\Delta L(t)| + |\Delta W(t)|\right)dt. \tag{18}$$

Аналогично из (13):

$$|\Delta x(1,t)| \le N_7 \int_0^t |\Delta y(1,\tau)|\, d\tau +$$
$$+ N_8 \int_0^T \left(|\Delta L(t)| + |\Delta W(t)|\right)dt. \tag{19}$$

Учитывая, что левые части (11) представляют производные вдоль характеристик (с точностью до постоянного множителя), и интегрируя вдоль этих характеристик, получим:

$$\begin{aligned}|\Delta x(l,t)| \le N_9 \int_{t_0'}^{t} \Big(|\Delta x(l(\tau),\tau)| + |\Delta y(l(\tau),\tau)| + \\ + |\Delta L(\tau)| + |\Delta W(\tau)| \Big) d\tau + |\Delta x(l_0',t_0')|, \end{aligned}$$

$$\begin{aligned}|\Delta y(l,t)| \le N_{10} \int_{t_0''}^{t} \Big(|\Delta x(l(\tau),\tau)| + |\Delta y(l(\tau),\tau)| + \\ + |\Delta L(\tau)| + |\Delta W(\tau)| \Big) d\tau + |\Delta x(l_0'',t_0'')|, \end{aligned} \qquad (20)$$

где (l_0',t_0'), (l_0'',t_0'') – начальные точки соответствующих характеристик, $0 \le t_0'$, $t_0'' \le t$, причем $t_0' = 0$ или $l_0' = 1$, $t_0'' = 0$ или $l_0'' = 0$.

Пусть

$$\alpha(t) = \max_{\substack{0\le\tau\le t,\\ 0\le l\le 1}} |\Delta x(l,\tau)|, \qquad \beta(t) = \max_{\substack{0\le\tau\le t,\\ 0\le l\le 1}} |\Delta y(l,\tau)|.$$

Тогда из (19) и (20) имеем

$$\begin{aligned}|\Delta x(l,t)| \le N_9 \int_0^t (\alpha(\tau)+\beta(\tau))\, d\tau + N_9 \int_0^T |\Delta L(t)|\, dt + \\ + N_7 \int_0^t \beta(\tau)\, d\tau + N_8 \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt, \end{aligned}$$

откуда

$$\alpha(t) \le N_{11} \int_0^1 (\alpha(\tau)+\beta(\tau))\, d\tau + N_{12} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt.$$

Аналогично можно получить оценку

$$\beta(t) \le N_{13} \int_0^t (\alpha(\tau)+\beta(\tau))\, d\tau + N_{14} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt.$$

Еще раз применим (18):

$$\alpha(t) \le N_{15} \int_0^t \beta(\tau)\, d\tau + N_{16} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt,$$

$$\beta(t) \le N_{17} \int_0^t \alpha(\tau)\, d\tau + N_{18} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt. \qquad (21)$$

Интегрируя от 0 до σ, учитывая, что при $f(\tau) \ge 0$

$$\begin{aligned}\int_0^\sigma dt \int_0^t f(\tau)\, d\tau = \int_0^\sigma d\tau \int_\tau^\sigma f(\tau)\, dt = \\ = \int_0^\sigma (\sigma - \tau) f(\tau)\, d\tau \le T \int_0^\sigma f(\tau)\, d\tau, \end{aligned}$$

и заменяя σ на t, получим

$$\int_0^t \alpha(\tau)\, d\tau \le N_{19} \int_0^t \beta(\tau)\, d\tau + N_{20} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt,$$

$$\int_0^t \beta(\tau)\, d\tau \le N_{21} \int_0^t \alpha(\tau)\, d\tau + N_{22} \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt.$$

Подставляя в (21) и еще раз применяя (17), приходим к оценкам вида

$$\alpha(t) \le P \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt,$$

$$\beta(t) \le P \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt,$$

откуда

$$\begin{aligned}|\Delta x(l,t)| \le P \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt, \\ |\Delta y(l,t)| \le P \int_0^T \Big(|\Delta L(t)| + |\Delta W(t)| \Big) dt. \end{aligned} \qquad (22)$$

Теперь мы можем оценить ε_1, ε_2. Обозначая $\delta(t) = |\Delta L(t)| + |\Delta W(t)|$; q – наибольшее количество слагаемых в формулах для ε_1, ε_2; δ_1 – максимум модулей коэффициентов при приращениях в этих формулах, получим:

$$\begin{aligned}|\varepsilon_1(l,t)| \le P^2 \left(\int_0^T \delta(t)\, dt \right)^2 \delta_1 q \delta(t) P \int_0^T \delta(t)\, dt, \\ |\varepsilon_2(l,t)| \le q \delta_1 \delta(t) P \int_0^T \delta(t)\, dt. \end{aligned} \qquad (23)$$

Сформулируем необходимые условия оптимальности в виде теоремы.

Теорема. Если $L(t)$, $W(t)$ – оптимальное управление, x, y, x_k – соответствующее решение задачи (1)…(4), то в точках непрерывности управлений

$$A\Delta L + B\Delta W \ge 0$$

для всех допустимых приращений ΔL, ΔW, или что то же самое,

$$\begin{aligned} L(t) &= \begin{cases} L_{\max} & \text{при } A(t) < 0, \\ L_{\min} & \text{при } A(t) > 0, \end{cases} \\ W(t) &= \begin{cases} W_{\max} & \text{при } B(t) < 0, \\ W_{\min} & \text{при } B(t) > 0. \end{cases} \end{aligned} \qquad (24)$$

Доказательство. Пусть t_0 – точка, в которой условия (19) нарушаются, например для $L(t)$. Возьмем $\varepsilon > 0$ так, что при $t_0 \le t \le t_0 + \varepsilon$

$$A(t) < \frac{1}{2} A(t_0),$$

$$L(t) < \frac{1}{2}\big(L(t_0) + L_{\max}\big) \text{ в случае } A(t_0) < 0,$$

$$A(t) > \frac{1}{2} A(t_0),$$

$$L(t) > \frac{1}{2}\big(L(t_0) + L_{\min}\big) \text{ в случае } A(t_0) > 0,$$

и выберем приращения

$$\Delta L(t) = \begin{cases} 0 \quad \text{при} \quad t \in [0,t_0] \cup [t_0+\varepsilon, T], \\ L_{\max} - L(t_0) \quad \text{при} \quad t_0 \le t < t_0 + \varepsilon \\ \text{в случае } A(t_0) < 0, \\ L_{\min} - L(t_0) \quad \text{при} \quad t_0 \le t < t_0 + \varepsilon \\ \text{в случае } A(t_0) > 0, \end{cases}$$

$\Delta W(t) = 0$.

Тогда при $t_0 < t < t_0 + \varepsilon$

$$A(t)\Delta L(t) < a < 0,$$

где

$$a = \begin{cases} \frac{1}{4} A(t_0)\left(L_{\max} - L(t_0)\right) & \text{при} \quad A(t_0) < 0, \\ \frac{1}{4} A(t_0)\left(L_{\min} - L(t_0)\right) & \text{при} \quad A(t_0) > 0, \end{cases}$$

$$\int_0^T \left[A\Delta L + B\Delta W\right] dt < a\varepsilon, \quad \left|\int_0^T \delta(t)\,dt\right| < b\varepsilon,$$

$$\left|\iint_\Omega \varepsilon_1 dt dl\right| < b_1\varepsilon^2, \quad \left|\int_0^T \varepsilon_2 dt\right| < b_2\varepsilon^2,$$

где b, b_1, b_2 – постоянные, не зависящие от ε; $\Delta I < a\varepsilon + b_1\varepsilon^2 + b_2\varepsilon^2$ при достаточно малом ε, что противоречит оптимальности управлений $L(t)$, $W(t)$. Аналогично рассматривается случай, когда (24) не выполняется для $W(t)$.

Как видим, доказательство теоремы указывает на способ улучшения управлений, если условия (24) не выполнены. Это позволяет построить итеративный процесс улучшения управлений при нарушении условий (24).

В качестве примера рассмотрим задачу оптимизации пускового режима для промышленной колонны К-34. Исходные данные: $F = 103{,}65$ кмоль/ч, $F_V = 0$, $D = 27{,}06$ кмоль/ч, $L_d = 45{,}21$ кмоль/ч, $W = 76{,}59$ кмоль/ч, $H_{x_d} = 50$ кмоль, $H_{x_k} = 30$ кмоль, $P_H = 3{,}6$ ат, $P_B = 3{,}5$ ат. Коэффициенты P_{1i}, P_{2i}, P_{3i} получены методом наименьших квадратов с использованием таблицы зависимости давления чистых компонентов от температуры. Состав сырья: $x_{f_1} = 0{,}029\,63$, $x_{f_2} = 0{,}222\,151$, $x_{f_3} = 0{,}143\,95$, $x_{f_4} = 0{,}044\,52$, $x_{f_5} = 0{,}043\,89$, $x_{f_6} = 0{,}420\,21$, $x_{f_7} = 0{,}096\,3$ [3].

Выберем целью управления достижение в выходных потоках концентраций компонентов $x_{di\,\text{ст}}$ и $x_{ki\,\text{ст}}$. В этой задаче возмущением является начальное состояние управляемого процесса, при котором концентрации компонентов по длине колонны равны концентрациям этих компонентов в сырье. Время управления T возьмем 20 ч.

Были проведены расчеты шести вариантов задачи оптимизации пускового режима.

В первых пяти вариантах задача оптимизации решается с одной управляющей функцией. Значения остальных четырех функций и начальное значение управляющей функции заданы такими же, как и при расчете статического режима (рис. 1, 2). При начальном управлении значение функционала для всех вариантов задачи одинаково: $F^0 = 34{,}0$.

В шестом варианте задачи в качестве управляющих взяты все пять параметров. Начальные значения для управляющих функций такие же, как в первых пяти вариантах (рис. 3).

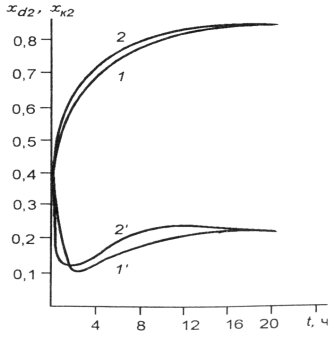


Рис. 1. Графики переходных процессов по концентрации бутана в дефлегматоре и кубе в пусковом режиме:
1, *2* – в дефлегматоре; *1′*, *2′* – в кубе; *1*, *1′* – при начальном управлении; *2*, *2′* – при оптимальном управлении; управляющий параметр – отбор вверху колонны

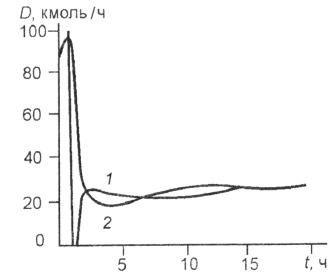


Рис. 2. Графики оптимальной функции управления (отбор вверху колонны):
1 – в пусковом режиме; *2* – совместно с другими оптимальными управляющими функциями L_d, F, l_f, x_{f2}

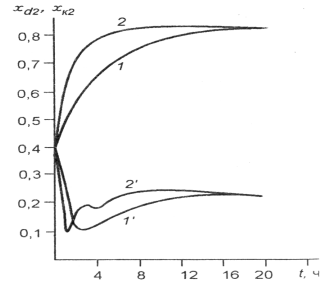


Рис. 3. Графики переходных процессов по концентрации бутана в дефлегматоре и кубе в пусковом режиме:
1, *2* – в дефлегматоре; *1′*, *2′* – в кубе; *1*, *1′* – при начальном управлении; *2*, *2′* – при оптимальном управлении; управляющие параметры: отбор вверху колонны, орошение, расход сырья, концентрация бутана в сырье, координата ввода сырья

В результате решения было получено, что наиболее эффективной управляющей функцией является L_d (поток орошения) (см. таблицу). Однако следует заметить, что увеличение потока орошения значительно увеличивает энергетические затраты.

Следующим по эффективности управляющим параметром является концентрация бутана в сырье x_{f_2}. Управление этим параметром связано со значительным перераспределением выходных потоков в предыдущей колонне, что не всегда допустимо.

Сравнительная оценка управляющих параметров при оптимизации пускового режима колонны

u_j	D	F_L	l_L	L_d	x_{f_2}
$F_{\min}$	28,5	29,6	30,4	20,9	26,2

Наименее эффективным управляющим параметром является координата ввода сырья l_L, однако изменение этого параметра осуществляется практически без дополнительных энергетических затрат.

Библиографические ссылки

1. Демиденко Н. Д. Управляемые распределенные системы. Новосибирск : Сиб. изд. фирма Сиб. отд-ния Рос. акад. наук «Наука», 1999.

2. Демиденко, Н. Д. Моделирование и оптимизация тепломассообменных процессов в химической технологии. М. : Наука, 1991.

3. Демиденко Н. Д., Потапов В. И., Шокин Ю. И. Моделирование и оптимизация систем с распределенными параметрами. Новосибирск : Изд-во Сиб. отд-ния Рос. акад. наук, 2006.

N. D. Demidenko, Ju. A. Tereschenko

OPTIMAL CONTROL OF PROCESSES WITH RECIRCULATION OF INTERACTING FLOWS

A mathematical model for the processes with recirculation of interactive flows is given. A problem of optimal control is formulated. Necessary optimal conditions in the form of Veyershtrass are obtained. A numerical method of optimization problems solving is developed.

Keywords: optimal control, systems with distributed constants, processes of multicomponent layering.

© Демиденко Н. Д., Терещенко Ю. А., 2010

УДК 681.5.01

Н. Д. Демиденко, Ю. А. Терещенко

ОПТИМИЗАЦИЯ ПРОЦЕССОВ РАЗДЕЛЕНИЯ МНОГОКОМПОНЕНТНЫХ СМЕСЕЙ

Предложена математическая модель процессов разделения многокомпонентных смесей. Сформулирована задача оптимального управления, получены необходимые условия оптимальности. Разработан численный метод решения задач оптимизации, проведены численные эксперименты.

Ключевые слова: многокомпонентная смесь, оптимальное управление, нелинейные процессы.

Исследование и проектирование систем химической технологии представляет собой сложную задачу, так как при ее постановке используются нелинейные системы дифференциальных уравнений в частных производных [1–5]. Математическая формулировка таких задач и вопросы их корректности, как правило, требуют специального рассмотрения. Трудности, прежде всего, связаны с нелинейностью уравнений и сложностью граничных условий, содержащих обыкновенные дифференциальные уравнения. С другой стороны, эти трудности обусловлены многомерностью задач, поскольку технологические процессы характеризуются довольно большим числом теплофизических и конструктивных параметров. Выбор эффективной методики решения задач моделирования и управления принято считать центральным вопросом в проблеме моделирования нестационарных режимов управляемых процессов разделения.

Основным инструментом проектирования технологических процессов и систем управления является вычислительный эксперимент, включающий в себя анализ многообразия технологических режимов, построение их физических и математических моделей, разработку и исследование вычислительных алгоритмов, их программную реализацию и проведение серии расчетов [3].

В настоящей работе формулируются и решаются задачи анализа математических моделей процессов разделения, задачи оптимального управления и численных экспериментов с целью проектирования энергосберегающих технологий разделения многокомпонентных смесей для промышленных объектов и создания соответствующих оптимальных систем управления.

Процесс разделения многокомпонентных смесей осуществляется в ректификационных колоннах на контактных устройствах (тарелках), распределенных по длине аппарата. Технологический процесс происходит в конечном числе точек объекта, однако его можно рассматривать непрерывным по длине, поэтому для моделирования возможно применение дифференциальных уравнений в частных производных. Оценка погрешности перехода от дискретной модели к непрерывной приведена в [1]. Концентрации целе-

вого продукта в жидкости $x = x(l,t)$ и паре $y = y(l,t)$ – управляемые параметры, они определяются в результате решения соответствующей краевой задачи.

Разделяемая смесь в количестве $F = F(t)$ с содержанием целевого продукта $x_F = x_F(t)$ подается в среднюю часть колонны. В нижней ее части (кубе) происходит испарение смеси, и паровой поток $V = V(l,t)$, поднимаясь вверх, контактируя со стекающей жидкостью $L = L(l,t)$ и обогащаясь целевым продуктом, конденсируется в верхней части колонны (дефлегматоре) и отбирается в количестве D с концентрацией целевого продукта $x_d = x_d(t)$. Часть сконденсированного пара $L_d = L_d(t)$ из дефлегматора возвращается в колонну для повышения качества конечного продукта. В кубе отбирается остаток в количестве $W = W(t)$ с содержанием целевого продукта $x_k = x_k(t)$ (рис. 1).

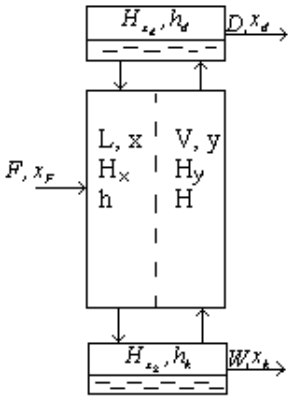


Рис. 1. Схема потоков пара и жидкости в колонне

Одним из требований к такому промышленному объекту является его способность к увеличению содержания целевого продукта в верхней части колонны и уменьшению в нижней. Важными параметрами объекта являются фазовые удерживающие способности: в колонне $H_x = H_x(l,t)$, $H_y = H_y(l,t)$, кубе $H_{x_k} = H_{x_r}(l,t)$ и дефлегматоре $H_{x_d} = H_{x_d}(l,t)$. Индексы x и y указывают на принадлежность параметра жидкости или пару, k и d – кубу или дефлегматору. В колонне происходит теплообмен между жидкой и паровой средами, которые характеризуются теплосодержаниями жидкости $h = h(l,t)$ и пара $H = H(l,t)$; аналогично, в кубе $h_k = h_k(t)$ и в дефлегматоре $h_в = h_в(t)$, $H_d = H_d(t)$. В куб подводится тепло Q_k, а из дефлегматора оно отводится – Q_d. Коэффициент массопередачи k_y характеризует процесс массообмена между жидкой и паровой фазами, а зависимость $y^* = y^*(x)$ – равновесную концентрацию в паре. Функции $x(l,t)$, $y(l,t)$, $x_k(t)$, $x_d(t)$ и $y_d(t)$ могут быть скалярными (для бинарных смесей) или векторными (для многокомпонентных). Более подробная физическая интерпретация математической модели содержится, например, в [5].

Используя законы сохранения массы и энергии, получим математическую модель нестационарных режимов процессов разделения для колонны, куба и дефлегматора:

$$\begin{gathered}\frac{\partial\left(H_x x\right)}{\partial t}-\frac{\partial\left(Lx\right)}{\partial l}=k_y\left(y-y^*\right)+F(t)\Phi_x(l)x_F,\\ \frac{\partial\left(H_y y\right)}{\partial t}+\frac{\partial\left(Vy\right)}{\partial l}=k_y\left(y^*-y\right),\\ \frac{\partial\left(H_y H\right)}{\partial t}+\frac{\partial\left(H_x h\right)}{\partial t}-\frac{\partial\left(Lh\right)}{\partial l}+\frac{\partial\left(VH\right)}{\partial l}=\Phi_h+\Phi_H,\\ \frac{\partial V}{\partial l}-\frac{\partial L}{\partial l}+\frac{\partial H_x}{\partial t}+\frac{\partial H_y}{\partial t}=\Phi_V+\Phi_L,\end{gathered} \tag{1}$$

$$\begin{gathered}x(l,0)=x_0(l),\ y(l,0)=y_0(l),\\ V(l,0)=V_0(l),\ \ L(l,0)=L_0(l);\end{gathered} \tag{2}$$

с граничными условиями при $l = 0$

$$\begin{gathered}\frac{\partial\left(H_{x_k}x_k\right)}{\partial t}=L(0,t)x(0,t)-V(0,t)y(0,t)-W(t)x_k(t),\\ y(0,t)=(y^*(x_k)-x_k)a+x_k(t),\\ \frac{\partial\left(H_{x_k}h_k\right)}{\partial t}=L(0,t)h(0,t)-V(0,t)H(0,t)-W(t)h_k(t)+Q_k,\\ \frac{\partial H_{x_k}}{\partial t}=L(0,t)-V(0,t)-W(t),\quad x_k(0)=x_{k,0};\end{gathered} \tag{3}$$

и при $l = 1$

$$\begin{gathered}\frac{\partial(H_{x_d}x_d)}{\partial t}=V_d y_d-(L_d+D)x_d(t),\quad x_d(0)=x_{d,0}\\ V_d(t)y_d(t)-V(1,t)y(1,t)=L_d(t)x_d(t)-L(1,t)x(1,t),\\ y_d(t)=y(1,t)+E_d(y^*(x(1,t)-y(1,t)),\\ \frac{\partial(H_{x_d}h_d)}{\partial t}=V_d(t)H_d-(L_d+D)h_d(t)-Q_d,\\ \frac{\partial H_{x_d}}{\partial t}=V_d(t)-(L_d(t)+D(t)),\quad V_d-V(1,t)=L_d-L(1,t),\\ H_dV_d-V(1,t)H(1,t)=L_dh_d-L(1,t)h(1,t).\end{gathered} \tag{4}$$

Решение ищется в области $\Omega=\left\{(l,t)\middle| l\in[0,1], t\in[0,T]\right\}$, где l, t – пространственная и временная независимые переменные соответственно; T – время управления.

Считаем в дальнейшем, что $H_x = \text{const}$, $H_y = \text{const}$, $k_y = k = \text{const}$, $H_{x_k} = \text{const}$, $H_{x_d} = \text{const}$, $x = x(l,t)$, $y = y(l,t)$, $L = L(l,t)$, $V = V(l,t)$, $\Phi_H = \Phi_H(l,t)$, $\Phi_h = \Phi_h(l,t)$, $\Phi_L = \Phi_L(l,t)$; $\Phi_V = \Phi_V(l,t)$ – известные функции, $H = \alpha_1 x + \alpha_2 y + \alpha_3 V + \alpha_4 L$, $h = \beta_1 x + \beta_2 y + \beta_3 V + \beta_4 L$, где $\alpha_i = \text{const}$, $\beta_i = \text{const}$, $i = \overline{1,4}$. При $E_d = 0$ пола-

гаем, что $y_d = y(1,t)$, $x_d = x(1,t)$, $x_d(0) = x(1,0) = x_{d,0}$, $L_d = L(1,t)$, $V_d = V(1,t)$.

Кроме того, $H_y\alpha_1 + H_x\beta_1 = A$, $H_y\alpha_2 + H_x\beta_2 = B$, $H_y\alpha_3 + H_x\beta_3 = C$, $H_y\alpha_4 + H_x\beta_4 = D^0$.

В качестве управляющих воздействий из технологических соображений выбираем потоки жидкости и пара на входах в управляемый аппарат $L(1,t)$ и $V(0,t)$. На эти управляющие потоки накладываются ограничения

$$L_{\min} \le L(t) \le L_{\max}, \qquad V_{\min} \le V(t) \le V_{\max}. \tag{5}$$

Поскольку в дальнейшем для решения задачи оптимального управления будет использован метод вариационного исчисления, введем дополнительные управляющие функции $u(t)$, $z(t)$, с помощью которых ограничения сводятся к равенствам:

$$\begin{aligned} &\big(L(1,t) - L_{\min}\big)\big(L_{\max} - L(1,t)\big) - u^2 = 0, \\ &\big(V(0,t) - V_{\min}\big)\big(V_{\max} - V(0,t)\big) - z^2 = 0. \end{aligned} \tag{6}$$

В качестве критерия оптимизации выбираем интеграл, характеризующий качество продуктов разделения на выходе управляемого объекта:

$$S = \int_0^T \int_0^1 \big(y(l,t) - \Theta^*(l,t)\big)^2 \, dldt \to \min, \tag{7}$$

где $\Theta(l,t)$ – заданный состав выходного продукта.

Сформулируем следующую задачу: во множестве кусочно-непрерывных функций, удовлетворяющих ограничениям (5) найти такие, которые в силу системы (1)–(4) минимизируют выражение (7).

Для получения необходимых условий оптимальности воспользуемся методами вариационного исчисления. Построим скалярные функции – гамильтонианы $\tilde{H}$ и $\tilde{h}$ в области Ω и на границе $\partial\Omega$ соответственно:

$$\tilde{H} = (y - \Theta^*)^2 + \sum_{i=1}^4 \lambda_i X_i + \sum_{i=1}^4 \mu_i \varsigma_i,$$

где λ_i, μ_i – множители Лагранжа;

$$\begin{aligned} \tilde{h} = {} & \frac{\lambda_k^{(1)}}{H_{x_k}}\big[L(0,t) - V(0,t)y(0,t) - W(t)x_k(t)\big] + \\ & + \lambda_k^{(2)}\big[y(0,t) - x_k - a(y^*(x_k) - x_k)\big] + \\ & + \lambda_k^{(3)}\big[L(0,t) - V(0,t) - W(t)\big] + \lambda_d^{(1)}\frac{V(1,t)}{H_{x_d}} \times \\ & \times \big[y(1,t) - x(1,t)\big] + \lambda_d^{(2)}\big[V(1,t) - L(1,t) - D(t)\big] + \\ & + \gamma\big[\big(L(1,t) - L_{\min}\big)\big(L_{\max} - L(1,t)\big) - u^2\big] + \\ & + \varepsilon\big[\big(V(0,t) - V_{\min}\big)\big(V_{\max} - V(0,t)\big) - z^2\big], \end{aligned}$$

где $\lambda_k^{(i)}$, $\lambda_в^{(i)}$, γ, ε – множители Лагранжа.

Будем считать, что $x = z_1$, $y = z_2$, $V = z_3$, $L = z_4$. Рассмотрим вспомогательный функционал:

$$\begin{aligned} J_1 &= \iint_\Omega \left[\big(y - \Theta^*\big)^2 + \sum_{i=1}^4 \lambda_i\left(X_i - \frac{\partial z_i}{\partial t}\right) + \sum_{i=1}^4\left(\varsigma_1 - \frac{\partial z_i}{\partial l}\right)\right] dldt = \\ &= \iint_\Omega \left[\tilde{H} + \sum_{i=1}^4\left(\frac{\partial\lambda_i}{\partial t} + \frac{\partial\mu_i}{\partial l}\right)z_i\right] dldt + \int_{\partial\Omega}\sum_{i=1}^4\big(\mu_i dt - \lambda_i dl\big)z_i. \end{aligned}$$

Пусть $t = \alpha(\sigma)$, $1 = \beta(\sigma)$ – параметрическое задание границы $\partial\Omega$. Тогда

$$\begin{aligned} J_1 = {} & \iint_\Omega \left[\tilde{H} + \sum_{i=1}^4\left(\frac{\partial\lambda_i}{\partial t} + \frac{\partial\mu_i}{\partial l}\right)z_i\right] dldt + \\ & + \int_{\partial\Omega}\sum_{i=1}^4\big(\mu_i\alpha'(\sigma) - \lambda_i\beta'(\sigma)\big)z_i d\sigma. \end{aligned}$$

Получим вариацию J_1, вызванную вариациями управлений $L(1,t)$ и $V(0,t)$:

$$\begin{aligned} \delta J_1 = {} & \iint_\Omega \left[\sum_{i=1}^4\left(\frac{\partial\tilde{H}}{\partial z_i} + \frac{\partial\lambda_i}{\partial t} + \frac{\partial\mu_i}{\partial l}\right)\delta z_i + \sum_{i=1}^4\frac{\partial\tilde{H}}{\partial u_i}\delta u_i\right] dldt + \\ & + \int_{\partial\Omega}\sum_{i=1}^4\big(\mu_i\alpha'(\sigma) - \lambda_i\beta'(\sigma)\big)z_i d\sigma. \end{aligned}$$

Вариация для вспомогательного функционала δJ_2 на границе $\partial\Omega$ вычисляется аналогично:

$$\begin{aligned} \delta J_2 = {} & \int_{\partial\Omega}\left[\frac{\partial\tilde{h}}{\partial x(0,t)}\delta x(0,t) + \frac{\partial\tilde{h}}{\partial y(0,t)}\delta y(0,t) + \frac{\partial\tilde{h}}{\partial L(0,t)}\delta L(0,t) + \right. \\ & + \frac{\partial\tilde{h}}{\partial V(0,t)}\delta V(0,t) + \frac{\partial\tilde{h}}{\partial u}\delta u + \frac{\partial\tilde{h}}{\partial z}\delta z + \frac{\partial\tilde{h}}{\partial y(1,t)}\delta y(1,t) + \\ & \left. + \frac{\partial\tilde{h}}{\partial L(1,t)}\delta L(1,t) + \frac{\partial\tilde{h}}{\partial V(1,t)}\delta V(1,t)\right] dt + \int_{\partial\Omega}\left(\frac{\partial\tilde{h}}{\partial x_k} + \frac{d\lambda_k^{(1)}}{dt}\right)\delta x_k dt + \\ & + \int_{\partial\Omega}\left(\frac{\partial\tilde{h}}{\partial x(1,t)} + \frac{d\lambda_d^{(1)}}{dt}\right)\delta x(1,t)dt - \big(\lambda_k^{(1)}\delta x_k + \lambda_d^{(1)}\delta x(1,t)\big)\Big|_{t=T}. \end{aligned}$$

Таким образом, вычислена вариация вспомогательного функционала $\delta J = \delta J_1 + \delta J_2$. Используя аргументацию вариационного исчисления, получим следующую сопряженную задачу относительно функций Лагранжа, на основе которой разработан численный алгоритм расчета оптимальных управляющих функций [1; 2; 5].

В области Ω имеет место следующее:

$$\begin{aligned} \frac{\partial\lambda_1}{\partial t} + \frac{\partial\mu_1}{\partial l} = {} & -\left[k(y^*)'_x\left(-\frac{\lambda_1}{H_x} + \frac{\lambda_2}{H_y} + \frac{\lambda_3}{C}\left(\frac{A}{H_x} - \frac{B}{H_y}\right)\right) + \right. \\ & \left. + u_4\left(\frac{\lambda_1}{H_x} - \frac{\lambda_3}{H_x}\frac{A}{C}\right) + \frac{u_4 + \Phi_V + \Phi_L}{C}\alpha_1 - \frac{u_4}{C}\beta_1\right], \\ \frac{\partial\lambda_2}{\partial t} + \frac{\partial\mu_2}{\partial l} = {} & -\left[2(y - \theta^*) + k\left(\frac{\lambda_1}{H_x} - \frac{\lambda_2}{H_y} - \frac{\lambda_3}{H_x}\frac{A}{C} + \frac{\lambda_3}{H_y}\frac{B}{C}\right) - \right. \\ & - u_4\left(\frac{\lambda_2}{H_y} - \frac{\lambda_3}{H_y}\frac{B}{C}\right) - \frac{\Phi_V + \Phi_L}{H_y}\left(\lambda_2 - \frac{B}{C}\lambda_3\right) - \\ & \left. - \lambda_3\left(\frac{u_4 + \Phi_V + \Phi_L}{C}\alpha_2 - \frac{u_4}{C}\beta_2\right)\right], \end{aligned}$$

$$\frac{\partial\lambda_3}{\partial t}+\frac{\partial\mu_3}{\partial l}=-\left[u_2\left(-\frac{\lambda_2}{H_y}+\frac{\lambda_3}{H_y}\frac{B}{C}-\frac{\lambda_3}{C}\alpha_2\right)-\right.$$

$$-\frac{u_4\lambda_3\left(-\alpha_3+\alpha_4\right)}{C}-\frac{\lambda_3\alpha_3}{C}\left(\Phi_V+\Phi_L\right)-$$

$$\left.-\frac{\lambda_3u_1\alpha_1}{C}-\lambda_3\left(\frac{u_4+\Phi_V+\Phi_L}{C}\alpha_3-\frac{u_4}{C}\beta_3\right)\right],$$

$$\frac{\partial\lambda_4}{\partial t}+\frac{\partial\mu_4}{\partial l}=-\left[\frac{u_1}{H_x}\left(\lambda_1-\lambda_3\frac{A}{C}\right)+\frac{\lambda_3}{C}\beta_2u_2+\right.$$

$$+\frac{\lambda_3}{C}u_4\left(\beta_3-\beta_4\right)+\frac{\lambda_3}{C}\beta_3\left(\Phi_V+\Phi_L\right)+$$

$$\left.+\frac{\lambda_3u_1\beta_1}{C}-\lambda_3\left(\frac{u_4+\Phi_V+\Phi_L}{C}\alpha_4-\frac{u_4}{C}\beta_4\right)\right].$$

В этой системе неизвестные μ_i или λ_i $(i=\overline{1,4})$ исключаются с помощью соотношений

$$\mu_1+\frac{\lambda_1}{H_x}L-\frac{\lambda_3}{H_x}\frac{A}{C}L-\frac{V\alpha_1-L\beta_1}{C}\lambda_3=0,$$

$$\mu_2-\frac{\lambda_2v}{H_y}+\frac{\lambda_3}{H_y}\frac{B}{C}V-\frac{\lambda_3}{C}\left(V\alpha_2-L\beta_2\right)=0,$$

$$\lambda_4-\frac{\lambda_3}{C}D^0=0,\qquad \mu_3+\mu_4+\frac{\lambda_1}{H_x}x-\frac{\lambda_2}{H_y}y-\frac{\lambda_3}{C}\times$$

$$\times\left(\frac{A}{H_x}x-\frac{B}{H_y}y+H-h-L\left(\beta_3+\beta_4\right)+V\left(\alpha_4-\alpha_3\right)\right)=0.$$

Граничные условия $0<t<T$ при $l=0$ имеют вид

$$\frac{d\lambda_k^{(1)}}{dt}-\frac{\lambda_k^{(1)}}{H_{x_k}}W\left(t\right)-\lambda_k^{(2)}\left[1+a\left(y^*\left(x_k\right)'-1\right)\right]=0,$$

$$\mu_1+\frac{\lambda_k^{(1)}}{H_{x_k}}L\left(0,t\right)=0,\quad \mu_2-\frac{\lambda_k^{(1)}}{H_{x_k}}V\left(0,t\right)+\lambda_k^{(2)}=0,$$

$$\mu_4+\frac{\lambda_k^{(1)}}{H_{x_k}}x\left(0,t\right)+\lambda_k^{(3)}=0,\quad 2\varepsilon z=0,$$

$$\text{grad}\,S_V\equiv\ \mu_3-\frac{\lambda_k^{(1)}}{H_{x_k}}y\left(0,t\right)-\lambda_k^{(3)}+$$

$$+\varepsilon\left(V_{\max}+V_{\min}-2V\left(0,t\right)\right)=0;$$

при $l=1$

$$\frac{d\lambda_d^{(1)}}{dt}-\frac{\lambda_d^{(1)}}{H_{x_d}}V\left(1,t\right)-\mu_1\left(1,t\right)=0,$$

$$\frac{\lambda_d^{(1)}}{H_{x_d}}V\left(1,t\right)-\mu_2\left(1,t\right)=0,$$

$$\frac{\lambda_d^{(1)}}{H_{x_d}}\left(y\left(1,t\right)-x\left(1,t\right)\right)+\lambda_d^{(2)}-\mu_3\left(1,t\right)=0,$$

$$\text{grad}\,S_L\equiv-\lambda_d^{(2)}+\gamma\left(L_{\max}+L_{\min}-2L\left(1,t\right)\right)-\mu_4\left(1,t\right)=0,$$

$$2\gamma u=0.$$

Начальные условия при $t=T,\quad 0<1<1$ выглядят так:

$$\lambda_1=\lambda_2=\lambda_3=\lambda_4=0,\quad \lambda_k^{(1)}(T)=0,$$

$$\lambda_d^{(1)}(T)=0.$$

Для вычисления оптимальных управляющих функций $V(0,t)$ и $L(1,t)$ применяется итерационный метод, который заключается в следующем:

1) задаются начальные приближения $V^0(0,t)$ и $L^0(1,t)$;

2) если известны $V^n(0,t)$ и $L^n(1,t)$, то находятся решения прямой и сопряженной задач;

3) полагаем, что $V^{n+1}(0,t)=V^n(0,t)-\tau_1\text{grad}S_V$, $L^{n+1}(1,t)=L^n(1,t)-\tau_2\text{grad}S_L$;

4) предельные значения $L(1,t)$ и $V(0,t)$ дают решение задачи оптимального управления.

Для численного решения краевых задач и задач оптимального управления разработан численный алгоритм с использованием метода центральных разностей и треугольных сеток [1]. В качестве примера приведены результаты расчетов оптимальной управляющей функции $L(l,t)$ при оптимизации пускового режима для промышленной колонны К-34 установки серно-кислотного алкилирования изобутана бутиленами (разделяемая многокомпонентная смесь сведена к бинарной). На рис. 2 показано изменение концентрации бутана в сырье (рис. 2, *а*), в дефлегматоре (рис. 2, *б*), в кубе (рис. 2, *в*) при управлении потоком жидкости. Оптимальная управляющая функция $L(l,t)$ изображена на рис. 3. При оптимальном управлении выход на заданное значение концентрации целевого продукта происходит быстрее, чем при неоптимальном. Подробное описание установки и экспериментальные значения основных параметров процесса приведены в [5].

Библиографические ссылки

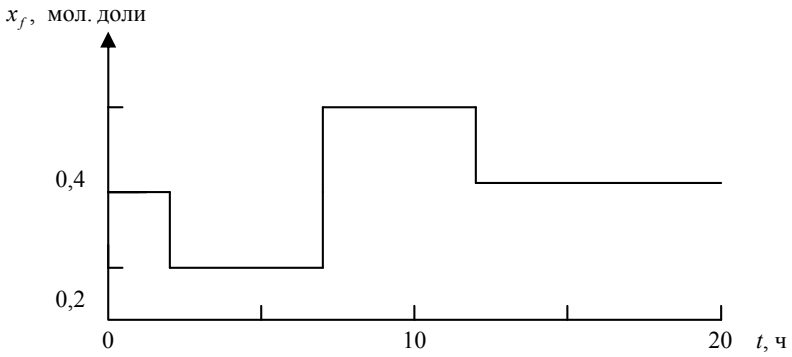
1. Демиденко Н. Д. Моделирование и оптимизация тепломассообменных процессов в химической технологии. М. : Наука, 1991.

2. Демиденко Н. Д. Управляемые распределенные системы. Новосибирск : Наука, 1999.

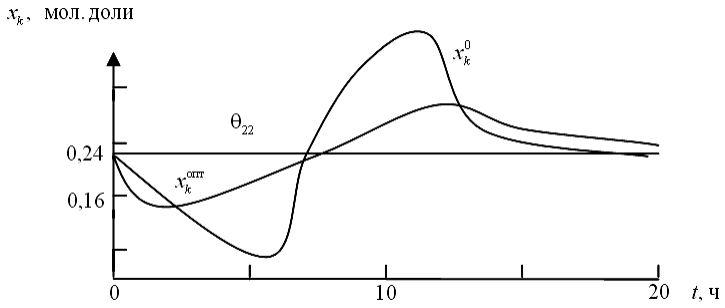
3. Вычислительный эксперимент в проблеме цунами / Ю. И. Шокин [и др.]. Новосибирск : Наука. Сиб. отд-ние, 1989.

4. Demidenko N. D. Modelling of Optimal Regimes in Chemical Engineering Objects with Interacting Flow Recirculation // Syst. Anal. Model. Simul. 1987. Vol. 4. P. 309–320.

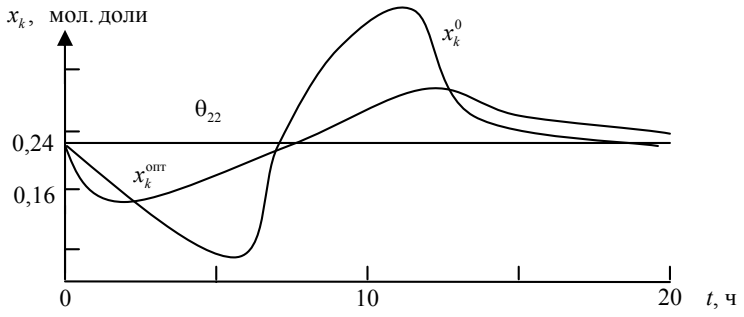
5. Демиденко Н. Д., Потапов В. И., Шокин Ю. И. Моделирование и оптимизация систем с распределенными параметрами. Новосибирск : Наука, 2006.



а



б



в

Рис. 2. Графики изменения концентраций бутана:
а – в сырье; *б* – в дефлегматоре; *в* – в кубе

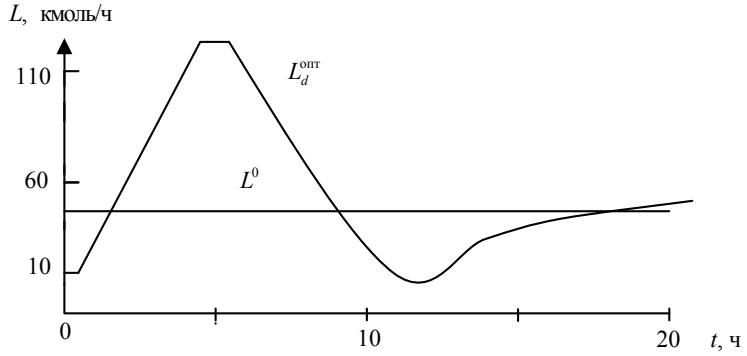


Рис. 3. График оптимального управляющего параметра

N. D. Demidenko, Ju. A. Tereschenko

OPTIMIZATION OF PROCESSES OF MULTICOMPONENT MIXTURES SEPARATION

A mathematical model for the processes of multicomponent mixtures separation is presented. A problem of optimal control is formulated. Necessary optimal conditions are obtained. A numerical method of optimization problems solving is developed.

Keywords: multicomponent mixture, optimal control, nonlinear processes.

© Демиденко Н. Д., Терещенко Ю. А., 2010

УДК 519.866

А. В. Медведев, П. Н. Победаш, Е. С. Семенкин

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ГЛОБАЛЬНОГО СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ*

Описана содержательная постановка задачи и предложена модель глобального социально-экономического развития на основе экономико-математической модели оптимального управления, учитывающей взаимодействие основных экономических агентов мировой социально-экономической системы – производителя, потребителя и управляющего центра (мирового правительства). Авторская концепция сравнивается с подходами других авторов, рассматриваются упрощающие предпосылки при математическом моделировании глобального развития.

Ключевые слова: глобальное социально-экономическое развитие, математическое моделирование, задача оптимального управления, мировая социально-экономическая система, многокритериальная многошаговая задача линейного программирования.

В настоящее время, в связи с глобальными изменениями климата, угрозой перенаселенности Земли, нарушением экологического равновесия и исчерпанием мировых запасов основных невозобновляемых ресурсов на планете, большое внимание уделяется вопросам прогнозирования тенденций развития мировой социально-экономической системы (МСЭС). Вследствие этого актуальной является разработка математических моделей глобальной динамики, описывающих вышеперечисленные аспекты мирового развития, а также создание на их основе систем поддержки принятия решений, которые могут позволить найти закономерности максимально долгого, по возможности бескризисного, функционирования мировой системы и принятия обоснованных управленческих решений.

При моделировании динамики МСЭС можно выделить два основных подхода – имитационный и оптимизационный. *Имитационные модели*, как правило, являются детализированными и позволяют подробно учесть законы функционирования мировой системы, однако не только не дают возможности относительно быстро выйти, но и не гарантируют выход на оптимальные или даже субоптимальные траектории развития. Кроме того, для анализа моделируемых процессов имитационные модели требуют обработки огромного количества независимых экспертных, статистических, экспериментальных данных. *Оптимизационные модели* без значительных затрат на имитационный перебор вариантов обеспечивают поиск наилучшего сценария развития, показатели эффективности которого в рамках выбранной модели превзойти невозможно. Кроме того, при наличии нескольких лиц, принимающих решения, оптимизационная постановка описывает наилучшие, в смысле заданных критериев, стратегии развития МСЭС, учитывающие компромисс интересов нескольких экономических агентов. Имитационный подход используется, например, в работах [1–3], а оптимизационный – в ряде моделей, представленных в [3]. Достаточно широкий обзор по тематике глобального моделирования приведен в статье [4] и монографии [5].

В публикации [6] рассмотрен оптимизационный подход к моделированию глобального социально-экономического развития, поскольку предполагается, как и у авторов работы [3], что у общества есть возможность осознанного выбора стратегии развития с помощью управляющих воздействий на МСЭС. Содержательный смысл рассматриваемой задачи можно сформулировать в следующем виде. Необходимо найти такое распределение инвестиций в основные производственные фонды (ОПФ) мировой экономической системы, количество произведенной в каждый момент t продукции каждой производственной отрасли, текущие инвестиции мирового производственного сектора на снижение уровня загрязнений (CO_2, фреон и др.) с учетом запасов продукции, количества невозобновляемых ресурсов, накопленных загрязнений, численности населения, интересов основных экономических агентов (максимизация их собственных средств) и времени бескризисного существования МСЭС, при которых достигаются Парето-точки в пространстве критериев экономических агентов.

*Работа выполняется в рамках проекта 2.1.1/2710 «Математическое моделирование инвестиционного развития региональных экономических систем» АВЦП «Развитие научного потенциала высшей школы в 2009–2010 гг.».

Таким образом, в [6] предлагается на основе единого экономического и математического ядра концепция математического моделирования МСЭС как задачи оптимального управления, описываемой в классе многокритериальных многошаговых задач линейного программирования (ММЗЛП). В моделях, представленных в работах [1–3], как правило, заранее не выделяются ограничения и критерии функционирования МСЭС, а исследуется вопрос, удовлетворяет ли фазовая траектория системы в своем развитии некоторым апостериорным экономическим, экологическим и социальным ограничениям, и если они нарушены, то можно ли ее вернуть в рамки этих ограничений (решается задача «возврата к пределам роста» по Медоузу [2]). В отличие от указанных работ, существенным концептуальным положением предлагаемого подхода является априорный учет наиболее значимых из перечисленных ограничений мировой системы, которые определяют допустимые варианты ее функционирования. В частности, одной из главных задач при изучении развития МСЭС, по нашему мнению, является построение ее простой и качественно адекватной математической модели, отражающей зависимости наиболее существенных параметров этой системы. При этом, в соответствии с ключевой идеей первых авторов глобальных моделей [1; 2] здесь не ставится задача получения и точного количественного прогноза изменения показателей МСЭС в динамике, а выясняются будущие тенденции этой системы и условия, при которых они реализуются.

Следует отметить мнение ряда исследователей, что для МСЭС (в отличие от ее подсистем) трудно сформулировать критерии ее функционирования как единой системы [2; 3; 7]. Вместе с тем, в современных условиях развития МСЭС существует хотя бы один, несомненно, значимый критерий – максимизация времени ее жизни (выживания). Это свидетельствует об актуальности и целесообразности применения оптимизационных, в особенности многокритериальных, моделей развития МСЭС, включающих, например, такие условия, как максимизация остатка ресурсов, минимизация загрязнений, максимизация собственных средств экономических агентов и т. п.

Рассмотрим математическую постановку задачи. Перечислим предварительно некоторые предпосылки, упрощающие моделирование развития МСЭС:

1) рассматриваются три агрегированных экономических агента МСЭС – производитель, потребитель и управляющий центр;

2) каждая отрасль мирового производственного сектора производит лишь один продукт (принцип чистых отраслей);

3) из всех видов налогов рассматриваются лишь налог на добавленную стоимость (НДС) производственного сектора и налог на доходы физических лиц (НДФЛ) потребительского сектора, которые предполагаются *едиными* для всей планеты, что обеспечивает равные экономические возможности для экономических агентов всех государств [8], из которых и складываются денежные средства управляющего центра;

4) учитываются лишь невозобновляемые ресурсы, поскольку в соответствии с [2] один из прогнозируемых в ближайшем будущем кризисов мировой экономики объясняется исчерпанием какого-либо из основных мировых ресурсов (нефти, железа и т. п.);

5) в первом приближении пренебрегаем запаздываниями влияния одних факторов на другие (например, влиянием накапливаемых производственных выбросов, в частности парниковых газов, на экологию [2]), хотя их учет и возможен за счет увеличения количества фазовых переменных модели;

6) рассматриваются линеаризованные (т. е. приближенные) уравнения, выражающие фактически нелинейные зависимости между основными показателями развития мировой системы, что позволяет создать простую качественную (феноменологическую) модель, отражающую их взаимное влияние. Например, с увеличением уровня загрязнений численность населения убывает, а с увеличением инвестиций на компенсацию загрязнителей уровень этих загрязнений снижается, что согласуется с качественными результатами, полученными по моделям из [1; 2].

С учетом приведенных предпосылок задачу управления глобальным социально-экономическим развитием можно описать в виде линейной задачи оптимального управления – ММЗЛП в стандартной постановке [9]:

$$
\begin{gathered}
x_k(t+1) = (1-1/T_k)x_k(t) + u_k(t)\ (k=1,...,K),\\
x_{K+k}(t+1) = (1-\eta_k)x_{K+k}(t) - a_k x_{3K+I+J+1}(t) -\\
-\sum_{s=1}^{K} d_{ks}u_{K+s}(t)\ (k=1,...,K),\\
x_{2K+k}(t+1) = x_{2K+k}(t) + u_{K+k}(t)\ (k=1,...,K),\\
x_{3K+i}(t+1) = x_{3K+i}(t) - \sum_{k=1}^{K} b_{ik}u_{K+k}(t)\ (i=1,...,I),\\
x_{3K+I+j}(t+1) = x_{3K+I+j}(t) -\\
-\sum_{k=1}^{K} c_{jk}u_{K+k}(t) - e_j u_{2K+j}(t)\ (j=1,...,J),\\
x_{3K+I+J+1}(t+1) = -\sum_{k=1}^{K} \chi_k x_{2K+k}(t) + \zeta x_{3K+I+J+1}(t), \quad (1)\\
x_{3K+I+J+2}(t+1) = -\sum_{k=1}^{K} \overline{P}_k x_{K+k}(t) +\\
+\left[(1-\rho_1-\beta)\sum_{k=1}^{K} a_k P_k - \sum_{i=1}^{I}\sum_{k=1}^{K} b_{ik}a_k \overline{P}_i\right]\times\\
\times x_{3K+I+J+1}(t) + x_{3K+I+J+2}(t) - \sum_{k=1}^{K} u_k(t) - \sum_{j=1}^{J} u_{2K+j}(t),\\
x_{3K+I+J+3}(t+1) = (\alpha+\beta-1)\sum_{k=1}^{K} a_k P_k x_{3K+I+J+1}(t) +\\
+(1-\rho_2)x_{3K+I+J+3}(t) + \psi x_{3K+I+J+4}(t),\\
x_{3K+I+J+4}(t+1) = (\rho_1-\alpha)\sum_{k=1}^{K} a_k P_k x_{3K+I+J+1}(t) + \rho_2 x_{3K+I+J+3}(t) +\\
+(1-\psi)x_{3K+I+J+4}(t)\ (t=0,...,T-1);\\
x_{3K+I+J+5}(t+1) = x_{3K+I+J+5}(t) + 1\ (t=0,...,T-1);\\
x_k(0) = x_k^0\ (k=1,...,3K+I+J+4),\ \ x_{3K+I+J+5}(0) = 0; \quad (2)\\
x_{3K+I+J+m}(t) \ge 0\ (m=1,...,4;\ \ t=0,...,T),
\end{gathered}
$$

$$u_{K+k}(t) \le a_k x_{3K+I+J+1}(t)\ (k=1,...,K;\ t=0,...,T-1),$$
$$u_{K+k}(t) \le \bar{\delta}_k x_k(t)\ (k=1,...,K;\ t=0,...,T-1),$$
$$x_{K+k}(t) \ge 0\ (k=1,...,K;\ t=0,...,T), \quad (3)$$
$$x_{3K+I+j}(t) \ge 0\ (j=1,...,J;\ t=0,...,T),$$
$$x_{3K+i}(T) \ge 0\ (i=1,...,I),$$
$$\sum_{k=1}^{K} c_{jk} x_{2K+k}(T) \le \bar{c}_j \Delta V\ (j=1,...,J);$$
$$u_k(t) \ge 0\ (k=1,...,2K+J;\ t=0,...,T-1); \quad (4)$$
$$J_i = x_{3K+i}(T) \to \max\ (i=1,...,I),$$
$$J_{I+j} = \sum_{k=1}^{K} c_{jk} x_{2K+k}(T) \to \min\ (j=1,...,J),$$
$$J_{I+J+1} = \sum_{t=0}^{T-1} \sum_{k=1}^{K} \frac{P_k (a_k x_{3K+I+J+1}(t) - u_{K+k}(t))}{(1+r_k(t))^t} \to \min, \quad (5)$$
$$J_{I+J+m} = \sum_{t=0}^{T-1} \frac{[x_{3K+I+J+m}(t+1) - x_{3K+I+J+m}(t)]}{[1+r(t)]^t} \to$$
$$\to \max\ (m=2,...,4),$$
$$J_{I+J+5} = x_{3K+I+J+5}(T) \to \max,$$

где $u_k(t)\ (k=1,...,K;\ t=0,...,T-1)$ – инвестиции в основные производственные фонды (ОПФ) *k*-й отрасли; $u_{K+k}(t)\ (k=1,...,K;\ t=0,...,T-1)$ – количество произведенной в момент *t* продукции *k*-го вида; $u_{2K+j}(t)\ (j=1,...,J;\ t=0,...,T-1)$ – текущие инвестиции мирового производственного сектора на снижение уровня *j*-го загрязнителя (метана, CO и т. п.); $x_k(t)$ – накопленная к моменту *t* стоимость ОПФ *k*-го типа; $x_{K+k}(t)\ (k=1,...,K;\ t=0,...,T)$ – количество запасов продукта *k*-го вида на момент *t*; $x_{2K+k}(t)\ (k=1,...,K;\ t=0,...,T)$ – количество произведенного продукта *k*-го вида на момент *t*; $x_{3K+i}(t)\ (i=1,...,I;\ t=0,...,T)$ – количество невозобновляемого ресурса *i*-го вида на момент *t*; $x_{3K+I+j}(t)\ (j=1,...,J;\ t=0,...,T)$ – количество (объем) загрязнителя *j*-го вида, накопленное к моменту *t*; $x_{3K+I+J+1}(t)\ (t=0,...,T)$ – численность населения, прогнозируемая к моменту *t*; $x_{3K+I+J+m}(t)\ (m=2,...,4;\ t=0,...,T)$ – текущие денежные средства производственного и потребительского секторов и управляющего центра в момент *t* соответственно; $x_{3K+I+J+5}(t)\ (t=0,...,T)$ – текущее время (время бескризисного существования) МСЭС; $x_k^0\ (k=1,...,3K+I+J+4)$ – значение *k*-й фазовой переменной в начальный момент времени; $a_k\ (k=1,...,K)$ – норма среднедушевого расхода продукта *k*-го типа; $b_{ik}\ (i=1,...,I; k=1,...,K)$ – норма расхода *i*-го ресурса при производстве единицы продукта *k*-го типа; $c_{jk}\ (j=1,...,J;\ k=1,...,K)$ – количество *j*-го загрязнителя, появляющееся при производстве единицы продукта *k*-го типа; $d_{ks}\ (k=1,...,K;\ s=1,...,K)$ – количество продукта *k*-го вида, затрачиваемое на производство единицы продукции *s*-го типа; $e_j\ (j=1,...,J)$ – количество убыли загрязнителя *j*-го вида, обусловленной инвестициями производственного сектора в одну денежную единицу на его компенсацию; $\bar{c}_j\ (j=1,...,J)$ – максимально допустимая (для выживания человечества как биологического вида) концентрация загрязнителя *j*-го типа; $\underline{P}_i, \overline{P}_k\ (i=1,...,I;\ k=1,...,K)$ – соответственно, стоимость добычи единицы ресурса *i*-го вида и хранения единицы продукции *k*-го типа (за 1 период); $\chi_k\ (k=1,...,K)$ – величина абсолютной убыли населения при суммарном росте уровня всех загрязнителей в связи с увеличением производства продукта *k*-го типа на единицу; $\zeta = 1+\nu$, где ν – темп прироста (убыли) населения; $\bar{\delta}_k = \frac{V_k}{c_k}\ (k=1,...,K)$ – относительная эффективность ОПФ *k*-го типа; $\eta_k \in [0;1]\ (k=1,...,K)$ – доля потери продукции *k*-го типа за 1 период в связи с хранением; V_k, T_k, c_k и P_k (k = 1, ..., K) – производительность, срок службы, стоимость единицы ОПФ и стоимость единицы продукции *k*-го типа соответственно; $\Delta V \approx 4\pi Rh[R+h]$ – объем слоя атмосферы Земли, в котором учитывается накопление загрязнителей, $h \approx 20...30$ км – средняя высота этого слоя [2], а $R \approx 6\,370...6\,400$ км – средний радиус Земли; *I*, *J*, *K* – соответственно, количество видов ресурсов, типов загрязнителей, выбрасываемых производственным сектором в окружающую среду, а также видов производимой им продукции (отраслей); $\alpha \in [0;1]$ – доля стоимости мирового валового продукта (выручки от реализации производственного сектора), оплачиваемая планетарным управляющим центром; $\beta \in [0;1]$ – доля выручки от реализации производственного сектора, выделяемая на фонд оплаты труда (ФОТ) потребительского сектора; $\psi \in [0;1]$ – доля доходов планетарного управляющего центра, выделяемая на ФОТ потребительского сектора; ρ_1, ρ_2 – ставки НДС и НДФЛ производственного и потребительского секторов соответственно; *T* и *r*(*t*) – соответственно, срок действия инвестиционного проекта и ставка дисконтирования в момент *t* (прогнозируемый уровень инфляции); $r_k(t)$ (k = 1, ..., K) – ставка дисконтирования, отражающая приоритетность продукта *k*-го типа в момент *t* в ассортименте производимой продукции.

Дадим краткую содержательную характеристику условий (1)–(5). Уравнения (1) являются, соответственно, уравнениями динамики стоимости ОПФ отраслей производственного сектора, количества запасов и произведенного продукта каждого вида, количества невозобновляемых ресурсов и загрязнителей всех типов к моменту *t*, а также численности населения и собственных средств производственного, потребительского секторов и управляющего центра, а также дискретного времени системы. Условия (2) выражают

начальное состояние МСЭС. Первое ограничение в условии (3) содержательно означает неотрицательность численности населения и денежных средств рассматриваемых мировых экономических агентов, а 2-е и 3-е условия – что количество произведенной продукции не превышает спроса МСЭС по каждому виду продукции в любой момент времени и производственных мощностей соответствующей отрасли; 4, 5 и 6-е ограничения в условии (3) означают неотрицательность текущих запасов продукции и объемов выбросов загрязнителей, а также запасов невозобновляемых ресурсов всех видов в конечный момент; смысл 7-го ограничения в условии (3) заключается в том, что количество загрязнителя каждого вида в момент T окончания проекта не превосходит предельно допустимого объема. Условия (4) отражают неотрицательность управляющих переменных в каждый момент времени.

Содержательный смысл 1-й группы из I условий в выражениях (5) заключается в необходимости оставить будущим поколениям максимальное количество ресурсов всех видов; 2-я группа из J соотношений означает минимизацию количества всех загрязнителей; 3-й критерий минимизирует суммарный дисбаланс между спросом и предложением МСЭС; 4-я группа из трех следующих условий состоит в максимизации чистого дисконтированного дохода (NPV) основных экономических агентов МСЭС, а 5-е – означает максимизацию времени жизни системы.

Существенным аргументом разработки модели глобального развития в виде ММЗЛП является возможность создания системы поддержки принятия управленческих решений, объединяющих математические модели МСЭС, методы их исследования и программное обеспечение автоматизированной информационной системы. В настоящее время нами разработан пакет прикладных программ, использование которого облегчает обработку потоков социально-экономической информации и анализ многокритериальных линейных динамических и статических задач экономического характера [10]. Использование указанного пакета для численного анализа модели (1)–(5) позволит повысить обоснованность принятия решений при управлении мировой системой с учетом интересов многих лиц.

Модель (1)–(5) является обобщением на макроэкономическом уровне модели эффективности регионального экономического развития [11]. Принципиальное отличие предлагаемой модели состоит в том, что спрос по каждому виду продукции уже нельзя трактовать как экзогенно заданную величину, поскольку он зависит от численности населения Земли – основной фазовой переменной модели, которая согласно [2] определяет пределы роста, потребности и необходимые пропорции производства и потребления мировой системы. В то же время, численность населения убывает с ростом загрязнений, что порождает уменьшение глобальной потребительской корзины, а значит, ведет к уменьшению доходов и сворачиванию активности производителя, в конце концов, уменьшая общий уровень загрязнений. Указанная цепочка взаимосвязей, реализуемая в предлагаемой модели, позволяет выявить саморегулирование в МСЭС (как имманентно присущее любой сложной системе свойство), которое может в полной мере проявиться лишь при разумном целенаправленном участии управляющего центра. Поэтому, как отмечается в [5; 12], вера ряда политических и экономических лидеров в то, что рыночная экономика способна самостоятельно справиться с кризисом, является необоснованной.

Необходимо подчеркнуть, что модель (1)–(5), в отличие от моделей в [1; 2], является оптимизационной, т. е. позволяет прогнозировать динамику именно оптимального (в смысле заданных критериев) поведения МСЭС, что сближает ее с некоторыми моделями, предложенными в [3]. Однако в отличие от последних, где критерии качества служат преимущественно лишь для удовлетворения тех или иных ограничений (т. е. построения допустимого управления), в ММЗЛП (1)–(5) изначально включены целевые функции, отражающие качество выбранной траектории развития с учетом интересов рассматриваемых экономических агентов. Более того, по мнению авторов работы [3], нет смысла говорить о каких-либо содержательных целевых критериях социально-экономической системы на макроуровне. Вместе с тем, мы убеждены, что такие критерии, указывающие обществу направления устойчивого развития, должны приниматься во внимание при моделировании глобальных процессов. Например, первые I соотношений в условиях (5) отражают тот факт, что в устойчиво развивающейся мировой экономике необходимо разумно тратить невозобновляемые ресурсы, стараясь оставить как можно более полную ресурсную базу будущим поколениям. Следующие J критериев в условиях (5) учитывают экологическую нагрузку, оказываемую человечеством на мировую экосистему, которую уже нельзя игнорировать в связи с современными масштабами производства. Последняя из целевых функций выражает «срок жизни» (или период бескризисного развития) МСЭС – показатель качества функционирования всей системы, являющийся глобальным аналогом такой характеристики, как средняя продолжительность жизни человека, рассмотренной в [2]. Остальные условия (5) выражают интересы основных экономических агентов мировой экономики.

Кроме того, отметим, что в отличие от модели в [11] часть переменных, ограничений и критериев качества ММЗЛП (1)–(5) представлены в натуральной, а не в универсальной стоимостной форме. Это объясняется тем, что в соответствии с [12] для понимания законов функционирования глобальной экономики важнее *физическое*, а не *денежное* представление некоторых показателей. Например, стоимость единицы невозобновляемого ресурса определяется не только затратами на его добычу (с учетом существующих технологий в добывающей отрасли), но и разведанными на сегодня его мировыми запасами. В свою очередь, некоторые ограничения (3) в нату-

ральной форме можно заменить на «стоимостное» представление и наоборот. Например, 2-е условие (3) в «стоимостном» формате примет вид: $P_k u_{K+k}(t) \le a_k P_k x_{3K+I+J+1}(t)$ $(k = 1, ..., K; t = 0, ..., T-1)$, а 3-е неравенство в вещественной форме эквивалентно соотношению $P_k u_{K+k}(t) \le \delta_k x_k(t)$ $(k = 1, ..., K;$ $t = 0, ..., T-1)$, где $\delta_k = P_k \bar{\delta}_k = \dfrac{P_k V_k}{c_k}$ – максимальная фондоотдача ОПФ *k*-го типа.

При моделировании глобального развития, на наш взгляд, целесообразно заложить в модель предпосылку о равенстве всех членов общества в удовлетворении своих потребностей в пище, одежде, жилье и т. п., так как в противном случае все достижения цивилизации (высокий валовой мировой продукт на душу населения, высокая производительность основных средств, научно-технический прогресс и др.) могут терять смысл из-за угрозы социальных конфликтов и нестабильности развития. В этой связи ММЗЛП (1)–(5) разрабатывалась как модель нормативного характера, в которой используются единые среднедушевые нормы a_k расхода продуктов. С одной стороны, эти нормы отражают равенство возможностей каждого человека в устойчиво развивающейся, социально справедливой экономике для удовлетворения своих насущных потребностей [12–14], а с другой – позволяют снизить либо вовсе исключить чрезмерное (надфизиологическое) потребление ресурсов и продуктов. Важно отметить, кроме того, что, по нашему мнению, любая математическая модель, описывающая социально-экономические взаимодействия, должна решать проблему соответствия предложения спросу как одну из краеугольных задач в социально-экономических системах. В модели (1)–(5) величины $a_k x_{3K+I+J+1}(t)$, $(k = 1, ..., K; t = 0, ..., T-1)$ отражают спрос потребительского сектора на *k*-й вид продукции в момент *t*, ограничивающий количество производимой в МСЭС продукции (2-е неравенство в условии (3)). Другим ограничителем роста продукции мирового производственного сектора, очевидно, является ограничение выпуска продукции технологическими возможностями производственного сектора (3-е неравенство в условии (3)).

Построенная модель является задачей линейного оптимального управления, хотя предлагаемая концепция позволяет обобщить ее и на нелинейный случай. Переход к нелинейной модели может повысить ее адекватность количественно, но не изменит качественных закономерностей, возникающих при взаимодействии мирового производителя и мирового потребителя под управлением мирового центра. Таким образом, для предварительного (качественного) анализа в первом приближении можно рассматривать линейный вариант модели развития МСЭС. Очевидно, что для уточнения полученных данных о пропорциях развития и времени жизни МСЭС потребуются значительные объединенные усилия специалистов различных областей знания – экономистов, экологов, социологов, демографов, математиков и др., а также использование высокоэффективных численных методов – генетических и эволюционных алгоритмов, параллельных вычислений и т. п. Отметим, что ММЗЛП (1)–(5) представляет собой лишь математический каркас МСЭС, который можно в дальнейшем агрегировать или детализировать в зависимости от целей исследования, дополняя и модифицируя ее уравнения, ограничения и целевые критерии.

Предлагаемый в данной работе подход позволяет с единых экономических (выбор, принципы описания деятельности, взаимодействий и целевых критериев экономических агентов) и математических (выбор класса ММЗЛП, структуры уравнений движения, ограничений и целевых критериев) позиций описать функционирование МСЭС, а также относительно просто рассматривать многие актуальные и потенциальные вопросы ее развития – ресурсные и экологические ограничения, бескризисность, взаимодействие нескольких экономических агентов и т. п.

Библиографические ссылки

1. Форрестер Дж. Мировая динамика. М. : Наука, 1978.
2. The Limits to Growth : A Report for the Club of Rome's Project on the Predicament of Mankind / D. L. Meadows [et al.]. N. Y. : Universe Books. 1972.
3. Математические модели глобального развития. Критический анализ моделей природопользования / В. А. Егоров [и др.]. М. : Гидрометеоиздат, 1980.
4. Махов С. А., Посашков С. А. Анализ стратегических рисков на основе математического моделирования : препринт / ИПМ РАН. М., 2007.
5. Прогноз и моделирование кризисов и мировой динамики / отв. ред. А. А. Акаев, А. В. Коротаев, Г. Г. Малинецкий. М. : Изд-во ЛКИ, 2010.
6. Моделирование стратегии мирового социально-экономического развития как задачи оптимального управления / М. А. Горбунов [и др.] // Вестник СибГАУ. Вып. 4(25). 2009. С. 71–75.
7. Махов С. А. Математическое моделирование мировой динамики и устойчивого развития на примере модели Форрестера : препринт / ИПМ РАН. М., 2005.
8. Solte D. Global financial system in balance. Crisis as opportunity for a sustainable future. Berlin : Terra Media Verlag, 2009.
9. Пропой А. И. Элементы теории оптимальных дискретных процессов. М. : Наука, 1973.
10. Конструктор и решатель дискретных задач оптимального управления («Карма») : пакет прикладных программ / А. В. Медведев [и др.]. Св-во о регистрации № 2008614387 от 11.09.2008.
11. Медведев А. В. Применение z-преобразования к исследованию многокритериальных линейных моделей регионального экономического развития : монография / Сиб. гос. аэрокосмич. ун-т им. акад. М. Ф. Решетнева. Красноярск, 2008.

12. Рандерс Й., Медоуз Д. Пределы роста. 30 лет спустя. М. : Академкнига, 2008.

13. Solte D. Weltfinanzsystem am Limit – Einblicke in den «Heiligen Gral» der «Globalisierung»). Berlin : Terra Media Verlag, 2007.

14. Радермахер Ф. Баланс или разрушение. Экосоциальная рыночная экономика как ключ к устойчивому развитию мира / Некоммерческое партнерство «За устойчивое информационное общество в России». Новосибирск, 2008.

A. V. Medvedev, P. N. Pobedash, E. S. Semenkin

MATHEMATICAL MODEL OF THE WORLDWIDE SOCIOECONOMIC DEVELOPMENT

A mathematical model of the worldwide socioeconomic development, based on the optimization approach, is suggested and analyzed in this article. The model takes into consideration the interests of producing, consuming, finance and control sectors of the worldwide socioeconomic system. The authors' conception is opposed with the other authors' approaches, and conditions which ease mathematic modeling of worldwide advancement are considered.

Keywords: world socioeconomic development, mathematical modeling, optimal control problem, worldwide socioeconomic system, multicriteria multistage linear programming problem.

© Медведев А. В., Победаш П. Н., Семенкин Е. С., 2010

УДК 539.3

Н. В. Молокова

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ НЕФТЕЗАГРЯЗНЕНИЯ ПОРИСТОЙ СРЕДЫ

Приведено построение и исследование математической модели двухфазной фильтрации, учитывающей движение углеводородных загрязнителей и воздуха в пористом грунте. Модель включает в себя систему уравнений в частных производных с дополнительными условиями. В число дифференциальных уравнений входит уравнение баланса массы в элементе пористой среды – уравнение неразрывности, а также дифференциальные уравнения движения. Для замыкания системы вводятся уравнения состояния рассматриваемого загрязнителя и среды. Начальные и граничные условия соответствуют фильтрационному процессу, начиная с поверхности грунта и начальной стадии разлива загрязнителя. Проводится сравнительный анализ результатов математического моделирования с экспериментами.

Ключевые слова: *модель двухфазной фильтрации, геофильтрация, нефтезагрязнение, пористая среда, математическое моделирование.*

Методы исследования проблем экологической безопасности при нефтяном загрязнении в настоящее время быстро развиваются, складываясь в новое научное направление – *нефтеэкологию*. Основой для выявления нефтеэкологической ситуации и выработки мероприятий по предупреждению и ликвидации нефтяного загрязнения является проведение комплексного мониторинга и детальной экспертизы нефтезагрязненных земель, унификация ряда параметров и методов, обеспечивающих очистку почв от нефтяного загрязнения, в том числе и построение компьютерных математических моделей для определения динамики распространения углеводородного загрязнения [1].

Создаются и исследуются геофильтрационные модели средствами математического моделирования. Использование теории фильтрации в сочетании с современными математическими методами и вычислительными машинами позволяет значительно сократить сроки перехода от лабораторных установок непосредственно к эффективному применению технологии экологической реабилитации.

Постановка задачи. Проблема исследования процессов миграции углеводородов, фильтрующихся с поверхности земли, является очень сложной и до сих пор остается областью активных научных исследований. В связи с этим объектом исследования является процесс нефтезагрязнения почвы в результате аварийных утечек и разливов, а предметом – прикладные аспекты моделирования фильтрации жидких углеводородов в пористой среде.

Цель исследования заключается в построении математической модели процесса нефтезагрязнения пористых сред, изучении на ее основе динамики распространения углеводородов с учетом действующих факторов, объединении разработанных методик и математической модели в программный комплекс.

Настоящее исследование основано на *гипотезе* сплошной среды. Для достижения поставленной цели в работе решаются следующие *задачи*:

1. Анализ проблемы исследования в научной отечественной и зарубежной литературе, анализ и сравнение между собой построенных ранее моделей дан-

ного объекта, сравнительный анализ методов исследования. Формирование модели предметной области.

2. Построение математической модели для оценки интенсивности и характера пространственно-временного распределения нефтепродуктов в почвенных средах. Модель должна быть адекватной естественным условиям.

3. Решение геофильтрационной задачи и проведение сравнительного анализа численных результатов с экспериментальными данными.

4. Создание программного комплекса; программы должны быть достаточно просты для их эффективного использования, а заложенные в них математические модели должны допускать дальнейшее развитие для решения нового круга задач.

Как и в любом методе научного познания в моделировании выделяют экспериментальную и теоретическую части. Исходная информация для обоснования системы основных уравнений модели была получена из анализа экспериментальных данных: 1) анализа механизма загрязнения в хорошо проницаемых и менее проницаемых средах; 2) анализа механизма загрязнения тундровых глеевых почв; 3) определения динамики миграции углеводородов в макрооднородной воздушно-сухой почве [2].

Проведенный анализ экспериментальных данных показал, что, попадая на поверхность земли, жидкие углеводороды начинают просачиваться по порам зоны аэрации, где преобладает вертикальная миграция. Когда нефтепродукты встречают на своем пути менее проницаемый слой, происходит их накопление и растекание в горизонтальном направлении. Процессы миграции углеводородов в почвах определяются их свойствами, параметрами среды и управляются законами теории фильтрации.

Рассмотрим построение модели процесса фильтрации загрязнителей в поровом пространстве и исследование на ее основе интенсивности и характера распределения углеводородного загрязнения с учетом действующих на этот процесс факторов.

При загрязнении нефтью и ее компонентами наблюдается совместная фильтрация несмешивающихся фаз. Область, в которой протекает процесс фильтрации, имеет вид, представленный на рис. 1.

Рис. 1. Реальное представление предметной области

В качестве наиболее используемого и разработанного метода количественного описания фильтрационных процессов в подземной гидромеханике применяется макроскопический метод, в основе которого лежит гипотеза сплошности, законы и методы механики сплошной среды.

Составим список параметров модели, от которых зависит степень и характер загрязнения геосреды. Для этого рассмотрим основные понятия геофильтрации. Под *геофильтрацией* понимают движение жидкости в пористой среде. Пористыми материалами являются многие естественные и искусственные тела: углеродное волокно, никель-алюминиевый сплав, древесина, грунты, горные породы и т. д. Структура этих материалов может быть разнообразной. Характерная особенность всех этих материалов – способность накапливать в себе жидкость и позволять ей двигаться под действием внешних факторов.

Основными свойствами жидких углеводородов являются вязкость μ и плотность ρ. Перечисленные зависимости определяют в лабораторных условиях путем анализа проб жидкого загрязнителя.

Пористая среда характеризуется двумя основными свойствами: пористостью m и проницаемостью k. Пористость (эффективная) определяется соотношением $m = V_n/V$, где V_n – объем сообщающихся пор; V – общий объем пористой среды. Таким образом, пористость – величина безразмерная и находится в пределах 0,1–0,4. Другой макроскопической характеристикой грунта является проницаемость, характеризующая свойство грунта пропускать через себя жидкость. Проницаемость измеряется в единицах площади ($м^2$). В процессе проникновения загрязнителей в грунт наблюдается многофазная фильтрация. Поэтому для характеристики проницаемости грунтов вводится понятие относительной фазовой проницаемости $k_f(s)$. Величина ее зависит не только от физических свойств пород, но также от степени насыщенности порового пространства жидкостями и газом. При исследовании возникает важный вопрос о том, какую часть пустот занимает каждая компонента. Насыщенность пористой среды определяется как относительная часть объема пустот среды, занятая этой фазой. Отсюда, обозначая насыщенность через s, получаем: $s = V_c/V$, где V_c – объем среды, занятый данной компонентой; V – общий объем пустот среды. Под скоростью фильтрации u понимают объем жидкости, протекающей в единицу времени через единичную площадку пористой среды, перпендикулярно направлению потока. Вектор скорости фильтрации направлен по нормали к той площадке, через которую происходит наибольший расход:

$$u = \frac{Q}{F} = \frac{k}{\mu}\frac{\Delta p}{L},$$

где u – скорость линейной фильтрации; Q – объемный расход жидкости в единицу времени; μ – динамическая вязкость жидкости; F – площадь фильтрации; Δp – внешний перепад давления; L – длина образца пористой среды.

Математическое моделирование, особенно геофильтрационных задач, всегда сопровождается некоторыми допущениями, необходимыми, с практической точки зрения, для того, чтобы сделать задачу разрешимой. Построение модели будем выполнять при следующих предположениях:

1) рассматривается несжимаемая жидкость (ρ = const, μ = const) в недеформируемом грунте (m = const, k = const);

2) боковые границы не влияют на процесс фильтрации;

3) в первом приближении рассматривается насыщение почвы только воздухом; давление газа считаем равным атмосферному, а воду капиллярно связанной;

4) сумма насыщенностей всегда равна единице, тогда для двухфазной фильтрации $s_1 + s_2 = 1$, поэтому из двух насыщенностей независима только одна, и вводится обозначение $s \equiv s_1$ – нефтенасыщенность, $s_2 = 1 - s$;

5) при двухфазном течении несмешивающейся жидкости давления в каждой из фаз не равны между собой ($p_1 \neq p_2$), $p_c(s) = p_2 - p_1$, индексы 1, 2 относятся к загрязнителю и газу соответственно.

На основании вышепринятых гипотез, для формализации модели используем известные из нефтегазовой подземной гидромеханики законы и методы механики сплошной среды. Основными уравнениями, описывающими движение жидкости в пористой среде, являются уравнение неразрывности и закон фильтрации Дарси. С учетом пористости закон сохранения массы в пористой среде или уравнение неразрывности в дивергентной форме и декартовой системе координат для однородной жидкости в недеформированном грунте имеет вид

$$m\frac{\partial s}{\partial t}+\left[\frac{\partial(\rho u_x)}{\partial x}+\frac{\partial(\rho u_y)}{\partial y}+\frac{\partial(\rho u_z)}{\partial z}\right]=0, \qquad (1)$$

где s – насыщенность пористой среды; m – пористость материала; ρ – плотность; $\vec{u}$ (u_x, u_y, u_z) – вектор скорости фильтрации; u_x, u_y, u_z – компоненты вектора фильтрации; x, y, z – координаты; t – время.

Обобщение закона фильтрации (закон Дарси) для трехмерного течения имеет следующий вид:

$$\vec{u}=-k\frac{k(s)}{\mu}(\nabla p-\rho\vec{g}), \qquad (2)$$

где k – проницаемость пористой среды; μ – динамическая вязкость; p – давление; ρ – плотность; $k(s)$ – относительная проницаемость; $\vec{g}$ – вектор ускорения свободного падения.

Для описания многофазного течения используются законы Дарси в обобщенном виде:

$$\vec{u}_l=-k\frac{k_l(s)}{\mu_l}(\nabla p_l-\rho_l\vec{g}), \qquad (3)$$

где k – проницаемость пористой среды; μ_l – динамическая вязкость l-й фазы; p_l – давление l-й фазы; ρ_l – плотность l-й фазы; $k_l(s)$ – относительная фазовая проницаемость l-й фазы; $\vec{g}$ – вектор ускорения свободного падения.

Рассмотрим пространственно-трехмерную постановку задачи геофильтрации, которая является исходной для большинства моделей, используемых в задачах геоэкологии. Движение углеводородного загрязнителя и воздуха с учетом взаимосвязи фильтрационных свойств пород с уровнем и интенсивностью их загрязнения в трехмерной постановке описывается системой уравнений в декартовой системе координат:

$$m\frac{\partial s}{\partial t}+\left[\frac{\partial(\rho_1 u_x)}{\partial x}+\frac{\partial(\rho_1 u_y)}{\partial y}+\frac{\partial(\rho_1 u_z)}{\partial z}\right]=0,$$

$$m\frac{\partial(1-s)}{\partial t}+\left[\frac{\partial(\rho_2 u_x)}{\partial x}+\frac{\partial(\rho_2 u_y)}{\partial y}+\frac{\partial(\rho_2 u_z)}{\partial z}\right]=0. \qquad (4)$$

В рассматриваемых условиях закон фильтрации (закон Дарси) записывается в следующем виде (индексы 1, 2 относятся к загрязнителю и воздуху):

$$u_1=-k\frac{k_1(s)}{\mu_1}(\nabla p_1-\rho_1\vec{g}),$$

$$u_2=-k\frac{k_2(s)}{\mu_2}(\nabla p_2-\rho_2\vec{g}). \qquad (5)$$

Капиллярное давление определяется соотношением $p_c(s) = p_2 - p_1$.

Ключевым моментом является наличие в выражении (5) двух членов. Первый из них – это фазовое давление p_l. Наличие этого члена означает, что в теории фильтрации загрязнителей в грунте должно учитываться различие фазовых давлений, обусловленное действием капиллярных сил. Другой важный параметр – относительная фазовая проницаемость $k_l(s)$. Функция относительной фазовой проницаемости отражает количественную потерю скорости фазы за счет того, что часть порового пространства занята другой фазой и является экспериментально измеряемой функцией насыщенности вытесняющей фазы. Проницаемость измеряется путем создания в образце одномерного потока и обработки результатов измерений при помощи соответствующей формы закона Дарси. При этом учитывают предельные значения насыщенности вытесняющей фазы s_* и s^*, такие, что при $s \le s_*$ и $s \ge s^*$ соответствующие фазовые проницаемости $k_l(s)$ обращаются в нуль. Из обобщенного закона Дарси тогда следует, что если $s \le s_*$, то неподвижна вытесняющая фаза, если же $s \ge s^*$, то неподвижна вытесняемая фаза. При этом s_* называют остаточной нефтенасыщенностью, $(1 - s^*)$ – остаточной газонасыщенностью. При $s \ge s^*$ капиллярными эффектами можно пренебречь. То есть, $k_1(0) = k_2(1) = 0$, $k_1(1) = k_2(0) = 1$. Типичный вид функций относительных фазовых проницаемостей $k_l(s)$ приведен в большинстве литературных источников.

В задачах, представляющих практический интерес, часто удобно пользоваться эмпирическими формулами Викова–Ботсета или Чень-Чжун-Сяна:

$$k_1(s)=\begin{cases}\left(\dfrac{s-s_*}{s^*}\right)^{\alpha_1}, & s_*<s\le 1\\ 0, & 0\le s\le s_*;\end{cases} \tag{6}$$

$$k_2(s)=\begin{cases}\left(\dfrac{s^*-s}{s^*}\right)^{\alpha_2}, & 0\le s<s^*\\ 0, & s^*\le s\le 1.\end{cases} \tag{7}$$

Соотношения (5), задающие связь p_l и $k_l(s)$ с фазовыми насыщенностями s_i, определяют характер модели.

Подставим уравнения (5) в уравнения неразрывности (4). В результате получим следующие:

$$m\frac{\partial s}{\partial t}=\operatorname{div}\left[k\frac{k_1(s)}{\mu_1}\left(\nabla p_1-\rho_1\vec{g}\right)\right], \tag{8}$$

$$m\frac{\partial(1-s)}{\partial t}=\operatorname{div}\left[k\frac{k_2(s)}{\mu_2}\left(\nabla p_2-\rho_2\vec{g}\right)\right], \tag{9}$$

$$p_2=p_1+p_c(s), \tag{10}$$

$s_1+s_2=1$, при $s\equiv s_1$ значение $s_2=1-s$.

Уравнения (8) и (9) после ряда преобразований с учетом выражения (10) запишем в безразмерной форме:

$$\operatorname{div}\left[k\left(\frac{k_1(s)}{\mu_1}+\frac{k_2(s)}{\mu_2}\right)\nabla p_1\right]= \\ =-\operatorname{div}\left[k\frac{k_2(s)}{\mu_2}\nabla p_c(s)\right]+k\frac{\partial k_1(s)}{\mu_1\partial z}\rho_1\vec{g}+k\frac{\partial k_2(s)}{\mu_2\partial z}\rho_2\vec{g}. \tag{11}$$

Примем, что отношение вязкости фаз $\mu_0=\dfrac{\mu_1}{\mu_2}$ и $p=p_1$, тогда уравнение (11) перепишем в виде

$$\operatorname{div}\left[\left(k_1(s)+\mu_0k_2(s)\right)\nabla p\right]=-\operatorname{div}\left[\mu_0k_2(s)\nabla p_c(s)\right]+ \\ +k_1(s)\rho_1\vec{g}+\mu_0k_2(s)\rho_2\vec{g}. \tag{12}$$

Сформулируем постановку задачи, т. е. зададим условия в начальный момент времени и краевые условия на границах исследуемой области.

Начальные условия:

При $t=0$:

$$s=1,\ \ (x,y,z)\in\Omega_0, \\ s=0,\ (x,y,z)\in\Omega\setminus\Omega_0, \tag{13}$$

$p=1,\ (x,y,z)\in\Omega\setminus\Omega_0,\ p=\rho_1\vec{g}z+p_{\text{атм}},\ (x,y,z)\in\Omega_0.$

Граничные условия:

Граница с атмосферой (Γ_1): на поверхности контакта с атмосферой имеем условие

$$p=p_{\text{атм}},\ \text{на}\ \Gamma_1, \tag{14}$$

где $p_{\text{атм}}$ – атмосферное давление.

На границе (Γ_2) между фазами давление нефти определяется соотношением

$$p=p_{\text{атм}}+p_c(s). \tag{15}$$

Непроницаемая граница (Γ_3): жидкость не может проходить через непроницаемую границу, поэтому условие непротекания можно записать в следующем виде:

$$k\frac{k_1(s)}{\mu_1}(\nabla p_1-\rho_1\vec{g})\cdot n\Big|_{\Gamma3}=0, \tag{16}$$

где n – вектор нормали к границе Γ_3.

Таким образом, математическая модель рассматриваемого процесса составлена из уравнений (5)–(16). Построенная модель, основанная на двухфазной фильтрации в физических переменных с учетом капиллярно-гравитационного взаимодействия, позволяет определить распределение скоростей фильтрации в почве в любой момент времени, распределение давления p_l, насыщенности s_l.

Метод решения. При численном решении приведенной системы уравнений была использована конечно-разностная аппроксимация. В результате дискретизации дифференциальных уравнений с помощью метода конечных разностей непрерывное распределение параметров заменяется дискретным. Чтобы построить разностную схему для исходной задачи, выберем равномерную сетку с шагом h_z по переменной z, с шагом h_x по переменной x, с шагом h_y по переменной y и шагом τ по времени t. Получим сеточную область:

$$w_{h,\tau}=\{z_{i1}=i_1h_z\ (0\le i_1\le n_z),\ x_{i2}=i_2h_x\ (0\le i_2\le n_x), \\ y_{i3}=i_3h_y\ (0\le i_3\le n_y),\ t_j=j\tau,\ 0\le j\le n_t\},$$

где $n_z=L_1/h_z$; $n_x=L_2/h_x$; $n_y=L_3/h_y$; $n_t=T/\tau$. Заменим производные во внутренних узлах $w_{h,\tau}$ конечно-разностными отношениями и с учетом представления переменного коэффициента получим разностную схему:

$$mS_t=\left(\lambda_\alpha(s)P_{\bar{x}_\alpha}-G_{\bar{x}_\alpha}\right)_{\hat{x}_\alpha}, \tag{17}$$

где $\lambda_{\alpha\pm0,5}(s)=k\left(\dfrac{s^{j+1}_{i_\alpha\pm1}+s^{j+1}_{i_\alpha}}{2}\right)$.

При $s\equiv1$ фазовая проницаемость $k_2(s)$ обращается в нуль, тогда разностную аппроксимацию уравнения для давления запишем в следующем виде:

$$P_{\bar{x}_\alpha x_\alpha}=0. \tag{18}$$

При $0<s<1$

$$P^{j+1}_{x_\alpha}=P_{\text{атм}}+P^{j+1}_c(s^{j+1}); \tag{19}$$

$$s^0_{i_1i_2i_3}=0,\ j=0,\ i_1=0,\ldots,n_z-1;$$

$$i_2=0,\ldots,n_x-1;\ i_3=0,\ldots,n_y-1;$$

$$s^j_{0i_2i_3}=1,\ i_1=0;\ 0\le i_2,\ i_3=3;$$

$$s^j_{0i_2i_3}=0,\ i_1=0;\ 3\le i_2,i_3\le n_{xy}-1;$$

$$P^j_{0i2i3}=\rho_1g,\ i_1=0;\ 0\le i_2,i_3=3;P^j_{0i2i3}=P_{\text{атм}},$$

$$i_1=0;\ 3\le i_2,i_3\le n_{xy}-1;$$

$$\frac{s_N^{j+1}-s_N^j}{\tau}=-\frac{2}{h}\lambda_{N-1/2}^j p_N^j-\rho_1 g,\quad j=1,\ldots,n_t-1.$$

Сформулированная постановка задачи и фильтрации двухфазной несжимаемой жидкости, где искомыми функциями являются P, S, позволяет привести схему решения.

1. По известным на j-м временном шаге P^j, S^j и краевым условиям для P определяем P^{j+1} из сеточного аналога уравнения (11) с помощью итерационного метода Гаусса–Зейделя.

2. По найденному P^{j+1}, известному S^j и краевым условиям для S^j определяем S^{j+1} из сеточного аналога уравнения (8). Уравнение (8) параболического типа. При этом целесообразно применять итерационные неявные разностные схемы.

3. Тогда по найденному S^{j+1} величина P^{j+1} определяется однозначно. Счет величин на $(n + 1)$-м шаге окончен, переходим к счету следующего шага и т. д.

Для реализации вычислительного алгоритма была построена компьютерная модель с использованием технологии визуального программирования Borland Delphi. Предлагаемая модель позволяет решать достаточно разнообразные задачи. В рамках проблемы исследования негативного воздействия углеводородного загрязнения на почвенную экосистему наиболее важными задачами являются изучение динамики распространения фронта загрязнения, анализ взаимосвязи фильтрационных свойств различного типа почв с интенсивностью их загрязнения и определение закономерности изменения загрязненной зоны.

Исследования, проводимые с помощью полученной модели и априорной информации. Исследование проводилось поэтапно [3–5]. На *первом этапе* осуществлялось изучение режима полного пропитывания нефтепродуктами пористой среды и выявление взаимосвязи фильтрационных свойств различного типа почв с интенсивностью их загрязнения. В этом случае система уравнений (5)–(16) решалась в предположении независимости решения от горизонтальных координат. То есть исходная трехмерная модель была сведена к одномерной.

При проведении расчетов принято следующее: пористость грунта m = 0,05–0,4; проницаемость k = $1{,}15\cdot10^{-3}$–11,5 мкм^2; плотность загрязнителя ρ_1 = 1 г/см^3; плотность воздуха ρ_2 = 0,00123 г/см^3; вязкость загрязнителя μ_1 = 0,54–0,9 сПз; высота области фильтрации загрязнителя L = 1–10 м; предельные значения насыщенности s^* = 0,9; s_* = 0,1; насыщенность на верхней границе области S_0 = 1,0; ускорение свободного падения g = 9,8 м/с^2.

По результатам моделирования и проведенным исследованиям можно отметить следующее:

1. Форму фронта определяют кривые $k(s)$ и члены дисперсионного типа. Эти параметры являются экспериментально измеряемыми функциями насыщенности. Тем самым возможность воспроизведения характеристик процесса фильтрации в модели напрямую связана с имеющимися экспериментальными данными. Как показывают расчеты, гравитация и капиллярное давление существенно влияют на распределение нефтенасыщенности. Увеличение гравитационного параметра ведет к уменьшению s_*, и решения приобретают вид, подобный бегущей ступеньке. Это говорит об увеличении фронтовой насыщенности. С другой стороны, ориентационный эффект снижается, фронт становится более сглаженным.

2. В результате расчетов максимальные скорости фильтрации определены для пород с минимальным содержанием песчаной фракции, а минимальные – для пород с содержанием песчаной фракции, близкой к максимуму. Следовательно, значительное количество нефтепродуктов мигрирует в направлении улучшения фильтрационных параметров почвы, и здесь следует ожидать максимального загрязнения.

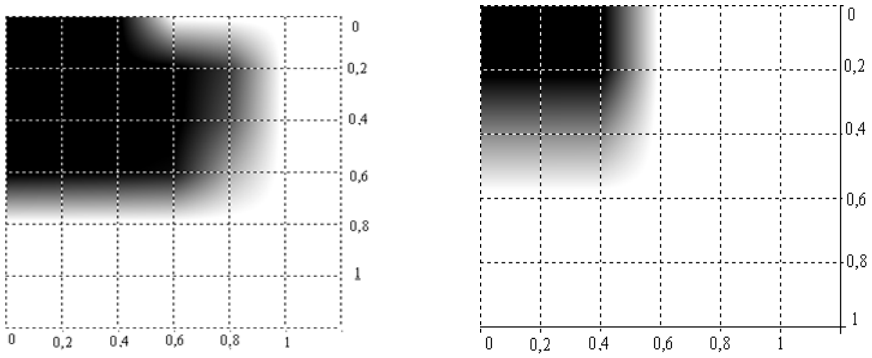
Экологическая реабилитация загрязненной зоны – это чрезвычайно сложный процесс. Поэтому прежде чем начать дорогостоящие экскавационные работы, важно определить не только глубину проникновения нефтезагрязнителя, но и установить закономерности изменения зоны нефтяного загрязнения.

В связи с этим *второй этап* исследований посвящен изучению интенсивности и характера пространственно-временного распределения нефтепродуктов. Система уравнений (5)–(16) решалась в предположении фиксированной одной горизонтальной координаты.

Численные результаты (рис. 2), показывают неоднозначный характер распределения нефтепродуктов. В случае, если разрез представлен суглинисто-глинистыми почвами, плановая фильтрация преобладает над вертикальной. Экспериментально установлено, что тяжелые углеводороды закрепляются в верхних слоях суглинистых и глинистых почв. Качественно это объясняется малой пропускной способностью пористой среды. В противном случае, когда разрез представлен менее дисперсными породами, преобладает вертикальная миграция нефтепродуктов.

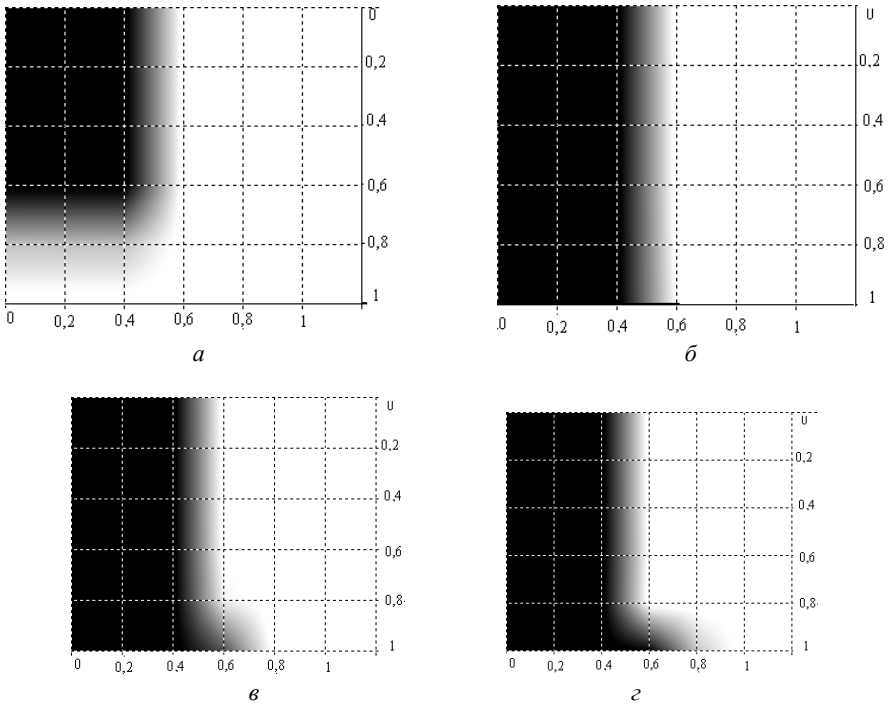
Результат расчета нефтезагрязнения в хорошо проницаемом песке при наличии капиллярной каймы (I – I) представлен на рис. 3. Загрязнитель распространяется почти в вертикальном направлении. Над I – I, практически непроницаемой для углеводородов, происходит их накопление. В зоне границы раздела фаз происходит горизонтальное распространение нефти. Объяснением тому является то, что гравитационный напор оказывается сильнее капиллярного противодействия в течение всего наблюдаемого времени.

Результаты сопоставления рассмотренной модели фильтрации с экспериментальными данными [2] позволяют сделать вывод, что построенная модель, основанная на двухфазной фильтрации, адекватна реальному процессу загрязнения. Проведенные исследования показали, что модель описывает при известных фильтрационных свойствах среды динамику распространения различных углеводородных загрязнителей и интенсивность их вертикальной и горизонтальной миграции.



а *б*

Рис. 2. Распространение нефтезагрязнения в вертикальном разрезе:
а – суглинок: $m = 0{,}25$, $k = 10^{-15}$ м^2; *б* – песок: $m = 0{,}25$, $k = 10^{-12}$ м^2



а *б* *в* *г*

Рис. 3. Распространение нефтезагрязнения в вертикальном разрезе в различное время T (сут):
а – 3 ч; *б* – 9 ч; *в* – 0,75 ч; *г* – 1,11 ч; *д* – 1,48 ч; *е* – 1,85 ч

Разработан программный комплекс, состоящий из трех модулей. Модуль 1 предназначен для анализа интенсивности загрязнения различного типа почв по глубине. Модуль 2 предназначен для определения характера пространственно-временного распределения зоны загрязнения. Мультимедийный модуль 3 предназначен для оценки области и степени загрязнения почвенных сред при фильтрации различных углеводородных загрязнителей в динамическом режиме.

Предложенная модель применима к однородным верхним слоям почвы и может быть использована для прогноза размеров области распространения углеводородов и оценки степени нефтезагрязнения.

В результате выполнения исследования были получены следующие научные и практические результаты:

1. Для решения актуальной задачи предложена проблемно ориентированная трехмерная модель в физических переменных с учетом гравитационно-капиллярного взаимодействия.

2. Представлен эффективный вычислительный алгоритм, учитывающий специфические особенности конкретной математической задачи и позволяющий существенно уменьшить вычислительные затраты при той же точности, что важно в чрезвычайных ситуациях.

3. Разработан специализированный моделирующий программный комплекс для компьютерной реализации модели при различных начально-граничных условиях.

4. Проведено комплексное исследование проблем экологической безопасности при нефтяном загрязне-

нии с применением современной технологии математического моделирования и вычислительного эксперимента.

Библиографические ссылки

1. Рогозина Е. А. Актуальные вопросы проблемы очистки нефтезагрязненных почв [Электронный ресурс] // Нефтегазовая геология. Теория и практика. 2006. № 1. URL: http://www.ngtp.ru.

2. Гриценко А. И., Акопова Г. С., Максимов В. М. Экология. Нефть и газ. М. : ВНИИЭГазпром, 1995.

3. Молокова, Н. В. Математическое моделирование распространения загрязнения в пористой среде // IV Всесибирский конгресс женщин-математиков : материалы Всерос. конф. Красноярск, 2006. С. 130–138.

4. Молокова, Н. В. Решение геофильтрационных задач средствами математического моделирования // Вестник СибГАУ. Вып. 3 (20). Красноярск, 2008. С. 31–34.

5. Молокова, Н. В. Численное исследование и анализ процесса загрязнения пористой среды // Вестник. Вып. 4 (21). Красноярск, 2008. С. 20–23.

N. V. Molokova

MATHEMATICAL MODELING OF PROCESSES OF OIL POLLUTION OF POROUS ENVIRONMENT

In this article the author presents a two-dimensional model of filtering, taking into account movement of air and hydrocarbon pollutants in porous soil. The model includes a system of partial differential equations with additional conditions. The differential equations include mass balance equation element in a porous medium – inseparability equation, and the differential equations of motion. For circuit system the author puts in the equation of state of pollutant and environment. Initial and boundary conditions correspond to the filtration process, beginning with the ground surface and initial stage of oil products spillage. A comparative analysis of the results of mathematical modeling with experiments is presented.

Keywords: *model of biphasic filtration, geofiltration*, *mathematical modeling, porous environment, oil pollution.*

© Молокова Н. В., 2010

УДК 517.977.1

А. Н. Рогалев

ГАРАНТИРОВАННЫЕ ОЦЕНКИ И ПОСТРОЕНИЕ МНОЖЕСТВ ДОСТИЖИМОСТИ ДЛЯ НЕЛИНЕЙНЫХ УПРАВЛЯЕМЫХ СИСТЕМ

Описывается применение гарантированных методов, позволяющих получать границы всех возможных фазовых состояний системы, учитывающих все воздействия на управляемую систему. Такие количественные оценки называются границами множеств достижимости. Приводятся примеры расчетов включений множеств достижимости.

Ключевые слова: верхние и нижние границы решений, гарантированные оценки множеств достижимости, постоянно действующие возмущения.

Пусть движение состояний системы *y* подчиняется следующему дифференциальному уравнению:

$$\frac{dy}{dt} = f(t, y; u, v), \qquad (1)$$

где t – текущее время; $u = u(t)$ – управляющее воздействие; $v = v(t)$ – неконтролируемое возмущение; $y(t) = \left(y_1(t), ..., y_n(t)\right)$ – вектор фазовых состояний.

Управляемая система общего вида подвержена действию управляющих и неопределенных факторов. Выбор возможных реализаций управляющих воздействий u стеснен ограничениями $u \in Q$, $t \in [t^0, T]$, отражающими особенности рассматриваемой задачи. Для многих задач ограничения на воздействия могут носить только геометрический характер. Это значит, что в каждый момент времени $t \in [t^0, T]$ значение $u(t)$ может быть любым из некоторого выпуклого компактного множества Q. В общем случае это множество может быть описано опорной функцией

$$Q = \left\{u(t) \mid \ q^T u(t) \le \delta(q, t), \|q\| = 1, t \in T\right\}. \qquad (2)$$

Опорная функция в каждой конкретной задаче назначается на основе априорной информации о свойствах возмущений, которые могут встречаться в задаче. При независимых ограничениях на модуль каждой составляющей $|u_j(t)| \le U_j(t)$ множество Q представляет собой параллелепипед с ребрами длины $2U_j$, ориентированными вдоль соответствующих координатных осей в пространстве возмущений. В этом слу-

чае опорная функция $\delta(q,t)$ задается своими значениями в n точках: $\delta(q^j,t)=\nu_j, j=1,2,...,n$. Кроме того, многие управляемые системы невозможно описать с допустимой точностью системой дифференциальных уравнений с точно заданными коэффициентами [1–3]. Это зависит от влияния многих факторов: внешних возмущающих сил, неконтролируемых вариаций параметров, ошибок в определении начальных условий. Во многих случаях вероятностные (стохастические) характеристики не могут эффективно описать поведение управляемой системы [1–3]. Это значит, что ограничения на неконтролируемые возмущения также могут носить только геометрический характер $\nu\in\Delta, t\in[t^0,T]$.

Задача гарантированного оценивания состоит в определении множества (или его границ):

$$Y(t)=\bigcup_{y^0\in Y^0,\,u\in Q,\,\nu\in\Delta} y(t,t^0,y^0,u,\nu). \qquad (3)$$

Среди математических описаний подобных задач, основанных на оценке множеств достижимости, мы выделим задачи проверки гарантированных условий безопасности и задачи построения «выживающих» траекторий.

Множеством достижимости $Y^*(t)$ в момент времени t [1; 2] называется множество всех точек из фазового пространства, в которые можно перейти на отрезке времени $[t^0,T]$ из всех возможных точек начального множества фазовых состояний M^0 по решениям системы (1) с начальным условием $y(t^0)\in M^0$ и с допустимым управлением $u(t)$. Если известно, что в начальный момент времени s всевозможные векторы состояний системы образуют некоторое множество M, то говорят об области достижимости $D(t,s;M)$ системы, исходя из множества M в момент s. Область достижимости, вообще говоря, зависит от реализации неконтролируемых возмущений. Понятие области достижимости предоставляет удобный язык, на котором могут быть описаны различные конкретные задачи теории управления. Кратко перечислим их.

Оценка достижимости. Если известны области достижимости $D(t,s;M)$, то можно легко оценить возможности управления. Рассмотрим задачу о приведении управляемой системы на заданное терминальное множество N в момент времени T. Ясно, что решение этой задачи эквивалентно выяснению того, пересекается ли терминальное многообразие N с множеством достижимости $D(t,s;M)$.

Управляемость. Рассмотрим множество $\Delta(s,M)$ – объединение областей достижимости $D(T,s;M)$ для всех моментов времени $T\ge s$. Если множество $\Delta(s,M)$ совпадает со всем пространством состояний системы, то такая система называется вполне управляемой. Выяснение того, управляема ли данная динамическая система, является классической задачей управления.

Оценка возмущений и практическая устойчивость. Рассмотрим случай неуправляемой системы. Это означает, что в уравнении (1) мы вводим параметр возмущений v, подчиненный геометрическим ограничениям, считаем его управляющим, тем самым полагаем, что возмущения отсутствуют. Тогда области достижимости приобретают смысл тех множеств, куда система может попасть под действием допустимых возмущений. Тем самым области достижимости характеризуют отклонения траекторий под действием возмущений или точность, с которой можно предсказать движение возмущаемой системы.

Задача управления с терминальным функционалом. Требуется найти допустимое управление $u(t)$ и начальный вектор из множества M, такие, что при этом достигается минимум терминального функционала, т. е. функционала вида $J=F(y(t))$, где F – заданная функция; T – заданный момент времени. Ясно, что сформулированная задача эквивалентна задаче минимизации функции F на области достижимости $D(T,s;M)$. Тем самым, знание областей достижимости позволяет свести сложную задачу оптимального управления к сравнительно более простой задаче нелинейного программирования.

Таким образом, полезность областей достижимости становится очевидной.

Существует несколько подходов к разработке методов построения и аппроксимации множеств достижимости. Однако получить надежную оценку множеств достижимости управляемых систем с нелинейной правой частью удается не всегда. Особенно это имеет место для задач с правыми частями, в которые входят управляющие воздействия произвольным образом, не только как аддитивный член. Поэтому описание этого нового подхода для использования в задачах оценки множеств достижимости может прояснить это направление и помочь специалистам в управлении. Используемые в данной работе термины «гарантированные методы», а также «гарантированные границы решений» использовались в задачах численного решения дифференциальных уравнений с неточно заданными данными. Они являются устоявшимися названиями в задачах гарантированного оценивания множеств достижимости, в задачах гарантированного поиска управляемых объектов.

Весьма часто рассматриваются те классы управляемых систем, для которых возможно выписать формулы решений и, следовательно, анализировать их верхние и нижние границы. Был разработан также метод эллипсоидов, суммирующий эллипсоиды неопределенности, включающие все решения управляемой системы ОДУ [2]. При этом в методе эллипсоидов строятся дифференциальные уравнения, которые описывают поведение центров и радиусов этих эллипсоидов [2; 4]. Методика сложения эллипсоидов дополняется решением этих дифференциальных уравнений. Анализ получаемых результатов сильно зависит от

знания свойств управляемых систем, так как переход к дифференциальным уравнениям эволюции эллипсоидов основан на бесконечно малых отклонениях, что не всегда строго соответствует реальным задачам.

Описание метода. В статье строятся включения области достижимости управляемых систем. Эти области определяются с помощью гарантированного метода оценивания множеств решений систем обыкновенных дифференциальных уравнений на основе символьных формул для аппроксимации оператора сдвига вдоль траектории [5–13]. Пусть в системе (1) f – непрерывно дифференцируемая функция по t ($f(t) \in C^1$), множество U компактно R^m, допустимые управления – это измеримые функции $u(t) \in U, t \in [s,T]$. Полагаем следующее: а) существует равномерная оценка $|y(t)| < b$ для всех решений системы (1) на интервале $t \in [s,T]$, где $b = \text{const} > 0$; б) множество $Y(y,t) = f(t,y,U)$ для всех y, $t \in [s,T]$ компактно и выпукло.

Тогда измеримое множество $D(t,s,M)$ управляемой системы (1) является компактным множеством $t \in [s,T]$.

Вообще говоря, возможные величины векторов управлений и возмущений (допустимых управлений и возмущений) могут быть сложным образом связаны с движением динамической системы. Например, часто рассматривают интегральные ограничения на управления или возмущения вида

$$\int_s^T \varphi(t,y,u,v)dt \le Q,$$

где $\varphi(t,y,u,v)$ – заданная функция; s, T – начальный и конечный момент времени движения; Q – имеет смысл предельной величины некоторого ресурса. В данной работе мы будем рассматривать лишь случай геометрических ограничений, имеющих вид

$$u \in U(t,y), v \in V(t,y).$$

Таким образом, возможные значения векторов управлений и возмущений полностью определяются текущим моментом времени и состоянием системы.

Такая ситуация, естественно, приводит к удобной переформулировке закона движения рассматриваемой динамической системы. Рассмотрим множество $F(t,y,v)$ значений вектор-функции f при всевозможных векторах управления $u \in U(t,y)$. Тогда уравнение движения (1) переписывается в виде дифференциального включения

$$\frac{dy}{dt} \in F(t,y,v),$$

где в правую часть помимо времени и фазового состояния входят неконтролируемые возмущения. Общее понятие дифференциального включения [1; 2; 14] определяется заданием для каждого момента времени и состояния системы некоторого множества, которое может также зависеть и от других параметров (в нашем случае – неконтролируемых возмущений). Это множество называется правой частью или индикатрисой дифференциального включения. После этого можно определить решение дифференциального включения как такую траекторию в пространстве состояний (фазовую траекторию), что вектор скорости в каждый момент лежит внутри индикатрисы. Решение дифференциального включения называют также допустимой траекторией. Таким образом, исследование динамических систем рассматриваемого класса сводится к изучению допустимых траекторий дифференциальных включений.

Построение оценок множеств достижимости. Включение множеств достижимости строится с помощью гарантированного метода, основанного на символьной формуле оператора сдвига вдоль траектории и вычислении множеств значений этой формулы по всем множествам управления и управляющим воздействиям [5–13].

Символьная формула (аналитическое выражение) – это запись символьной формы, включающей знаки операций, имена функций и констант, вычисление по которой позволяет получить значение решения.

Пусть K_i – это последовательность нормированных пространств, зависящая от некоторого параметра $u \in U$. Пусть $Y^i, i = 1,...,n-1$ – последовательность символьных формул непрерывных операций F^i, определенных на прямом произведении $K_1 \cdot K_2 \cdots K_n$, отображающих произведение этих пространств в пространство K_{n+1} и устанавливающих зависимость между значениями решений в каждой точке области определения. Тогда результат последовательного исполнения преобразования формул

$$\begin{aligned}
Y^1 &= F^1(t^0,t^1,Y^0,Y^1) = S^1(Y^0),\\
Y^2 &= F^2(t^0,t^1,t^2,Y^0,Y^1,Y^2) = S^2(Y^0) \circ S^1(Y^0),\\
Y^i &= F^i(t^0,\ldots,t^i,Y^0,Y^1,\ldots,Y^i) =\\
&= S^i(Y^0) \circ \ldots \circ S^2(Y^{i-1}) \circ S^1(Y^i), \qquad (4)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
Y^m &= F^m(t^0,\ldots,t^m,Y^0,Y^1,\ldots Y^m) =\\
&= S^m(Y^0) \circ \ldots \circ S^2(Y^{i-1}) \circ S^1(Y^i)
\end{aligned}$$

будет называться символьной формулой метода сдвига вдоль траектории (опорной траекторией).

Приведем простые примеры подобных символьных формул:

$$y^m = y^{m-1} + \varphi(h, y^{m-1}),$$
$$y^m = y^{m-1} + \varphi(h, y^m, y^{m-1}).$$

Поскольку управление – это измеримая функция, то метод рядов Тейлора не применим для вычисления опорной траектории.

Выполнение гарантированного метода состоит из следующих этапов:

1. Начало работы алгоритма – присвоение значений переменным, идентифицирующим систему: размерность, правая часть, начальные данные.

2. Построение композиции символьных формул как векторной функции с символьными компонентами $s(t^k)$, зависящими от символьных начальных данных $y_1^0, ..., y_n^0$. Каждая компонента символьного вектора определяется заново как функция, зависящая от символьных начальных данных values $y_1^0, ..., y_n^0$ в каждой точке сетки t^k. Эта формула описывает сдвиг вдоль траектории приближенного решения, сходящегося к решению системы. Сдвиг вдоль траектории задан построенной символьной формулой решений.

3. Последовательное исполнение метода хранения и переработки символьной информации при продвижении вдоль траектории решений производится на основе статичного хранения этой информации, работы с адресацией памяти с помощью функций поточной обработки.

4. Символьная формула приближенного решения преобразуется к виду, который позволяет эффективно и быстро вычислять оценки областей значений приближенных решений, соответствующие изменениям параметров задачи. Для этого используется кусочно-полиномиальное представление символьных формул и опорные функции для многозначных функций, описывающих области значений.

5. Символьная формула (вектор с символьными компонентами) используется для нахождения экстремальных значений (максимального и минимального) области значений решений.

6. Оценки глобальной ошибки строятся на основе теорем о неподвижной точке в виде алгоритмов гомотопии. Этот алгоритм состоит в поиске «серых» симплексов согласно алгоритму Ивза [15] и их модификаций к многозначным функциям. В алгоритме определяются *2n* функций, являющихся границами множеств точных решений. Построенные для этих функций графики образуют граничные гиперплоскости множеств решений управляемых систем.

7. Нижняя граница вектора глобальной ошибки (вектора глобальной ошибки, состоящего из 2*n* компонент) определяется путем накапливания локальных ошибок вдоль траектории точного решения. Аналогично вычисляется нижняя граница глобальной ошибки в каждой точке t^k. В итоге выполняются два независимых процесса: построение символьных формул решений и накапливание величин локальной ошибки вдоль траектории решения.

8. Граница глобальной ошибки прибавляется к вычисленным на шаге 5 границам вектора решений, происходит объединение этих множеств. Все вычисления числовых значений границ решений отложены на самый последний этап исполнения алгоритма. При вычислениях применяется арифметика с направленными округлениями. При вычислении границ множества решений используется тот факт, что построенные символьные формулы решений – это кусочно-гладкие полиномиальные функции, зависящие от $y_1^0, ..., y_n^0$.

Множество приближенных решений будет выпуклым в силу свойств полиномиальных функций и их вариаций. Это множество будет также компактным. Выполнение преобразований символьных формул – это первый этап гарантированного метода оценивания множеств достижимости. Его можно было бы реализовать, используя одну из многих систем аналитических преобразований. Однако при этом катастрофически быстро растет объем получаемых символьных формул. Это означает, что прямые расчеты по этим формулам неэкономичны, особенно при использовании их в циклах.

Кроме того, компьютерная система символьных преобразований производит вычисления с длинными по разрядности числами, что приводит к большим затратам памяти. Поэтому для выполнения гарантированного метода вначале организуется процесс хранения и преобразования символьных формул, чтобы организовать вычисление границ экономным образом на финальном этапе метода. Символьные формулы будут представлены как рекурсивные структуры, величина которых растет. Для записи такой формулы в компьютере используются линейные динамические структуры.

Преобразование и вычисление по этим формулам выполняется без явного выписывания компонент композиции этих формул на каждом шаге. Передача компонент от шага к шагу выполняется с использованием операций с адресами в памяти, хранящимися в стеке. Генерация кода для вычисления по символьным формулам организуется путем обхода по дереву, начиная с вершин дерева.

Примеры применения. Рассмотрим управляемую систему:

$$\begin{aligned} \frac{dy_1}{dt} &= 2y_2 + y_4, \\ \frac{dy_2}{dt} &= 2y_1 + y_3 + u(t), \\ \frac{dy_3}{dt} &= y_2 + 10y_4, \\ \frac{dy_4}{dt} &= 2y_1 - 2y_3 + u_2(t). \end{aligned} \qquad (5)$$

На возмущение $u(t)$ в системе (5) наложено ограничение $|u| \le 1$. Начальные условия подчинены ограничению $\max_i |y_i(t^0)| \le 1$. В [2], где рассматривается эта система, показано, что объем эллипсоида, аппроксимирующего множество достижимости, растет очень быстро. Для предотвращения такого катастрофического роста оценок в [2] предложено применять непрерывные наблюдения за фазовым состоянием динамической системы и фильтрацию ошибок наблюдений. При этом объем эллипсоида неопределенности становится меньше, чем в методе без наблюдений [2].

Применение метода гарантированного оценивания позволяет получить хорошие результаты оценки множества достижимости без применения дополнительных измерений и фильтрации ошибок измерений (рис. 1, 2).

Рис. 1

Рис. 2

Оценим множество достижимости простейшей модели движения объекта на горизонтальной плоскости, описываемой нелинейной системой третьего порядка:

$$\begin{aligned} \frac{dy_1}{dt} &= V\cos\varphi, \\ \frac{dy_2}{dt} &= V\sin\varphi, \\ \frac{d\varphi}{dt} &= \frac{k}{V}u, |u| \le 1, \\ V &= \text{const}, k = \text{const} > 0, \end{aligned} \tag{6}$$

где y_1, y_2 – координаты объекта, отождествляемого с точкой на плоскости; φ – угол между вектором скорости объекта и осью x; u – управляющий параметр, удовлетворяющий указанному ограничению и характеризующий скорость изменения угла φ; k – максимальное боковое ускорение; V – величина скорости, $\alpha = \frac{k}{V}$. Неравенство (6) ограничивает радиус кривизны траектории объекта, а именно, радиус кривизны не может больше единицы. Рассматриваемая система использовалась Р. Айзексом при постановке задачи «шофер-убийца» [16]. Слова «автомобиль», «пешеход», «шофер-убийца» оказались на редкость удачными, хотя в качестве реальных объектов Р. Айзекс подразумевал управляемую торпеду и увертывающийся от нее небольшой катер. Состояние $z_0 = (y_1(t_0), y_2(t_0), \varphi(t_0))$ объекта в начальный момент времени предполагается заданным. Без ограничения общности имеем $y_1(t_0) = y_2(t_0) = \varphi(t_0) = 0$.

Система (6) функционирует на конечном достаточно большом промежутке времени $T = [0, t^*]$. В качестве множества допустимых управлений выберем множество U всех измеримых по Борелю функций $U : T \to [-1, 1]$. Каждое управление $u \in U$ порождает движение, исходящее из начальной позиции, которое будем обозначать через

$$\left(y_{1_u}, y_{2_u}, \varphi_u\right) = \left(\left(y_{1_u}(t), y_{2_u}(t), \varphi_u(t)\right), t \in T\right).$$

В силу стационарности системы (6) выбор начального момента времени не существенен. Кроме того, специфика системы (6) такова, что начальное состояние влияет на множество достижимости лишь с точностью до поворота и переноса. Множество достижимости $G(T)$ в момент времени $T \ge 0$ есть совокупность всех точек фазового пространства, в каждую из которых возможен перевод системы (6) в момент времени $T \ge 0$ при помощи некоторого допустимого управления на промежутке $T = [0, t^*]$ из начальной точки z_0. Из общих результатов математической теории управления следует, что множество $G(T)$ замкнуто и ограниченно. Совмещая начало относительной системы координат с движущимся объектом и направляя ось y по вектору его скорости, перейдем к системе

$$\begin{aligned} \frac{dy_1}{dt} &= -y_2 u + v_x, \\ \frac{dy_2}{dt} &= y_1 u - 1 + v_y, \\ |u| \le 1, v &= (v_x, v_y)^T, |v| \le \nu. \end{aligned} \tag{7}$$

Для системы (7), используя гарантированный метод, были получены оценки множеств достижимости на моменты времени $t \in [0, 25]$ (рис. 3).

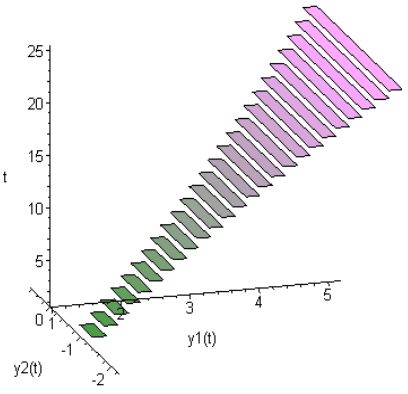


Рис. 3

Вывод, к которому можно прийти, внимательно исследуя проблемы построения гарантированных оценок управляемых систем, описываемых обыкновенными дифференциальными уравнениями, заклю-

чается в необходимости установления связи между начальными значениями и значениями в каждой конкретной точке решения системы. Установление связи достигается на основе аппроксимации оператора сдвига вдоль траектории. Описанные в этой статье гарантированные методы используют такую аппроксимацию, что позволяет вычислять верхние и нижние границы множеств решения (включения) достаточно широкого класса управляемых систем. При этом для гарантированных оценок можно оценить их близость к множеству точных решений.

Библиографические ссылки

1. Куржанский А. Б. Управление и наблюдение в условиях неопределенности. М. : Наука, 1977.

2. Черноусько Ф. Л. Оценивание фазового состояния динамических систем. М. : Наука, 1988.

3. Овсеевич А. И., Шматков А. М. Сравнение вероятностного и гарантированного подходов прогноза фазовых состояний динамических систем // Известия РАН. Теория и системы управления. 1997. № 4. С. 11–16.

4. Черноусько Ф. Л. Эллипсоидальные оценки множеств достижимости линейных систем с недоопределенной матрицей // Прикладная математика и механика. 1996. № 6 (60). С. 940–950.

5. Новиков В. А., Рогалев А. Н. Построение сходящихся верхних и нижних оценок решений систем обыкновенных дифференциальных уравнений // Журнал вычислительной математики. 1993. № 33(2). С. 219–231.

6. Рогалев А. Н. Исследование практической устойчивости при постоянно действующих возмущениях // Вычислительные технологии. 2002. № 7(5). С. 148–150.

7. Рогалев А. Н. Гарантированные методы решения систем обыкновенных дифференциальных уравнений на основе преобразования символьных формул // Вычислительные технологии. 2003. № 8(5). С. 102–116.

8. Рогалев А. Н. Гарантированные оценки безопасного функционирования технических и электроэнергетических систем // Современные методы математического моделирования природных и антропогенных катастроф : тр. Всерос. конф. с междунар. участием. Красноярск : ИВМ СО РАН. 2003. Т. 3. С. 42–48.

9. Рогалев А. Н. Границы множеств решений систем обыкновенных дифференциальных уравнений с интервальными начальными данными // Вычислительные технологии. 2004. 9(1). С. 86–93.

10. Рогалев А. Н. Гарантированные границы решений дифференциальных уравнений // Тихонов и современная математика : Междунар. конф. МГУ им. М. В. Ломоносова ; Рос. акад. наук. М. : ВМиК МГУ им. М. В. Ломоносова. 2006. С. 100–103.

11. Рогалев А. Н. Символьные вычисления в гарантированных методах, выполненные на нескольких процессорах // Вестник НГУ. Сер. Информационные технологии. 2006. № 1 (4). С. 56–62.

12. Рогалев А. Н. Вопросы устойчивости ансамблей дифференциальных уравнений // Вычислительные технологии. 2008. № 3(13). С. 111–117.

13. Rogalyov A. N. Computation of reachable sets guaranteed bounds // Control, Diagnostics, and Automation (ASIT – CDA 2010) : Proc. of the IASTED Inern. Conf. on Automation, Control, and Information Technology / ACTA Press. B6. Calgary. Alberta T3E 709. Canada, 2010. P. 132–139.

14. Булгаков Б. В. Накопление возмущений линейных осциллирующих систем // Доклады Академии наук. 1946. № 51. С. 339–342.

15. Eaves R. Saigal R. Homotopies for computation of fixed points on unbounded regions // Mathematical Program. 1972. № 3(2). P. 225–237.

16. Айзекс Р. Дифференциальные игры. М. : Мир, 1967.

A. N. Rogalyov

GUARANTEED ESTIMATIONS AND CONSTRUCTION OF SETS OF ATTAINABILITY FOR NONLINEAR CONTROLLABLE SYSTEMS

In the article the author describes application of the guaranteed methods, allowing to receive bounds of all possible phase state conditions considering all influences on controllable system. Such quantitative estimations are called bounds of sets of attainability. Examples of computations of sets of attainability inclusions are given.

Keywords: the upper and lower bounds of the solutions, the guaranteed estimations of sets of attainability, constantly operating perturbations.

© Рогалев А. Н., 2010

УДК 004.42:004.056

Л. К. Птицына, А. В. Птицын

ПРЕОДОЛЕНИЕ НЕОПРЕДЕЛЕННОСТИ ОТНОСИТЕЛЬНО ДИНАМИЧЕСКИХ ПРОФИЛЕЙ КОМПЛЕКСНЫХ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ

Выделяются типовые приемы комплексирования программно-аппаратных средств в комплексных системах защиты информации, формируется математический базис для определения их динамических характеристик.

Ключевые слова: комплексные системы защиты информации, принципы параллельности, профили, динамические характеристики, точность оценивания, функция распределения.

Применение принципа мультиплексирования при организации комплексных систем защиты информации предоставляет широкие возможности совершенствования архитектуры с целью повышения эффективности их функционирования. Преимущества мультиплексирования могут раскрыться в наибольшей степени, если параллельная работа программных и аппаратных средств обеспечения информационной безопасности организуется не только при обнаружении возможных угроз, но и при проведении необходимых мероприятий по защите информации. Планирование эффекта от подобных действий становится возможным в случае преодоления неопределенности относительно динамических профилей комплексных систем защиты информации в условиях параллельной работы средств. В связи с этим появляется объективная необходимость определения динамических характеристик систем защиты информации при комплексировании программно-аппаратных средств. Распространенные в настоящее время приемы преодоления указанной неопределенности сводятся к оцениванию динамических характеристик в процессе функционирования комплексных систем защиты информации, когда изменения в их конфигурации ограничены возможностями выбранной архитектуры. В то же время, неопределенность игнорируется при выборе их архитектуры. Для устранения подобной проблемной ситуации предлагается формальный подход к определению динамических характеристик комплексных систем защиты информации, который опирается на математический аппарат анализа распределенных процессов в части моделирования механизмов их синхронизации [1].

Ориентируясь на практическую значимость, в качестве динамических характеристик выбираются средние времена защиты информации.

Возможные способы комплексирования отдельных средств обеспечения защищенности данных основываются на различных сочетаниях типовых вариантов их согласованного функционирования:

– параллельное функционирование аппаратных средств;

– параллельное функционирование программных средств;

– параллельное функционирование программных и аппаратных средств.

Применительно к перечисленным типовым вариантам осуществляется определение выбранных показателей качества защиты информации.

Среди известных принципов объединения результатов параллельного функционирования отдельных средств широко распространены принципы И-параллельности и ИЛИ-параллельности, выбор каждого из которых обусловливается спецификой функционального назначения комплексируемых средств. По этой причине при определении среднего времени защиты информации учитывается характер процедур объединения частных результатов, описываемый с помощью логических функций «И» ($\wedge$), «ИЛИ» ($\vee$).

Выполним определение искомых показателей на случай параллельного функционирования аппаратных средств. Пусть каждое *n*-е аппаратное средство обеспечивает такое выполнение возлагаемых задач, при котором среднее время защиты информации характеризуется величиной $T_{n,P}$. Принимая гипотезу об экспоненциальном распределении времени защиты, находим вероятность того, что все необходимые мероприятия будут реализованы в промежутке времени $(0, t_k)$:

$$W_n(t_k) = 1 - e^{-t_k / T_{n,P}}. \qquad (1)$$

После нахождения функции распределения времени окончания *N* параллельно выполняемых функциональных процессов в условиях применения принципа И-параллельности, получим следующее соотношение для среднего времени защиты информации T_P:

$$T_P \approx \Delta t \sum_{k=1}^{K} k \left[\prod_{n=1}^{N} W_n(t_k) - \prod_{n=1}^{N} W_n(t_{k-1}) \right], \qquad (2)$$

$$\Delta t = t_k - t_{k-1}, \qquad k = 1, 2, ..., K,$$

где Δt – интервал обновления результатов работы аппаратных средств защиты информации.

В соотношении (2) переменная *K* представляет наименьшее значение, удовлетворяющее неравенству

$$1 - \sum_{k=1}^{K} \left[\prod_{n=1}^{N} W_n(t_k) - \prod_{n=1}^{N} W_n(t_{k-1}) \right] \le \varepsilon, \qquad (3)$$

где *N* – число параллельно работающих аппаратных средств.

В соотношении (3) величина ε представляет собой вероятность события, при котором время защиты информации превышает граничное значение, равное $K\Delta t$.

Через функцию распределения времени окончания N параллельно выполняемых функциональных процессов в условиях применения принципа ИЛИ-параллельности находим среднее время защиты информации для второго типового случая:

$$T_P \approx \Delta t \sum_{k=1}^{K} k \left[\prod_{n=1}^{N} \left(1 - W_n(t_{k-1})\right) - \prod_{n=1}^{N} \left(1 - W_n(t_k)\right) \right], \quad (4)$$

где K – наименьшее целое значение, удовлетворяющее неравенству

$$1 - \sum_{k=1}^{K} \left[\prod_{n=1}^{N} \left(1 - W_n(t_{k-1})\right) - \prod_{n=1}^{N} \left(1 - W_n(t_k)\right) \right] \le \varepsilon. \quad (5)$$

Проанализируем случай параллельного действия программных средств защиты информации. В соответствии с характерными особенностями функционирования программных средств, раскрытыми в [2], n-й подпроцесс защиты информации, включающий M этапов, описывается полумарковской моделью, позволяющей получить выражение для определения среднего времени защиты информации:

$$T_{n,P} = \frac{\left[\prod_{m=1}^{M-1} (1 - \mathrm{P}_{n,m}) \right] \sum_{m=1}^{M} t_{n,m}}{1 - \prod_{m=1}^{M} (1 - \mathrm{P}_{n,m})} + \frac{t_{n,1} \left[1 - \prod_{m=1}^{M-1} (1 - \mathrm{P}_{n,m}) \right] + \sum_{m=2}^{M-1} \left\{ t_{n,m} \left[\prod_{j=1}^{m-1} (1 - \mathrm{P}_{n,j}) \left(1 - \prod_{K=m}^{M-1} (1 - \mathrm{P}_{n,\kappa}) \right) \right] \right\}}{1 - \prod_{m=1}^{M} (1 - \mathrm{P}_{n,m})} + \frac{\sum_{m=1}^{M} \mathrm{P}_{n,m} t_{\acute{K},n,m} \prod_{k=1}^{m-1} (1 - \mathrm{P}_{n,k})}{1 - \prod_{m=1}^{M} (1 - \mathrm{P}_{n,m})}. \quad (6)$$

В указанной модели случайные длительности отдельных этапов контроля информационных процессов при функционировании n-го программного средства защиты представляются математическими ожиданиями $t_{n,m}$, $m = \overline{1, M}$. Причем на каждом из этапов предусматривается возможное обнаружение появляющейся угрозы с вероятностью $P_{n,m}$, $m = \overline{1, M}$.

После подстановки соотношения (6) в формулу (1) математическое ожидание времени защиты информации в комплексной системе определяется по формулам (2), (3) в условиях И-параллельности и по формулам (4), (5) в случае ИЛИ-параллельности.

Рассмотрим порядок определения динамических характеристик систем защиты информации при параллельной работе программно-аппаратных средств. По найденным функциям распределения $W_n(t_k)$, $n = 1, 2, ..., N$ на основании соотношений (2), (3) образуется следующее представление искомой оценки при реализации принципа И-параллельности:

$$T_P \approx \Delta t \sum_{k=1}^{K} k \left[\left(\prod_{n=1}^{N_S} W_n(t_k) \right) \left(\prod_{n=1}^{N_H} W_n(t_k) \right) - \left(\prod_{n=1}^{N_S} W_n(t_{k-1}) \right) \left(\prod_{n=1}^{N_H} W_n(t_{k-1}) \right) \right], \quad (7)$$

$$1 - \sum_{k=1}^{K} \left[\left(\prod_{n=1}^{N_S} W_n(t_k) \right) \left(\prod_{n=1}^{N_H} W_n(t_k) \right) - \left(\prod_{n=1}^{N_S} W_n(t_{k-1}) \right) \left(\prod_{n=1}^{N_H} W_n(t_{k-1}) \right) \right] \le \varepsilon, \quad (8)$$

где N_S – число программных средств; N_H – число аппаратных средств; $N = N_S + N_H$ – общее число комплексируемых средств; $W_n(t_k)$, $n = 1, 2, ..., N_S$ – функции распределения, характеризующие качество защиты информации с применением программных средств; $W_n(t_k)$, $n = 1, 2, ..., N_H$ – функции распределения, представляющие качество защиты информации при использовании аппаратных средств.

Неравенство (8) предназначается для выбора значения K по заданному уровню ε.

В соответствии с формулами (4), (5) среднее время защиты информации в случае параллельного функционирования программных и аппаратных средств, результаты работы которых объединяются с применением логической функции «ИЛИ», выражается следующим образом:

$$T_0 \approx \Delta t \sum_{k=1}^{K} k \left[\left(\prod_{n=1}^{N_S} \left(1 - W_n(t_{k-1})\right) \right) \left(\prod_{n=1}^{N_H} \left(1 - W_n(t_{k-1})\right) \right) - \left(\prod_{n=1}^{N_S} \left(1 - W_n(t_k)\right) \right) \left(\prod_{n=1}^{N_H} \left(1 - W_n(t_k)\right) \right) \right], \quad (9)$$

$$1 - \sum_{k=1}^{K} \left[\left(\prod_{n=1}^{N_S} \left(1 - W_n(t_{k-1})\right) \right) \left(\prod_{n=1}^{N_H} \left(1 - W_n(t_{k-1})\right) \right) - \left(\prod_{n=1}^{N_S} \left(1 - W_n(t_k)\right) \right) \left(\prod_{n=1}^{N_H} \left(1 - W_n(t_k)\right) \right) \right] \le \varepsilon. \quad (10)$$

Верхняя граница для переменной k представляет собой наименьшее значение K, которое удовлетворяет неравенству (10). В данном случае задаваемая величина ε является вероятностью того, что время защиты информации при комплексировании программно-аппаратных средств по принципу ИЛИ-параллельности превышает верхнюю границу $K\Delta t$.

Благодаря использованию неравенств (3), (5), (8), (10) обеспечивается возможность регулирования степени приближения оценок (2), (4), (7), (9) к реальным значениям математических ожиданий времен защиты информации.

Выражения (1)–(10) образуют базис соотношений для получения приближенной количественной оценки показателей качества систем защиты информации при комплексировании программно-аппаратных средств по принципу И- и ИЛИ-параллельности с регулируемой степенью приближения.

Таким образом, формализация процесса преодоления неопределенности относительно динамических профилей комплексных систем защиты информации, основанная на системе выведенных аналитических соотношений, обеспечивает планирование их качества при выборе архитектуры.

Библиографические ссылки

1. Птицына Л. К., Соколова Н. В. Программное обеспечение компьютерных сетей. Моделирование механизмов синхронизации параллельных вычислительных процессов в системах мониторинга и управления : учеб. пособие. СПб. : Изд-во Политехн. ун-та, 2010.

2. Птицына Л. К., Птицын А. В. Архитектура ЭВМ и систем. Модели и методы анализа динамических характеристик программных систем защиты информации : учеб. пособие. СПб. : Изд-во Политехн. ун-та, 2007.

L. K. Ptitsyna, A. V. Ptitsyn

OVERCOMING OF UNCERTAINTY OF DYNAMIC PROFILES OF COMPLEX SYSTEMS OF PROTECTION OF THE INFORMATION

Typical modes of hardware-software means complexing in complex systems of protection of the information are allocated, the mathematical basis for definition of their dynamic characteristics is formed.

Keywords: complex systems of protection of the information, parallelism principles, profiles, dynamic characteristics, accuracy of estimation, distribution function.

© Птицына Л. К., Птицын А. В., 2010

УДК 681.3.004.8

В. А. Филимонов

УЧЕБНО-ИССЛЕДОВАТЕЛЬСКИЙ СИТУАЦИОННЫЙ ЦЕНТР – ПОЛИГОН ДЛЯ КОМАНДЫ СИСТЕМНЫХ АНАЛИТИКОВ

Рассматриваются учебно-исследовательские ситуационные центры как инфраструктура для реализации процессов коллективного исследования, проектирования и обучения. Предложены варианты построения прототипов, схема «4 уровня» для рассмотрения объектов. Описана технология подготовки сервисных команд для ситуационных центров. Приведен обзор некоторых проектов, реализованных в Омске в 2005–2010 гг.

Ключевые слова: ситуационный центр, сервисная команда, рефлексивный анализ.

Рассматриваемой проблемой является отсутствие теории подготовки команд системных аналитиков и соответствующих систем подготовки таких команд. Объектом исследований являются системы коллективной деятельности: исследовательской, проектной и учебной.

Комплекс задач и подходы к решению. Указанный выше коллектив будем далее называть проектной группой. Нами сформулирован следующий комплекс задач [1]. Проектная группа решает *задачу 1*: создает проект, которым, в частности, может быть представление (теория, модель и т. п.) о некотором объекте исследования. Предполагая, что работу проектной группы обеспечивают технические средства и сервисная команда, приходим к *задаче 2*: создание (виртуальной) технологии оптимальной поддержки всего жизненного цикла постановки и решения задачи 1. Далее возникает *задача 3* – создание инфраструктуры (машины, комплекса), в которой формируются технологии, указанные в задаче 2. Создание технологии требует постановки *задачи 4* – подготовки специалистов по решению перечисленных задач (как учить) и *задачи 5* – создания собственной технологии для решения задачи 4 (как учить учителей).

В качестве инфраструктуры нами используются учебно-исследовательские ситуационные центры (СЦ). Сам термин «ситуационный центр» уже является достаточно распространенным. В частности, в поисковых машинах на соответствующий запрос выдается от 60 000 до 600 000 ссылок. Мы будем понимать СЦ как пространство, предназначенное для динамического коллективного формирования образа ситуации, объекта, процесса, обеспеченное ключевыми, т. е. критическими относительно решаемой задачи, ресурсами [2]. Основные отличия учебного аспекта СЦ от исследовательского заключаются в следующем:

– преимущественное внимание уделяется изучению методов, а не рассмотрению информации, относящейся к определенной задаче;

– для рассматриваемых задач существуют решения, заведомо квалифицированные как правильные;

– одной из основных задач, решаемых в учебно-исследовательском центре, является анализ и оценка компетентности, а также приращения компетентности индивидуально для каждого участника учебно-проектной группы;

– качество проектов, разрабатываемых учебно-проектными группами, является вторичным показателем относительно роста индивидуального уровня квалификации.

Основное отличие учебно-исследовательского СЦ от учебной аудитории, оснащенной аналогичным оборудованием, состоит в применении компонентов технологии СЦ, в частности, относящихся к компетенции сервисной команды (планшетист, методолог, игротехник), для разработки достаточно долгосрочного учебного проекта. Иными словами, учебная аудитория может быть превращена в СЦ за счет реализации соответствующего регламента работы. Наоборот, при отсутствии такого регламента СЦ фактически превращается в конференц-зал, студию для ток-шоу и другие объекты.

Системные инструменты. Опишем два инструмента, используемые нами дополнительно к известным. Для детализации анализа мы применим схему «4 уровня», позволяющую структурировать различные описания объектов (процессов). Термин «форма» используется в широком смысле как внешнее проявление функций, отношений, связей и законов (геометрическая форма, форма описания и т. п.). Термин «функция» использован для обозначения всего набора функций, которые может выполнять объект в рамках интересующей нас задачи. Термином «фундамент» обозначается все то, что делает принципиально возможным выполнение функций. В большинстве случаев это требует указания связей, отношений и законов. Такое рассмотрение позволяет понять, какие деформации (нарушения формы в использованном смысле) лишают объект возможности исполнить определенные функции. Также можно предварительно оценить, какие функции являются критическими в том смысле, что их отсутствие лишает объект возможности соответствовать своему имени. Пример для термина «ситуационный центр» приведен в таблице.

Процесс анализа направлен от имен к фундаменту. Появление новых возможностей (новых физических эффектов, новых функций) позволяет переходить от анализа к конструированию новых объектов (от фундамента – к новым функциям, формам и именам). Есть и более простые случаи конструирования, например, использование нового имени для существующего объекта, а также объединение нескольких объектов в один, выполняющий функции каждого из составляющих.

В качестве общего подхода мы руководствуемся так называемой промежуточной технологией (intermediate technology). Под этим термином, введенным Э. Ф. Шумахером, обычно понимают методы производства, которые «избегают новой технологии…, но являются усовершенствованием местных методов». Мы используем этот подход как универсальный системный принцип, в соответствии с которым могут быть сформированы прототипы моделей, проектов и тому подобных структур для различных объектов и процессов. Для задач быстрого прототипирования формируется «экран»-прототип (иначе – рабочая модель) в соответствии с принципом: «наиболее простой вариант целого, содержащий наиболее сложный компонент». Используемое нами понятие простоты/сложности ориентировано на две характеристики формирования объекта в заданном базисе:

– количество элементов (операций и операндов);

– стоимость и дефицитность требуемых ресурсов (в частности времени, квалификации персонала, ограничений и т. п.).

Здесь следует упомянуть следствие из распределения Парето: «20 % ключевых ресурсов обеспечивают 80 % результата».

История применения и проекты. На основе учебно-исследовательских СЦ как инфраструктуры строятся различные варианты информационных технологий, ориентированных на решение перечисленных выше задач. Началом использования технологий учебно-исследовательских СЦ в вузах Омска можно считать 2001 г., когда в Омском государственном институте сервиса (ОГИС) был сформирован первый набор студентов, обучающихся по специальности «Прикладная информатика (в сфере сервиса)». На основе докторской диссертации автора «Методология и технология компьютерной поддержки работы коллектива экспертов» (1999 г.) был спроектирован прототип учебного курса подготовки сервисных команд для СЦ. За прошедшие 10 лет фрагменты курса были прочитаны также в Омском государственном университете и Омской государственной медицинской академии (ОмГМА).

Анализ термина «ситуационный центр»

Уровни описания	Примеры описания объекта на заданных уровнях
Имена	Ситуационный центр (ситуационная комната, зал, интеллектуальный кабинет, визионариум и т. д.)
Формы	Пространство: помещения, автомобили, самолеты и т. п. Оборудование: полиэкраны, компьютеры, телекоммуникация, интеллектуальный интерфейс и т. п. Программное обеспечение: контент-анализ, статистический анализ, управление проектами и т. п. Регламент работы: коллективное обсуждение, анонимный опрос
Функции	Представление большого объема существенно разнородной информации для понимания сложной ситуации в целом и принятия решений, исследование последствий для вариантов принятых решений и т. п.
Фундамент	Мобилизация сенсорных и интеллектуальных возможностей человека, наличие методов глубокой переработки информации (Data Mining), когнитивный формат представления информации

Использованные подходы и полученные результаты описаны в [2–6]. Перечислим наиболее интересные проекты.

1. Рефлексивный анализ. Рефлексивные процессы для коллективных систем являются ключевыми. В связи с этим мы активно используем разработанный В. А. Лефевром формальный аппарат рефлексивного анализа, последняя версия которого представлена в [7]. По инициативе В. А. Лефевра было написано учебное пособие по математике для старшеклассников [8]. Дисциплина «Рефлексивный анализ» включена в учебный план подготовки студентов ОГИС по специальности «Прикладная информатика (в сфере сервиса)» и аспирантов специальности 05.13.01 «Системный анализ, управление и обработка информации». Подготовлено учебное пособие [9].

2. Рефлексивный театр ситуационного центра. Формальный аппарат рефлексивного анализа В. А. Лефевра был дополнен нами «правополушарными» компонентами [3]. Опишем способ подготовки материала для реализации предложенным способом. Схема, которую должны усвоить студенты, представляется им в нескольких сценах (вариантах ситуации). Важно, что при сохранении схемы (т. е. структуры связей и отношений действующих лиц) может изменяться реквизит, персонажи, актеры. Комментатор находится вне пространства как актеров, так и зрителей. Его комментарий может быть провокационным, противоречащим ситуации и схеме. Считается, что процесс обучения завершен успешно, если студенты в состоянии реконструировать схему, являющуюся инвариантом представленных ситуаций, и реализовать свой вариант ситуации, соответствующий данной схеме. Такая реализация процесса может быть конструктивным дополнением других способов организации коллективной работы и сама может быть дополнена аналитикой.

3. Конференции. Начиная с 2007 г. проведены три конференции «Рефлексивный театр ситуационного центра». В рамках конференций работали проектные группы. В 2009 г. работа проектной группы была организована в режиме телемоста «Омск – Владивосток». Сервисная команда в Омске, в составе которой были доцент О. М. Куликова и профессор В. С. Чернявская, организовала работу студентов-дизайнеров ОГИС и Владивостокского государственного института сервиса (ВГУЭС). Студенты выступили в роли экспертов, работающих методом анализа иерархий. Результатом явилась модель формирования компетенции дизайнера. По результатм совместных исследований в 2010 г. в ОГИС была издана коллективная монография авторов из Абакана, Владивостока, Москвы, Новосибирска и Омска [2].

4. Лавина. Одним из примеров стал проект «Лавина» – игровое моделирование чрезвычайной ситуации в экспериментальном учебно-исследовательском СЦ Омского филиала Института математики СО РАН. Мероприятие было разработано для студентов ОмГМА и реализовано в два этапа. На первом этапе студенты ОГИС, осваивающие технологии СЦ в рамках специальности «Прикладная информатика (в сфере сервиса)», сформировали сценарий ликвидации последствий схода лавины, сыграли роли в соответствии со сценарием (МЧС, медицинская служба и т. п.), записали на видео, а затем смонтировали материалы в виде учебно-игрового фильма. После соответствующего анализа материалы как прототип были показаны двум группам студентов-медиков, которые создали и проиграли в СЦ свои версии развития чрезвычайной ситуации.

5. Планетонавтика. В 2010 г. начат проект «Планетонавтика–Омск» – фрагмент проекта «Планетонавтика» Н. Ф. Сайфуллина (http://ru.vlab.wikia.com/wiki/Планетонавтика). Одна из особенностей этого проекта состоит в том, что проектная группа является сетевой, широкомасштабной и динамической. В рамках данного проекта О. М. Куликова приняла участие в проектной игре «Автономное подводное поселение-шельф», проводимой под руководством Н. Ф. Сайфуллина командой студентов Казанского государственного технического университета (группа профессора Р. Т. Сиразетдинова) и дистанционно – командой студентов ВГУЭС (группа профессора В. С. Чернявской).

В настоящее время идет подготовка конференции «Рефлексивный театр ситуационного центра – 2010», на которой будут подведены итоги работы и определены перспективы дальнейших исследований.

В презентации к докладу, включающей фрагменты видеоматериалов, приведены примеры реализации предложенных подходов на материале проектов, проводимых с нашим участием в Омске в 2005–2010 гг. Подготовлен компакт-диск с материалами по данной тематике, в частности, с электронными вариантами авторских монографий и других полезных публикаций. Возможна организация работы проектной группы.

Таким образом, предложена технология организации коллективных процессов исследования, проектирования и обучения, ориентированная на использование проектными группами учебно-исследовательских ситуационных центров, описана технология подготовки сервисных команд для таких ситуационных центров.

Библиографические ссылки

1. Филимонов В. А. Исследовательский комплекс «Ген-Гуру» (эскиз многодисциплинарного проекта) // Знания–Онтологии–Теории (ЗОНТ-07) : материалы Всерос. конф. с междунар. участием (14–16 сент. 2007, г. Новосибирск). Новосибирск : Ин-т математики им. С. Л. Соболева СО РАН. Т. 1. С. 24–31.

2. Информационные технологии для ситуационных центров [Электронный ресурс] / О. С. Анисимов [и др.]. Омск : Изд-во ОГИС, 2010. URL: http://www.ofim.oscsbras.ru/~filimono (дата обращения 30.06.2010).

3. Рефлексивный театр ситуационного центра (РТСЦ-2007) : Всерос. конф. с междунар. участием

[Электронный ресурс]. Омск : Омс. гос. ин-т сервиса. 2007. URL: http://www.ofim.oscsbras.ru/~filimono (дата обращения 30.06.2010).

4. Филимонов В. А. Технологии ситуационного центра для социальной инженерии // Проблемы управления в социальных системах. 2009. Т. 1. Вып. 2. С. 63–74.

5. Филимонов В. А. «Спинной мозг» ситуационного центра // Ситуационные центры и перспективные информационно-аналитические средства поддержки принятия решений : сб. материалов Междунар. науч.-практ. конф. (7–9 апреля 2008 г.). М. : Рос. акад. гос. службы при Президенте РФ, 2009. С. 361–366.

6. Дедюлина Н. В., Лучко О. Н., Филимонов В. А. Взаимодействие учебно-исследовательских центров: омский вариант // Ситуационные центры и перспективные информационно-аналитические средства поддержки принятия решений : сб. материалов Междунар. науч.-практ. конф. (7–9 апреля 2008 г.). М. : Рос. акад. гос. службы при Президенте РФ, 2009. С. 359–362.

7. Лефевр В. А. Лекции по теории рефлексивных игр. М. : Когито-центр, 2003.

8. Филимонов В. А. Алгебра логики и совести. Омск : Омский гос. ин-т сервиса, 2006.

9. Филимонов В. А. Экран-сервис технологии (Винтсервинг: Винт 3) : метод. указания к практ. занятиям. Омск : ОГИС, 2005.

V. A. Filimonov

TRAINING AND RESEARCH SITUATIONAL CENTER AS A RANGE AND A PROVING GROUND FOR A TEAM OF SYSTEM ANALYSTS

Training and research situational centers are considered as an infrastructure for realization of processes of collective research, design and training. Versions of construction of prototypes, the diagram « 4 levels » for consideration of objects are offered. Technology of training of service teams for the situational centers is described. Review of some projects 2005–2010 is made.

Keywords: a situational center, service team, reflexive analysis.

© Филимонов В. А., 2010

УДК 681.3.004.8

В. В. Сугоняк, В. А. Филимонов

УПРАВЛЕНИЕ ЧЕЛОВЕЧЕСКИМИ РЕСУРСАМИ ДЛЯ КОЛЛЕКТИВНОГО ИССЛЕДОВАНИЯ И ОБУЧЕНИЯ

Рассматриваются опыт и перспективы применения системного подхода к управлению человеческими ресурсами. Результаты эксперимента по системному проектированию системы обучения в организации дополнены предложениями по применению других подходов, связанных с особенностями человека как биологического и социального субъекта. Инструментами рассмотрения являются синергетика, рефлексивный театр и тайцзы как пример подхода, основанного на восточной философии.

Ключевые слова: системное проектирование, синергетика, рефлексивный театр.

В данной статье представлены начальные шаги по проектированию «машины обучения» [1] в коммерческом банке, предпринятые командой разработчиков. Взяв на вооружение методологию прикладного системного анализа [2], мы предприняли попытку пройти цикл системного исследования и сконструировать улучшающее вмешательство, направленное на развитие человеческих ресурсов банка [3].

Системное проектирование технологий обучения с позиции развития человеческих ресурсов предприятия. Нами была запланирована последовательность прохождения системного исследования, которая предложена в [2]. Работа началась с первоначальной формулировки проблемы: целостность, связность информации отсутствует; неоднократные попытки построения системы обучения успехом не увенчались. Перейдя к диагностике проблемы, разработчики приняли вариант вмешательства в сложившееся положение дел. Составление списка заинтересованных участников (стейкхолдеров) проблемной ситуации вызвало ожидаемые трудности. В связи со значительной объемностью процесса обучения персонала предприятия был создан прототип – обучение специалистов РКО головного офиса и филиальной сети. Следующий этап – выявление «проблемного месива». Группой был сформирован первоначальный перечень отношений к проблеме и интересов стейкхолдеров, исходя из общих соображений. Этап формирования конфигуратора предполагает создание модели ситуации, посредством которой можно будет испытывать и сравнивать варианты решений. Было сделано описание моделей, которые использовались разработчиками при обсуждении программы эксперимента. Следующим шагом должно было бы стать

связывание моделей посредством конфигуратора и построение формализуемого описания ситуации для формирования критериев. Но эта работа была отложена до обсуждения итогов первого эксперимента. Этапы целевыявления и определения критериев были пропущены в силу явной недостаточности информации для качественного проведения этих работ. Было принято решение провести в рамках проекта-прототипа обучение выбранной группы сотрудников по сокращенной программе.

Проведенный эксперимент был направлен на достижение двух целей:

– выявить и зафиксировать текущие способы, методы в обучении, компетенции обучающих, объем и качество транслируемой информации;

– собрать недостающую информацию для проведения системного анализа.

Реализация экспериментального исследования заняла около двух месяцев и позволила выяснить следующее:

1) запуск обучения даже в «пилотном» варианте мобилизующе подействовал на сотрудников, участвовавших в эксперименте, вызвал интерес и отклик среди сотрудников, напрямую не задействованных в этом; была заложена база двустороннего взаимодействия;

2) в ходе эксперимента стало ясно, что создание интегрированной, встроенной в структуру банка, органичной системы обучения – задача непростая, но решаемая.

Применение прикладного системного анализа дает возможность создания адекватных моделей, выступает языком межпрофессионального, межпредметного взаимодействия. Однако это не исчерпывает всех возможностей управления человеческими ресурсами.

Синергетика для управления человеческими ресурсами. Синергетический подход [4] позволяет согласовать частнотеоретические и полидисциплинарные представления различных научных направлений (управленческих наук, социологии, психологии, информатики и др.) через теорию самоорганизации и мягкой редукции в иерархии уровней, через коммуникацию и неустойчивость развивающихся систем. При этом выделяется два типа сложности: детальная (множественность объектов и взаимосвязей между ними) и динамическая (разнесенность во времени причины и следствия воздействий). Сложность системы выражается в многообразии функций (подбор, расстановка, адаптация, развитие т. д.), ее целевых объектов (результативность труда, трудовое поведение, отношение к труду), множественности «петель» положительной (усиливающей) и отрицательной (стабилизирующей) обратной связи, что приводит к циклически развивающимся нелинейным внутри- и межгрупповым взаимодействиям в коллективе.

В динамично меняющейся социальной обстановке управление должно носить опережающий, превентивный характер, что предполагает передачу управленческих функций, в том числе и стратегических, из административного центра на уровень структурных подразделений в целях повышения оперативности управленческих решений. Возникает так называемая обогащенная внутренняя среда с поливариантным выбором, живущая по законам самоорганизации и креативного динамического хаоса, порождающего новые цели (смыслы), ценности и творческие импульсы. Задачу управления можно представить как комплекс задач двух уровней: тактических (управление спонтанно сложившимся многообразием) и стратегических (управление самим процессом возникновения многообразия).

Учет коллективных взаимодействий приводит к нелинейным откликам на административные воздействия: сетевые коммуникации отдельных сотрудников и групп создают собственные иерархические уровни, живущие параллельно с административными центрами и дополняющие их, которые, зачастую, более оперативные, чем последние.

Условием развития самоорганизации, творческой инициативы, создания инновационной среды является открытость системы. Возникновение нового качества системы по горизонтали, т. е. на одном уровне, когда медленное изменение управляющих параметров мегауровня приводит к бифуркации, неустойчивости системы на макроуровне и перестройке ее структуры. Это указывает на инновационные механизмы и явления смены доминант, рождение коллективных инициатив, сплоченности сотрудников.

Нам видятся следующие формы реализации методологии синергетики для управления человеческими ресурсами [5; 6]. Контексты выделены, исходя из частоты встречаемости их в практике.

Одна из форм – описание ограничений и определение диапазонов применимости (релевантности) традиционных равновесных поведенческих моделей. Исследования Н. Д. Кондратьева показывают, что недооценка иррациональной природы человека сводит на нет результативность методов оценки и предсказания социальных процессов. Занимающийся с 1975 г. исследованиями в этой области Ч. Маркетти, оперируя статистическими данными, наглядно демонстрирует, что природа человеческой деятельности, в основном, иррациональна [7].

Следует упомянуть о теме управляемых катастроф как управленческого инструмента. Если воспользоваться медицинской аналогией, то прививка провоцирует болезнь в легкой форме, при этом вырабатывает иммунитет, позволяющий организму выжить в столкновении с реальной угрозой. Осмысление многообразия и эффективности парадоксального инструментария способно принести значимые плоды (пример тому – множество книг по управлению в нашей стране и за рубежом, посвященных стратагемному, нелинейному подходу).

Практическая роль синергетики проявляется и в том, что она позволяет ослабить психологические барьеры, страх перед сложными, масштабными социообразованиями. Утверждение о фрактальности, самоподобности реальности позволяет использовать простые, ограниченные в числе и объеме фундаментальные модели и закономерности.

Известно, что в некоторых случаях кооперативность (теснота связи, когерентность) элементов в сложных системах вблизи точек бифуркации достигает невероятных значений. Это так называемые «режимы с обострением» (режимы сверхбыстрого развития), которые описывают происходящее внутри нелинейных образований. Создание (или предотвращение) условий для лавинообразного роста входит в арсенал мастеров управления, среди которых упомянем В. К. Тарасова [8]. Выявление паттернов мастерства подобного рода, эксплицитное научение этому специалистов-практиков даст значительный эффект.

Существование аттракторов наводит на мысль о том, что управление (как процесс) может иметь несплошной (прерывистый) характер. Малые воздействия в критических моментах (бифуркациях) могут определить желаемую траекторию эволюции системы. Аналогичным образом действуют гомеопатические методы в медицине. Выделим два инструментальных направления:

1. Использование механизма дискретных критических воздействий в текущем управлении социосистемой. Используя синергетический аппарат, можно выделить текущую траекторию развития системы, вектор устремленности, локализовать окрестности точки бифуркации. Определившись с критической точкой, можно оценить поле выборов, определиться с предпочтениями и сконструировать точечное воздействие, ведущее к желаемому исходу. Управленец при этом не вмешивается в функционирование системы, а зафиксировав точку бифуркации, проводит разовое воздействие.

2. Осмысленное конструирование целей-аттракторов, определяющих динамику системы (создание тунеля желаемого развития). Это стратегический аспект конструирования реальности.

Развитие системы определяется не столько прошлым, историей, традициями системы, сколько будущим, структурами-аттракторами эволюции. Данное положение позволяет моделировать спектры структур-аттракторов, спектры «целей» саморазвития социальных систем и спектры социальных сред. Любые сложные системы имеют несколько альтернативных путей развития. Пути эволюции определяются спектрами структур-аттракторов нелинейных сред. Изменения социальных сред приводят к перестройке спектров структур-аттракторов, изменению спектров возможных путей в будущее.

Применение принципов синергетики дает возможность исследовать систему управления человеческими ресурсами как сложную нелинейную систему со встроенными механизмами обратной связи и отложенной во времени и пространстве реакцией на воздействие.

Ингерентность и базис, адекватный задаче. Как правило, рекомендации по управлению человеческими ресурсами пишутся для некоторого эксперта – «универсального солдата» и формируют, в итоге, достаточно абстрактную профессиональную компетентность. В то же время у каждого эксперта есть вполне определенный пол, возраст, состояние здоровья, опыт социализации, должность и т. п., что в значительной мере определяет процесс взаимодействия субъектов в процессе управления. Учет этих поистине человеческих ресурсов необходим для грамотного применения общих принципов различных подходов.

Одно из направлений дальнейшего развития управления человеческими ресурсами связано с построением «машин» и технологий, ингерентных субъектам управления [2]. Это требует, в свою очередь, наличия технологий оперативного выбора и построения базиса, адекватных задачам, которые ставит и решает субъект. В качестве таких «машин» мы видим ситуационные центры [9]. В качестве технологий построения базиса нами используется, в частности, рефлексивный театр и элементы китайской философии тай-цзы.

Отличия рефлексивного театра от аналогов, использующих идею театра и мобилизацию креативности, следующие [10]:

– основной режим работы – длительное сопровождение сложного проекта, в первую очередь учебного, который реализуется достаточно стабильным коллективом, причем каждая учебная сессия – не изолированное мероприятие, а часть «проектного сериала», история которого сохраняется и анализируется;

– одновременно и параллельно используются аналитические инструменты, в особенности схематизация, а также художественные средства (фрагменты художественных произведений, притчи, коаны, аудио- и видеофрагменты фильмов и т. п.);

– актеры (методологи и игротехники сервисной команды ситуационного центра) могут использовать театральные технологии и реквизит;

– строятся «зеркала» (в частности психологические портреты и характеристики компетентности) всех участников процесса;

– наконец, театр является именно рефлексивным, поскольку все действия, рефлексивные в том числе, сопровождаются рефлексивными, логическими и (или) художественными комментариями на соответствующих полиэкранах.

Использование параметров, измеряемых с помощью медицинских приборов [11], дает возможность дополнительного исследования и мобилизации человеческих ресурсов.

Еще одним примером создания базиса, адекватного субъекту как индивидуальному, так и коллективному, а также адекватного задаче, решаемой этим субъектом, является тай-цзы – китайская философия, позволяющая ставить и решать задачи различного уровня. Схемы тай-цзы являются достаточно универсальными инструментами достижения ингерентности, поскольку всегда ориентированы на использование естественных механизмов функционирования и развития различных систем.

В презентации к докладу приведены примеры реализации предложенных подходов на материале проектов, выполненных в Омске в 2005–2010 гг.

Таким образом, в работе описан опыт практического применения прикладного системного анализа, намечены направления исследований, ориентированных на управление человеческими ресурсами в процессе коллективной работы. Соответствующие исследования реализуются в учебно-исследовательских ситуационных центрах г. Омска.

Библиографические ссылки

1. Щедровицкий Г. П. О системе педагогических исследований (методологический анализ) // Избранные труды. М. : Школа культурной политики, 1995.

2. Тарасенко Ф. П. Прикладной системный анализ (Наука и искусство решения проблем) : учебник. Томск : Изд-во Том. ун-та, 2004.

3. Армстронг М. Стратегическое управление человеческими ресурсами. М. : Инфра-М, 2002.

4. Князева Е. Н., Курдюмов С. П. Основания синергетики. Синергетическое мировидение. М. : КомКнига, 2005.

5. Катунина И. В., Сугоняк В. В. Синергетический подход в управлении человеческими ресурсами // Межвуз. сб. тр. Омск : СибАДИ, 2008. Вып. 5. Ч. 2. С. 110–116.

6. Катунина И. В., Сугоняк В. В. Форма реализации синергетического подхода в управлении человеческими ресурсами // Проблемы и перспективы развития экономики труда : сб. материалов 8-й межрегион. конф. Иркутск : БГУЭП, 2008, С. 163–169.

7. Marchetti C. Anthropological Invariants in Travel Behavior // Technological Forecasting and Social Change. 1994. № 47. С. 75–88.

8. Тарасов В. К. Искусство управленческой борьбы. Технологии перехвата и удержания управления. М. : Добрая книга, 2007.

9. Информационные технологии для ситуационных центров / О. С. Анисимов [и др.]. Омск : Изд-во ОГИС, 2010.

10. Филимонов В. А. Рефлексивный театр в проекте многодисциплинарного исследовательского комплекса «Ген-Гуру» // Рефлексивные процессы и управление : сб. материалов VI Междунар. симп. М. : Когито-Центр, 2007. С. 121–123.

11. Повстяная А. Н., Филимонов В. А. Театр медицины и медицина в театре // Рефлексивный театр ситуационного центра (РТСЦ-2007) : Всерос. конф. с междунар. участием / Омс. гос. ин-т сервиса. Омск, 2007. С. 102–110.

V. V. Sugonyak, V. A. Filimonov

HUMAN RESOURCES MANAGEMENT FOR COLLECTIVE RESEARCH AND TRAINING

In the article we consider experience and prospects of application of the system approach to management of human resources. Results of experiment on system design of system of training in organization are elaborated with offers on application of other approaches connected with features of a person as the biological and social subject. Tools of consideration are the synergetics, reflexive theatre and Tai-Chi as an example of approach based on oriental philosophy.

Keywords: system design, synergetics, reflexive theatre.

© Сугоняк В. В., Филимонов В. А., 2010

УДК 51.77

Т. Б. Юргельян, Е. С. Кирик, Д. В. Круглов

О ВАЛИДАЦИИ МОДЕЛИ ДВИЖЕНИЯ ЛЮДЕЙ SIgMA.CA ПО ДАННЫМ ФУНДАМЕНТАЛЬНЫХ ДИАГРАММ*

Проводится валидация дискретной стохастической модели движения людей SIgMA.CA. Приводятся фундаментальные диаграммы (зависимость потока от плотности) для некоторых геометрий пространства в зависимости от параметров модели. Модельные данные сравниваются с натурными.

Ключевые слова: дискретно-стохастическая модель движения людей, валидация модели, фундаментальная диаграмма.

В наши дни вопрос безопасной эвакуации людей из зданий и сооружений становится все более актуальным. Поэтому возникает потребность в математических моделях движения людей, например, с целью определения наилучшей геометрии пространства, где предполагается скопление и задержка движения, определения времени эвакуации и т. д.

Несмотря на то, что существует достаточное количество как коммерческих предложений подобных моделей, так и научных исследований в этой области [1; 2], продолжение исследований остается востребованным из-за сложности рассматриваемого процесса (движение людей) и вариабельности натурных данных.

*Работа выполнена при поддержке ФЦП «Пожарная безопасность в Российской Федерации на период до 2012 г.», ГК № 09.0708.11.014, Лаврентьевского конкурса молодежных проектов 2010 г.

Вопросы валидации и верификации моделей являются одними из самых важных. На сегодняшний день единственным качественным способом верификации моделей движения людей является построение фундаментальной диаграммы (ФД). ФД – это кривая, отображающая зависимость скорости человека от плотности или потока от плотности. В то же время, диаграммы, полученные разными учеными в разных странах и в различных условиях, могут отличаться друг от друга в силу того, что движение человека обусловлено многими факторами, зависит от окружающей обстановки и психологического состояния. Тем не менее, у всех ФД есть одно общее свойство: с ростом плотности поток растет и, достигнув максимума при определенном значении плотности, начинает убывать; скорость уменьшается с ростом плотности.

Обычно данные замеряются на прямом участке пути. В литературе утверждается, что повороты не влияют на движение людей, однако не приводятся ни качественные, ни количественные доказательства этого утверждения [3]. При малых плотностях потока или небольших скоростях это, действительно, так. Но с другой стороны, если скорость достаточно велика, то при совершении поворота она уменьшится. Также средние и большие плотности влияют на динамику движения, и поворот не может не сказаться на величине потока. Поэтому возникает вопрос, будут ли ФД, построенные для различных геометрий пространства, совпадать.

В работе рассматриваются ФД, полученные с помощью дискретной стохастической модели SIgMA.CA (Stochastic fIeld Movement of Artificially People Intelligent Cellular Automata model) для двух различных геометрий пространства, и приводится сравнение модельных данных с некоторыми натурными [4] (см. Kirik E., Yurgel'yan T., Krouglov D. On realizing the shortest time strategy in a CA FF pedestrian dynamics model // Cybernetics and Systems. 42:01. 2011 (to appear)). Цель работы – выявить, влияет ли вид пути на ФД, и насколько модельные данные соответствуют натурным.

Модель движения людей SIgMA.CA. Пусть известны геометрия рассматриваемого пространства и начальное расположение частиц (людей) в этом пространстве. Пространство представляет собой плоскость, разбитую на ячейки 40 × 40 см, которые могут быть либо свободными, либо занятыми только одной частицей. В ячейках также могут располагаться стены и другие недвижимые препятствия.

Целью передвижения частиц является ближайший выход. Информацию о местоположении любого выхода дает статическое поле S, которое можно интерпретировать как карту местности [5]. Поле S радиально возрастает от точки выхода, не изменяется со временем и под действием частиц.

На каждом временном шаге частица может переместиться в одну из четырех свободных соседних ячеек или остаться на месте (окрестность фон Неймана). Направление частицы предполагается случайным и определяется на основе вероятностей переходов в каждом направлении для каждой частицы в каждый дискретный шаг времени и набором правил переходов частиц [4].

Правила перехода частиц следующие: сначала рассчитывается поле S, затем на каждом временном шаге вычисляются переходные вероятности и выбирается направление для перехода. Если две или более частицы претендуют на одну и ту же ячейку, применяется процедура разрешения конфликтов: либо все частицы остаются на прежних местах, либо только одной будет позволено переместиться. Стоит отметить, что вероятность перехода в занятую ячейку в общем случае не равна нулю. Частице, выбравшей занятое направление, предоставляется возможность либо выбрать другое направление для перехода, либо остаться на месте и подождать, пока выбранное направление освободится. Правила применяются ко всем частицам в один и тот же момент времени.

При вычислении переходных вероятностей учитывается «карта» расстояний до выходов S, желание частиц двигаться по направлению к выходу и желание держаться на некотором расстоянии от людей и от стен, что обеспечивается соответствующими параметрами модели $k_S >= 0$, $k_P >= 0$, $k_W >= 0$ и радиусом видимости $r > 0$ (расстояние в количестве ячеек).

Исследования. Рассматривались два примера (рис. 1). Было организовано движение частиц с постоянной плотностью p [1/м^2], частицы равномерно располагались по всему пространству. Замерялось время T (в шагах), за которое через сечение пройдет определенное количество частиц N (для прямого коридора N = 1 000, для коридора с поворотами N = 200) для заданной плотности потока.

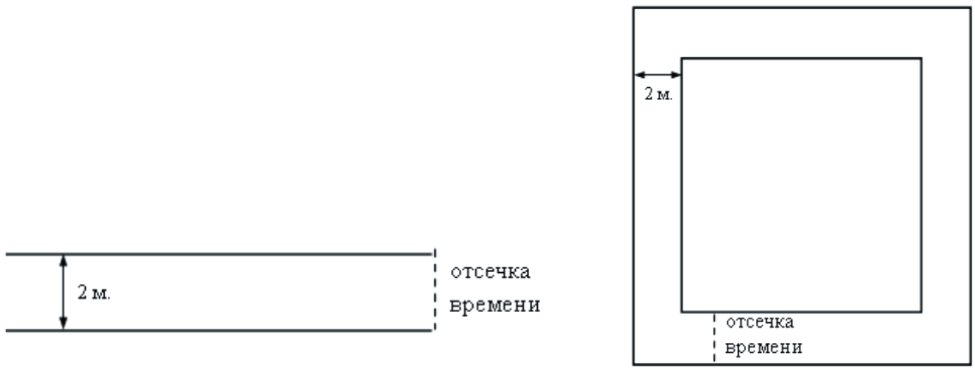
Примечание. Эксперименты проводились Т. Б. Юргельян; программный комплекс, реализующий модель SIgMA.CA, разработан Д. В. Кругловым, постановка задачи и научное руководство – Е. С. Кирик.

Параметры модели k_S и k_W были фиксированными, и $k_S = k_W = 4$ (параметр $k_S = 4$ соответствует направленному движению частиц), параметры k_P и r изменялись. Рассмотрим следующие пары: $k_P = 2$ и $r = 1$, $k_P = 4$ и $r = 1$, $k_P = 4$ и $r = 10$, $k_P = 12$ и $r = 10$.

По полученным данным был вычислен поток $J_{мод} = N/T/2$ [1/(м·шаг)] и построены ФД (рис. 2), которые показывают, что модель чувствительна к изменению вида пути (прямой или с поворотами).

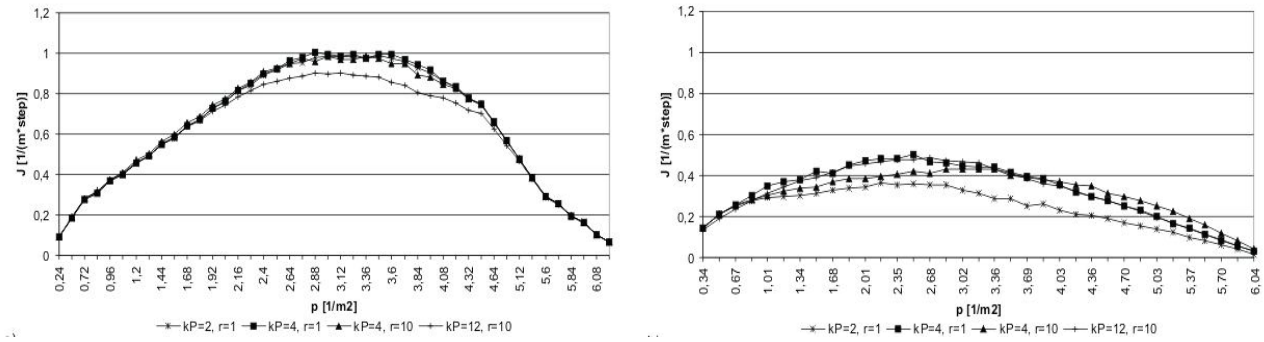
На малых плотностях (при $p < 1$ [1/м^2]) ФД совпадают, но на средних – поток для примера 2 (рис. 2, *б*) в два раза меньше потока для примера 1 (рис. 2, *а*). Это означает, что в прямом коридоре скорость движения выше, чем в коридоре с поворотами.

Плотность, при которой $J_{мод}$ достигает максимального значения, для примера 1 больше, чем для примера 2 (почти всегда) (см. таблицу). Значение $p(J_{max})$ интерпретируется как величина плотности, после которой передвижение людей становится затруднительным и падает скорость движения. Таким образом, в прямом коридоре частицы двигаются быстрее при большей плотности.



а *б*

Рис. 1. ФД, построенные по модельным данным:
а – пример 1; *б* – пример 2



а *б*

Рис. 2. Фундаментальные диаграммы:
а – для примера 1; *б* – для примера 2

Значения $p(J_{\max})$ для примеров 1, 2

$p(J_{\max})$, 1/м^2	$k_P = 2$, $r = 1$	$k_P = 4$, $r = 1$	$k_P = 4$, $r = 10$	$k_P = 12$, $r = 10$
Для примера 1	3,44	2,88	2,96	2,88
Для примера 2	2,18	2,52	3,19	2,68

Параметры модели определяют динамику модели и оказывают влияние на величину потока [4]. Например, параметры, которые соответствуют максимальному по величине потоку для примера 1 ($k_P = 2$, $r = 1$), для примера 2 дают минимальное значение потока (см. таблицу). Это объясняется влиянием параметров на динамику движения, но этот вопрос в данной статье обсуждаться не будет. Отметим только, что для прямого коридора наилучшими параметрами с точки зрения реалистичности динамики движения будет пара $k_p = 2$, $r = 1$, для коридора с поворотами – $k_P = 4$, $r = 10$ и $k_P = 12$, $r = 10$, на малых плотностях достаточно хорошие результаты могут быть и при $k_P = 4$, $k_P = 2$ и $r = 1$.

Сравнение ФД, полученных по модельным и натурным данным. В качестве натурных использовались данные Предтеченского и Милинского (PM) [6], Вейдманна (WM) [7] и Зейфрида (Seyfried) [8]. В первом и во втором случае данные получены на прямых участках пути, данные Зейфрида – на замкнутом контуре формы овала. Отметим, что в экспериментах Вейдманна и Зейфрида люди двигались медленнее (поток ниже), чем у Предтеченского и Милинского.

Для сравнения модельных и натурных данных требуется приведение к единым единицам измерения. Напомним, что за один расчетный шаг частица может передвинуться на 40 см, поэтому единица измерения потока 1/(м·шаг) переводится в 1/(м·с) по формуле $J = N/(T{\cdot}0{,}4/V)/2$, где V – скорость движения потока (экспериментальные данные), соответствующая данной плотности.

На рис. 3 отражены полученные результаты, приведены потоки для экспериментальных данных и модельных расчетов. В случае прямого коридора наблюдается совпадение расчетных и экспериментальных значений потоков для малых и средних плотностей (примерно до плотности 2,88 1/м^2). Поэтому можно утверждать, что на таких плотностях в целом динамика потока, воспроизводимая моделью, совпадает с имеющей место в натурных экспериментах.

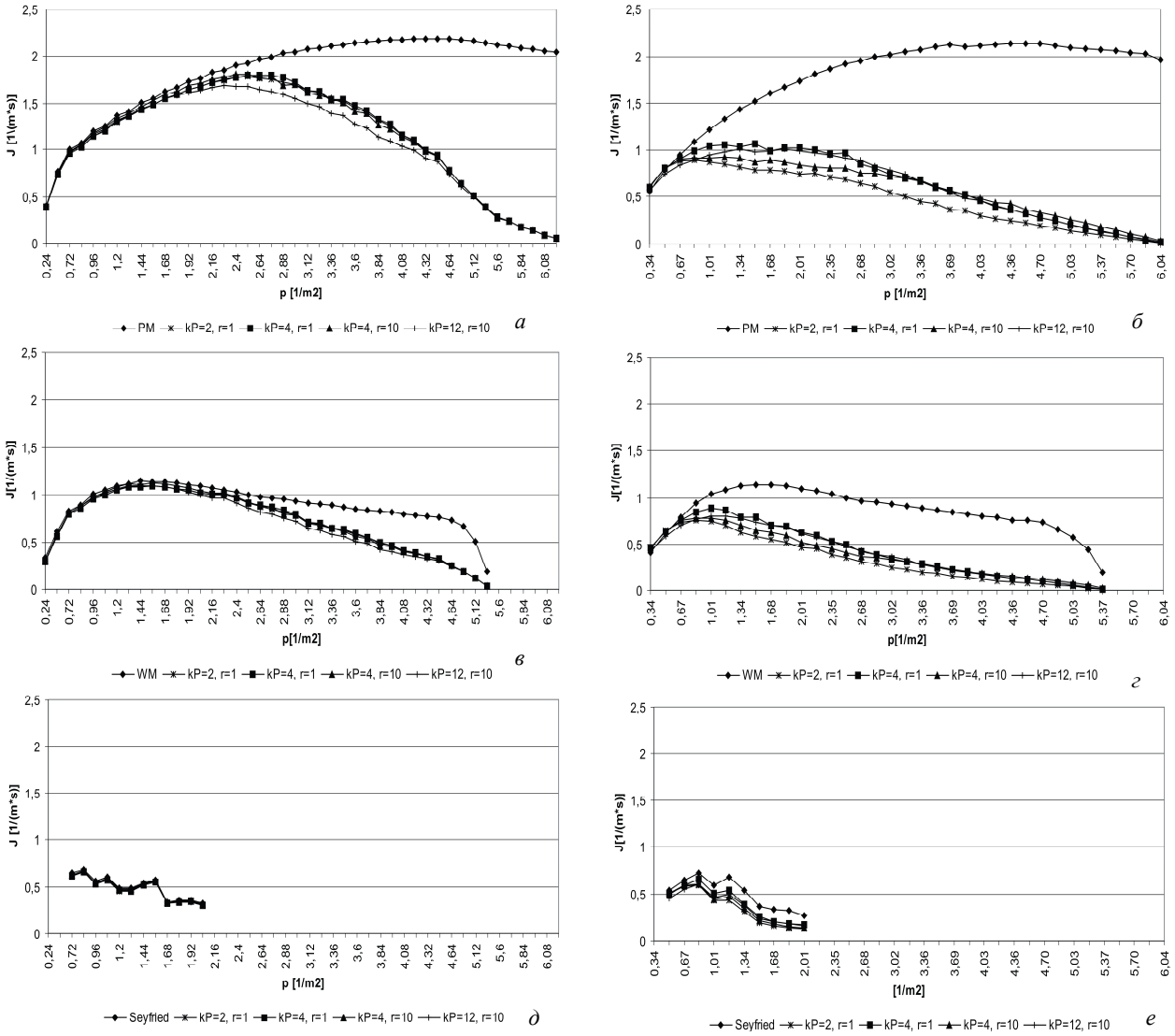


Рис. 3. Фундаментальные диаграммы, вычисленные с помощью различных натурных данных:
PM – данные Предтеченского и Милинского, WM – Вейдманна и Seyfried – Зейфрида;
а, *в*, *д* – для примера 1; *б*, *г*, *е* – для примера 2

С дальнейшим ростом плотности динамика модели замедляется. Причем, чем медленнее экспериментальная скорость, тем при больших значениях плотности начинаются существенные расхождения модельных и экспериментальных кривых.

В случае движения по замкнутому кругу, чем медленнее экспериментальная скорость, тем меньше величина расхождения между модельными и экспериментальными данными. Причем кривые начинают расходиться уже при плотности 0,84 1/м^2, т. е. при достаточно свободном движении. Наилучшее совпадение для данной формы пути наблюдается с данными Зейфрида, которые получены также на пути замкнутой формы.

Анализ и сравнение результатов для разных форм пути (рис. 3, *а*, *б*, *в*, *г* соответственно) показывают следующее. Экспериментальные и модельные кривые, полученные в схожих условиях (рис. 3, *а*, *в*), совпадают при малых и средних плотностях, но отличаются на тех же плотностях на рис. 3, *б*, *г*. Из этого факта можно сделать вывод, что наличие поворотов все же имеет влияние на динамику людского потока. Этот тезис подтверждает рис. 3, *е*, где происходит сравнение с натурными данными, полученными на замкнутом контуре, и имеется наибольшее совпадение.

Таким образом, можно сделать следующие выводы:

– модель чувствительна к виду пути: при наличии поворотов поток уменьшается примерно в два раза по сравнению с прямым участком;

– поток для модели SIgMA.CA согласуется с натурными данными на прямых участках пути при плотностях до 2,88 1/м^2;

– отсутствие натурных данных для геометрий, подобных примеру 2, не позволяет делать полноценные выводы об адекватности модели;

– необходимы дополнительные натурные эксперименты.

Библиографические ссылки

1. Kuligowsky E. D., Peacock R. D. A rewiew of evacuation models / National Institute of Standards and Technology, U. S. Department of Commerce, Technical note. 1471. 2005.

2. Evacuation Dynamics: Empirical Results, Modeling and Applications / A. Schadschneider [et al.] // Encyclopedia of Complexity and System Science. Springer. 2009.

3. Холщевников В. В., Самошин Д. А. Эвакуация и поведение людей при пожарах : учеб. пособие. М. : Академия ГПС МЧС России, 2009.

4. Kirik E., Yurgel'yan T., Krouglov D. The Shortest Time and/or the Shortest Path Strategies in a CA FF Pedestrian Dynamics Model // Журн. СФУ. Сер. Математика и физика. 2009. С. 271–278.

5. Schadschneider A., Seyfried A., Validation of CA models of pedestrian dynamics with fundamental diagrams // Cybernetics and Systems. 2009. Vol. 40 (5). P. 367–389.

6. Предтеченский В. М., Милинский А. И. Проектирование зданий с учетом организации движения людских потоков. М., 1969.

7. Weidmann U. Transporttechnik der Fubganger. Schriftentreihe des IVT. Nr. 90. ETH-Zurich, 1993.

8. The Fundamental Diagram of Pedestrian Movement Revisited / A. Seyfried [et al.] // J. Stat. Mech. Theory Exp. P10002. 2005.

T. B. Yurgel'yan, E. S. Kirik, D. V. Krouglov

ON VALIDATION OF SIGMA.CA MODEL OF PEDESTRIAN DYNAMICS ACCORDING TO DATA OF FUNDAMENTAL DIAGRAMS

Validation of SIgMA.CA model of pedestrian dynamics is carried out. Fundamental diagrams (the relation between density and flow) for a certain geometries are presented for different way shapes and model parameters. The simulation data are compared with experimental.

Keywords: discrete stochastic model of pedestrian dynamics, validation, fundamental diagram.

© Юргельян Т. Б., Кирик Е. С., Круглов Д. В., 2010

УДК 519.8

Я. И. Демченко, А. С. Орлова

ОБ ИССЛЕДОВАНИИ НЕКОТОРЫХ НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК ФУНКЦИИ РЕГРЕССИИ ПО НАБЛЮДЕНИЯМ

Рассматривается задача восстановления функции регрессии по наблюдениям со случайными ошибками. Применение для этих целей классических непараметрических оценок кривой регрессии не всегда дает удовлетворительные результаты на выборках небольшого объема. Предлагается новый класс непараметрических оценок, позволяющих повысить качество восстановления кривой регрессии по наблюдениям.

Ключевые слова: непараметрические оценки, ядерные оценки, квадратичное отклонение, регрессия, статистическое моделирование, сходимость, дискретно-непрерывные процессы.

В настоящее время известно большое число работ, посвященных непараметрическому восстановлению функций регрессии по наблюдениям с ошибками [1; 2]. Однако при обработке реальных данных приходится сталкиваться с ограниченными неравномерными выборками переменных $\{x_i, y_i, I = 1, \ldots, s\}$, содержащих сгущения, пустоты и разреженности в выборочном пространстве. В таких ситуациях классические непараметрические оценки кривой регрессии дают неудовлетворительные результаты, поэтому возникает необходимость введения новых непараметрических оценок функции регрессии, которые отличаются от известных использованием специальных ядерных функций. Исследуются также асимптотические свойства полученных оценок и приводятся результаты статистического моделирования.

Непараметрические оценки функции регрессии. Пусть (x, y) – случайная величина со значениями в пространстве $\Omega(x, y) \subset R^2$, $p(x, y) > 0$ – плотность распределения двумерной случайной величины (x, y), она неизвестна, кроме того $p(x) > 0$. Дана выборка из s статически независимых наблюдений двумерной случайной величины $(x, y) – (x_1, y_1), (x_2, y_2), \ldots, (x_s, y_s)$.

Обычно за непараметрическую оценку функции регрессии принимают статистику [1; 2]

$$y_s(x) = \sum_{i=1}^{s} y_i \Phi\left(\frac{x - x_i}{C_s}\right) \Big/ \sum_{i=1}^{s} \Phi\left(\frac{x - x_i}{C_s}\right), \qquad (1)$$

где интегрируемая с квадратом функция $\Phi(C_s^{-1}(x-x_i))$ и параметр C_s (коэффициент размытости) удовлетворяют некоторым условиям сходимости [3].

Новые непараметрические оценки кривой регрессии. В качестве непараметрической функции регрессии предлагается класс статистик:

$$\overline{y}_s(x)=\frac{\sum_{i=1}^{s} y_i \Phi_1\left(\frac{x-x_i}{C_s}\right)\Phi_2\left(\frac{x-x_i}{C_s}\right)}{\sum_{i=1}^{s} \Phi_1\left(\frac{x-x_i}{C_s}\right)\Phi_2\left(\frac{x-x_i}{C_s}\right)}, \qquad (2)$$

где функции $\Phi_1(C_s^{-1}(x-x_i))$, $\Phi_2(C_s^{-1}(x-x_i))$ и параметр C_s также удовлетворяют условиям сходимости [3] и свойству

$$C_s^{-1}\int_{\Omega(x)} \Phi_1(C_s^{-1}(x-x_i))\Phi_2(C_s^{-1}(x-x_i))dx=1. \qquad (3)$$

В случае, когда $(x,\ldots, x_n)\subset R^n$, непараметрическая оценка кривой регрессии (2) принимает вид

$$\overline{y}_s(\vec{x})=\frac{\sum_{i=1}^{s} y_i \prod_{j=1}^{n} \Phi_1\left(\frac{x_j-x_j^i}{C_s}\right)\Phi_2\left(\frac{x_j-x_j^i}{C_s}\right)}{\sum_{i=1}^{s}\prod_{j=1}^{n} \Phi_1\left(\frac{x_j-x_j^i}{C_s}\right)\Phi_2\left(\frac{x_j-x_j^i}{C_s}\right)}. \qquad (4)$$

Возможный вид функций $\Phi_1(\cdot)$, $\Phi_2(\cdot)$ представлен на рис. 1.

Также может быть введена оценка следующего вида:

$$\tilde{y}_s(x)=\frac{\sum_{i=1}^{s} y_i \left(\Phi_1\left(\frac{x-x_i}{C_s}\right)+\Phi_2\left(\frac{x-x_i}{C_s}\right)\right)}{\sum_{i=1}^{s}\left(\Phi_1\left(\frac{x-x_i}{C_s}\right)+\Phi_2\left(\frac{x-x_i}{C_s}\right)\right)}, \qquad (5)$$

где интегрируемые с квадратом, ограниченные, четные, дельтообразные функции $\Phi_1\left(C_s^{-1}(x-x_i)\right)$, $\Phi_2(C_s^{-1}(x-x_i))$ и параметр C_s (коэффициент размытости) удовлетворяют условиям сходимости:

$$\lim_{s\to\infty} C_s^{-1}(\Phi_1(C_s^{-1}(x-x_i))+ \\ +\Phi_2(C_s^{-1}(x-x_i)))=\delta(x-x_i), \qquad (6)$$

$$C_s^{-1}\int_{\Omega(x)} (\Phi_1(C_s^{-1}(x-x_i))+\Phi_2(C_s^{-1}(x-x_i)))dx=1. \qquad (7)$$

Если $(x,\ldots, x_n)\subset R^n$, то непараметрическая оценка кривой регрессии (5) принимает вид:

$$\tilde{y}_s(\vec{x})=\frac{\sum_{i=1}^{s} y_i \prod_{j=1}^{n} \left(\Phi_1\left(\frac{x-x_i}{C_s}\right)+\Phi_2\left(\frac{x-x_i}{C_s}\right)\right)}{\sum_{i=1}^{s}\prod_{j=1}^{n} \left(\Phi_1\left(\frac{x-x_i}{C_s}\right)+\Phi_2\left(\frac{x-x_i}{C_s}\right)\right)}. \qquad (8)$$

Непараметрические оценки функции регрессии имеют непосредственное отношение к задаче идентификации дискретно-непрерывных процессов по наблюдениям «входных–выходных» переменных. При этом выборочное пространство наблюдения соответствующих переменных может иметь некоторые особенности, в частности, сгущения, пустоты, разреженности.

Для непараметрических оценок функции регрессии $\overline{y}_s(x)$ (2) и $\tilde{y}_s(x)$ (5) имеют место следующие теоремы.

Теорема 1. Пусть $y(x)$ дважды дифференцируема и с вероятностью 1 $p(x)>0$, $\forall x\in\Omega(x)$, а функции $\Phi_1(C_s^{-1}(x-x_i))$, $\Phi_2(C_s^{-1}(x-x_i))$ и параметр размытости C_s удовлетворяют условию сходимости (3), тогда:

$$\lim_{s\to\infty} \mathrm{M}\{(y(x)-\overline{y}_s(x))^2\}=0, \quad \forall x\in\Omega(x).$$

Теорема 2. Пусть $y(x)$ дважды дифференцируема и с вероятностью 1 $p(x)>0$, $\forall x\in\Omega(x)$, а функции $\Phi_1(C_s^{-1}(x-x_i))$, $\Phi_2(C_s^{-1}(x-x_i))$ и параметр размытости C_s удовлетворяют условиям сходимости (6) и (7), тогда:

$$\lim_{s\to\infty} \mathrm{M}\{(y(x)-\tilde{y}_s(x))^2\}=0, \\ \forall x\in\Omega(x).$$

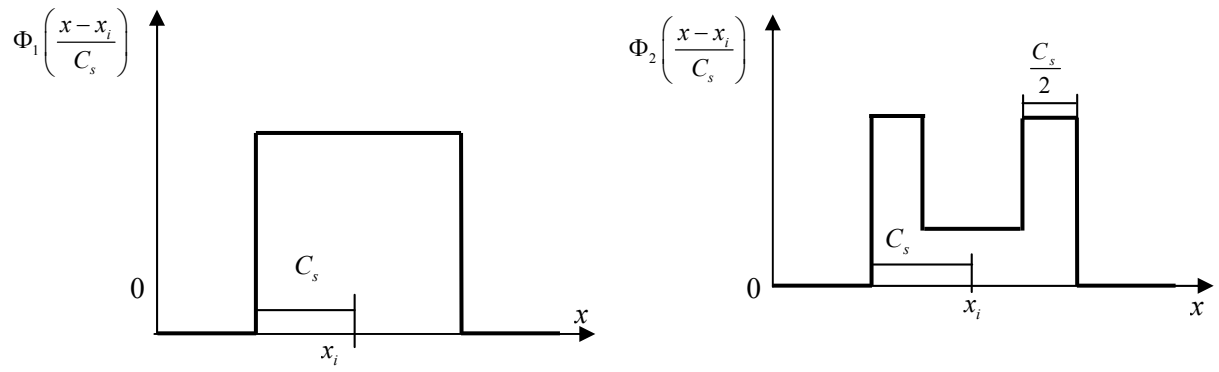


Рис. 1. Вид функций $\Phi_1\left(\frac{x-x_i}{C_s}\right)$ и $\Phi_2\left(\frac{x-x_i}{C_s}\right)$

При доказательстве теорем использованы приемы, предложенные в [1; 4]. Более подробные доказательства приведены в [5; 6].

Исследование непараметрических оценок методом статистического моделирования. В процессе исследований использовалась истинная зависимость вида $y(x_1, x_2) = x_1^2 + \frac{1}{2}e^{x_2} + 2x_1x_2 + h$, где h – уровень помех, распределенных по нормальному закону с нулевым математическим ожиданием и ограниченной дисперсией, а значения (x_1, x_2) генерировались из интервала [0, 4] таким образом, что стохастически зависимые переменные x_1 и x_2 образуют «трубчатый» процесс в выборочном пространстве (y, x_1, x_2). Зависимость $y(x_1, x_2)$ необходима для получения выборочных данных, но при восстановлении регрессии информация о ее параметрической структуре считается неизвестной. Эксперименты проводились при различных объемах выборок «входных–выходных» переменных s, с различным уровнем помех h. Для восстановления $y(x_1, x_2)$ использовались непараметрические оценки кривой регрессии $\bar{y}_s(x_1, x_2)$ вида (4), где $n = 2$, и $\tilde{y}_s(x_1, x_2)$ вида (8), где $n = 2$. При построении непараметрических оценок функции регрессии использовались функции $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, приведенные на рис. 1. Параметры размытости C_s выбирались путем минимизации критерия оптимизации:

$$w(C_s) = \frac{1}{s}\sum_{j=1}^{s}(y_j - y_s(x_{1j}, x_{2j}))^2 \to \min_{C_s},$$

где

$$y_s(x_1, ..., x_n) = \frac{\sum_{i=1}^{s} y_i \prod_{j=1}^{n} \Phi_1\left(\frac{x_j - x_j^i}{C_s}\right)}{\sum_{i=1}^{s}\prod_{j=1}^{n} \Phi_1\left(\frac{x_j - x_j^i}{C_s}\right)}, \; n = 2, \qquad (9)$$

либо $\bar{y}_s(x_1, x_2)$ вида (4), либо $\tilde{y}_s(x_1, x_2)$ вида (8).

В качестве точности аппроксимации использовалась квадратичная ошибка. Для оценки кривой регрессии $y_s(x_1, x_2)$ вида (9) ошибку обозначим как w_s, для оценки $\bar{y}_s(x_1, x_2)$ вида (4) – как $\bar{w}_s$, для оценки $\tilde{y}_s(x_1, x_2)$ вида (8) – как $\tilde{w}_s$.

На рис. 2 график оценки кривой регрессии $\tilde{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 80 точек, на интервалах [0,5; 1], [1,7; 2,1], [2,4; 2,8], [3,4; 3,6] и [3,9; 4] присутствуют сгущения, на интервалах [1; 1,7], [2,9; 3,3] и [3,7; 3,8] имеются разреженности, и на интервалах [2,1; 2,4], [2,8; 2,9], [3,3; 3,4] и [3,8; 3,9] отсутствуют элементы выборки. Уровень помех h = 10 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,11 и C_{s2} = 0,14. Ошибки аппроксимации $\tilde{w}_s = 0{,}05$, $w_s = 0{,}15$. Таким образом, новая оценка кривой регрессии $\tilde{y}_s(x_1, x_2)$ вида (8) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

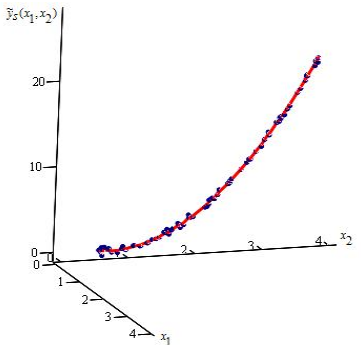


Рис. 2

На рис. 3 график оценки кривой регрессии $\tilde{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 60 точек, на интервалах [0,5; 2,1], [2,3; 2,7], [2,8; 2,9], [3,4; 3,5] и [3,7; 3,8] присутствуют сгущения, на интервалах [3; 3,3], [3,6; 3,7] и [3,9; 4] имеются разреженности, и на интервалах [2,1; 2,3], [2,7; 2,8], [2,9; 3], [3,3; 3,4] и [3,8; 3,9] отсутствуют элементы выборки. Уровень помех h = 0 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,13 и C_{s2} = 0,18. Ошибки аппроксимации $\tilde{w}_s = 0{,}03$, $w_s = 0{,}12$. Таким образом, новая оценка кривой регрессии $\tilde{y}_s(x_1, x_2)$ вида (8) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

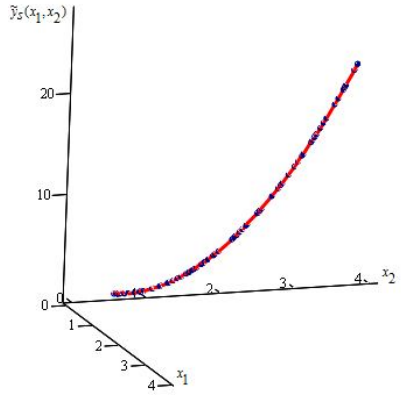


Рис. 3

На рис. 4 график оценки кривой регрессии $\tilde{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 100 точек, на интервалах [0,5; 2], [2,7; 3,1] и [3,4; 3,6] присутствуют сгущения, на интервалах [2,1; 2,6] и [3,1; 3,2] имеются разреженности, и на интервалах [2,6; 2,7] и [3,3; 3,4] отсутствуют элементы выборки. Уровень помех h = 25 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,07 и C_{s2} = 0,08. Ошибки аппроксимации $\tilde{w}_s = 0{,}08$, $w_s = 0{,}19$. Таким образом, новая оценка кривой регрессии $\tilde{y}_s(x_1, x_2)$

вида (8) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

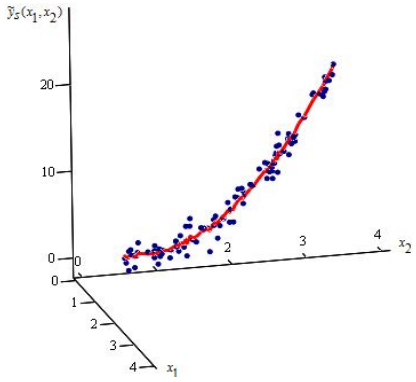


Рис. 4

На рис. 5 график оценки кривой регрессии $\overline{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 50 точек, на интервалах [1; 1,3], [2,1; 2,5] и [3,4; 3,5] присутствуют сгущения, на интервалах [0,6; 1], [1,3; 2,1], [2,9; 3,4] и [3,6; 4] имеются разреженности, и на интервалах [2,5; 2,7] и [2,7; 2,9] отсутствуют элементы выборки. Уровень помех h = 0 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,25 и C_{s2} = 0,29. Ошибки аппроксимации $\overline{w}_s = 0{,}06$, $w_s = 0{,}17$. Таким образом, новая оценка кривой регрессии $\overline{y}_s(x_1, x_2)$ вида (4) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

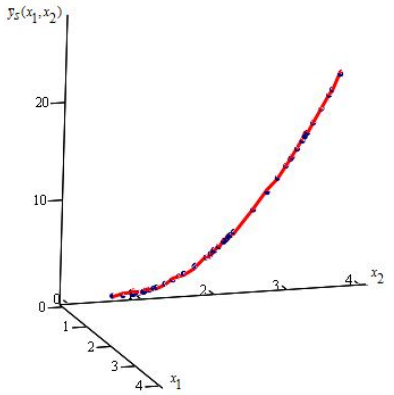


Рис. 5

На рис. 6 график оценки кривой регрессии $\overline{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 90 точек, на интервалах [2,9; 3,1] и [3,2; 3,3] присутствуют сгущения, на интервалах [0,7; 1,7], [1,8; 2,9] и [3,3; 4] имеются разреженности, и на интервалах [1,7; 1,8] и [3,1; 3,2] отсутствуют элементы выборки. Уровень помех h = 20 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,09 и C_{s2} = 0,08. Ошибки аппроксимации $\overline{w}_s = 0{,}1$, $w_s = 0{,}22$. Таким образом, новая оценка кривой регрессии $\overline{y}_s(x_1, x_2)$ вида (4) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

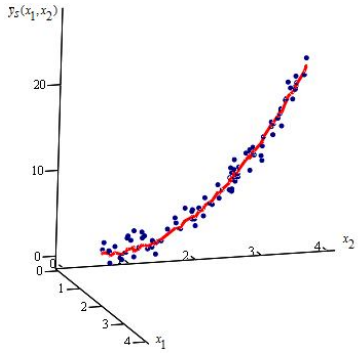


Рис. 6

На рис. 7 график оценки кривой регрессии $\overline{y}_s(x_1, x_2)$ представлен линией, исходная выборка процесса – точками. Объем выборки s = 120 точек, на интервалах [0,6; 1,5] и [2,3; 3] присутствуют сгущения, на интервалах [1,5; 2,3], [3; 3,4], [3,5; 3,7] и [3,8; 3,9] имеются разреженности, и на интервалах [3,4; 3,5] и [3,7; 3,8] отсутствуют элементы выборки. Уровень помех h = 8 %. Значения параметров размытости функций $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, соответственно, C_{s1} = 0,05 и C_{s2} = 0,05. Ошибки аппроксимации $\overline{w}_s = 0{,}09$, $w_s = 0{,}2$. Таким образом, новая оценка кривой регрессии $\overline{y}_s(x_1, x_2)$ вида (4) более эффективна на данной выборке, чем оценка $y_s(x_1, x_2)$ вида (9).

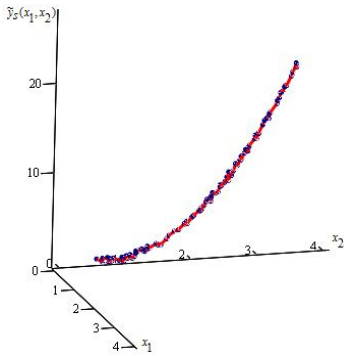


Рис. 7

Если $(x_1, ..., x_5) \in R^5$, использовалась истинная зависимость вида $y(x_1, ..., x_5) = x_1^2 + \frac{1}{2}e^{x_2} + 5\sin(x_3) - 3\lg(x_4) + x_5 + h$, где h – уровень помех, распределенных по нормальному закону с нулевым математическим ожиданием и ограниченной дисперсией, а значения $(x_1, ..., x_5)$ генерировались из интервала [0, 4] таким образом, что стохастически зависимые переменные $x_1, ..., x_5$ образуют «трубчатый» процесс в выборочном пространстве $(y, x_1, ..., x_5)$. Зависимость $y(x_1, ..., x_5)$ необходима для получения выборочных данных, но при восстановлении регрессии информация о ее параметрической структуре считается неиз-

вестной. Эксперименты проводились при различных объемах выборок «входных–выходных» переменных *s*, с различным уровнем помех *h*. Для восстановления $y(x_1,...,x_5)$ использовались непараметрические оценки кривой регрессии $\bar{y}_s(x_1,...,x_5)$ вида (4), где $n = 5$, и $\tilde{y}_s(x_1,...,x_5)$ вида (8), где $n = 5$. При построении непараметрических оценок функции регрессии использовались функции $\Phi_1(\cdot)$ и $\Phi_2(\cdot)$, приведенные на рис. 1. Параметры размытости C_s выбирались путем минимизации критерия оптимизации:

$$w(C_s) = \frac{1}{s}\sum_{j=1}^{s}(y_j - y_s(x_{1j},...,x_{5j}))^2 \to \min_{C_s},$$

где $y_s(x_1,...,x_5)$ вида (9), $n = 5$, либо $\bar{y}_s(x_1,...,x_5)$ вида (4), либо $\tilde{y}_s(x_1,...,x_5)$ вида (8).

В качестве точности аппроксимации использовалась квадратичная ошибка. Для оценки кривой регрессии $y_s(x_1,...,x_5)$ вида (9) ошибку обозначим как w_s, для оценки $\bar{y}_s(x_1,...,x_5)$ вида (4) – как $\bar{w}_s$, для оценки $\tilde{y}_s(x_1,...,x_5)$ вида (8) – как $\tilde{w}_s$ (см. таблицу).

Таким образом, при восстановлении функции регрессии по результатам экспериментальных данных в выборочном пространстве наблюдений могут иметь место сгущения, разреженности, пропуски данных. В этом случае классические непараметрические оценки восстановления стохастической зависимости типа (9) могут оказаться недостаточно эффективными. Предложен прием формирования ядерных функций некоторым специальным образом (рис. 1). Для новых непараметрических оценок функции регрессии доказаны соответствующие теоремы сходимости.

Проведение численных исследований показывает более высокую эффективность непараметрических оценок функции регрессии $\bar{y}_s(x_1,...,x_n)$ вида (4) и $\tilde{y}_s(x_1,...,x_n)$ вида (8). При использовании функций (4) и (5) квадратичная ошибка в среднем уменьшается в два раза. Следует обратить внимание, что исследование непараметрических оценок проводилось для процессов «трубчатой» структуры.

Библиографические ссылки

1. Надарая Э. А. Некоторые вопросы теории вероятностных процессов // Вычислительный центр / АН ГрузССР. 1965. Вып. 5. С. 56–68.

2. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М. : Наука, 2004.

3. Медведев А. В. Непараметрические системы адаптации. Новосибирск : Наука, 1983.

4. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее применение. 1969. Т. 14. Вып. 1. С. 156–162.

5. Демченко Я. И. О некоторых непараметрических оценках плотности вероятности и кривой регрессии. Ч. 1 // Молодой ученый. 2010. № 11(22). Т. 1. С. 10–20.

6. Демченко Я. И. О некоторых непараметрических оценках плотности вероятности и кривой регрессии. Ч. 2 // Молодой ученый. 2010. № 11(22). Т. 1. С. 21–32.

Экспериментальные данные

Кривая регрессии	Объем выборки, s	Уровень шума *h*, %	Значение C_{s1}	Значение C_{s2}	Ошибка аппроксимации $\bar{w}_s$ или $\tilde{w}_s$	Ошибка аппроксимации w_s
$\bar{y}_s(x_1,...,x_5)$	50	0	0,15	0,13	0,1	0,24
$\bar{y}_s(x_1,...,x_5)$	80	5	0,1	0,09	0,12	0,28
$\bar{y}_s(x_1,...,x_5)$	100	10	0,09	0,08	0,09	0,19
$\tilde{y}_s(x_1,...,x_5)$	65	5	0,12	0,13	0,14	0,31
$\tilde{y}_s(x_1,...,x_5)$	90	0	0,1	0,1	0,08	0,2
$\tilde{y}_s(x_1,...,x_5)$	110	10	0,09	0,08	0,07	0,22

Ya. I. Demchenko, A. S. Orlova

ABOUT RESEARCHE OF SOME NONPARAMETRIC ESTIMATIONS OF REGRESSION FUNCTION ACCORDING OBSERVATIONS

Task of regression function restoration according observations with random errors is considered in the article. Usage of the classical nonparametric estimations of regression curve for such purposes doesn't always give satisfactory results on small samples. The new class of nonparametric estimations is proposed, it allows to increase the regression curve restoration quality of according observations.

Keywords: nonparametric estimators, kernel estimators, squared deviation, regression, statistic simulation, convergence, discrete-continuous process.

© Демченко Я. И., Орлова А. С., 2010

УДК 62.501

Д. А. Игнатьев, А. Н. Сергеев

НЕПАРАМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОРГАНИЗАЦИОННЫХ СИСТЕМ

Рассматриваются некоторые задачи моделирования и управления организационными системами в условиях неопределенности. Отмечены некоторые особенности организационных процессов по сравнению с другими, происходящими на различных технических и технологических объектах. Предлагается непараметрическая модель организационных систем и алгоритмы принятия решений в условиях неполной информации.

Ключевые слова: непараметрическая идентификация, модель, идентификация, априорная информация, организационные процессы, принятие решений.

При разработке систем управления дискретно-непрерывными процессами часто возникает необходимость управления не отдельным технологическим объектом, а всем производственным комплексом. При этом представляется целесообразным рассматривать иерархическую схему управления промышленным комплексом. Разработка соответствующих обучающихся моделей и алгоритмов управления существенно определяется наличием априорной информации о технологических процессах, каналах измерения и, в сущности, предопределяет использование тех или иных разделов теории идентификации и управления. Основной акцент сделаем на наличии минимальной, как это представляется сегодня, априорной информации об исследуемых объектах, системах, т. е. будем рассматривать задачи идентификации и управления в условиях непараметрической неопределенности. В этом случае отсутствует этап определения модели или алгоритма управления с точностью до вектора параметров. Это обусловлено тем, что чаще всего информация, необходимая для предварительной параметризации или выбора параметрической структуры модели, недостаточная. Это делает подобный подход более реалистическим, чем общепринятая теория. Конечный существенный интерес представляет случай, когда некоторые объекты или каналы связи промышленного комплекса соответствуют одновременно как уровню параметрической неопределенности, так и уровню непараметрической. Это отдельная проблема, которая требует специального рассмотрения. Таким образом, настоящее исследование посвящено только случаю, когда априорной информации об исследуемой системе или процессе соответствует уровень непараметрической неопределенности. Отметим, что верхний иерархический уровень управления подобным комплексом представляет собой организационную систему управления с лицом, принимающим решения (ЛПР).

Моделирование дискретно-непрерывных процессов. Дадим краткую характеристику дискретно-непрерывного процесса, который является предметом исследования в традиционной задаче идентификации (рис. 1).

На рис. 1 введены следующие обозначения: A – оператор процесса; $u(t)$ и $\mu(t)$ – управляемые и неуправляемые входы процесса, соответственно; $x(t)$ – выходная переменная процесса; $\hat{x}_t$ – выходная переменная модели; H^{μ}, H^{u} и H^{x} – каналы измерения соответствующих переменных; $h^{\mu}(t)$, $h^{u}(t)$ и $h^{x}(t)$ – помехи, действующие в каналах измерения; μ_t, u_t и x_t – переменные процесса, измеренные в дискретные моменты времени (подразумевается, что они зашумлены соответствующей помехой h). Следует также отметить, что и на сам процесс оказывает влияние некоторая помеха $\xi(t)$.

В дальнейшем исследуется дискретно-непрерывный процесс, представленный на рис. 2.

Здесь введена переменная выхода $q(t)$, измеряемая через интервал времени $\Delta T >> \Delta t$, где Δt – дискретность измерения переменных $u(t)$, $\mu(t)$, $x(t)$. Чаще всего эти переменные измеряются электромагнитными средствами. Поэтому Δt в технических системах достаточно мало. K – ключ, символизирующий канал контроля.

Обратим внимание на то, что ΔT может значительно превышать постоянную времени объекта, что приводит к необходимости рассматривать объект как статический с запаздыванием (рис. 3).

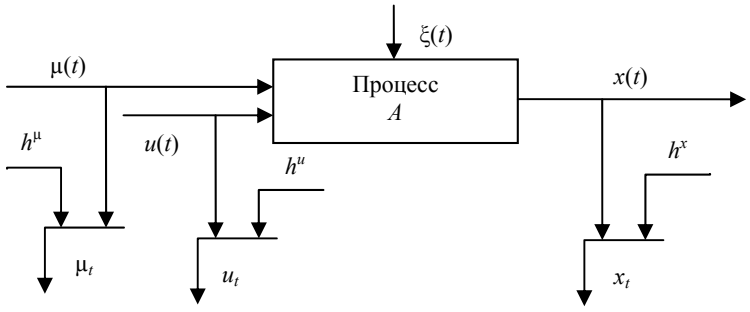


Рис. 1. Традиционная схема дискретно-непрерывного процесса

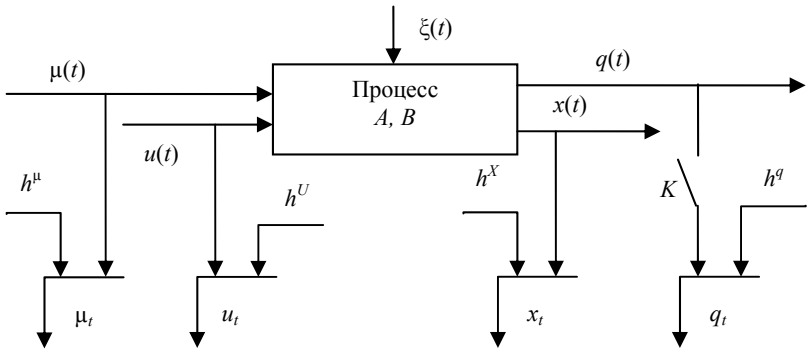


Рис. 2. Схема дискретно-непрерывного процесса при различной дискретности контроля

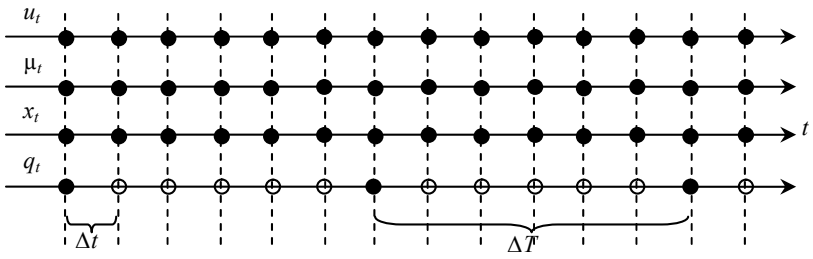


Рис. 3. Измерение переменных процесса производится с различной дискретностью $\Delta T >> \Delta t$:
● – моменты измерений по каждой из переменных; ○ – пропущенные значения

Процесс, представленный на рис. 1 и 2, описывается следующим образом:

$$x(t) = A\big(u(t), \mu(t), \xi(t), t\big), \qquad (1)$$

$$q(t) = B\big(u(t), \mu(t), \xi(t), t\big), \qquad (2)$$

где A, B – оператор процесса – класс операторов, который определяется на основании имеющейся априорной информации. При существенном влиянии на $x(t)$ случайных факторов $\xi(t)$, удовлетворительную модель построить нельзя, даже при условии удачного выбора класса операторов B [1].

Непараметрическая модель процесса (1) может быть представлена как

$$\hat{x}(u, \mu) = \frac{\sum\limits_{i=1}^{S} x_i \prod\limits_{v=1}^{m} \Phi\left(\frac{u^v - u_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu^v - \mu_i^v}{C_S}\right)}{\sum\limits_{i=1}^{S} \prod\limits_{v=1}^{m} \Phi\left(\frac{u^v - u_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu^v - \mu_i^v}{C_S}\right)}, \qquad (3)$$

где $\hat{x}(u, \mu)$ – непараметрическая оценка регрессии $x(u, \mu)$; $u \in R^m$, $\mu \in R^m$; S – объем выборки; $\Phi(\cdot)$ – колоколообразная функция, или как ее еще называют ядерная функция; C – параметр размытости ядерной функции $\Phi(\cdot)$. Функция $\Phi(\cdot)$ обладает следующими свойствами:

а) $0 \le \Phi(u) < A < \infty$;

б) $\Phi(u) = \Phi(-u)$;

в) $\int\limits_{-\infty}^{\infty} \Phi(u)\, du = 1$;

г) $\int\limits_{-\infty}^{\infty} \Phi(u) u^m du < \infty$, при $m > 0$, $u = \dfrac{y - y_i}{C_S}$;

д) $\lim\limits_{S \to \infty} C_S^{-1} \Phi(u) = \delta(y - y_i)$ – свойство δ-образности, а коэффициенты размытости C_S ядер эмпирической плотности вероятности зависят в общем случае от объема выборки S, причем $C_S \to 0$ и $sC_S \to \infty$ при $s \to \infty$, y – произвольный аргумент.

Непараметрическую модель процесса (2) можно представить в следующем виде:

$$\hat{q}(u, \mu, x) = \frac{\sum\limits_{i=1}^{S} q_i \prod\limits_{v=1}^{m} \Phi\left(\frac{u^v - u_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu^v - \mu_i^v}{C_S}\right) \prod\limits_{v=1}^{n} \Phi\left(\frac{x^v - x_i^v}{C_S}\right)}{\sum\limits_{i=1}^{S} \prod\limits_{v=1}^{m} \Phi\left(\frac{u^v - u_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu^v - \mu_i^v}{C_S}\right) \prod\limits_{v=1}^{n} \Phi\left(\frac{x^v - x_i^v}{C_S}\right)}, \qquad (4)$$

где $\hat{q}(u, \mu, x)$ – непараметрическая модель процесса $q(u, \mu, x)$, $x \in R^n$, в которую включена выходная переменная $x(t)$, измерения которой доступны через интервалы времени Δt; ядерная функция $\Phi(\cdot)$ удовлетворяет показанным выше условиям.

Непараметрическое управление дискретно-непрерывными процессами. Управление процессом, представленным на рис. 1, осуществляется по схеме, представленной на рис. 4.

Обозначения на рис. 4 те же, что и выше, а x^* – задающее воздействие. Непараметрический алгоритм управления при активном накоплении информации имеет вид

$$\hat{u}_{S+1}^k = \frac{\sum\limits_{i=1}^{S} u_i \prod\limits_{v=1}^{n} \Phi\left(\frac{x^{*v}_{S+1} - x_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu_{S+1}^v - \mu_i^v}{C_S}\right)}{\sum\limits_{i=1}^{S} \prod\limits_{v=1}^{n} \Phi\left(\frac{x^{*v}_{S+1} - x_i^v}{C_S}\right) \prod\limits_{v=1}^{l} \Phi\left(\frac{\mu_{S+1}^v - \mu_i^v}{C_S}\right)} + \Delta u_{S+1}^k, \qquad (5)$$

где Δu_{S+1}^k – поисковые шаги, $k = \overline{1, m}$.

Управление УУ^x (рис. 5) – устройство управления по выходной переменной $x(t)$, УУ^q – устройство управления по выходной переменной $q(t)$, M_1 – макрообъект первого уровня, M_2 – макрообъект второго уровня. Роль управляемой переменной на этом уровне иерархии выполняет x^*. Алгоритм управления на следующем уровне иерархии имеет вид

$$\hat{x}^{*k}_{S+1} = \frac{\sum_{i=1}^{S} x_i \prod_{v=1}^{w} \Phi\left(\frac{q^{*v}_{S+1} - q^v_i}{C_S}\right) \prod_{v=1}^{l} \Phi\left(\frac{\mu^v_{S+1} - \mu^v_i}{C_S}\right)}{\sum_{i=1}^{S} \prod_{v=1}^{w} \Phi\left(\frac{q^{*v}_{S+1} - q^v_i}{C_S}\right) \prod_{v=1}^{l} \Phi\left(\frac{\mu^v_{S+1} - \mu^v_i}{C_S}\right)} + \Delta x^{*k}_{S+1}, \quad (6)$$

где Δx^{*k}_{S+1} – поисковые шаги, $k = \overline{1, n}$; $q \in R^w$. Следует заметить, что на этом уровне иерархии предполагается наличие обучающей выборки с последующим ее пополнением в процессе функционирования замкнутой схемы. Дальнейшее обобщение системы управления сложным промышленным комплексом может быть представлено в виде схемы (рис. 6).

На рис. 6 СПР – система принятия решений с ЛПР; $Q^*(q, z)$ – заданная целевая функция, зависящая от q и от производственных показателей z, оказывающих существенное влияние на характер вырабатываемых воздействий СПР; отдельно выделена «Информация», указывающая на возможность получения блоком СПР или оптимизаторами любой доступной информации о системе, требуемой для функционирования.

Общая постановка задачи имеет следующий вид:

$$M\{Q(q,z)\} \to \min_{q \in \Omega(q)}. \quad (7)$$

Ограничения в общем случае имеют вид:

$$M\{\varphi(q,z)\} = 0,\ M\{\psi(q,z)\} \geq 0. \quad (8)$$

Тогда непараметрический алгоритм оптимизации может быть представлен следующим образом:

$$\hat{q}^{*k}_{S+1} = \frac{\sum_{i=1}^{S} q_i \omega_{s,i}}{\sum_{i=1}^{S} \omega_{s,i}} + \Delta q^{*k}_{S+1}, \quad (9)$$

$$\omega_{s,i} = \prod_{v=1}^{w} \Phi\left(\frac{q^{*v}_{S+1} - q^v_i}{C_s^{q^v}}\right) \prod_{v=1}^{t} \Phi\left(\frac{z^v_{S+1} - z^v_i}{C_s^{z^v}}\right) \times \times \prod_{v=1}^{r} \Phi\left(\frac{\varphi_v(u_i, \mu_i)}{C_s^{\varphi^v}}\right) \prod_{v=1}^{p} \operatorname{sgn}\left(\psi_v(u_i, \mu_i)\right), \quad (10)$$

где Δq^{*k}_{S+1} – поисковые шаги, $k = \overline{1, w}$; $z \in R^t$; функция sgn(y) с произвольной переменной y следующая:

$$\operatorname{sgn}(y) = \begin{cases} 1, \text{ если } y \geq 0, \\ 0, \text{ если } y < 0. \end{cases} \quad (11)$$

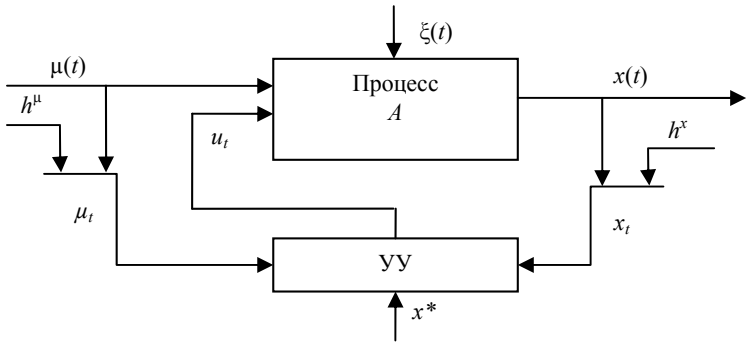


Рис. 4. Система управления комбинированным объектом

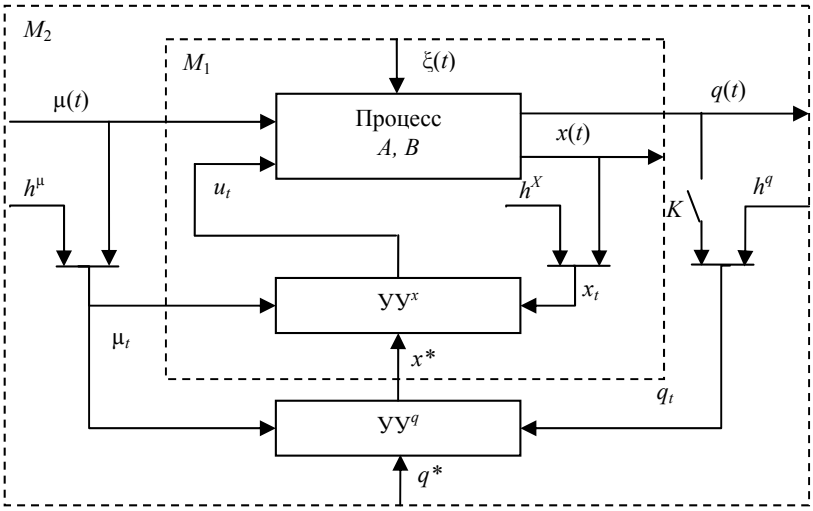


Рис. 5. Иерархическая система управления

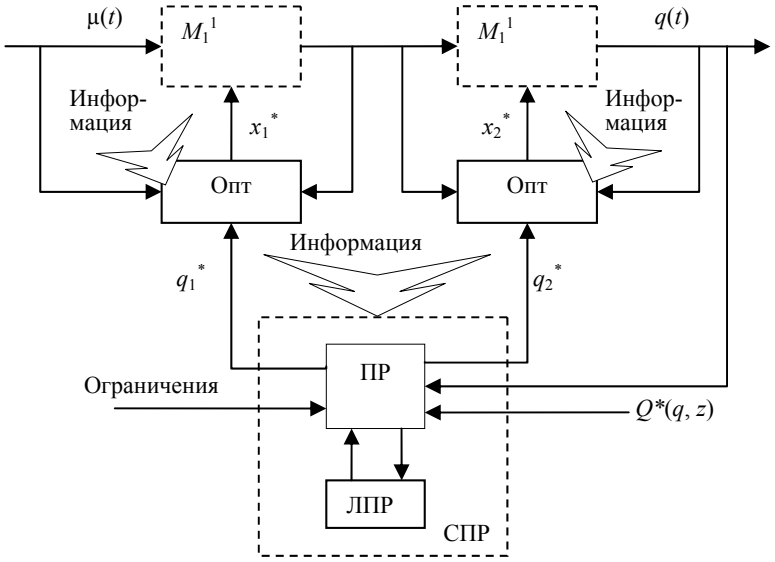


Рис. 6. Иерархическая система принятия решений с ЛПР

Параметр размытости C_S – отдельная задача теории оптимизации. Вследствие трудности постановки экспериментов на реальном объекте в организационных системах, функция оптимизатора отводится человеку. ЛПР располагает областью определения параметра и, в зависимости от состояния системы, устанавливает то или иное значение, анализируя отклик организационной системы на выбранное значение параметра. В случае необходимости значение параметра изменяется, и вновь анализируется оклик. Область определения параметра C_S, в свою очередь, обусловлена как естественными ограничениями, так и опытом и знаниями ЛПР в исследуемой области.

Обучающаяся система управления с идентификатором. Естественно считать, что первым этапом на пути создания подобной системы является построение модели (идентификатора). Причем требования к идентификатору достаточно высоки. Такая система, представленная на рис. 7, определена высокой степенью важности объекта (жизнеобеспечивающей и пр.), экспериментирование на котором крайне затруднено.

Поэтому в таких случаях используют не автоматические системы управления, а автоматизированные, т. е. системы с участием человека. В частности, сначала в диалоговом режиме вырабатывается управленческое решение с использованием идентификатора. Ключ K_1 разомкнут, ключи K_2 замкнуты. Конечно же, в этом случае идентификатор должен с достаточно высокой степенью точности описывать объект, включая внештатные ситуации или выход переменных процесса за пределы технологического регламента. После определения управляющего воздействия оно подается на исполнительный механизм (ключ K_1 замкнут).

Работа системы управления с идентификатором включает в себя следующие действия:

– ключ K_1 разомкнут, ключи K_2 замкнуты;

– выбирается некоторое заданное в рамках данной задачи значение выхода $Q^*(x, z)$;

– значение $Q^*(x, z)$ подставляется в алгоритм расчета управляющего сигнала;

– полученное значение $\hat{u}^l$ подается на идентификатор; отклик $\hat{x}_l$ идентификатора возвращается на управляющее устройство и подвергается анализу ЛПР;

– находится значение $\hat{x}_l^* = \hat{x}_l + \alpha(C_S, l)$, где $\alpha(C_S, l)$ – убывающая по мере увеличения S функция; l – счетчик алгоритма;

– работа алгоритма продолжается с шага 2 до тех пор, пока не будет найдено некоторое значение $u^* = \hat{u}^l$, удовлетворяющее цели управления;

– после того как u^* найдено, ключи K_2 размыкаются, а ключ K_1 замыкается, в результате чего, управляющий сигнал подается непосредственно на объект.

Таким образом, были рассмотрены задачи управления технологическими объектами и их производственными образованиями в условиях малой априорной информации, в частности, в условиях непараметрической неопределенности. При этом уделяется специальное внимание типичным для практического случая процессам, когда дискретность измерения выходной переменной процесса промышленного комплекса происходит через существенные интервалы времени, что обуславливает особенности при постановке задач идентификации и управления. Приводятся алгоритмы управления и принятия решений в иерархических системах управления в условиях неполной информации. Рассматривается также схема построения контура управления с идентификатором, которую естественно применять для процессов, имеющих существенное значение.

Следует обратить внимание на эффективность использования непараметрических моделей и алгоритмов в интеллект-системах управления и принятия решений для предприятий и отраслей с непрерывным характером технологического процесса.

Библиографическая ссылка

1. Медведев А. В. О моделировании организационных процессов // Вестник Сиб. гос. аэрокосмич. ун-т им. акад. М. Ф. Решетнева : сб. науч. тр. / под ред. проф. Г. П. Белякова. Вып. 1. Красноярск, 2000.

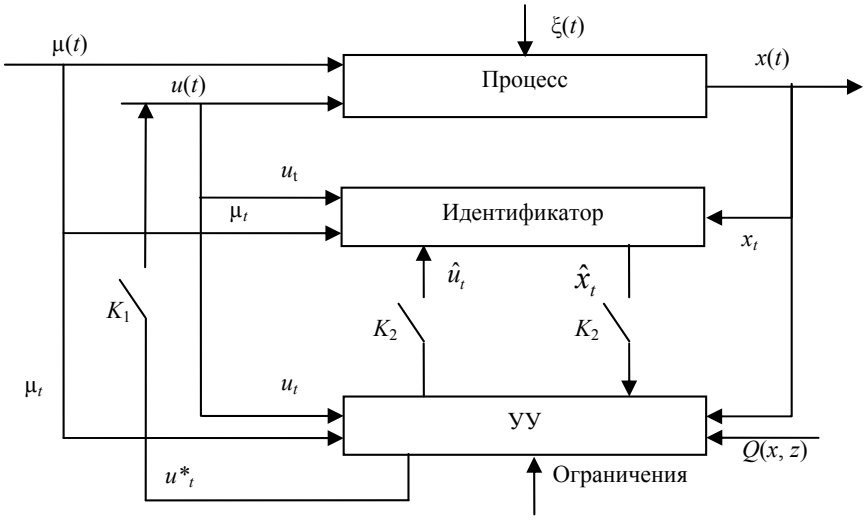


Рис. 7. Система управления с идентификатором

D. A. Ignatyev, A. N. Sergeev

NONPARAMETRIC MODELING OF ORGANIZATIONAL PROCESSES

Some modeling and managing problems of organizational systems under conditions of uncertainty are discussed in the article. Some peculiarities of organizational processes which take place in different technical and technological objects are mentioned in comparison with other processes. Nonparametric model of organizational systems and decision-making algorithms when information is not complete is offered.

Keywords: non-parametric estimation, model, identification, a priory information, organizational processes, decision making.

© Игнатьев Д. А., Сергеев А. Н., 2010

УДК 62.501

Р. Е. Козин, А. В. Медведев

О НЕПАРАМЕТРИЧЕСКОМ УПРАВЛЕНИИ ЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМОЙ

Исследуется задача управления линейной динамической системой в условиях непараметрической неопределенности. Вид уравнения, описывающего процесс, неизвестен. Предлагается оценка обратного оператора динамического процесса как алгоритм управления. Построение непараметрической модели динамической системы осуществляется на основании его переходных характеристик. Приводятся результаты исследования непараметрического алгоритма управления линейной динамической системой методом статистического моделирования.

Ключевые слова: *непараметрическая оценка, оценка обратного оператора, регрессия, моделирование, управление, идентификация.*

При управлении динамическими системами обычно предполагается известным уравнение объекта либо его передаточная функция с точностью до вектора параметров. В этом случае достаточно хорошо разработаны методы анализа и синтеза систем автоматического регулирования и управления. В теории автоматического регулирования и управления, когда нет математического описания линейного динамического

объекта, часто используются П-, ПИ- и ПИД-регуляторы. Настройка параметров последних является предметом анализа систем автоматического регулирования. В условиях непараметрической неопределенности, т. е. когда вид у уравнения, описывающего процесс, неизвестен, а известны лишь качественные характеристики, например, сведения о линейности процесса, классическая теория управления не может быть применена. Однако в рамках непараметрического подхода могут быть построены как модели исследуемого объекта, так и алгоритмы управления. В основе этого подхода лежат оценки переходных и весовых функций процесса. При этом этап выбора их с точностью до вектора параметров отсутствует. Для их восстановления используются соответствующие непараметрические оценки по результатам наблюдений переходных характеристик, «снятых» на исследуемом объекте.

Постановка задачи. Пусть динамический процесс относится к классу линейных. Кроме того, имеется возможность постановки эксперимента по «снятию» переходных характеристик объекта. Ясно, что эксперимент по «снятию» весовой функции системы провести, чаще всего, невозможно, из-за трудностей «подачи» на вход объекта дельта-функции. Заметим, что весовая функция объекта $h(t)$ – производная переходной функции $k(t)$, т. е. $h(t) = k'_t(t)$.

Известно, что для линейного динамического объекта выход объекта $x(t)$ можно представить в виде оператора $x(t) = A[u(t)]$, где $u(t)$ – входное воздействие; A – линейный оператор [1]. Если для линейного оператора A существует обратный оператор A^{-1}, то $A^{-1}A = I$, где I – единичный оператор. В этом случае $u(t) = A^{-1}[x(t)]$. Если $x(t) = x^*(t)$, где $x^*(t)$ – желаемая траектория $x(t)$, то алгоритм управления примет вид

$$u^*(t) = A^{-1}[x^*(t)], \qquad (1)$$

где $u^*(t)$ – управление, которое приводит объект в состояние $x^*(t)$. Можно считать, что в этом случае $u^*(t)$ – идеальное управление. В дальнейшем такой регулятор будем называть *u*-регулятором.

Непараметрический *u*-регулятор. Известно, что линейный оператор A можно описать следующим функционалом при нулевых начальных условиях (интеграл Дюамеля) [1]:

$$x(t) = A[u(t)] = \int_0^t h(t-\tau)u(\tau)d\tau, \qquad (2)$$

где $h(t)$ – весовая функция системы.

Известно, что оператор, обратный линейному оператору (2), имеет вид [2]:

$$u(t) = A^{-1}[x(t)] = \int_0^t \nu(t-\tau)x(\tau)d\tau, \qquad (3)$$

где $\nu(t)$ – весовая функция в направлении «выход–вход». Таким образом, если $x(t) = x^*(t)$, то *u*-регулятор будет иметь вид

$$u^*(t) = A^{-1}[x^*(t)] = \int_0^t \nu(t-\tau)x^*(\tau)d\tau. \qquad (4)$$

Однако ясно, что оператор A неизвестен, соответственно, не может быть найден и оператор A^{-1}. В данной ситуации целесообразно использовать в качестве модели (2) непараметрическую модель линейной динамической системы [2]:

$$x_N(t) = \int_0^t h_N(t-\tau, \vec{x}_N, \vec{t}_N)u(\tau)d\tau, \qquad (5)$$

где $h_N(t-\tau, \vec{x}_N, \vec{t}_N)$ – оценка весовой функции системы, а $\vec{x}_N = (x_1, ..., x_N)$ и $\vec{t}_N = (t_1, ..., t_N)$ – временные векторы.

Переходную характеристику в направлении «выход–вход» будем «снимать» на модели (5). По видимому, впервые эта идея была высказана в [3], т. е. $x_N(t) = 1(t)$. В итоге, получим выборки $\vec{u}_n = (u_1, ..., u_n)$ и $\vec{t}_n = (t_1, ..., t_n)$ – временные векторы. Для удобства дальнейшего изложения введем обозначение $\vec{u}_n = \vec{\omega}_n = (\omega_1, ..., \omega_n)$, подчеркивая тем самым, что $\vec{\omega}_n$ – значение управляющего воздействия при условии, что $x_N(t) = 1(t)$. Тогда, оценка *u*-регулятора примет вид

$$u_n(t) = \int_0^t \nu_n(t-\tau)x^*(\tau)d\tau. \qquad (6)$$

В качестве оценок весовых функций $h_N(t)$ и $\nu_N(t)$ в дальнейшем используются непараметрические статистики.

Непараметрические оценки *u*-регулятора. Пусть после эксперимента имеем выборку наблюдений $\{x_i = k_i, t_i\}$, $i = \overline{1, N}$, где N – объем выборки; k_i – наблюдаемое значение выходной переменной объекта $x(t)$ при $u(t)=1(t)$ в момент времени t_i.

Для оценки переходной функции по выборке наблюдений воспользуемся непараметрической оценкой вида

$$\hat{k}_N(t) = \sum_{i=1}^N k_i \Phi\left(\frac{t-t_i}{C_N}\right) \Big/ \sum_{i=1}^N \Phi\left(\frac{t-t_i}{C_N}\right), \qquad (7)$$

где $\hat{k}_N(t)$ – оценка переходной функции в момент времени t по выборке объемом N; $\Phi(\cdot)$ – колоколообразная функция и параметр размытости C_N удовлетворяют некоторым условиям сходимости [4].

В качестве оценки весовой функции может быть использована разностная оценка весовой функции:

$$\hat{h}_N(t) = \sum_{i=1}^N \hat{h}_i \Phi\left(\frac{t-t_i}{C_N}\right) \Big/ \sum_{i=1}^N \Phi\left(\frac{t-t_i}{C_N}\right), \qquad (8)$$

где $\hat{h}_i = \dfrac{\hat{k}(t_i + \Delta t) - \hat{k}(t_i)}{\Delta t}$, $i = \overline{1, N}$, и непараметрическая оценка весовой функции:

$$h_N(t)=\hat{k}_N'(t)=\left[\left(\sum_{i=0}^{N}x_i\Phi'\left(\frac{t-t_i}{C_N}\right)\right)\sum_{i=0}^{N}\Phi\left(\frac{t-t_i}{C_N}\right)-\right.$$
$$\left.-\left(\sum_{i=0}^{N}x_i\Phi\left(\frac{t-t_i}{C_N}\right)\right)\sum_{i=0}^{N}\Phi'\left(\frac{t-t_i}{C_N}\right)\right]\Bigg/\left(\sum_{i=0}^{N}\Phi\left(\frac{t-t_i}{C_N}\right)\right)^2. \quad (9)$$

Для выражения (9) критерий оптимальности выбора параметра размытости принимает следующий вид:

$$W(C_N)=\frac{1}{N}\sum_{i=1}^{N}\left(k_i-\int_0^{t_i}\hat{h}_N(t_i,C_N)\right)^2\to\min_{C_N}. \quad (10)$$

Оценив весовую функцию объекта, мы построили оценку оператора, т. е. модель объекта в виде (5). Используя построенную модель, найдем обратную переходную функцию объекта. Для этого запишем модель вида (5) в дискретном виде:

$$x[i]=\sum_{j=0}^{i_i}h[i-j]u[j]\Delta t, \quad (11)$$

где $x[i]$, $h[i]$, $u[i]$ – значения функций $x(t)$, $h(t)$, $u(t)$ в момент времени t_i. Фиксируя на выходе функцию Хэвисайда ($x[i]=1[i]$), найдем реализацию $u[i]=\omega[i]$:

$$1[i]=\sum_{j=0}^{i}h[i-j]\omega[j]\Delta t=\sum_{j=0}^{i-1}h[i-j]\omega[j]\Delta t+h[0]\omega[t]\Delta t, \quad (12)$$

$$\begin{cases}\omega[0]=1/h[0]\Delta t,\\ \omega[i]=\left[1-\sum_{j=0}^{i-1}h[i-j]\omega[j]\Delta t\right]\Big/h[0]\Delta t.\end{cases} \quad (13)$$

Вычислительные соотношения (13) дают выборку переходной характеристики объекта в направлении «выход–вход». Непараметрическая оценка переходной функции в направлении «выход-вход» или «обратной» переходной функции имеет вид:

$$\hat{\omega}_n(t)=\sum_{i=1}^{n}\hat{\omega}_i\Phi\left(\frac{t-t_i}{C_n}\right)\Bigg/\sum_{i=1}^{n}\Phi\left(\frac{t-t_i}{C_n}\right). \quad (14)$$

Производная от этой функции есть весовая функция в направлении «выход–вход», или «обратная» весовая функция. Оценка производной «обратной» переходной функции через конечные разности имеет вид

$$\hat{\nu}_n(t)=\sum_{i=1}^{n}\hat{\nu}_i\Phi\left(\frac{t-t_i}{C_n}\right)\Bigg/\sum_{i=1}^{n}\Phi\left(\frac{t-t_i}{C_n}\right), \quad (15)$$

где $\hat{\nu}_i=\dfrac{\hat{\omega}(t_i+\Delta t)-\hat{\omega}(t_i)}{\Delta t}$, $i=\overline{1,n}$.

Непараметрическая оценка «обратной» весовой функции $\nu(t)$ имеет вид

$$\nu_n(t)=\hat{\omega}_n'(t)=\frac{\left(\sum_{i=1}^{n}\omega_i\Phi'\left(\frac{t-t_i}{C_n}\right)\right)\sum_{i=1}^{n}\Phi\left(\frac{t-t_i}{C_n}\right)-}{\left(\sum_{i=1}^{n}\Phi\left(\frac{t-t_i}{C_n}\right)\right)^2}$$
$$\frac{-\left(\sum_{i=1}^{n}\omega_i\Phi\left(\frac{t-t_i}{C_n}\right)\right)\sum_{i=1}^{n}\Phi'\left(\frac{t-t_i}{C_n}\right)}{\left(\sum_{i=1}^{n}\Phi\left(\frac{t-t_i}{C_n}\right)\right)^2}. \quad (16)$$

Непараметрическая оценка (6) (u_n-регулятор) будет иметь вид

$$u(t_k)=\sum_{j=0}^{k}\hat{\nu}_N(t_k-t_j)x^*(t_j)\Delta t. \quad (17)$$

Численные исследования. В качестве объекта исследования рассмотрим динамический объект, поведение которого описывается дифференциальным уравнением первого порядка:

$$a_1\frac{dx(t)}{dt}+a_0x(t)=\alpha u(t), \quad (18)$$

где $a_1=1$, $a_0=1$, $\alpha=1$.

Переходная функция системы – $k(t)=1-e^{-t}$, а весовая функция имеет вид $h(t)=k'(t)=e^{-t}$. «Обратная» переходная функция – $w(t)=\delta(t)+1(t)$. Ее дискретный аналог с шагом дискретизации Δt = 0,05 представлен на рис. 1. Обратная весовая функция – $\nu(t)=w'(t)=\delta'(t)+\delta(t)$, дискретный аналог которой с шагом дискретизации Δt = 0,05 представлен на рис. 2.

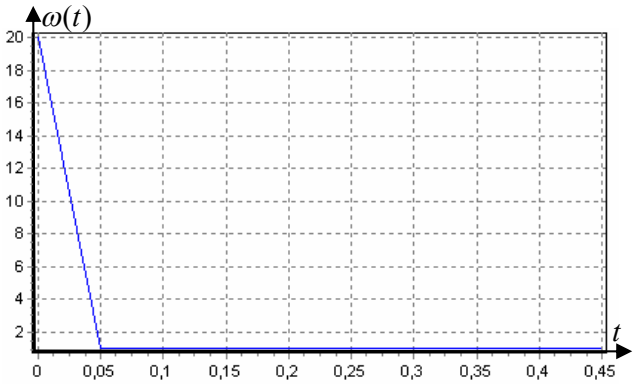


Рис. 1. $w(t)$ в дискретном виде, первые 10 тактов

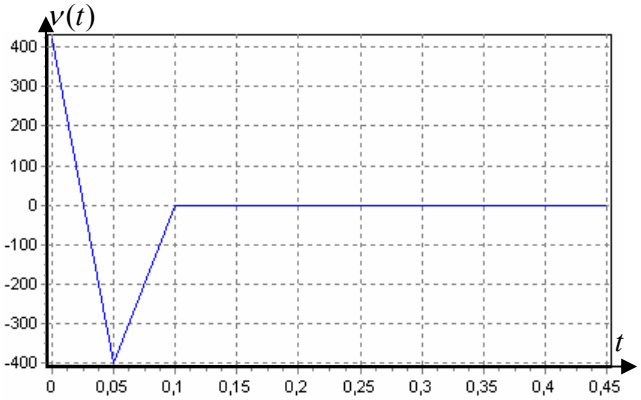


Рис. 2. $\nu(t)$ в дискретном виде, первые 10 тактов

Управление для задания $x^*(t)$ = sign(sin(t/4)) представлено на рис. 3, реакция объекта на это управление, совпадающая с заданием, представлено на рис. 4.

Управление для случайного задания (случайный сигнал в области [–1; 1]) и реакция объекта на это

управление, совпадающая с заданием, представлены на рис. 5 и рис. 6, соответственно.

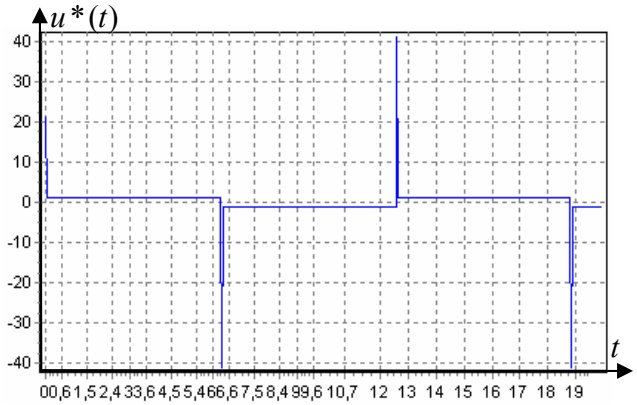


Рис. 3. Управление для задания $x^*(t)$= sign(sin(t/4))

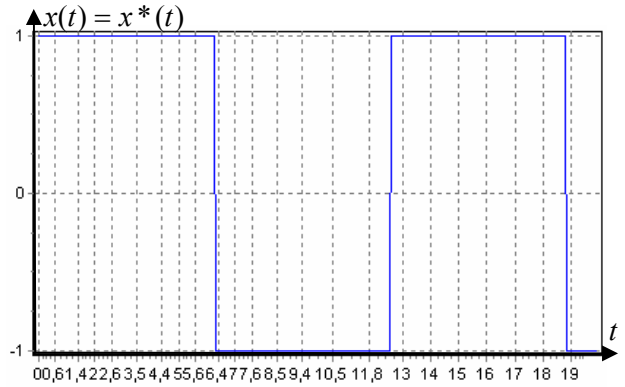


Рис. 4. Выход объекта

Теперь вернемся к исходной постановке задачи, где в качестве исходной информации используем выборку наблюдений входной и выходной величины. Решая численным методом (методом Рунге–Кутта) уравнение (18), сгенерируем выборку наблюдений, подав на вход функцию Хэвисайда.

Получим выборку переходной характеристики объекта объемом N. В данном случае объем выборки составил N = 400 точек на временном интервале [0; 20], шаг дискретизации Δt = 0,05. Для получения переходной функции по выборке используется непараметрическая оценка (7). Для решения «проблемы концов» при непараметрическом оценивании функции, исходная выборка дополняется десятью точками на концах интервала. Полученный результат представлен на рис. 7.

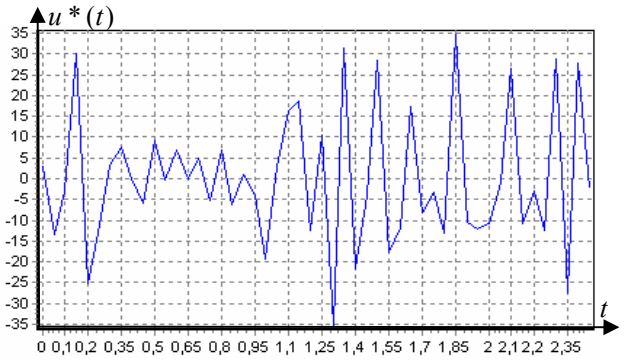


Рис. 5. Управление для случайного задания

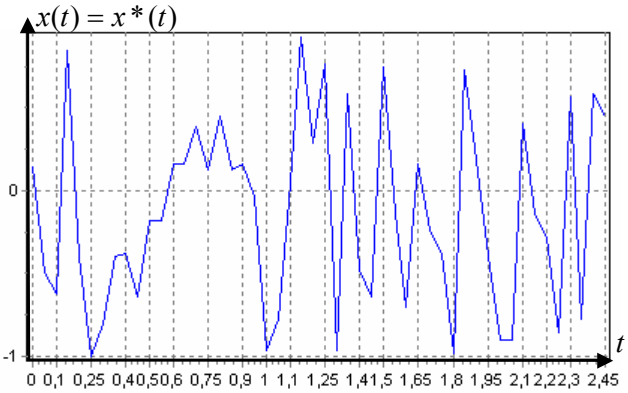


Рис. 6. Выход объекта

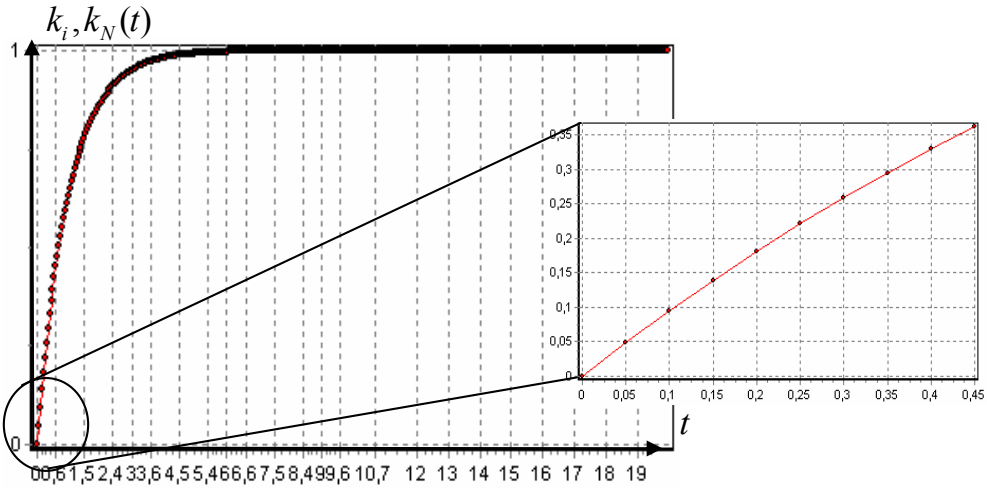


Рис. 7. Исходная выборка наблюдений и построенная непараметрическая оценка переходной функции

Рассмотрим два способа: оценивание через конечные разности и производную непараметрической оценки. В дальнейшем, будем приводить вычислительные результаты, полученные в случаях использования выражений (8) и (9), в сравнительной форме.

Оценки весовой функции для выражений (8) и (9) представлены на рис. 8. «Обратная» переходная функция представлена рис. 9.

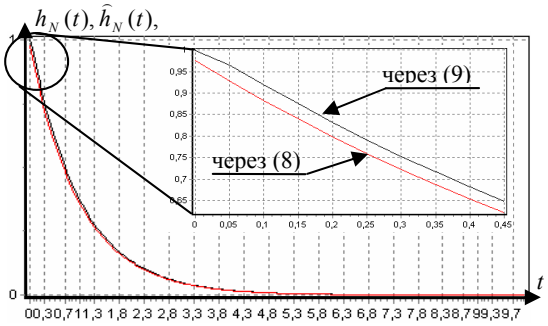


Рис. 8. Весовая функция системы. Первые 200 тактов

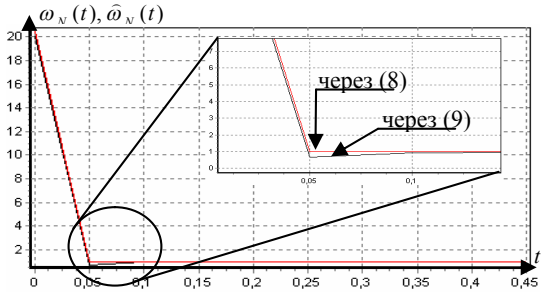


Рис. 9. Обратная переходная функция системы. Первые 10 тактов

Имея обратную переходную функцию системы, найдем обратную весовую функцию системы (рис. 10).

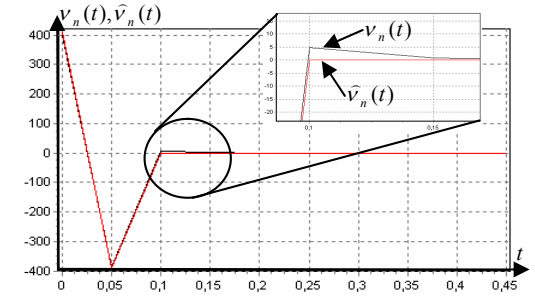


Рис. 10. Обратная весовая функция системы, первые 10 тактов

Получив оценку обратной весовой функции системы, мы построили оценку обратного оператора системы в виде (8). Оценка обратного оператора системы представляет собой регулятор системы. В качестве задания возьмем три траектории: $x^*(t) = \sin(t)$, $x^*(t) = \text{sign}(\sin(t/4))$, случайный сигнал в диапазоне $[-1; 1]$. Графические результаты рассчитанного управляющего сигнала и реакции объекта в сравнении с заданием приведены на рис. 11–16.

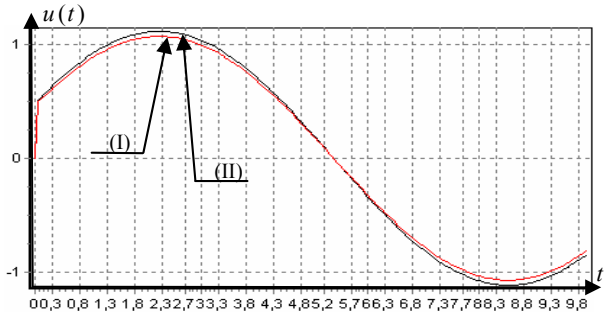


Рис. 11. Управление для задания $x^*(t) = \sin(t)$. Первые 200 тактов:
I – алгоритм через аналитические оценки производных;
II – алгоритм через конечные разности

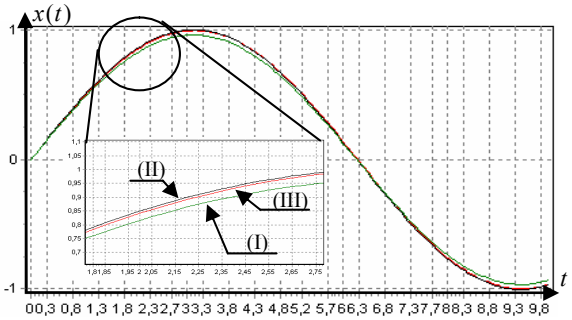


Рис. 12. Выход объекта и задание. Первые 200 тактов:
I – алгоритм через аналитические оценки производных;
II – алгоритм через конечные разности; *III* – задание $x^*(t) = \sin(t)$

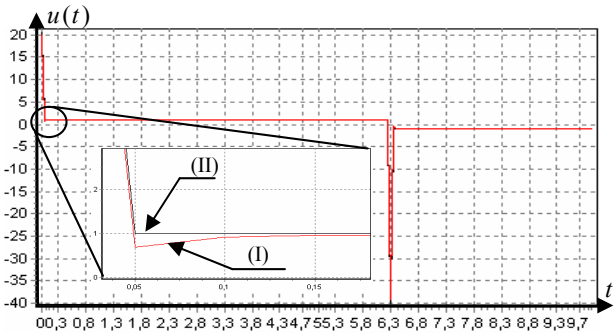


Рис. 13. Управление для задания $x^*(t) = \text{sign}(\sin(t/4))$. Первые 200 тактов:
I – алгоритм через аналитические оценки производных;
II – алгоритм через конечные разности

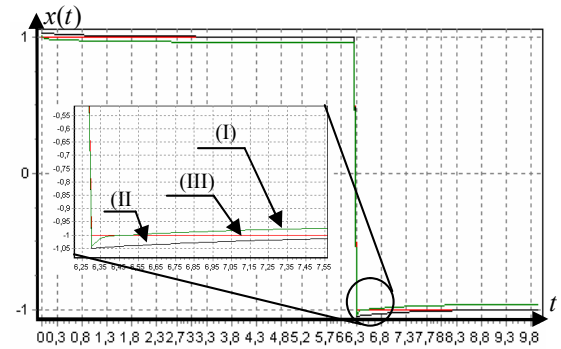


Рис. 14. Выход объекта и задание. Первые 200 тактов:
I – алгоритм через аналитические оценки производных;
II – алгоритм через конечные разности; *III* – задание $x^*(t) = \text{sign}(\sin(t/4))$

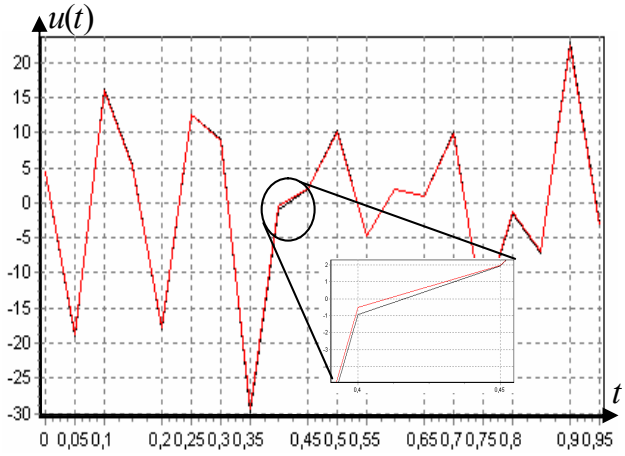


Рис. 15. Управление для случайного задания. Первые 20 тактов

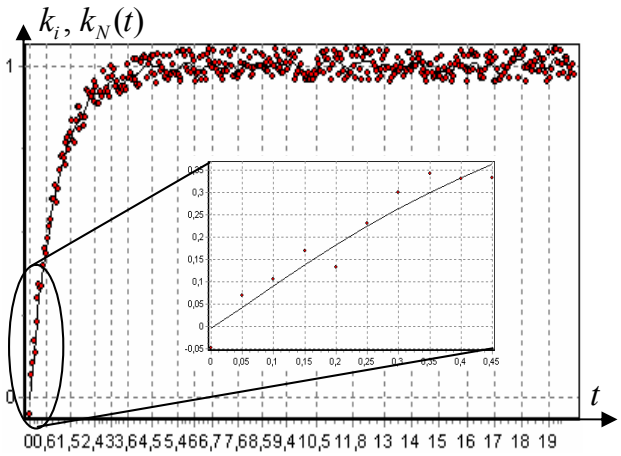


Рис. 17. Точки выборки и восстановленная переходная функция

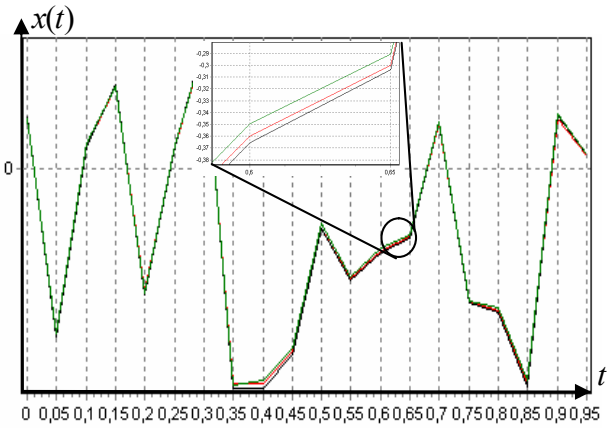


Рис. 16. Выход объекта и задание. Первые 20 тактов

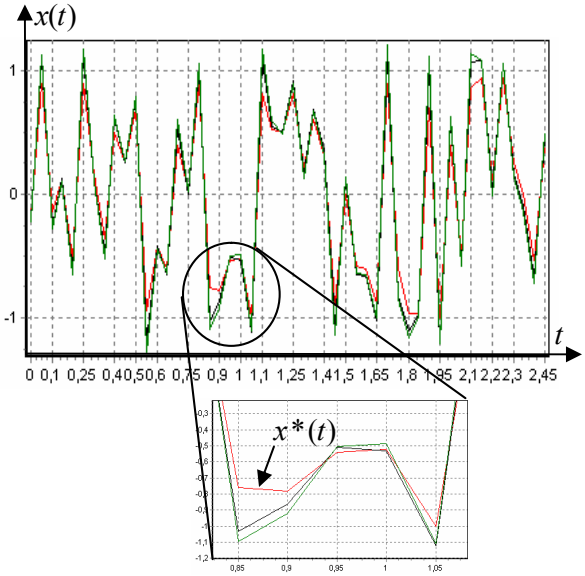


Рис. 18. Выход объекта и задание. Первые 50 тактов

Результаты статистического моделирования в случае наличия аддитивной помехи 5 % с нормальным законом распределения в выходном канале измерения приведены на рис. 17–19.

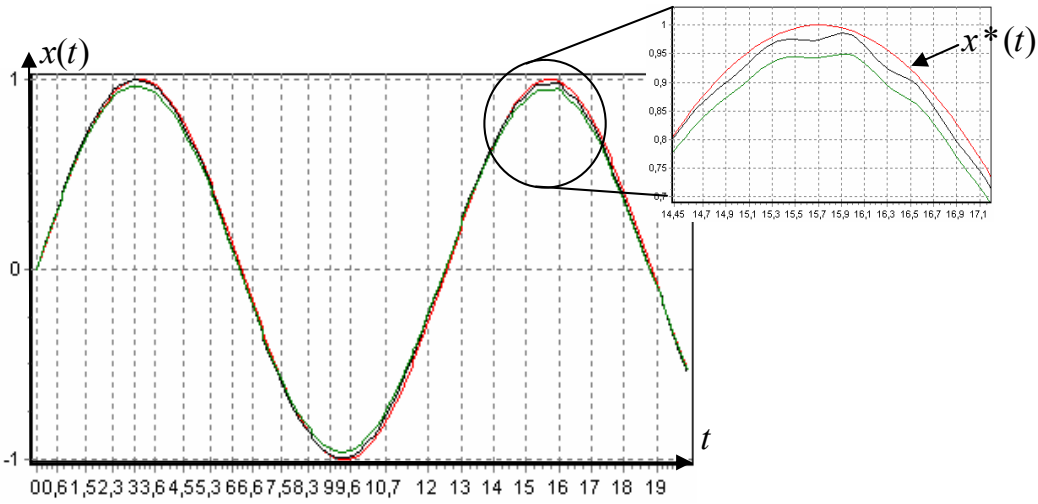


Рис. 19. Выход объекта и задание $x^*(t) = \sin(t)$

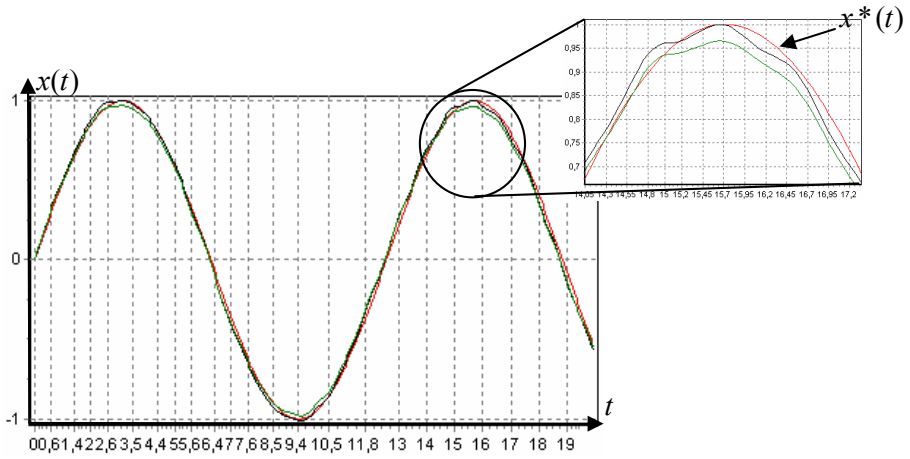


Рис. 20. Реакция объекта в сравнении с заданием $x^*(t) = \sin(t)$

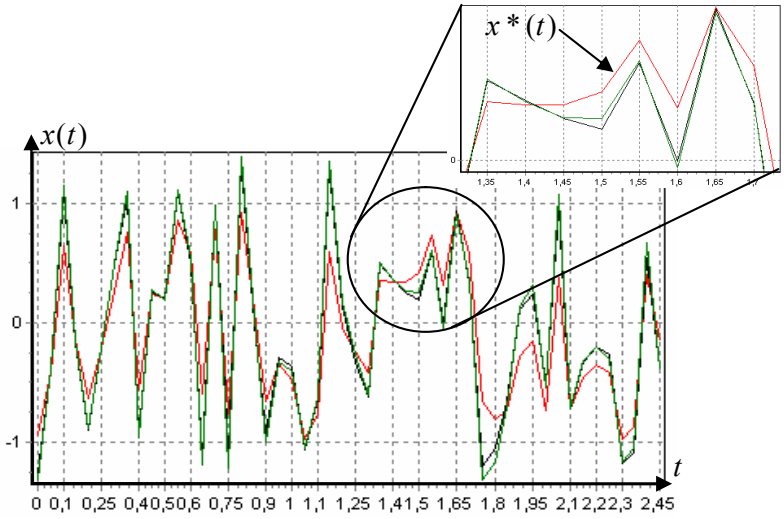


Рис. 21. Выход объекта и задание. Первые 50 тактов

Результаты статистического моделирования в случае наличия аддитивной помехи 10 % с нормальным законом распределения в выходном канале измерения приведены на рис. 20, 21.

Библиографические ссылки

1. Заде Л., Дезоер Ч. Теория линейных систем. М. : Наука, 1970.

2. Куликовский Р. Оптимальные адаптивные процессы. М. : Наука, 1967.

3. Medvedev A. V. Identification and control for linear Dynamic System of unknown Order // Optimization Techniques : IFIP Technical Conf. Berlin–Heidelberg–New York, 1975. P. 48–55.

4. Медведев А. В. Теория непараметрических систем. Общий подход // Вестник СибГАУ. Вып. 3. Красноярск, 2008.

R. E. Kozin, A. V. Medvedev

ABOUT NONPARAMETRIC CONTROL OF A LINEAR DYNAMIC SYSTEM

The paper considers a problem of control of linear dynamic system in conditions of nonparametric uncertainty. A type of equation describing the process, is indeterminate. Estimation of reverse operator of the dynamic process as the algorithm of control is offered. A nonparametric model of a dynamic system is realized on the basis of its transfer features. The results of the research of the nonparametric algorithm of control by a linear dynamic system using the method of statistic modeling.

Keywords: *nonparametric estimation, estimation of the reverse operator, regression, modeling, control, identification.*

© Козин Р. Е., Медведев А. В., 2010

УДК 62.501

А. В. Медведев, А. В. Фаустов

О НЕПАРАМЕТРИЧЕСКОЙ ОЦЕНКЕ НЕСТАЦИОНАРНОЙ ФУНКЦИИ РЕГРЕССИИ ПО НАБЛЮДЕНИЯМ

Рассматривается проблема восстановления функции регрессии по наблюдениям со случайными ошибками. Предлагаются непараметрические оценки нестационарной функции регрессии. Приводятся результаты статистического моделирования непараметрических алгоритмов. Рассматривается задача моделирования статических многомерных нестационарных систем с запаздыванием.

Ключевые слова: непараметрические оценки, дискретно-непрерывные процессы, нестационарность, регрессия, идентификация, моделирование.

В многочисленных приложениях часто возникает необходимость в восстановлении функции регрессии $M\{x|u\}$ по наблюдениям случайных величин (u_i, x_i), $i=\overline{1,s}$, где s – объем выборки наблюдений. В классе непараметрических оценок принята статистика, которая будет представлена далее [1]. Предполагается, что плотность вероятности $p(u)$, $\forall u$, $u \in \Omega(u)$, $p(u) > 0$ неизвестна. При этом предполагается, что $x(u)$ не меняется с течением времени. Поскольку при моделировании дискретно-непрерывных процессов широко используются регрессионные модели, то существенное значение имеет факт дрейфа во времени характеристик исследуемого объекта. Среди многих факторов, влияющих на последнее, отметим только безусловное старение оборудования, других технических средств, где и протекает интересующий нас процесс. Однако укажем кратко и на другие причины, влияющие на дрейф характеристики процесса, а также отметим некоторые его проявления при моделировании:

– наличие неизвестных переменных либо переменных, не поддающихся измерению (естественно, что для успешного решения задачи идентификации с присутствием неконтролируемых входных переменных процесса их значения должны меняться медленно; в противном случае, и тем более, когда значения таких переменных случайны и приводят к осциллирующему процессу, ожидать хороших моделей не приходится);

– изменение параметров модели объекта во времени (речь в данном случае идет не об изменении параметров модели, как это происходит в ряде рекуррентных алгоритмов идентификации, когда наблюдается сходимость параметров некоторым оценкам, а о дрейфе параметров, связанных именно с изменчивостью объекта идентификации);

– изменение структуры объекта во времени (редко встречающаяся на практике ситуация, тем не менее, интересная с точки зрения теории, побуждающая к построению более универсальных, адаптивных алгоритмов идентификации).

Приведем примеры изменения характеристики нестационарного процесса. Из соображений простоты рассмотрим одномерные статические объекты (рис. 1). На рис. 1, *а* и 1, *в* представлены истинные характеристики стационарного и нестационарного объекта. Характеристика последнего изменилась за некоторый интервал времени $[t_1, t_2]$. Точки выборки, полученные с обоих объектов, показаны на рис. 1, *б* и 1, *г*. Объемы выборки наблюдений одинаковы, на выход объекта наложена одинаковая аддитивная центрированная помеха.

На рис. 1 приняты следующие обозначения: u – входная переменная процесса; x – выходная переменная. Приведенная иллюстрация, помимо разъяснения, о каких процессах идет речь, подводит к следующей задаче: необходимо определить, имеем ли мы дело со стационарным процессом и присутствуют ли помехи в каналах измерений, или имеющаяся выборка отражает поведение нестационарного процесса, а не просто несет в себе искажения, вносимые помехами измерений.

Далее рассмотрим непараметрические оценки функции регрессии по наблюдениям при неизвестном дрейфе изучаемого процесса.

Непараметрические оценки функции регрессии. Непараметрическая оценка функции регрессии по наблюдениям с ошибками $\{x_i, u^{(1)}{}_i, u^{(2)}{}_i, \ldots, u^{(m)}{}_i\}$, $i=\overline{1,s}$, где m – размерность вектора входных переменных, имеет вид [1; 2]

$$x_s(u_1,\ldots,u_m) = \sum_{i=1}^{s} x_i \left[\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)\right] \Bigg/ \sum_{i=1}^{s}\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right), \quad (1)$$

где финитная колоколообразная функция $\Phi(\cdot)$ удовлетворяют некоторым условиям сходимости [2]:

$$\lim_{s\to\infty} \frac{1}{c_s} \Phi\left((u-u_i)/c_s\right) = \delta(u-u_i),$$

$$\frac{1}{c_s} \int_{\Omega(u)} \Phi\left((u-u_i)/c_s\right) du = 1, \quad (2)$$

где $\delta(u–u_i)$ – дельта-функция Дирака. Параметр размытости c_s удовлетворяет следующим условиям сходимости [1; 2]:

$$c_s > 0, \quad \lim_{s\to\infty} c_s = 0, \quad \lim_{s\to\infty} s \cdot c_s{}^m = \infty. \quad (3)$$

Асимптотические свойства непараметрической оценки (1) достаточно хорошо исследованы, в частности, сходимость оценки x_s доказана в среднеквадратичном и почти наверное [3; 4]. Однако эта непараметрическая оценка (1) непригодна для восстановления регрессионных характеристик нестационарного объекта.

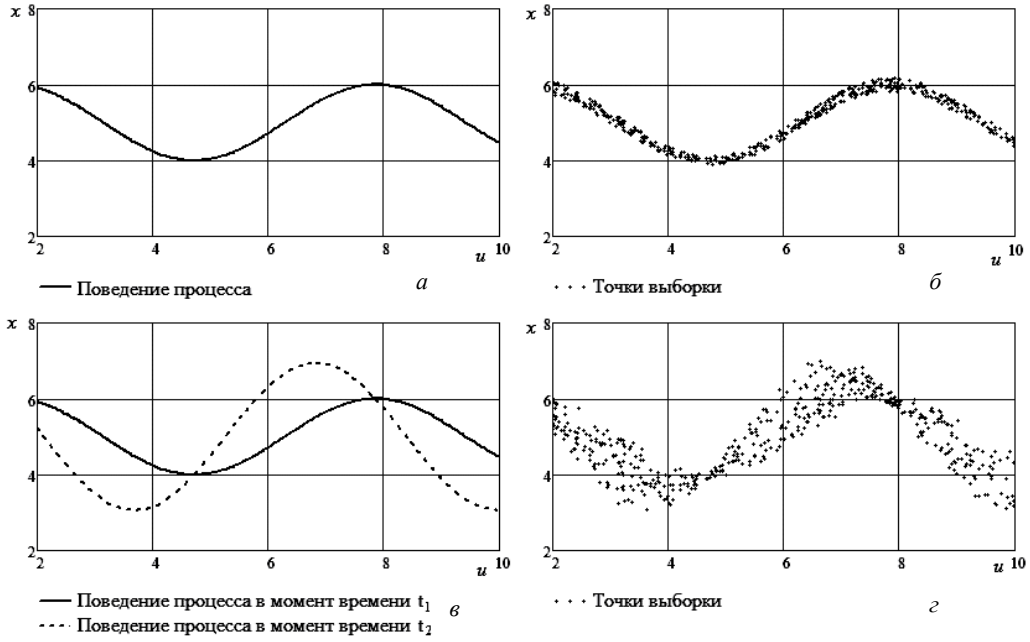


Рис. 1

Нестационарная непараметрическая оценка функции регрессии. Пусть истинная характеристика процесса изменяется во времени в пространстве входных и выходных переменных заранее непредвиденным образом. В этом случае использование всей выборки наблюдений, проводимых в оценке (1), будет искажать оценку функции регрессии, и обусловлено это будет тем, что наблюдения, ранее полученные на объекте, не отражают свойств объекта в настоящий момент времени. Естественно считать, что «старая» информация в виде элементов выборки $\{x_i,\ u^{(1)}{}_i,\ u^{(2)}{}_i,\ \ldots,\ u^{(m)}{}_i,\ I = 1,\ 2,...\}$ должна использоваться с меньшим весом.

Идея непараметрического алгоритма, учитывающего временной дрейф характеристик исследуемого процесса, состоит в том, что «старая» информация в непараметрической оценке функции регрессии получает меньший вес [2].

Для восстановления функции регрессии в нестационарных условиях $x(u, t)$ предлагается ввести в непараметрическую оценку функции регрессии (1) некоторую функцию «памяти», убывающую с увеличением аргумента $\rho = (s–i)$, где i – дискретное время поступления информации, $I = 1, 2, \ldots s$; s – текущее время [2]. Тогда непараметрическая оценка нестационарной функции регрессии по наблюдениям (x, u) будет иметь вид

$$x_s^*(u_1,...,u_m) = \frac{\sum_{i=1}^{s} x_i \left[\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)\right] \rho(s-i)}{\sum_{i=1}^{s}\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right) \rho(s-i)}, \qquad (4)$$

где $\rho = (s–i)$ – функция «памяти», удовлетворяющая следующим свойствам:

1) $i = s \Rightarrow \rho(s-i) = \rho(0) = 1$;

2) $\forall i, j \in [1, s],\ i > j \Rightarrow \rho(s-i) \geq \rho(s-j)$;

3) $i = 1 \Rightarrow \rho(s-i) = \rho(s-1) \geq 0$.

Приведем варианты вида функции памяти $\rho(s – i)$:

$$\rho(s-i, \alpha_1, \alpha_2) = \left[\frac{-(s-i)+\alpha_1}{|(s-i)-c_1|+\alpha_2}\, 0{,}5 + 0{,}5\right]\alpha_3, \qquad (5)$$

$$\alpha_3 = \frac{0{,}5\cdot\alpha_1}{\alpha_1+\alpha_2} + 0{,}5, \qquad (6)$$

где параметр $\alpha_1 \in (0, s)$ определяет положение точки перегиба функции; параметр $\alpha_2 \in (0, \infty)$ определяет «крутизну» функции; параметр α_3 нормирует функцию «памяти» на единицу. Поведение функции (5) (функции «памяти») при различных значениях параметров α_1 и α_2 приведено на рис. 2.

Также функция «памяти» может быть кусочно-постоянной, приводя к скользящему аналогу непараметрической оценки (1).

Введем показатель изменчивости $\delta(t_1, t_2)$ характеристики исследуемого процесса в следующем виде:

$$\delta(t_1, t_2) = \int_{\Omega(u)} |x(u, t_1) - x(u, t_2)|\, du. \qquad (7)$$

Обозначим дискретность измерения переменных (x, u) через Δt. Величина Δt, естественно, будет зависеть от показателя изменчивости $\delta(t_1, t_2)$, где $t_1 < t_2$. Можно считать, что при медленно меняющейся во времени характеристике процесса, т. е. при малых значениях $\delta(t_1, t_2)$, дискретность измерений Δt «входных–выходных» переменных процесса (x, u) больше, чем в случае больших значений $\delta(t_1, t_2)$. В последнем случае параметр Δt – дискретность измерения «входных–выходных» переменных, следует уменьшить. Анализ этого вопроса требует дополнительных исследований имеющейся выборки наблюдений $\{x_i,\ u^{(1)}{}_i,\ u^{(2)}{}_i,\ \ldots,\ u^{(m)}{}_i,\ I = 1, 2, ...\}$.

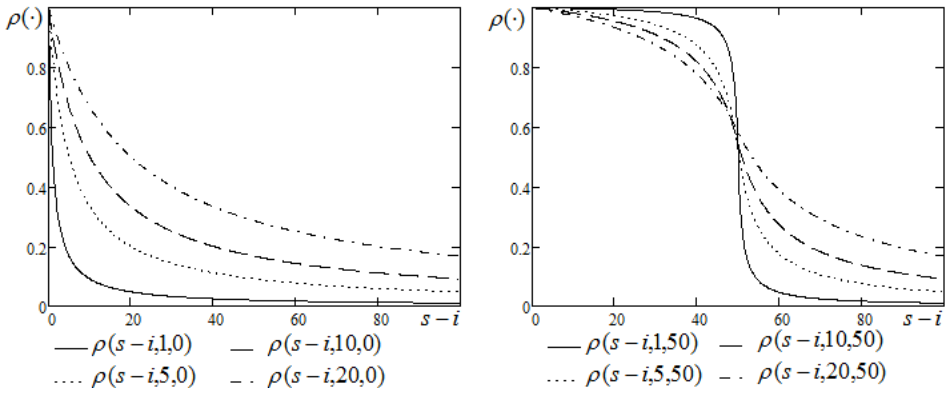


Рис. 2

Другой подход к попытке оценить изменчивость исследуемого процесса состоит в выявлении подобластей в пространстве входных и выходных переменных, которые могут отражать нестационарность процесса. Он основан на применении условной дисперсии $D\{x|u\} = M\{x^2|u\} - [M\{x|u\}]^2$ [5]. Условная дисперсия выходной переменной может быть оценена следующей непараметрической статистикой:

$$D_s(x \mid u) = \frac{\sum_{i=1}^{s} x_i^2 \left[\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)\right]}{\sum_{i=1}^{s}\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)} - \left[\frac{\sum_{i=1}^{s} x_i \left[\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)\right]}{\sum_{i=1}^{s}\prod_{j=1}^{m} \Phi\left(\frac{u^j - u^j{}_i}{c_s{}^j}\right)}\right]^2. \quad (8)$$

Большие значения оценки (8) (к примеру, выше некоторого порогового значения) могут сигнализировать об изменчивости объекта в рассматриваемых подобластях пространства наблюдений.

Нестационарные адаптивные модели и алгоритмы. Рассмотрим нестационарные адаптивные модели.

Распознавание образов. Оценка (4) с использованием функции «памяти», помимо применения в задаче оценки нестационарной функции регрессии, также может применяться и к непараметрическим алгоритмам классификации, а необходимость в них возникает, когда обучающая выборка накапливалась в течение некоторого времени, за которое расположение классов в признаковом пространстве изменилось, при этом классические алгоритмы классификации, не учитывающее изменение классов во времени, естественно, будут давать плохие результаты.

Рассмотрим случай, когда некоторая ситуация S может быть отнесена к одному из двух классов V_1 или V_2 и характеризуется значениями вектора признаков $\nu = (\nu_1, \ldots, \nu_s)$. Задача классификации сводится к построению решающего правила на основании обучающей выборки $\left\{\vec{v}_s = (v_1,\ldots,v_m), \vec{U}_s\right\}$ объемом s, где $\vec{U}_s$ – указание учителя о принадлежности S к классу V_1 или V_2. В этом случае модификация непараметрического алгоритма классификации, учитывающая изменчивость классов, имеет вид

$$U_s(v_1,\ldots,v_m) = \frac{\sum_{i=1}^{s} U_i \left[\prod_{j=1}^{m} \Phi\left(\frac{v^j - v^j{}_i}{c_s{}^j}\right)\right]\rho(s-i)}{\sum_{i=1}^{s}\prod_{j=1}^{m} \Phi\left(\frac{v^j - v^j{}_i}{c_s{}^j}\right)\rho(s-i)}. \quad (9)$$

Управление. Аналогичную предыдущей модификацию можно предложить и для непараметрического алгоритма управления статическим нестационарным объектом:

$$u_s(x) = \frac{\sum_{i=1}^{s} u_i \Phi\left(\frac{x^* - x_i}{c_s}\right)\rho(s-i)}{\sum_{i=1}^{s} \Phi\left(\frac{x^* - u_i}{c_s}\right)\rho(s-i)}, \quad (10)$$

где x^* – требуемое значение выхода объекта.

Статистическое моделирование нестационарных непараметрических оценок функции регрессии. При исследовании непараметрической оценки нестационарной функции регрессии (4), а также при сравнении результатов с использованием оценки (1) использовался метод статистического моделирования. В качестве уравнения, описывающего моделируемый объект, было выбрано следующее:

$$x(\bar{u}_i) = \sin u_i^{(1)} + \frac{i}{250}\cdot\cos\left(u_i^{(2)}\right) - u_i^{(3)}\cdot u_i^{(4)} + \sin\left(u_i^{(5)} + 0{,}5\right), \quad (11)$$

где $i \in [1, s]$, s – объем выборки; $u^{(1)} \in [0,5]$, $u^{(2)} \in [0,5]$, $u^{(3)} \in [0,5]$, $u^{(4)} \in [0,5]$, $u^{(5)} \in [0,3]$. Значения входных переменных u генерировались случайно в соответствии с равномерным законом распределения в заданных интервалах.

Для имитации функционирования объекта в условиях помех различной интенсивности на имитируемый выход объекта (11) накладывается центрированная аддитивная помеха по следующему правилу:

$$x_i = x_i^* + \xi_i, \quad (12)$$

где x_i – выход объекта с наложенной на него помехой; x_i^* – выход объекта без помехи; ξ – случайная величина, подчиняющаяся нормальному закону распределения $N(0, \sigma)$, а значение среднеквадратичного отклонения σ определяется в зависимости от заданного уровня помехи e (%) согласно выражению

$$\sigma = \frac{e\left(\begin{matrix}\max(x(u \mid i=0), x(u \mid i=s)) - \\ -\min(x(u \mid i=0), x(u \mid i=s))\end{matrix}\right)}{6 \cdot 100\ \%}, \qquad (13)$$

в этом случае интервал 6σ будет соответствовать заданному в процентах интервалу от изменения выхода объекта.

Оценка качества моделирования проводилась с использованием следующего критерия:

$$W = \sqrt{\frac{1}{s} \cdot \sum_{l=1}^{s} \left(x(u_l) - x_s(u_l, c_s)\right)^2}, \qquad (14)$$

где $x(u_l)$ – значение выходной переменной в l-й точке выборки в момент времени s; $x_s(u_l, c_s)$ – значение выхода модели (1), построенной по имеющейся выборке в той же точке. Аналогично будем проводить оценку качества моделирования при использовании непараметрической модели (4):

$$W' = \sqrt{\frac{1}{s} \cdot \sum_{l=1}^{s} \left(x(u_l) - x_s^*(u_l, c_s)\right)^2}, \qquad (15)$$

где $x_s^*(u_l, c_s)$ – значение выхода оценки (4).

Оценку ошибок моделирования по нескольким экспериментам, проводимым в одних и тех же условиях, будем проводить усреднением по серии экспериментов.

Результаты исследования качества моделирования в зависимости от уровня шума e (%) приведены в табл. 1, усреднение проводилось по 50 экспериментам при каждом уровне шума, объем выборки s = 3 000.

Сравнение зависимости среднеквадратичных оценок ошибок моделирования (14) и (15) для стандартной (1) и модифицированной (4) непараметрических оценок функции регрессии от уровня шума e (%) приведено на рис. 3.

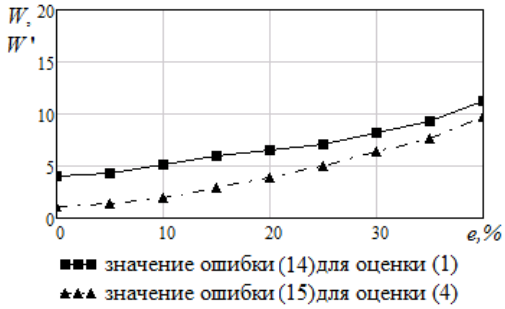


Рис. 3

Можно сделать вывод, что исследуемый алгоритм чувствителен к помехам, а также к точкам выбросов, ввиду того, что больший вес придается последним наблюдениям, даже если они оказываются ошибочными.

Приведем результаты моделирования при увеличении объема выборки за счет уменьшения шага дискретизации снятия наблюдений. То есть увеличим число наблюдений при условии, что последнее наблюдение будет происходить при одинаковых параметрах объекта (табл. 2). Усреднение проводилось по 50 экспериментам при каждом объеме выборки, на выход объекта наложена 5%-я помеха.

Графически сравнение зависимостей среднеквадратичных ошибок моделирования (14) и (15), указанных в табл. 2, представлено на рис. 4.

Как видно из рис. 4, при уменьшении шага дискретизации снятия наблюдений по времени (а в данном случае именно за счет этого увеличивался объем выборки) заметна некоторая сходимость модифицированного алгоритма (4).

Увеличим объем наблюдений, оставив шаг дискретизации снятия наблюдений постоянным, сохранив исходный характер нестационарности (11). Объем выборки наблюдений s будет равен 3 000, 5 000, 7 000, 9 000, 12 000. Результаты приведены в табл. 3.

Сравнение зависимостей относительных ошибок моделирования (14) и (15) для стандартной и модифицированной непараметрических оценок функции регрессии по проведенной серии экспериментов представлено на рис. 5.

Таблица 1

Значение усредненных оценок ошибок моделирования в зависимости от уровня помех

e, %	0	5	10	15	20	25	30	35	40
W	4,08683	4,35962	5,1858	5,97146	6,50943	7,11131	8,16638	9,34669	11,26575
W'	1,10454	1,42067	1,91831	2,9245	3,90674	4,93265	6,40269	7,70785	9,69883

Таблица 2

Значение усредненных оценок ошибок моделирования в зависимости от объема выборки

s	3 000	5 000	7 000	9 000	11 000	20 000
W	4,35962	4,35114	4,28764	4,25306	4,20386	4,18571
W'	1,42067	0,71623	0,30293	0,15271	0,14397	0,11291

Таблица 3

Значение усредненных оценок ошибок моделирования в зависимости от объема выборки

S	3 000	5 000	7 000	9 000	12 000
$E_{отн}$	4,35962	5,65906	8,00766	11,24024	17,08449
$E^*_{отн}$	1,42067	1,56532	1,62685	1,7869	2,05778

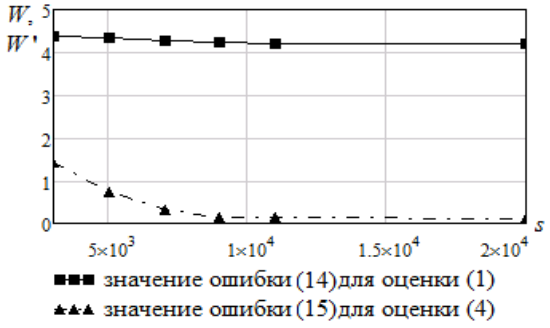


Рис. 4

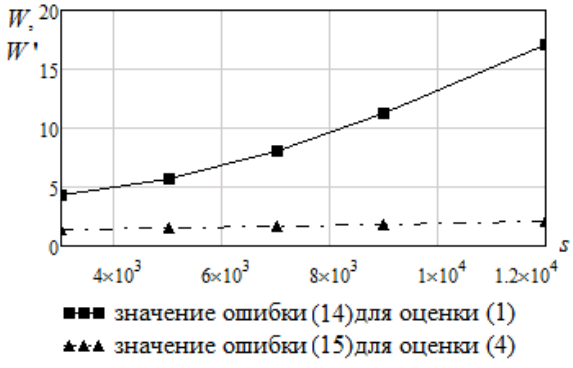


Рис. 5

Последний рисунок, отражает наиболее существенный момент, отмеченный ранее: непараметрическая оценка (1) непригодна для восстановления регрессионных характеристик нестационарного объекта.

В ходе численных исследований определено, что использование предлагаемой непараметрической оценки нестационарной функции регрессии дает более точные результаты оценивания, чем известная непараметрическая оценка, кроме случаев с большим значением помехи. Рассматриваемый алгоритм более чувствителен к помехам.

Однако в любом случае адекватность получаемой модели зависит от реального объекта, скорости изменения характеристик объекта, частоты дискретизации снятия наблюдений, наличия априорной информации и качества получаемых данных.

При исчерпывающей выборке наблюдений позволительно использовать лишь ее часть для получения более точных оценок. Предлагаемый вид функции «памяти» рекомендуется использовать в тех ситуациях, когда невозможно определить вес каждого наблюдения конкретно. В противном случае, необходимо видоизменять функцию «памяти» в соответствии с имеющейся информацией.

Библиографические ссылки

1. Медведев А. В. Элементы теории непараметрических систем управления // Актуальные проблемы информатики, прикладной математики и механики. Ч. 3. Информатика. Новосибирск ; Красноярск : Изд-во СО РАН. 1996. С. 87–112.

2. Медведев А. В. Адаптация в условиях непараметрической неопределенности // Адаптивные системы и их приложения. Новосибирск : Наука, 1978. С. 4–34.

3. Кошкин Г. М., Пивен И. Г. Непараметрическая идентификация стохастических объектов : науч. пособие // Хабаровск : ДВО РАН, 2009. С. 336.

4. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей / отв. ред. Н. А. Кузнецов. М. : Наука, 2004. С. 508.

5. Дисперсионная идентификация / Н. С. Райбман [и др.] ; под ред. проф. Н. С. Райбмана. М. : Наука, 1981. С. 336.

A. V. Medvedev, A. V. Faustov

ABOUT NONPARAMETRIC ESTIMATION OF NONSTATIONARY REGRESSION FUNCTION ACCORDING OBSERVATIONS

A problem of backup of mathematical models of nonstationary processes in a non-parametric uncertainty is considered in the article. Estimations of nonstationary regression function are offered. Results of numerical simulation are given. A problem of static multi-dimensional nonstationary systems modeling with delay is considered as well.

Keywords: nonparametric estimations, discrete-continuous processes, nonstationarity, regression, identification, modeling.

© Медведев А. В., Фаустов А. В., 2010

УДК 62-506.1

О. В. Шестернева, Т. В. Мальцева

О ВЫБОРЕ ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ В ЗАДАЧЕ НЕПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Рассмотрена задача получения передаточной функции объекта управления по математической модели замкнутой линейной динамической системы с целью настройки параметров управляющего устройства (регулятора). Структурный синтез модели замкнутого контура осуществляется методами непараметрического моделирования.

Ключевые слова: настройка регулятора, структурный синтез, передаточная функция, оценка Надарая–Ватсона, коэффициент размытости.

Несмотря на то, что в последнее время все большее значение приобретают цифровые управляющие устройства, до сих пор управление многими технологическим процессами осуществляется посредством или при помощи параметрических аналоговых регуляторов различных типов. Поэтому задача настройки параметров таких регуляторов не теряет своей актуальности. Ежегодно разрабатываются различные методики, рекомендации и инструкции по выбору управляющего устройства и настройке его параметров, которая может быть произведена согласно критериям устойчивости А. Гурвица (1895 г.) и А. В. Михайлова [1] при наличии явного вида передаточных функций объекта и корректирующих звеньев. При этом предполагается знание передаточной функции объекта управления или, по меньшей мере, возможность получения его переходной характеристики. В тех случаях, когда структура объекта управления неизвестна, настройка параметров производится эмпирически, что представляет определенную сложность, требует временных и финансовых затрат, а в ряде случаев и вовсе нежелательна. Знание явного вида передаточной функции объекта позволяет довольно просто и качественно настраивать параметры регулятора не на самой системе управления, а на ее модели [2].

В данной работе рассматривается метод, позволяющий получить передаточную функцию объекта управления в условиях непараметрической неопределенности при любом задающем воздействии (не требует возможности снятия переходной характеристики). Идея метода заключается в построении модели объекта путем предварительного определения порядка дифференциального уравнения замкнутой системы методами непараметрического моделирования.

Постановка задачи. Имеется замкнутая линейная динамическая система (ЛДС), корректирующим устройством которой является параметрический регулятор: пропорциональный (П-типа), пропорционально-интегральный (ПИ-типа) или пропорционально-интегро-дифференциальный (ПИД-типа), структура которого известна, а значения параметров устанавливаются проектировщиком или иным лицом, контролирующим работоспособность системы. Сведения о структуре объекта управления отсутствуют, известны только некоторые качественные свойства: стационарный, линейный динамический объект, на вход которого поступает управляющее воздействие, выработанное параметрическим регулятором (уровень непараметрической неопределенности). Датчики фиксируют значения сигнала $x^*(t)$ (задающее воздействие), поступающего на замкнутую систему (макрообъект), и реакцию системы $x(t)$ на задающее воздействие (рис. 1). Измерения производятся в моменты времени $t(i)$. В каналах измерения действует центрированная помеха с ограниченной дисперсией, сведения о законе распределения помех отсутствуют. Данные измерений формируют обучающую выборку $\left\{x^*(i), x(i)\right\}$, $i=\overline{1,n}$, некоторого объема n.

Ставится задача получения явного вида передаточной функции объекта управления, которая сводится, в свою очередь, к задаче построения математической модели замкнутого контура в случае малой априорной информации. Существуют различные методы, решающие эту задачу [3; 4], однако сам процесс моделирования зачастую достаточно трудоемкий (в основном за счет сложности выбора структуры модели) и требует больших затрат. Задача моделирования усложняется еще и тем, что измерению поддаются только задающее воздействие и выходной сигнал макрообъекта, и нет возможности измерить то управление, которое подается на сам объект. В связи с последним обстоятельством предлагается первоначально получить параметрическую модель замкнутой системы, а затем по ней определить передаточную функцию объекта управления.

Задача построения модели замкнутой системы представляет неменьшую сложность. При отсутствии каких-либо сведений о структуре объекта управления построение параметрической модели весьма проблематично. При включении корректирующего устройства и замыкании отрицательной обратной связью структура исследуемой системы становится сложнее структуры объекта управления, что, в свою очередь, увеличивает число определяемых параметров, а значит и сложность параметрического моделирования.

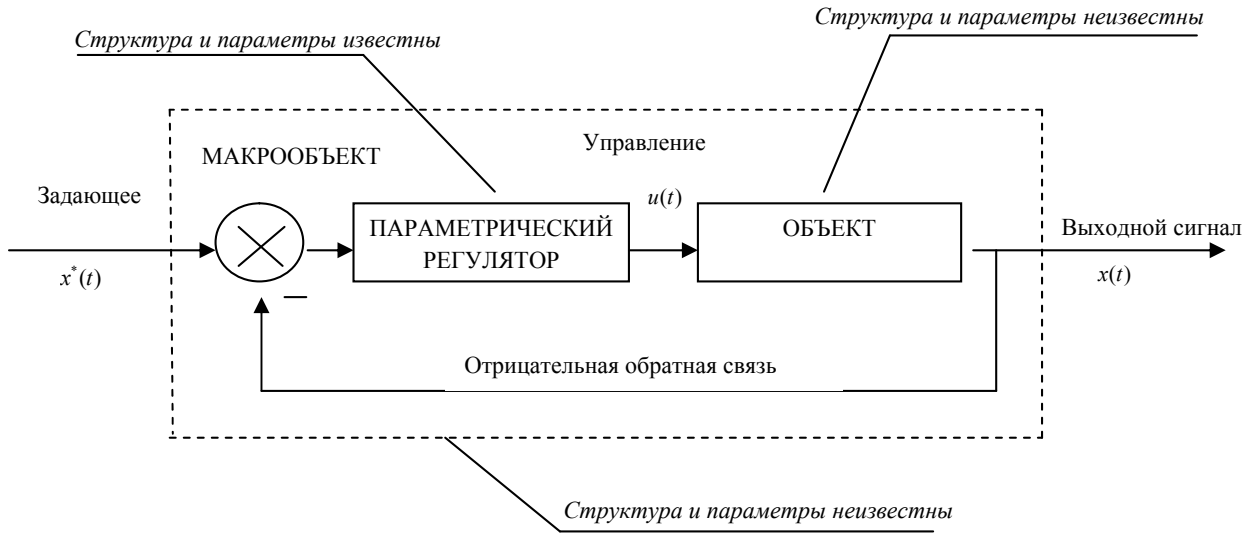


Рис. 1. Схема замкнутой динамической системы как макрообъекта

В настоящей работе рассматривается подход, позволяющий избежать сложности подбора динамической структуры посредством сочетания непараметрических и параметрических методов математического моделирования [5]. Идея заключается в предварительном определении порядка дифференциального уравнения, описывающего объект, и последующем использовании полученной информации в создании параметрической модели. Порядок уравнения предлагается определять путем построения регрессионной непараметрической модели между входными и выходными сигналами объекта, после чего задача моделирования сводится к определению значений параметров параметрической модели известными методами, например, методом наименьших квадратов.

Структурный синтез модели замкнутой ЛДС. Первая часть предлагаемого метода основана на определении порядка дифференциального уравнения, описывающего макрообъект при помощи методов непараметрической аппроксимации стохастической зависимости входного (в общем случае входных) и выходного (выходных) сигналов. По данным обучающей выборки $\left\{x^*(i), x(i)\right\}$, $i=\overline{1,n}$, строится многомерная непараметрическая оценка регрессии Надарая–Ватсона [6; 7]:

$$\hat{x}(t)=\frac{\sum\limits_{i=1}^{n} x_i \cdot \Phi\left(\frac{x^*(t)-x^*(i)}{h_{x^*}}\right)\cdot\prod\limits_{j=1}^{s}\Phi\left(\frac{x(t-j)-x(t(i)-j)}{h_{x(t-j)}}\right)\times \\ \times\prod\limits_{k=1}^{r}\Phi\left(\frac{x^*(t-k)-x^*(t(i)-k)}{h_{x^*(t-k)}}\right)}{\sum\limits_{i=1}^{n}\Phi\left(\frac{x^*(t)-x^*(i)}{h_{x^*}}\right)\cdot\prod\limits_{j=1}^{s}\Phi\left(\frac{x(t-j)-x(t(i)-j)}{h_{x(t-j)}}\right)\times \\ \times\prod\limits_{k=1}^{r}\Phi\left(\frac{x^*(t-k)-x^*(t(i)-k)}{h_{x^*(t-k)}}\right)}, \quad (1)$$

где в качестве аргументов используется как входное воздействие на текущем шаге, так и значения входного и выходного сигналов на предыдущих шагах. Такой подход позволяет учитывать динамику объекта, так как значения выходного сигнала объекта на нескольких шагах, являясь аргументами оценки регрессии (1) на последующих шагах, влияют на оценку выхода.

Число s предыдущих выходных (r входных) сигналов (которые выступают в оценке (1) в качестве аргументов), включаемых в модель, является аналогом порядка дифференциального уравнения: чем выше порядок, тем длиннее период функционирования объекта, влияющий на его последующее поведение, и тем больше данных, полученных на предыдущих шагах, мы должны учитывать.

Функция $\Phi(\cdot)$ – ядро (колоколообразная, дельтообразная функция) – удовлетворяет некоторым условиям сходимости [6; 7], влияние же вида ядра на точность оценивания незначительно. В данной работе использовалось параболическое ядро [8]:

$$\Phi(z)=\begin{cases}0,75\cdot(1-z)^2, & |z|\le 1,\\ 0, & |z|\ge 1.\end{cases}$$

Параметры $h=\left(h_{x^*}, h_{x^*(t-j)}, h_{x^*(t-k)}\right)$, $j=\overline{1,s}$, $k=\overline{1,r}$ в формуле (1) – коэффициенты размытости, настройка которых производится согласно условию минимума среднеквадратичного критерия методом скользящего экзамена. Заметим, что с ростом h сглаживающие свойства оценки нарастают, по h для каждого конечного объема выборки существует некоторый оптимум (при малых h оценка представляет собой набор непересекающихся или слабопересекающихся дельтообразных функций и теряет свой смысл, а при больших h оценка становится сильно сглаженной и не отражает индивидуальных особенностей оцениваемой зависимости) [9].

Настройка значений коэффициентов размытости в формуле (1) осуществляется одним из методов оптимизации путем минимизации среднеквадратичного критерия:

$$W(\bar{c}) = \frac{1}{n} \cdot \sum_{i=1}^{n} (x(i) - \hat{x}(t, u, x(t-j), \bar{c}))^2 \to \min_{\bar{c}}. \quad (4)$$

В настоящей работе использовался метод случайного спуска, где в качестве алгоритма поиска локального минимума был выбран последовательный симплексный метод [10].

Последовательно строятся непараметрические модели (1), включающие все большее и большее число аргументов. Модель, значение критерия (4) для которой оказывается минимальным, считается наилучшей, а число предыдущих измерений выходного *s* и входного *r* сигналов, включенных в эту модель, определяют структуру параметрической модели. Работоспособность описанного алгоритма уже проверялась ранее [5]. В настоящей работе рассмотрим влияние таких факторов, как частота дискретных измерений входного и выходного сигналов и величина помехи на точность определения порядка.

Первоначально разберем вопрос о влиянии шага дискретизации на процесс моделирования. Рассмотрим незашумленную выборку. Объект управления описывался дифференциальным уравнением второго порядка, в качестве управляющего устройства использовался ПИ-регулятор. В этом случае дифференциальное уравнение, описывающее поведение замкнутой системы управления, имело третий порядок. Графические результаты работы непараметрического алгоритма в случае, когда шаг дискретизации $\Delta t = 0,1,$ приведены на рис. 2.

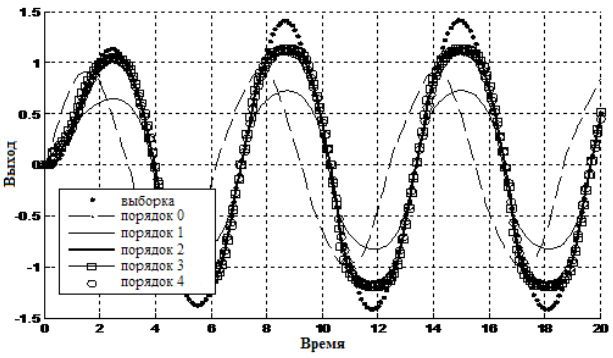


Рис. 2. Результаты непараметрического моделирования

В случае отсутствия, помехи при включении в модель двух предыдущих шагов происходит резкое улучшение качества модели (назовем этот эффект переломным моментом), которое, тем не менее, начинает незначительно ухудшаться с дальнейшим увеличением порядка (численные значения критерия (2) при разном значении Δt представлены в табл. 1).

Из данных табл. 1 можно увидеть, что в случае чрезмерно малого значения Δt непараметрический алгоритм имеет тенденцию к занижению порядка. При увеличении Δt до определенного значения сначала наблюдается точность в определении порядка. При чрезмерно большом значении Δt динамика прослеживается хуже, что приводит к снижению точности оптимизационной процедуры настройки параметров моделей, а это, в свою очередь, влияет на точность определения порядка дифференциального уравнения. Малая представительность выборки приводит к тому, что число измерений становится недостаточным для оптимальной настройки параметров моделей, и переломный момент пропадает. Наблюдается тенденция «мнимого» улучшения качества непараметрических моделей за счет введения дополнительных членов, а следовательно, завышение порядка.

Далее проведем исследование влияния уровня помехи на точность определения порядка. Для чистоты эксперимента используем шаг дискретизации $\Delta t = 0,4,$ при котором, как было установлено ранее, для рассматриваемой системы порядок определяется правильно. По выборочным данным было проведено непараметрическое исследование путем построения моделей (1), графические результаты которого в случае 10%-й помехи приведены на рис. 3, а численные значения критерия представлены в табл. 2.

Примечание. Помеха накладывалась следующим образом: измерялся интервал изменения сигнальной части Δ, задавался уровень помех ρ (от 0 до 1). С помощью генератора случайных чисел формировался вектор (размерность вектора совпадала с объемом выборки) значений равномерно распределенной на интервале [–Δρ; Δρ] случайной величины, который впоследствии складывался с вектором значений сигнальной части.

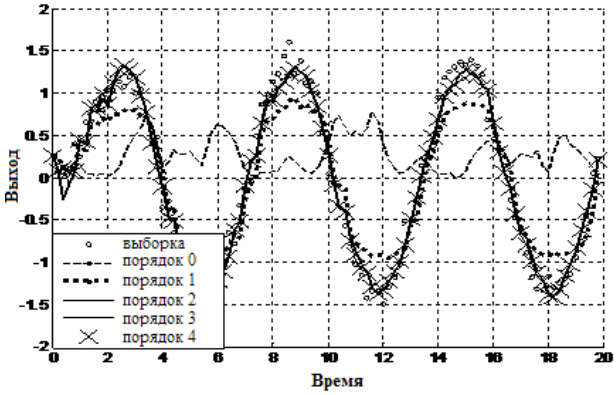


Рис. 3. Результаты непараметрического моделирования

В случае нулевой помехи (см. рис. 2) при включении в модель двух предыдущих шагов происходит резкое улучшение качества модели (назовем этот эффект переломным моментом), минимум же критерия достигается при включении в модель трех предыдущих шагов. Увеличение помехи приводит сначала к тому, что пропадает «очевидность» выбора структуры, а затем к завышению порядка, которое тем больше, чем выше уровень помехи. Отметим, что при этом пропадает переломный момент, наблюдаемый при небольших помехах: значение критерия достаточно равномерно уменьшается с добавлением в модель все большего числа предыдущих шагов. Таким образом, в качестве вывода отметим общую тенденцию к завышению порядка при больших помехах.

Таблица 1

Зависимость среднеквадратичной ошибки моделирования от шага дискретизации

Шаг дискретизации измерений	Число измерений выхода в модели				
	0	1	2	3	4
0,1	0,5560	0,1352	0,0129	0,0158	0,0166
0,2	0,9889	0,1404	0,0250	0,0171	0,0244
0,4	1,1575	0,1451	0,0301	0,0174	0,0290
0,5	1,2870	0,4270	0,2202	0,1975	0,2266
1	1,0954	0,5676	0,4871	0,4019	0,2506
2	0,9265	0,8976	0,7481	0,5797	0,3915

Таблица 2

Зависимость среднеквадратичной ошибки моделирования от уровня помехи

Уровень помехи, %	Число измерений выхода в модели				
	0	1	2	3	4
0	1,1575	0,1451	0,0301	0,0174	0,0290
5	1,1535	0,1283	0,0502	0,0373	0,0438
10	1,1759	0,1559	0,0651	0,0530	0,0705
20	1,1690	0,1923	0,0853	0,0573	0,0779
30	1,2662	0,2120	0,1256	0,1127	0,1142
40	1,3112	0,2706	0,1983	0,1789	0,1834
50	1,4257	0,4531	0,3861	0,3724	0,3649
100	1,5020	0,9355	0,9206	0,8883	0,8461

Рассматриваемый непараметрический алгоритм структурного синтеза модели имеет тенденцию к завышению порядка при увеличении помехи и шага дискретизации и тенденцию к занижению порядка при слишком маленьком шаге дискретизации. Тем не менее, следует отметить, что завышение порядка модели происходит при плохом качестве выборочных данных – очень высокий уровень помехи или малая представительность выборочных данных, тогда как занижение порядка происходит вследствие неверного выбора Δt даже при качественной выборке (достаточный объем, низкий уровень помехи). В ряде случаев неточности определения порядка могут быть устранены на этапе параметрического синтеза за счет исключения из модели тех составляющих, для соответствующих коэффициентов которых подтверждается гипотеза равенства нулю (при завышении порядка). Однако при заниженном порядке проверка указанной гипотезы не приведет к уточнению структуры. В данной работе нас интересует не столько сама модель, сколько то, как ошибки в структурном синтезе модели могут повлиять на настройку параметров регулятора, а следовательно, на качество управления.

Получение передаточной функции объекта. Так как структура модели определена ранее методами непараметрического моделирования, задача сводится к нахождению оценок неизвестных параметров модели макрообъекта. Представим структуру модели в виде разностного уравнения [4]. Применяя метод наименьших квадратов (в работе рассматривался наиболее простой случай некоррелированных равноточных измерений), получим уравнение для вектора оптимальных оценок параметров модели [4]. В качестве характеристики точности параметрической модели регрессии использовалось так называемое значение R^2 [3]. Проверка значимости коэффициентов осуществлялась на основании t-критерия [3] при уровне значимости $\alpha = 0,05$. Результаты параметрического моделирования представлены на рис. 4.

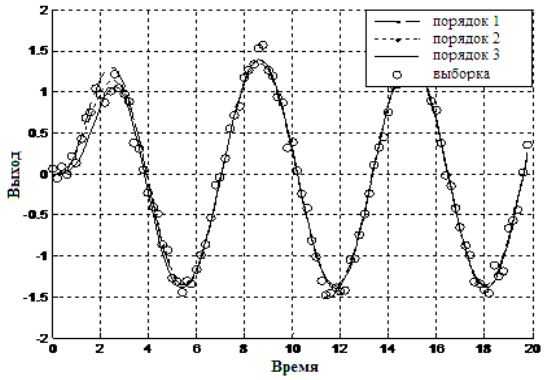


Рис. 4. Результаты параметрической идентификации моделей

На основании полученной параметрической модели замкнутой системы управления при известной модели управляющего устройства можно получить модель объекта управления.

Пример. Получение передаточной функции объекта управления по уравнению замкнутого контура. Пусть получена модель макрообъекта в следующем виде:

$$a_4^{\text{МО}} \cdot \frac{d^4x}{dt^4} + a_3^{\text{МО}} \cdot \frac{d^3x}{dt^3} + a_2^{\text{МО}} \cdot \frac{d^2x}{dt^2} + a_1^{\text{МО}} \cdot \frac{dx}{dt} + a_0^{\text{МО}} \cdot x = b_1^{\text{МО}} \cdot \frac{du}{dt} + b_0^{\text{МО}} \cdot u. \quad (3)$$

Применяя преобразование Лапласа, получим передаточную функцию замкнутой системы

$$W_з(p) = \frac{b_1^{\text{МО}} p + b_0^{\text{МО}}}{a_4^{\text{МО}} p^4 + a_3^{\text{МО}} p^3 + a_2^{\text{МО}} p^2 + a_1^{\text{МО}} p + a_0^{\text{МО}}}. \quad (4)$$

Разомкнутая система представляет собой последовательное соединение объекта управления и регулятора [11], а следовательно, передаточная функция разомкнутой системы имеет вид (в соответствии с правилом преобразования цепей)

$$W_{\text{раз}}(p) = W_{Obj}(p) \cdot W_{\text{ПИ}}(p). \quad (5)$$

Используя правило Мейсона для замкнутого контура с единичной отрицательной обратной связью [11]

$$W_{\text{з}}(p) = W_{\text{раз}}(p) / 1 + W_{\text{раз}}(p) \quad (6)$$

и учитывая тот факт, что в качестве управляющего устройства используется ПИ-регулятор, передаточная функция которого имеет вид [2; 11]

$$W_{\text{ПИ}}(p) = K_1 p + K_2 / p, \quad (7)$$

получаем

$$W_{\text{з}}(p) = \frac{b_1^{\text{МО}} p + b_0^{\text{МО}}}{a_4^{\text{МО}} p^4 + a_3^{\text{МО}} p^3 + a_2^{\text{МО}} p^2 + a_1^{\text{МО}} p + a_0^{\text{МО}}} = \frac{b_1 K_1 p + b_0 K_2}{a_3 p^4 + a_2 p^3 + a_1 p^2 + (a_0 + b_1 K_1) p + b_0 K_2}, \quad (8)$$

где $a = [a_0, a_1, a_2, a_3]$, b_0 – коэффициенты объекта управления.

Таким образом, передаточная функция объекта управления имеет вид

$$W_{Obj}(p) = \frac{b_0}{a_3 p^3 + a_2 p^2 + a_1 p + a_0}. \quad (9)$$

Зная передаточную функцию объекта управления, можно получить значения параметров конкретного типа регулятора исходя из требования устойчивости замкнутой системы [2; 11]. Таким образом, получение адекватной модели объекта как минимум позволяет избежать проблем, связанных с «выпадением» системы управления из области устойчивости, и как максимум обеспечивает возможность выбора таких значений параметров регулятора, которые бы обеспечивали желаемый результат управления.

Влияние выбора структуры на настройку регулятора. Возникает закономерный вопрос, как влияют установленные ранее тенденции завышения или занижения порядка при непараметрическом структурном синтезе модели замкнутой системы на применимость получаемой при этом модели объекта управления к настройке параметров регулятора.

Для того чтобы исследовать это влияние, по построенным ранее параметрическим моделям замкнутой системы (см. рис. 4) определим передаточные функции вида (9) соответствующих моделей объекта управления (см. пример) и рассмотрим, как согласуется настройка параметров регулятора на каждой из моделей с тем, что мы при этом получим на истинном объекте. Следует отметить, что в связи со спецификой задачи идентификации параметров дифференциального уравнения, а также с тем фактом, что в каналах измерений действует помеха, коэффициенты передаточной функции объекта (9) получаются неоднозначными и сильно зависят от параметров модели замкнутой системы, полученных на этапе параметрического синтеза модели. Тем не менее, характер поведения выхода моделей в зависимости от структуры можно проследить (табл. 3).

Таким образом, можно сделать вывод о том, что использование модели, порядок дифференциального уравнения которой превосходит порядок дифференциального уравнения истинного объекта, при настройке параметров регулятора значительно «сужает» область устойчивости системы и, следовательно, может не привести к получению качественного управления (в случае, если оптимальные значения параметров регулятора для истинной системы не принадлежат области устойчивости системы управления моделью). В то же время, «занижение» порядка модели может привести к тому, что при качественном управлении моделью истинная система даже не окажется устойчивой. При совпадающей структуре модели и истинного объекта область устойчивости оказывается примерно одинаковой, и результаты настройки параметров регулятора на такой модели могут привести к значительному повышению качества управления истинным объектом. В связи с вышесказанным следует отметить, что при выборе структуры лучше завысить порядок модели, чтобы избежать неустойчивости истинной системы при полученных на модельной системе параметрах регулятора.

Библиографические ссылки

1. Михайлов А. В. Метод гармонического анализа в теории регулирования // Автоматика и телемеханика. 1938. № 3. С. 27–81.

2. Ротач В. Я. Теория автоматического управления : учеб. пособие. М. : МЭИ, 2004.

3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ : пер с англ. 3-е изд. М. : Вильямс, 2007.

4. Пащенко Ф. Ф. Введение в состоятельные методы моделирования систем : учеб. пособие : в 2 ч. Ч. 2. Идентификация нелинейных систем. М. : Финансы и статистика, 2007.

5. Мальцева Т. В. Об одном методе построения математической модели линейного динамического объекта // Молодой ученый, 2008. С. 40–48.

6. Надарая Э. А. О непараметрических оценках плотности вероятности и регрессии // Теория вероятностей и ее применение. 1965. Т. 10(1). С. 199–203.

7. Watson G. Smooth regression analysis // Sankhya, ser. A. 1965. Vol. 26. Part 4. P. 359–372.

8. Лагутин М. Б. Наглядная математическая статистика : учеб. пособие. М. : БИНОМ : Лаборатория знаний, 2007.

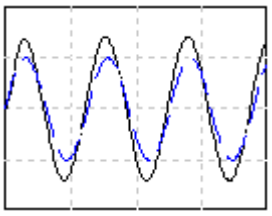
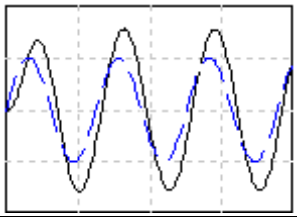
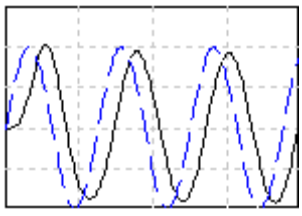
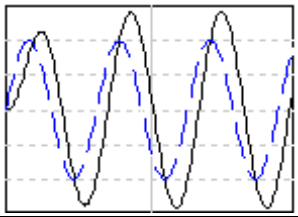
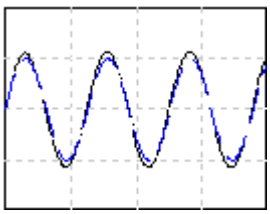
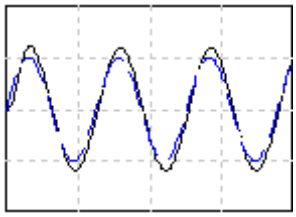
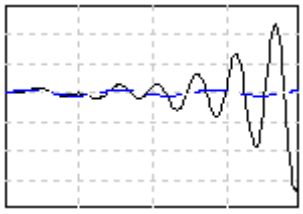
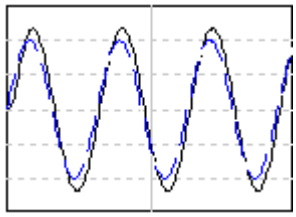
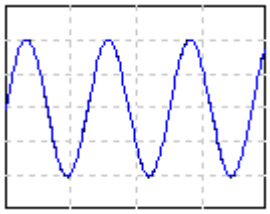
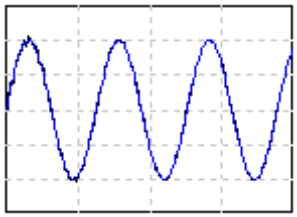
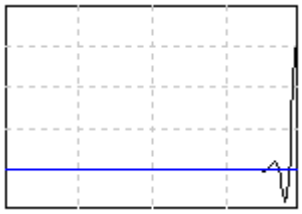
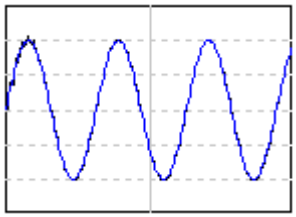
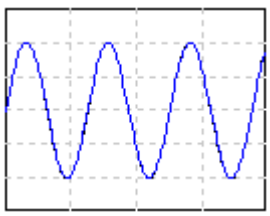
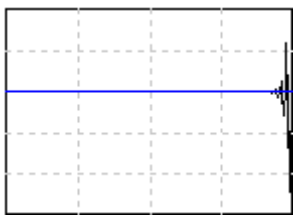
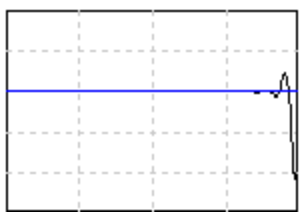
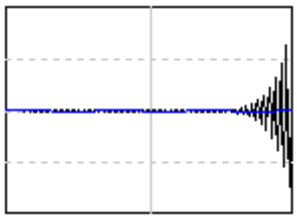
9. Рубан А. И. Идентификация стохастических объектов на основе непараметрического подхода // Автоматика и телемеханика. 1979. № 11. С. 106–118.

10. Химмельблау Д. Прикладное нелинейное программирование. М. : Мир, 1975.

11. Юревич Е. И. Теория автоматического управления. СПб. : БХВ-Петербург, 2007.

Таблица 3

Настройка параметров регулятора на различных моделях объекта

Порядок модели ниже порядка объекта	Структуры модели и объекта совпадают	Порядок модели выше порядка объекта	Истинный объект
Случай 1			
			
При исходных настройках регулятора выход замкнутой системы в случае управления истинным объектом и всеми полученными моделями расходится с задающим воздействием: требуется корректировка параметров регулятора. Система при этом устойчива во всех случаях			
Случай 2			
			
Изменение параметров регулятора привело к повышению качества управления при использовании моделей более низкого и совпадающего порядка, однако система управления моделью более высокого порядка, нежели порядок объекта, оказалась неустойчивой. Управление истинным объектом при этом улучшилось. Там не менее, все еще наблюдается расхождение выхода системы с задающим воздействием при управлении как моделями с более низким и совпадающим порядком, так и истинным объектом			
Случай 3			
			
Оптимальные параметры регулятора для системы управления моделью совпадающего порядка привели к улучшению качества управления моделью более низкого порядка, а также истинным объектом. Система управления моделью более высокого порядка неустойчива			
Случай 4			
			
Оптимальные параметры регулятора для системы управления моделью более низкого, чем порядок объекта, порядка приводят к тому, что система управления моделью совпадающего порядка перестает быть устойчивой. Истинная система также становится неустойчивой			

Примечание. На всех графиках, приведенных в таблице, сравнивается выход замкнутой системы с задающим воздействием; по оси абсцисс – время, по оси ординат – выход системы

O. V. Shesterneva, T. V. Maltseva

ABOUT CHOICE OF PARAMETRIC MODEL IN NONPARAMETRIC IDENTIFICATION TASK

In the paper the authors consider a task of creation of a control object transfer function in accordance with a mathematical model of closed linear dynamic system for a regulator parameters tuning. Structure synthesis of a closed loop model is carried out with nonparametric modeling methods.

Keywords: regulator tuning, structure synthesis, transfer function, Nadaraya–Watson estimation, bandwidth (smoothing parameter).

© Шестернева О. В., Мальцева Т. В., 2010

УДК 510.9872

П. В. Густов, Е. С. Ильин, В. А. Терсков

ОЦЕНКА ОПТИМАЛЬНОГО БЫСТРОДЕЙСТВИЯ ЭЛЕМЕНТОВ УПРАВЛЕНИЯ СЛОЖНЫМИ ТЕХНИЧЕСКИМИ СИСТЕМАМИ

Исследуется зависимость производительности многопроцессорных вычислительных комплексов от быстродействия процессоров, разработан алгоритм оценки эффективного быстродействия процессоров.

Ключевые слова: многопроцессорные системы, оптимизация, генетический алгоритм, нейросеть.

Аппаратная реализация функций упрощает процесс программирования, дает возможность более эффективно распараллеливать сложные алгоритмы, реализуемые на многопроцессорных вычислительных комплексах (МВК), обеспечивает более естественное распределение частей алгоритмов между параллельно работающими процессорами [1]. Следовательно, для повышения производительности МВК необходимо включать в их структуру спецпроцессоры, ориентированные на реализацию одной или нескольких независимых функций. При этом снижается интенсивность обращения процессоров к шине и, следовательно, средняя длина очереди, значения которых зависят от быстродействия процессоров.

К аппаратным методам вычисления элементарных функций предъявляются следующие требования:

– получение заданной точности вычисления функций на всем диапазоне изменения аргумента;

– обеспечение высокой скорости вычисления значения функций;

– использование малого объема дополнительного оборудования, необходимого для реализации метода (объема памяти для хранения программ, микропрограмм, констант и пр.);

– метод должен быть универсальным (применимость для вычисления заданного набора функций, независимость функциональной схемы от изменения требований точности);

– должна учитываться структурная реализация АЛУ (использование принципа микропрограммирования, БИС-технологии и т. п.).

При определении эффективного быстродействия спецпроцессоров учитываются следующие условия функционирования МВК:

– при увеличении быстродействия процессора он чаще обращается к шине, что увеличивает среднюю длину очереди, а следовательно, и время ожидания в очереди другими процессорами, что приводит к уменьшению производительности МВК;

– при уменьшении быстродействия процессора средняя длина очереди уменьшается, но увеличивается и время решения задачи, что также приводит к уменьшению производительности МВК.

В [2] показано, что производительность МВК зависит от количества спецпроцессоров и их быстродействия нелинейно. Это означает, что с ростом быстродействия процессоров производительность МВК сначала повышается, а затем перестает расти из-за конфликтов при обращении к памяти. Поэтому при проектировании оптимальных МВК нельзя ориентироваться на максимальное быстродействие спецпроцессоров, а надо выбирать его так, чтобы обеспечить максимальную производительность всего МВК в целом. Для формальной постановки задачи это означает, что величины относительных времен выполнения операций процессорами i-го типа (T_{0i}) не могут быть постоянными, а должны также быть включены в число переменных оптимизации. Более того, они становятся «главными» переменными, определяющими параметры системы (ν_i, μ_i), т. е. эти параметры становятся функциями от T_{0i} – $\nu_i(T_{0i})$, $\mu_i(T_{0i})$. Это приводит к значительному усложнению оптимизационной задачи, превращая ее в двухуровневую иерархическую задачу:

$$(\Pi^*(T_{01}, ..., T_{0i}, ..., T_{0N}),$$
$$C^*(T_{01}, ..., T_{0i}, ..., T_{0N})) \to \text{extr},$$

где $П^*$ и $С^*$ – решение задачи выбора оптимальной структуры МВК при

$$\nu_i = \nu_i(T_{01},\ldots, T_{0i}, \ldots, T_{0N}),\ \mu_i = \mu_i(T_{01},\ldots, T_{0i}, \ldots, T_{0N}).$$

Переменные задачи верхнего уровня иерархии могут быть непрерывными (задача оптимального проектирования структуры) или дискретными (задача выбора эффективного варианта структуры).

Однако, как уже было замечено выше, зависимости параметров потоков ν_i и μ_i от величин T_{0i} носят довольно сложный характер. В результате этого, с увеличением числа компонентов системы зависимости будут все усложняться, задача станет громоздкой и может перестать быть решаемой. Чтобы свести задачу к допустимому уровню сложности, был предложен следующий подход: предположить параметры T_{0i} известными и зафиксированными. Разумнее всего было бы взять эти параметры из реальной, уже работающей системы, решить для этих T_{0i} задачу выбора оптимальной структуры МВК, а уже затем для выбранной структуры подобрать T_{0i}. Таким образом, задача разбивается на две. Первая – это задача выбора оптимальной структуры МВК при фиксированных T_{0i}:

$$(П_в(m_i, n_j), С_в(m_i, n_j)) \to \text{extr},$$
$$m_i^- \le m_i \le m_i^+,\ i = 1, \ldots, N,$$
$$n_j^- \le n_j \le n_j^+,\ j = 1, \ldots, n,$$

а вторая – это задача выбора оптимальной структуры МВК при выбранных параметрах m_i, n_j, n и N.

$$(П^*(T_{01}, ..., T_{0i}, T_{0N}), C^*(T_{01}, ..., T_{0i}, T_{0N})) \to \text{extr}.$$

Кажущееся громоздким такое разделение способно, тем не менее, помочь справиться с комбинаторным взрывом, возникшим при введении нового уровня оптимизации.

Однако данный подход имеет несколько существенных недостатков. Оба этапа предложенного подхода сами по себе являются сложными задачами оптимизации. Кроме того, нет никакой гарантии, что полученная на первом этапе работы оптимальная структура МВК при фиксированных параметрах быстродействия процессоров и шин будет оптимальна и после настройки параметров быстродействия.

Таким образом, более перспективным представляется подход, при котором происходит одновременный выбор и оптимальной структуры МВК, и оптимального быстродействия процессоров и шин. Реализация такого подхода возможна только при наличии такой процедуры, которая способна одновременно учитывать все аспекты нашей оптимизационной задачи, а именно: переменные задачи выражены в различных шкалах измерения (целочисленные и вещественные переменные), оптимизация производится по нескольким экстремальным критериям одновременно, априорные сведения о свойствах целевых функционалов отсутствуют (оптимизация производится только по измерениям функционалов в фиксированных точках). Метод оптимизации должен учитывать условия, накладываемые на критерии. Это означает, что при выборе эффективной конфигурации МВК возникает необходимость решения задачи многокритериальной условной оптимизации с алгоритмически заданными функциями разношкальных переменных высокой размерности.

В качестве приемлемого средства решения такой задачи может выступать генетический алгоритм [3].

Проверка работоспособности системы поддержки принятия решений (СППР) проводилась на задачах выбора эффективного быстродействия процессоров МВК ПС-2000 и «Эльбрус-2».

Телеметрический вычислительный комплекс центра управления космическими полетами (ЦУП) использовал с 1986 г. и вплоть до 1997 г. систему предварительной обработки телеметрической информации на базе ПС-2000, связанную в единый комплекс с центральной системой обработки на базе многопроцессорного вычислительного комплекса «Эльбрус-2». Высокий параллелизм обработки информации в ПС-2000 позволил реализовать новые алгоритмы обработки телеметрической информации. ПС-2000 можно представить как МВК со следующими характеристиками: количество шин n от 1 до 4, количество однотипных процессоров 8, 16, 32 или 64, их быстродействие T_{0i} = 0,32 мкс, скорость обработки запроса в памяти τ_i = 0,96 мкс. Однако технически мы можем увеличивать время выполнения операций процессорами, тем самым незначительно снижать общее быстродействие системы, но если при этом мы сможем разрешить конфликты, возникающие при одновременном обращении процессоров к памяти, то сумеем избежать простоя запросов в очередях и увеличить относительную производительность системы. Подбирая быстродействие сразу для всех процессоров, мы имеем до 8 типов процессоров, отличающихся быстродействием. Количество процессоров каждого типа будет неизменным и равняется 8.

В случае, когда известно количество типов процессоров, количество процессоров каждого типа и количество шин, отпадает необходимость выбора оптимальной структуры МВК. Решая данную задачу, достаточно определить оптимальное быстродействие для всех типов процессоров. Это позволит значительно сократить поисковое пространство. Решение в генетическом алгоритме содержало информацию лишь о быстродействии процессоров. Решение кодировалось в виде бинарной хромосомы фиксированной длины 40 бит. Даже при небольшом количестве циклов работы генетического алгоритма удалось получить такой вариант набора параметров для МВК, который превосходил изначальный, предложенный разработчиками ПС-2000, по относительной производительности более чем в 2 раза и не уступал при этом по показателю стоимости.

В таблице представлены быстродействия процессоров всех типов, а также рассчитанная стоимость и производительность. В первой строке находятся данные по модели МВК ПС-2000. Во второй и третьей строке представлены данные по двум моделям МВК, предложенным СППР в качестве наиболее предпочтительных решений.

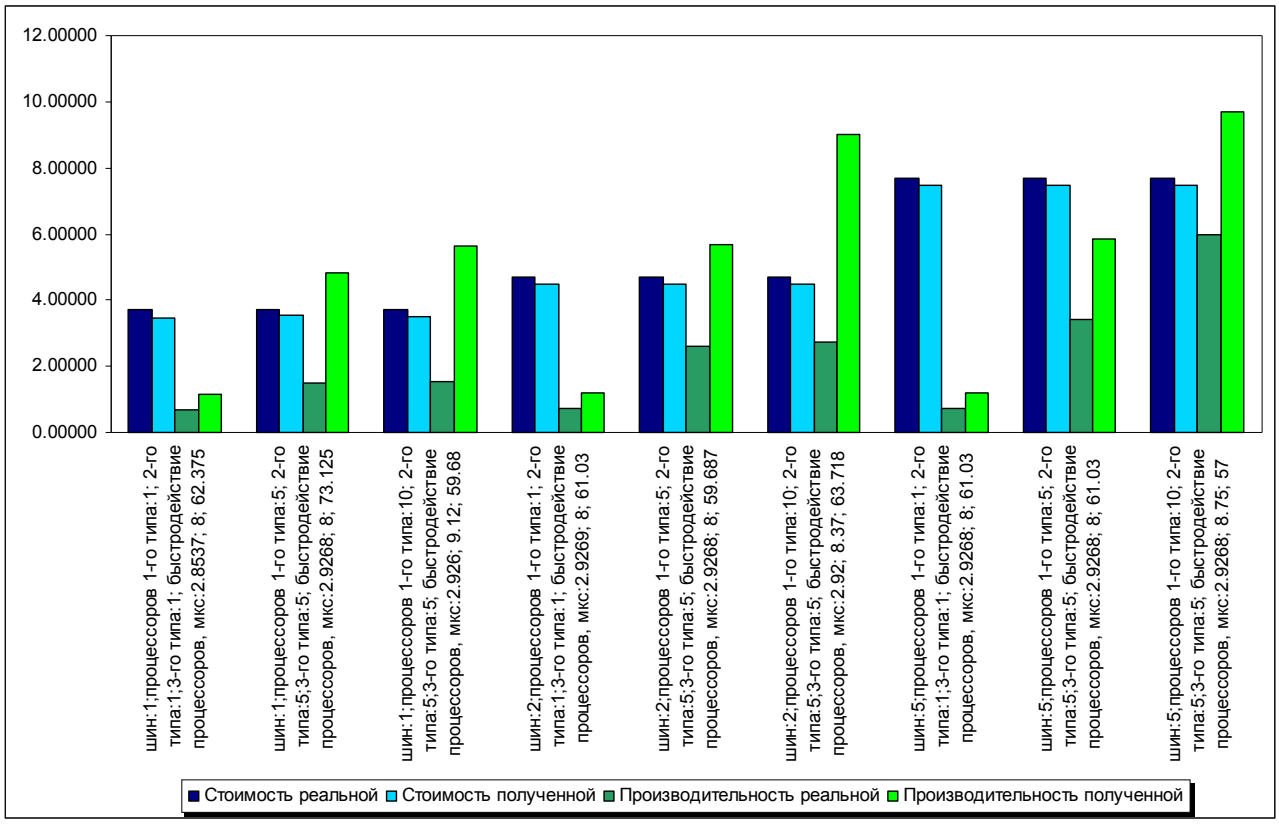
Для проверки работоспособности СППР решалась также и задача выбора оптимального быстродействия спецпроцессоров для различных вариантов структуры МВК «Эльбрус-2». Как и в случае выбора оптимального быстродействия спецпроцессоров, для многопроцессорной системы ПС-2000 было известно количество типов процессоров, количество процессоров каждого типа и количество шин. То есть, решая данную задачу, СППР достаточно определить оптимальное быстродействие для каждого типа процессоров. Решение кодировалось в виде бинарной хромосомы фиксированной длины 15 бит. Модифицированный генетический алгоритм работал с оптимальным набором параметров, найденных ранее.

Результаты расчетов оптимального быстродействия для МВК «Эльбрус-2» различных конфигураций представлены на рисунке. Для всех вариантов МВК быстродействие шин 1,2 мкс и три типа процессоров. Для реальной МВК «Эльбрус-2» быстродействие процессоров различных типов равно, соответственно, 0,66, 8 и 57 мкс. На рисунке приведены параметры, однозначно определяющие структуру МВК: количество шин и количество процессоров каждого типа. Также представлены значения производительности и стоимости как для вариантов МВК с реальными значениями быстродействия спецпроцессоров, так и для лучших вариантов МВК, полученных в результате настройки быстродействия спецпроцессоров. В последнем столбце указаны оптимальные значения быстродействия спецпроцессоров для МВК с соответствующими структурами.

Обучив искусственную нейронную сеть [4] примерам оптимального быстродействия многопроцессорных систем различных конфигураций, подавая на вход нейросети характеристики архитектуры МВК, на выход – стоимость и производительность МВК, получаем модель, позволяющую с требуемой точностью определить быстродействие многопроцессорных вычислительных комплексов для выбранной архитектуры. После проведения этапов по упрощению обученной нейронной сети получаем возможность извлечь правила функционирования нейросети, которые в дальнейшем используются экспертами для оценки оптимального быстродействия элементов МВК.

Данные по ПС-2000 и решениям, полученным СППР

Стоимость	Производительность	T_{01}	T_{02}	T_{03}	T_{04}	T_{05}	T_{06}	T_{07}	T_{08}
12,2358	1,64394	0,32	0,32	0,32	0,32	0,32	0,32	0,32	0,32
15,045	4,69568	5,86	2,91	9,33	1,63	0,64	0,45	9,33	5,21
11,1412	3,65461	1,15	0,51	8,309	2,31	1,29	0,72	2,31	1,15



Данные по МВК «Эльбрус-2» и решениям, полученным СППР

Таким образом, показано, что реализация разработанных алгоритмов и методов в виде единой программной системы позволяет генерировать многопроцессорные системы оптимальной структуры и с эффективным быстродействием специализированных процессоров. Апробация созданной системы поддержки принятия решений проводилась на моделях реальных, хорошо зарекомендовавших себя многопроцессорных систем.

Было показано, что использование СППР для выбора эффективного быстродействия процессоров многопроцессорной вычислительной системы позволяет не только значительно облегчить и ускорить процесс разработки МВК, а значит сократить затраты на него, но и получить такие варианты многопроцессорных вычислительных комплексов, которые по некоторым параметрам превосходят МВК, разрабатываемые экспертами.

Библиографические ссылки

1. Kuhl J., Reddy S. Fault-Tolerance Considerations in Large, Multiple-Processor Systems // Computer. 1986. March. P. 56–67.

2. Лебедев В. А., Терсков В. А. Моделирование и оптимизация многопроцессорных систем оперативного управления. М. : МАКС Пресс, 2002.

3. Goldberg D. E. Genetic Algorithms in Search, Optimization, and Machine Learning. Reading : Addison-Wesly, 1989.

4. Интеллектуальная система анализа данных на основе нейронных сетей / ВНТИЦ ; Ильин Е. С., Семенкин Е. С. М., 2004. № ГР 50200400996.

P. V. Gustov, E. S. Iluin, V. A. Terskov

EVALUATION OF OPTIMUM EXECUTION SPEED OF COMPLEX TECHNICAL SYSTEMS CONTROL ELEMENTS

In the article the authors investigate dependence of multiprocessor systems productivity on processors execution speed and develop algorithm for estimation of effective execution speed of processors.

Keywords: multiprocessor systems, optimization, genetic algorithm, neural network.

© Густов П. В., Ильин Е. С., Терсков В. А., 2010

ЧТО ДЕЛАТЬ? КТО ВИНОВАТ?
ДИСКУССИЯ ОБ ОБРАЗОВАНИИ

Лучший пророк для будущего – прошлое.
Д. Н. Байрон

Важнейшая задача цивилизации – научить человека мыслить.
Т. А. Эдисон

Успешно готовить творчески активных исследователей в науке и технике могут лишь те учёные, которые сами непосредственно занимаются – и не только занимаются, но и увлекаются исследовательской работой. Никакой пересказ учебников и даже новейших статей из научных журналов людьми, которые сами не ведут научно-исследовательской работы, не решит задачи: ведь «ум юноши – не сосуд, который надо наполнить, а факел, который надо зажечь».
И. Е. Тамм

А. В. Медведев (профессор, СибГАУ, Красноярск). Сейчас образовательному процессу уделяется большое внимание. Вы посмотрите, сколько конференций посвящено тому, как надо учить. Тем не менее образовательный уровень студентов сегодня значительно ниже, чем 20 лет тому назад. Многочисленные публикации посвящены тому, как учить, используя современные информационные технологии. Я не помню, чтобы в давние времена было такое количество конференций по этому поводу.

Я бы хотел обратить ваше внимание на то, что образование в СССР было на достаточно высоком уровне. Те изменения, которые начались в новое время в России, в конечном счете привели к ухудшению образовательного процесса, несмотря на стремление к лучшему. Вряд ли можно считать, что изменения, которые происходили в науке и образовании в России, были так уж необходимы. Конечно, следовало бы учитывать те новые веяния, которые появились в связи с вхождением России в международное образовательное пространство (я имею в виду Болонский процесс). Мне представляется, что более целесообразным было бы «надеть фрак международного покроя на тело российского образования».

Несколько слов хочу сказать по поводу оценки качества научного сотрудника, поскольку современный образовательный процесс тесно связан с проведением научных исследований во время обучения. Хочу напомнить одну ситуацию, которая была описана в книге Г. И. Марчука*. Цитирую: «Оценка труда учёного – дело непростое. Однажды я дал поручение заведующим отделами – проанализировать работу своих сотрудников. В отделе члена-корреспондента М. М. Лаврентьева одного из сотрудников хотели было признать профессионально непригодным: за три года работы в Вычислительном центре он не опубликовал ни одной статьи и не сделал ни одного доклада! Узнав об этом, „обвиняемый“ очень разволновался и наутро принёс М. М. Лаврентьеву свою теоретическую работу, просмотрев которую, коллеги установили, что автор получил результат мирового класса. Все удивлялись: почему не показал, не рассказал никому? На что автор скромно ответил: „Боялся, что не поверят. А уж раз собрались увольнять, терять было нечего…“ Довольно скоро он блестяще защитил диссертацию.

Этот пример свидетельствует, что формальный метод оценки труда в науке непригоден, он может привести к серьёзным ошибкам. Число публикаций – показатель не слишком объективный: один человек может опубликовать десяток статей, в которых нет ничего существенного, а другой напишет за три года одну работу, но она станет заметным вкладом в науку.

Из этого следует, что при организации соревнования в науке необходимо учитывать специфику глубокого научного поиска. Всякое крупное научное исследование – длительный процесс. Только через несколько лет, после выхода публикаций, получения авторских свидетельств или патентов, после признания результатов научной общественностью можно оценить, насколько крупный вклад в науку или новую технику представляет та или иная работа, каково её место среди всесоюзных и мировых научных и технических достижений. Поэтому представляется, что правильно было бы сравнение (оценку) фундаментальных результатов, полученных исследовательскими институтами, академическими и ведомственными, и конструкторскими бюро, проводить с периодичностью раз в пять лет…» [1].

В этой связи хотелось бы упомянуть о следующем. Недавно, читая некоторые заметки о науке, я обратил внимание на то, что у А. Н. Колмогорова было всего порядка 150 публикаций, причем треть из них – заметки в БСЭ, а у П. Л. Капицы за его долгую творческую жизнь – менее ста публикаций.

Что же происходит в настоящее время? Приведу высказывание профессора А. Клёмина, опубликованное в газете «Поиск» (№ 22, 2010 г.): «Вспоминаются слова молодого казанского преподавателя, только что защитившего кандидатскую: „Мне всё равно, что читать, лишь бы выйти на заработок в 1 000 долларов“. И правда – многим всё равно, что читать. За кафедрами – поколение безликих, всеядных и всечитающих лекторов. В погоне за своей тысячей таким не только научные статьи писать – размышлять некогда.

*Г. И. Марчук – академик АН СССР, лауреат Ленинской и Государственных премий, президент АН СССР.

Посмотрим на сборники статей. Да, они дают возможность авторам получить нужное количество публикаций. Содержание же… Это статьи-однодневки, публикации ради публикаций. Преподаватели вынуждены приспосабливаться к системе, которая, к сожалению, к лучшему не меняется. Приличная научная статья с гипотезой, доказательствами, эмпирикой, обобщениями не менее двадцати страниц. Но стоит ли утруждать себя большим объёмом? Ведь в „послужном списке“ учёного качество публикаций никак не учитывается. Всякое считается за одно – „штуку“. Написал 47 страниц или две – в любом случае засчитается как одна статья. Монография в 600 страниц или в 48 – всё одна монография. И пустое словоблудие, и гениальная статья – счёт тот же: одна публикация. В итоге в России выходят тысячи „научных“ статей и книг, не вызвавших ни отклика, ни пользы».

Вышесказанное имеет далеко идущие последствия. В самом деле, преподавателю университета, да и просто научному сотруднику не выгодно браться за серьёзные задачи. Очевидно, что это потребует значительных усилий, времени. А как же тогда с рейтингом? Ведь показатели, требуемые сегодня, могут оказаться более чем скромными. Отсюда естественный вывод – берись за наиболее простые задачи, опубликовывай полученные результаты, пиши методические пособия и с рейтингом будет всё в порядке. Пока я закончу. Кто хочет высказаться?

Ф. П. Тарасенко (профессор, ТГУ, Томск). В своём докладе я уже говорил, что человек как субъект представляет из себя то, что он знает, какие модели у него есть. И образование – это как раз процесс создания мира моделей данного субъекта. Я уже говорил, что мой любимый автор – Рассел Акофф. Я недавно перевел его очередную книжку, которая при переводе на русский была названа «Преобразование образования». Её английское название очень длинное, и мы договорились с Расселом, что по-русски надо сказать покороче. По-английски название этой книги звучит так: «Перевернуть образование с головы на ноги. Вернуть обучение на правильный путь» – вот такое название книги. Её основная идея состоит в том, что система образования, которая на самом деле сложилась до сих пор не только у нас – в мире, исходила из той самой механистической парадигмы, с которой мы занимались вообще всеми вопросами управления социальными системами. Мы готовили людей в системе образования для того, чтобы они исправно, хорошо исполняли ту функциональную роль, которая им отводится в этой механистической системе. Готовили людей быть полезными членами общества, и всё. А теперь, если мы начали понимать, что человек – это не винтик в системе, а сам тоже целеполагающая система, у которой свои собственные интересы, собственные возможности, что все разные от рождения, – то и образование должно измениться. Главной целью образования должно стать раскрытие тех способностей, таланта, того индивидуального потенциала, который уникален у каждого человека. В этой книжке изложены не только идеи, там предлагаются методы реализации этих идей. В США уже несколько десятилетий работают средние школы, на практике реализующие идею индивидуализации обучения. Сам Р. Акофф в университете штата Пенсильвания даже программу магистерскую сделал такую, которая не предусматривала ни вступительных экзаменов, ни жёстких учебных программ (магистранты сами формируют свою программу обучения). Важно лишь, чтобы человек хотел учиться тому, что способствует его развитию, изучал только то, что человек хочет изучать. Ну, а главные идеи, которые там излагаются, можно сформулировать так: в образовательном процессе самая главная роль принадлежит не учителю, а учащемуся. Учитель только призван создавать условия, помогать раскрытию потенциала данного субъекта: у каждого свои способности, свои возможности, свои модели, свои наклонности. Один хочет свои художественные способности проявлять, другой – изобретательские, третий, допустим, писательские, на что человек, как говорится, от природы настроен, потому что у него, когда он родился, уже заложен весь потенциал. Образование должно быть направлено на то, чтобы этот потенциал реализовать. Как это сделать? Там тоже эти вопросы обсуждаются. Главная идея, например, заключается в следующем: человека можно чему-нибудь научить только в том случае, если он хочет этому научиться. А если он не хочет, ты хоть что тут делай, хоть какие методики, хоть кого приводи, самых умных, приноси всякие материалы, если он не хочет этому научиться, бесполезно все это.

А. В. Медведев. Совершенно с вами согласен. Мы как-то стремительно перешли в новое время. Другой стала страна. Утрачен престиж науки, образования, а значит, и тех, кто работал в этой области. Мы столкнулись с тем, что стали не очень-то востребованными в новых условиях.

Ф. П. Тарасенко. Многие особенности нашей системы образования, особенно недостатки, мы просто не замечаем. Но человеку со стороны они бросаются в глаза. На факультете, которым я руковожу, ежегодно преподают профессора зарубежных вузов. Каждый раз при расставании я задаю им вопрос: в чём разница между нашей и их образовательной системами? Поразительно, но все как один говорят об одном и том же. Всё сводится к четырём недоумённым вопросам ко мне. Первый: «Студенты сидят в аудитории по 36 часов в неделю! А когда же они учатся?» У них (в зарубежных университетах) в неделю бывает 12–14 аудиторных занятий. Остальное время они самостоятельно изучают рекомендованные источники, выполняют самостоятельные работы. Второй: «А почему они на экзамене списывают? Ведь это обман и некачественное овладение знаниями!» За списывание на экзамене следует отчисление из университета без права восстановления. Третий: «А почему вы позволяете пересдавать экзамен? Ведь то, что следовало упорно изучать целый семестр, невозможно освоить за несколько дней!» Это к вопросу нашего отношения к качеству подготовки специалистов. Четвёртый: «А почему они пропускают занятия, даже если оплачивают обуче-

ние? Ведь это всё равно как прийти в магазин, заплатить за товар, и не забрать его!».

Вот поэтому и идея состоит в том, чтобы все образование было построено на желании чему-то учиться, основанном на раскрытии своего потенциала. Там расписывается, в том числе и то, как оценивать усвоение знаний, и так далее. Короче говоря, обсуждая проблемы современной постановки образования, надо мировой опыт решения этих проблем изучать и перенимать.

А. В. Медведев. Безусловно это следует делать, но стоит ли копировать? Я напомню одно из высказываний Б. Трентовского, приведенного в книге Ф. П. Тарасенко: «При одной и той же политической идеологии кибернет (руководитель) должен управлять различно в Австрии, России или Пруссии. Точно так же в одной стране он должен управлять завтра иначе, чем сегодня». Здесь речь идет об управлении государством, но в равной степени это применимо и для управления в стране образовательным процессом, равно как и другими областями деятельности.

Ф. П. Тарасенко. Ну, а что касается рейтинга научно-педагогических работников, Александр Васильевич предлагал, чтобы я вам про рейтинг рассказал. Наверное, это потому, что я в свое время проблемой упорядочивания в многомерных пространствах специально занимался. Мы рассматриваем рейтинг именно как процедуру упорядочивания в многомерном пространстве, и это процедура принципиально не однозначная. Решающую роль играет выбор упорядочивающей функции. Поэтому надо искать не единственное наилучшее решение, надо просто находить паретовское множество в этом многомерном пространстве…

В. А. Филимонов (профессор, Омский филиал Института математики СО РАН, Омск). Я задаю вопрос и задам его каждому, кто возьмет микрофон. Что из того, что вы говорите, вы будете делать в ближайшем году?

А. В. Медведев. Мы не просто говорим, мы это делали и делаем сегодня. Главное состоит в том, что уже с третьего курса бакалавриата студент начинает вести исследовательскую работу с избранным им научным руководителем. В последующем она переходит в выпускную работу и завершается защитой магистерской диссертации. Ясно, что за этот период появляются соответствующие научные публикации, доклады на конференциях, гранты и т. д.

Ф. П. Тарасенко. Дело в том, что с нами уже происходит реформа образования, переходы на многоступенчатое образование, включая аспирантуру и докторантуру, послевузовское и непрерывное образование, дистанционное образование, кое-где уже началось включение средних специальных учебных заведений в состав профильных вузов. Но следует признать, что всё это ещё не решает проблем индивидуализации обучения.

А. В. Медведев. Кстати, о вопросах обучения в аспирантуре и докторантуре. Приведу некоторые выдержки из статьи профессора Г. Ибрагимова «Требуются неудобные», опубликованной в газете «Поиск» (№ 44, 2010 г.). «Сложившаяся сегодня практика такова, что нередко членами диссертационных советов… оказываются доктора наук, не имеющие в своём активе уже защитившихся под их руководством кандидатов наук. Среди них, к сожалению, попадаются люди случайные, кое-как написавшие и кое-как защитившие свои собственные диссертации и, как правило, неизвестные в научном мире, поскольку у них нет сколько-нибудь заметных работ... Такие доктора наук, становясь членами совета, начинают воспроизводить себе подобных. Отсюда низкий уровень экспертизы диссертаций. <…> Ещё одна проблема, связанная с формированием состава советов. Практически в каждом диссертационным совете есть члены, „неудобные“ для руководства и некоторых членов совета. „Неудобство“ их проявляется в том, что они задают слишком много вопросов, они критически выступают, строго следят за соблюдением процедуры защиты и т. п. Казалось бы, такое поведение находится в полном соответствии с требованиями „Положения…“, и таких членов совета надо только поддержать. Однако на практике зачастую председатели советов стараются от них избавиться, и это им легко удаётся».

Н. Г. Загоруйко (профессор, ИМ СО РАН, Новосибирск). Ну, я не считаю себя специалистом в образовании. Просто попытаюсь сказать, что, мне кажется, можно было сделать и что я сам собираюсь сделать. Уже было сказано, что сейчас, кого ни спроси, говорят, что уровень подготовки специалистов стал хуже. В вузах приходится восполнять некачественное школьное образование. Но даже при написании дипломного проекта прямо в самом тексте столько ошибок! Не знают русского языка. Вот приходится им запятые расставлять, и так далее. «Нынче молодежь совсем другая: старших не уважают, учиться не хотят», – так писал Платон за 4 века до нашей эры. Так что оценки, что сейчас плохо, могут быть обусловлены различными жизненными взглядами, опытом и так далее. Ну, а так, в общем, какой я вижу недостаток – студенты сейчас могут черпать знания из Интернета, могут на машине считать сложные вещи, даже не понимая, что они при этом делают. И вот это плохо. Мой знакомый, уехавший во Францию, рассказывал, как его студентка, получив задание посчитать диаметр земного шара по исходным данным, пришла и со светлыми глазами сказала, что у нее получилось 6,3 км. И она даже не задумалась. А у меня в опыте было, что одна моя студентка пришла и сказала, что у нее получилась надежность распознавания 130 %. Значит, иногда студенты не очень вникают в суть того, что они делают. Вот когда сам процесс обучения был ближе к ручному, не компьютеризованному труду, то студент, пользуясь микроощущениями, когда он это все дело считал, он задумывался и уделял больше времени этой задаче. Мне кажется, понимали суть задач лучше. Может, это просто частное наблюдение. Вот, что мне кажется, куда надо двигать образование в нашей области, скажем, в области информатики – активизировать студента в процессе передачи знаний, сделать учебный процесс активным, а это делается, например, с помощью каких-либо обучающих игровых систем, игровых программ. Ну, например, мне понравилось, что даже для школьников есть програм-

ма «История Англии». Показана такая деловая игра: ты Карл Великий, ты решаешь, как расставить войска. Вот он расставлял их так, и дальше события происходили вот так-то. Вот тут уже приходится не только выучить и узнать из истории какие-то факты, но и просто модель события воспроизвести. А вот в нашей области я читаю курс, который называется в разные годы по-разному, но, все равно, «Интеллектуальный анализ данных». Значительная часть курса посвящена алгоритмам, а каждый алгоритм можно себе представить вполне в доступной визуальной форме и с возможностью участия студентов в формулировке гипотез и данных исходных, и отслеживать: как идет процесс, что у него там делается. Мне кажется, это очень сильно может ускорить изучение и позволит лучше понимать. Сам я собираюсь сделать такой курс и сделать его доступным через Интернет. Идея дистанционного образования представляется мне очень интересной, вот я и хочу в этом направлении поработать.

Если же говорить о рейтинге… (*Его прерывает А. В. Медведев*).

А. В. Медведев. Я бы хотел заметить, что рейтинг (оценка, порядок, классификация, субъективная оценка какого-либо плохо измеримого явления по заданной шкале) – «функция» не убывающая, а то, что сегодня принято называть рейтингом – это, скорее, показатели работы текущего года.

Ф. П. Тарасенко. У меня создалось впечатление, что наш спор вызван типичной причиной всех споров и разногласий – различным пониманием одних и тех же слов. По-моему, те из нас, кто считают, что упорядочивать можно (и нужно!) без критериев, понимают под критериями некоторые измеримые, количественные характеристики оцениваемого объекта. И тогда они правы в том, что во многих случаях важные качества объекта просто невозможно оценить количественно. Гениальность Пушкина или Эйнштейна, например. Со своей стороны, я, отстаивая точку зрения, что упорядочивание без критериев невозможно, имел в виду, что критерии есть способ оценивания, и могут быть не только числовыми, но и качественными (представимыми в слабых измерительных шкалах), и (о, ужас!) неосознаваемыми, интуитивными, эмоциональными (в последнее время всё больше говорят о роли ***неявных*** знаний в поведении человека). Если с этим пониманием термина «критерий» согласиться, то наш спор затихнет. Остаётся только вопрос о различиях использования упорядочивания в целях познавательных, проектных и управленческих, но это уже совсем другая тема.

Н. Г. Загоруйко. Я по поводу рейтинга, Феликс Петрович. Задача упорядочивания: все объекты с их характеристиками нужно установить в некий порядок, их расставить по рангу, от самого лучшего до самого плохого. Это касается оценок работы вузов, и из этого идут всякие последствия в смысле грантов, признания и так далее. Это касается сравнения любых многомерных объектов. Ну, со студентами я такого не встречал, а вот что касается сравнения кафедр и вузов, да, я встречал и принимал участие. В чем там трудность? Задача формулируется всегда в незаданном виде, не излагающем ясных критериев. Вот как упорядочить кафедры? Я слышал, вот такие-то признаки есть и такие-то. Возьмешь другие признаки, или даже эти, но с другими весами – получишь другой порядок. Чем замкнуть? Как проверить, что хорошо, что плохо? Вот в одной работе я предложил сделать так: давайте мы сначала попросим достаточно разумных, трезво мыслящих профессоров, хорошо знающих вуз и все кафедры, чтобы они упорядочили кафедры, как эксперты, и пытались, когда они это будут хором обсуждать, отстаивать свое виденье. Может, кто-то не знал чего-то, а он знает, надо это еще обсуждать, вот если они в дискуссии установят некий порядок, тогда задача становится замкнутой. Значит, мы берем некое исходное множество признаков, применяем методы выбора информативных признаков, определяем коэффициенты взвешиваний, а критерий у нас уже обеспечен, нужно, чтобы по этим признакам с этими весами кафедры упорядочились так же, как и упорядочили их эксперты. При этом станет прозрачным, почему именно такой вес, а не другой. Совершенно очевидно, что если изменить весовые коэффициенты показателей, то положение кафедры в рейтинге существенно изменится.

Что касается упорядочения региональных субъектов, пришлось мне как-то попросить Феликса Петровича прорецензировать одну работу, значит, способ упорядочивания субъектов Российской Федерации по показателям общим. Там применяются модели без вот такого замыкания, а только процессы того, что если я разумно выберу признаки, и эксперты одобрят это, то тогда, что бы ни получилось – это оно и есть. При этом так получалось в этой статье, чтоб повысить благосостояние данного региона, нужно уменьшить значение некоторых характеристик, в частности, продолжительность жизни, вот тогда оставшиеся будут жить хорошо. В такой интерпретации этого не было в статье, но сам факт наличия в результате вот такой вещи говорит о том, что вся система ранжировалась неправильно. Я глубоко убежден, что нужно каким-то образом обеспечивать замыкающий критерий. Экспертами теми же или, может быть, коллективом решающих правил, множество вариантов, а потом это как-то их усреднять, как-то я не уверен, что это спасет положение. В общем, задача очень трудная, трудная потому, что эксперты неохотно этим занимаются. Возможно, субъективизм играет роль – какой-то заведующий кафедрой больше нравится, и хочется поставить эту кафедру повыше. От этого никак не отделаешься, это всегда признак оценивания, влияния на эксперта могут плохо сказываться. В общем, до конца эту задачу я не знаю, как решать строго, потому что, если брать на душу грех и решать эти задачи, то надо все-таки побеспокоиться о каком-то способе оценки: правильное получили решение или нет. Спасибо.

Ф. П. Тарасенко. Я попробую ответить на вопрос Николая Григорьевича. Значит, задача ранжирования –

это, действительно, задача упорядочивания. А упорядочивание – это значит определить, что лучше, что хуже. А когда мы говорим «лучше–хуже» – это оценочные слова, и, для того чтобы сказать, что лучше, что хуже, надо сказать, какой критерий мы используем для упорядочивания. Так вот, оказывается, что есть задачи, которые совершенно очевидно и правильно решаются этим способом, если это однокритериальная задача. Если у нас есть один критерий, вот (*пишет на доске*). Давайте скажем, что чем больше это значение, тем лучше. Вот мы берем и измеряем этот критерий. Вот у нас есть несколько, допустим, три объекта, которые надо упорядочить. Вопрос совершенно очевидный. Надо на первое место поставить того, у кого критериальная функция больше. Как только мы переходим к многокритериальной задаче, то ситуация сразу совершенно меняется и усложняется. Есть такой математический анекдот: кибернетика спросили, может ли он написать алгоритм получения чая. Он ответил, что нет ничего проще: «Операция первая: берем чайник. Операция вторая: наливаем воду. Операция третья: ставим на огонь. Операция четвертая: дожидаемся кипения. Операция пятая: кидаем заварку. Операция шестая: ждем 2 минуты. Чай готов». А теперь ему говорят: «Можешь ли ты написать алгоритм получения чая, если я тебе дан чайник с уже налитой водой?». Тот говорит: «Так ничего проще нет: выливаем воду и сводим задачу к предыдущей». Что это означает? Это означает, что если человек сталкивается с задачей, которую он не решал и не знает, как решить, то он пытается ее свести к той задаче, которую знает, как решить. А мы знаем, как упорядочивать однокритериальные варианты. Поэтому все начали придумывать, какую однокритериальную задачу сформулировать, которая бы учитывала много критериев. А сделать это можно по-разному. И много вариантов есть. И, более того, есть даже такое сообщество, вроде нашего, но международное. Называется «Многокритериальное принятие решений». Каждые два года они собираются где-нибудь. Я в своё время был на нескольких таких симпозиумах, а последнее время я туда перестал ездить, потому что там занимаются вопросом «как построить глобальный критерий для большого количества параметров». В самом деле, многокритериальные задачи имеют большую размерность. А на самом деле, предположим, что я принял линейный критерий. Это означает, что я беру линейную упорядочивающую функцию. Стоит мне изменить коэффициенты, сразу упорядочивающая функция меняется. И теперь нам ясно, что по этому критерию эта будет лучше. Получается так, что если я строю глобальную функцию упорядочивания, то я найду порядок качества в смысле этого глобального критерия. И там все зависит от того, какие я параметры у этой функции задал. Потом начинают спрашивать: «Почему именно линейная комбинация? У нас на одном, как раз таком, совещании, как сегодняшнее, принимал участие Я. З. Цыпкин[*]. Он говорит: «Вот вы мне скажите, как вы хотите, чтобы я упорядочил эти точки? Вот первая у вас должна быть самая лучшая, а остальные хуже. Я бы придумал такую функцию упорядочивания, которая это и сделает. Только она будет замысловатая». Становится ясно, что упорядочивание в многомерном пространстве одномерной функцией – дело произвольное. А между тем давно, ещё в прошлом столетии, эта задача решена правильно. Что значит правильно упорядочить? На этот вопрос ответил Парето[**]. Многомерное упорядочивание нужно производить на порядке доминируемых и доминирующих альтернатив. Давайте назовем альтернативу, которая по всем критериям лучше другой, доминирующей. А ту, которая по всем критериям хуже – доминируемой. Тогда многомерное упорядочивание сводится к одной простой операции: давайте выбросим все доминируемые альтернативы. Оставим недоминируемые, потому что нет альтернатив лучше каждой из них по всем критериям. Они обе хороши в смысле этих критериев. А между собой они несравнимы. Поэтому выбор в многомерном пространстве таков: давайте найдем множество недоминируемых альтернатив. В случае континуального множества альтернатив найти паретовское множество, во-первых, означает, что все внутренние точки нужно выкинуть, потому что они все доминируемые. Значит, у нас останутся только граничные, и то не все. После удаления граничных доминируемых точек, все альтернативы, которые остались, и есть лучшие. Но не зря же говорят, что нужно упорядочить всех. Тогда, как сделать выбор на паретовском множестве? Ответ: эти точки между собой несравнимы. Поэтому выбор на паретовском множестве может быть любым. А на практике? Мне же надо выбрать что-то одно. Как? А методов сколько угодно. Хотите, случайный выбор делайте. Хотите, критерии применяйте. А есть еще метод волевого выбора. Вот как-то, когда всех заведующих кибернетическими кафедрами раз в пять лет вызывали в Москву, там перед нами выступали самые квалифицированные люди в нашей области. И вот однажды перед нами выступал генеральный конструктор самолетов-истребителей. Рассказывал: «Я должен построить истребитель. Самый лучший. То есть он должен быть лучшим по таким, таким и таким-то критериям. Есть критерии экономические, есть военные, и так далее. Около двухсот критериев. Нужно найти паретовское множество на одной группе критериев. Потом на другой, и так, пока не дойдем до конца. Все равно получим паретовское множество. А вот тут как раз и наступает моя роль, как генерального конструктора. Я ни перед кем не отчитываюсь. И выбираю сам модель самолёта». Я спрашиваю: «А что будет, если самолеты встретятся в небе?» Он ответил, что это уже не зависит от конструкции самолета, так как он не может заложить в конструкцию, храбрый летчик или трусливый, расстрелял весь боеприпас или нет, и т. д. То есть исход будет зависеть от обстоятельств, но они никак не связаны с тем, какие критерии самолета мы выбрали. Вот что означает многокритериальный выбор.

[*]Я. З. Цыпкин – академик АН СССР, лауреат Ленинской премии.

[**]В. Парето – итальянский экономист и социолог.

А теперь, если вернуться опять к ранжированию учебных заведений, то тогда тут ситуация интересная получается. Вот, например, я буду делать выбор с помощью линейных комбинаций. Это означает, что я каким-то критериям придаю больший вес, а каким-то меньший. Что это означает? Это означает то, что я хочу, чтобы подчинённые мне подразделения усиливали работу в определённом направлении. Вот, например, рейтинги университетов. Почему российские университеты где-то там очень далеко, даже самые лучшие наши университеты? Потому что, например, в мировом рейтинге огромный вес придаётся наличию нобелевских лауреатов среди преподавателей. Если у данного университета нет нобелевских лауреатов ни среди выпускников, ни среди преподавателей, он никогда не получит высокий рейтинг. Или, скажем, сейчас очень важным критерием считается цитируемость. Если цитируемость ваших публикаций по мировым индексам цитируемости большая, значит, у вас есть большой шанс. Сейчас Томский университет получил статус национального исследовательского университета с мощными поощрениями, деньги будут в ближайшие годы даны университету на развитие. Но правительство наше сразу установило критерии. Поэтому сейчас публикации ценятся только в цитируемых и, более того, в зарубежных цитируемых журналах. И если этот показатель будет слабенький – всё, через какой-то срок скажут: «Вы не годитесь в исследовательский университет».

Вот так представляется проблема упорядочивания в многомерном пространстве. Главное: в многомерном пространстве упорядочивание неоднозначно, принципиально неоднозначно. А если мы хотим однозначное упорядочивание, то тогда возникает вопрос уже об относительной важности критериев, а это политический вопрос, а не вопрос измерений.

А. В. Добровидов (профессор, ИПУ РАН, Москва). Проблема объективной оценки научной деятельности ученых и преподавателей, занимающихся кроме обучения научной работой, действительно очень важная и актуальная проблема со многими последствиями для качества жизни людей и всей страны в целом. Я полностью согласен, что существующий подход к оценке труда ученого при помощи числа публикаций никак не соответствует истинной ценности (рейтингу) научного работника. С другой стороны, Ф. П. Тарасенко совершенно прав, напомнив в дискуссии, что в изначальном документе Минобразования предлагалось учитывать те работы, на которые есть ссылки у нас и за рубежом. Эта составляющая критерия оценки так и осталась неиспользованной, поскольку, с одной стороны, индекс цитируемости вычисляется только на Западе, а с другой – русские работы, даже хорошие, никто на Западе не читает, так что откуда взять этот индекс? Альтернативу ему не нашли и критерий вылился в поток слабой и ненужной научной продукции. К примеру, я совсем недавно при рецензировании вынужден был просматривать работы одного в прошлом советского специалиста в нашей области, который сейчас работает в США. Его последняя большая книга была написана в 1995 г. в Штатах. Он, конечно, мастак по рекламе. Но тем не менее его индекс цитируемости в 2009 г. после 14 лет его основной публикации выражается громадным числом. Это поразительно. В то же время индекс цитируемости, равный 0,5–1, считается вполне приличным. Так что, как правильно говорит Ф. П. Тарасенко, а до этого еще в 60-е гг. говорил Я. З. Цыпкин, все дело в критерии. Правда, Я. З. Цыпкин добавлял, что критерий не выбирают. В рассматриваемой же проблеме критерий надо выбирать, проверять и стремиться к разумным результатам, получаемым из этих критериев.

А. В. Медведев. Я вспоминаю доклад Якова Залмановича по идентификации, как раз связанный с выбором критериев. На слайде был изображен конькобежец, бегущий по льду к кромке, за которой льда уже не было. Это была запретная область для принятого критерия идентификации. Там бегун должен упасть. Это была иллюстрация того, что о критериях не спорят.

А. В. Добровидов. Вы спросите, какой же критерий я предлагаю? Ну, конечно же, я не знаю. Могу сказать только, какой разумный вывод (результат) я бы хотел получить из этого критерия. Я думаю, так же, как и в университетах, в научных институтах надо повышать зарплаты достойных (по выбранному критерию) научных сотрудников. Откуда брать деньги? Конечно, от государства. Но вместе с тем вы вряд ли будете возражать, что и в университетах, и особенно в научных институтах после всех сокращений все равно полно либо просто бездельников, либо бесполезных для науки людей. Конечно, их надо сокращать и дальше, если зарплаты уволенных оставлять в фонде лаборатории, кафедры или научного учреждения. Такой опыт оказался очень полезным и продуктивным в нашем институте в 90-е гг.

Н. Г. Загоруйко. Я говорил, что принимал участие в конкретной работе по упорядочиванию кафедр. 17 кафедр, нет ни одной доминируемой и ни одной доминирующей.

Ф. П. Тарасенко. Так не может быть.

Н. Г. Загоруйко. Но хоть по какому-то критерию да лучше, хоть по какому-то да хуже.

Ф. П. Тарасенко. Да, правильно.

Н. Г. Загоруйко. Но отбрасывать некого.

Ф. П. Тарасенко. Паретовское множество надо найти. В паретовском множестве единственная альтернатива бывает только в очень экзотическом случае, когда само множество уникальный вид имеет. Понятное дело, что это будет единственная альтернатива, которая доминирует над всеми. Но обычно паретовские множества – это множества. Я думаю, что сейчас какой-то проект в образовании можно делать только где-то в зазоре существующей системы. Есть деревня в Томской области, откуда регулярно появляются ребята, хорошо знающие информатику, потому что там есть один замечательный учитель, и там их можно воспроизводить.

А. В. Медведев. Я бы попросил, чтобы выступили по этому поводу студенты, ради которых всё это делается. А пока, давайте заглянем в прошлое. Мне очень понравилось движение, которое происходило в 1999 г. Была сделана попытка определить человека второго тысячелетия. По телевидению было много круглых столов в преддверии третьего тысячелетия. Видимо, по инициативе журналистов, пытались проанализировать вопрос: кто же мог быть человеком второго тысячелетия? Что получилось? Естественно, была принята какая-то процедура отбора, я её не помню, но важен итог. На верхнюю планку были подняты А. Эйнштейн и У. Шекспир. Я не помню, кто был избран из них. Важно другое: и та и другая фигура достойны, если они действительно оказались на верхней планке, то это вполне разумно. Кто это решал? А это решили люди без всяких инструкций и рецептов, к которым мы часто стремимся. Если бы по нынешним рейтингам оценивали А. Эйнштейна в 1940 г., то он мог бы оказаться на задворках.

Студенты третьего курса (СибГАУ). Прежде всего, следует отметить довольно слабую подготовку в средней школе. Одна из проблем состоит в том, что преподаватель сравнительно равнодушно относится к студентам, неинтересно читает лекции, тем самым отбивая желание учиться. Также было бы полезно включить в учебные планы дополнительные практические занятия.

Студенты четвёртого курса (СибГАУ). На наш взгляд, образование – это процесс непрерывного плодотворного сотрудничества студентов и преподавателей. Для повышения мотивации студентов к получению новых знаний и применению их в практической деятельности недостаточно только желания преподавателей. Студенту необходимо обладать собственным сознанием необходимости, полезности образования, что, безусловно, будет способствовать как эффективному усвоению учебного материала, так и применению новых знаний в профессиональной деятельности.

Студенты пятого курса (СФУ). Ещё одна проблема – отдалённость преподаваемых предметов от реальных ситуаций, в которых окажутся выпускники, устроившиеся на работу. Технологии, о которых рассказывают преподаватели, не имеют реального применения на производстве. В настоящее время выпускнику трудно устроиться на работу по специальности, поэтому большинству всё равно, какой диплом получать – лишь бы безболезненно. Поэтому цель обучения студента, к сожалению, диплом, а не получение навыков и знаний.

Аспиранты (СибГАУ). Хотелось бы видеть в лице преподавателей сложных комплексных дисциплин не новоиспечённых кандидатов и аспирантов, а опытных практиков, которые могут не только прочитать лекцию, но и компетентно ответить на вопросы аудитории.

А. В. Медведев. Вернёмся, однако, к тому, как следовало бы учить студентов в диалоге «учитель–ученик». Я думаю, что нужно делать то, что делали в прошлом. Как-то давно мы с Феликсом Петровичем обсуждали книгу профессора Р. Юнга «Ярче тысячи солнц». Там описана атмосфера обучения студентов в 20-е гг. прошлого столетия в старинном университете Георгии Августы в Геттингене. Привожу выдержку из этой замечательной книги: «В годы после Первой мировой войны такую магнетическую силу приобрела атомная физика. Поскольку в этой области оказалось много нового и неопределённого, учителя и ученики здесь сплотились гораздо теснее, чем при изучении других научных дисциплин. Прежние заслуги ценились не очень высоко. Старость и молодость становились равноправными товарищами в этом походе внутрь материи. И та и другая одинаково гордились своими успехами и проявляли одинаковые скромность и смущение перед лицом неведомого.

Профессора не делали секрета из своих ошибок и сомнений. Они знакомили учеников с частной корреспонденцией, в которой обсуждали с иностранными коллегами нерешённые проблемы. Всем этим они вдохновляли молодёжь на новые поиски.

Джеймс Франк, обладавший к тому времени Нобелевской премией, проводя сложнейшие вычисления и утеряв путь дальнейших выкладок, мог, например, отвернуться от доски и спросить у одного из своих студентов: „Может быть, вам удалось увидеть следующий шаг?“

В течение семестра кульминационным пунктом каждой недели бывали „Семинары о материи“, проводившиеся в институте М. Борном, Дж. Франком и Д. Гильбертом. Для Д. Гильберта стало почти традицией открывать работу семинара фразой „Итак, господа, подобно вам, я хотел бы, чтобы мне сказали точно, что такое атом?“. И каждый раз студенты старались просветить профессора. За проблему энергично принимались сызнова и пытались найти новое решение. Но каждый раз, когда кто-нибудь из молодых гениев начинал искать спасение в доступных лишь избранным высотах усложнённых математических толкований, Д. Гильберт прерывал его: „Я совершенно не могу вас понять, молодой человек. Не угодно ли вам будет рассказать всё снова?“ Таким образом, каждый был вынужден высказываться с максимальной ясностью и строить прочные мосты через провалы в знаниях, а не перепрыгивать через них путём поспешных умозаключений» [2].

Достаточно сказать, что многие студенты впоследствии стали нобелевскими лауреатами.

В. А. Филимонов. Александр Васильевич, так если ж вы никаких проблем не видите, так зачем вы этот разговор-то затеяли?

А. В. Медведев. Я не просто вижу проблему. Я ее, если хотите, ощущаю. Вот одна из характеристик, показывающих качество нашего нынешнего образования. Следующая ниже фраза журналиста А. Любимова взята из книги «Имя Россия. Исторический выбор»: «Моя знакомая – декан факультета журналистики одного уважаемого университета – принимает зачёт у первокурсников. Их первая сессия.

– Скажите, а когда в России появилось периодическое печатное издание?

– При Петре I.

Моя знакомая-декан уже заносит ручку, чтобы поставить „зачёт“, но не может отказать себе в удовольствии вступить с выдающимся студентом-первогодкой в интеллектуальный диалог.

– А в каком году это было?

– В 1917!

Моя знакомая в шоке. И стыдно, и горько. Да и сверху разнарядка – „не валить“ первокурсников. И она решила студенту помочь наводящим вопросом.

– То есть вы хотите сказать, что Пётр I был вдохновителем и организатором Октябрьской революции?

– Совершенно верно» [3].

Ф. П. Тарасенко. В чём проблема-то?

А. В. Медведев. В образовании. Проблема состоит в том, что качество образования упало, и то, что делается сегодня – мягко говоря, странно. Это касается и рейтингов, и критериев оценки знаний, и многого другого.

В. А. Филимонов. Александр Васильевич, некоторое время назад в Омске появился большой плакат. На этом плакате симпатичный такой мальчик, и там написано: «Школа должна растить гармоническую личность (с этим я согласен – *В. Ф.*), а не специалиста». Подписано: А. Эйнштейн. Так вот, противопоставлять процессы формирования личности и специалиста ошибочно. Мне кажется, при формулировании проблем образования это надо рассматривать системно, но раздельно.

А. В. Медведев. Так мы и не противопоставляем.

Ф. П. Тарасенко. Нет, я всё-таки не понимаю, какую проблему видит Александр Васильевич. Ну, например, он говорит: вот А. Эйнштейн у нас вообще ни по какому критерию не был бы первым. Ещё раз повторяю: если ты чего-то упорядочиваешь, значит, ты упорядочиваешь по какому-то критерию или критериям. К себе в аспирантуру Л. Д. Ландау* принимал студента, только если он ему сдаст экзамен по десяти предметам теоретической физики. Не сдал – значит, ты не годишься. Так же и здесь. Вот эта девочка – ей какую отметку поставить? Ну, понятно, какую. Зависит от того, какой критерий. Например, если критерий «за находчивость» – она не смутилась («я не знаю»), а сказала «в 1917» – всё, за находчивость поставить зачёт. Если у тебя критерий не тот, что теряется или не теряется человек, а знает он точно или нет, то у тебя совсем другая будет оценка. Критерий есть всё равно, что бы ты ни делал, он может быть не количественным.

А. В. Медведев. Приведу еще один пример из истории, имеющий отношение к критериям и их оценке. Это касается искривления светового луча, предсказанного в общей теории относительности. Эксперимент был поставлен в мае 1919 г. во время полного солнечного затмения. Итог подводился на объединенном заседании Британского и Лондонского астрономического королевских обществ. Председательствовал «Джи-Джи»**. Он начал читать результаты обработки данных экспедиций, «его рука, державшая бумажный лист, дрожала» (критерий!). После прочтения результатов эксперимента (теория А. Эйнштейна подтвердилась) «глубокое молчание воцарилось в зале» (критерий!). Вот они критерии – неколичественные, но они-то и являются определяющими.

Ф. П. Тарасенко. Николай Григорьевич рассказывал интересную историю. Как-то они сравнивали разные колхозы в Новосибирской области. Председатель лучшего колхоза – Герой труда, жуткие деньги дают – в чём дело? Почему он самый лучший? Смотрят: да просто потому что он друг самого большого начальника, который решает, кому дать деньги, а кому не дать. Вот и все критерии. Понимаешь, какая штука: без критериев упорядочивания быть не может, и весь вопрос об упорядочивании: либо мы, как бы сказать, просто сравниваем между собой альтернативы, тогда надо искать паретовское множество. А если мы упорядочиваем из политических соображений, то есть мы хотим, чтобы вся ситуация менялась в нужном направлении, то мы говорим: мне вот надо, чтоб ты делал вот это, если сделаешь, то получишь поощрение, не сделаешь – наказание. То есть мне требуется не объективное измерительное сравнение, а субъективное, управленческое. Поэтому тут я особых проблем просто не вижу.

А. В. Медведев. В общем, я не буду объяснять Феликсу Петровичу существо проблемы, если он не видит, то, значит, не видит, что поделаешь. Я привёл достаточно примеров, по-моему, достаточно убедительных.

Ф. П. Тарасенко. Ну и какие критерии были?

А. В. Медведев. Да не было никаких критериев, которые можно количественно измерить.

Ф. П. Тарасенко. Да так не бывает, чтобы не было критериев, иначе не может быть сравнения.

А. В. Медведев. Продолжайте фокусироваться на критериях. Я просто приведу для вас высказывание Джеймса Франка из упоминавшейся уже книги Р. Юнга «Ярче тысячи солнц»: «Единственным критерием, по которому я могу судить о действительной важности новой идеи, является чувство ужаса, которое охватывает меня». Я не против использования критериев. Хочу отметить другое, что есть критерии, которые можно как-то представить численно, представить в виде каких-то показателей, но есть многие другие, которые совершенно невозможно измерить.

В частности, сейчас даже оценка диссертационных работ осуществляется в значительной степени по формальным показателям. На совете часто возникает вопрос к соискателю о наличии научного результата докторского ранга, и уже в течение многих лет я не слышу вразумительного ответа. Ответ обычно сводится к наличию совокупности результатов. Таким образом, на пьедестале оказываются формальные параметры, учебные пособия, которых в Интернете великое множество, но не научный результат.

---

*Л. Д. Ландау – академик АН СССР, лауреат Ленинской и Государственной премий, нобелевский лауреат.

**Томсон Джозеф Джон – английский физик, нобелевский лауреат.

В конечном счёте это привело к тому, что качество диссертационных работ резко снизилось. По этому поводу я уже говорил и упоминал статью Г. Ибрагимова. Но эта ситуация обязательно будет иметь продолжение, которое состоит в том, что все эти доктора и кандидаты наук окажутся членами диссертационных советов, различного рода экспертных советов, фондов и др. Рано или поздно именно им придётся принимать или готовить решения различных уровней сложности. Но преподавателей побуждают писать методички, делать как можно больше публикаций, т. е. возникли какие-то требования, которые, я думаю, играют пагубную роль.

Ф. П. Тарасенко. Пагубное – это оценочное слово. По какому критерию пагубный?

А. В. Медведев. По критерию качества образования. Посмотрите: ввели показатели для оценки рейтинга преподавателя – началась погоня за показателями. Диссертационных советов стало значительно больше, а один из показателей работы совета – число защищенных диссертаций в год. Результат: началась погоня за соискателями ученых степеней. Общий итог: уровень образования ухудшился, уровень диссертаций снизился.

Нынешний подход к оценке рейтинга преподавателя по существующим критериям побуждает его не браться за серьезные проблемы.

Ф. П. Тарасенко. Тогда и не говорите об упорядочивании.

А. В. Медведев. Почему?

Ф. П. Тарасенко. Потому что вы не хотите даже знать, что такое критерий.

А. В. Медведев. По какому критерию Пушкин гениальный поэт? По какому?

Ф. П. Тарасенко. Объясняю. Потому что он оказался автором литературного русского языка. Это один критерий. Второй критерий: потому что он гениальный поэт.

А. В. Медведев. Вот именно – гениальный. Приведу очень простой пример – вот они (студенты) сидят и слушают нас. До четвёртого курса они будут учиться нормально, на пятом, поступив в магистратуру, они начнут искать работу, а на шестом курсе учатся уже единицы. У них нет времени, чтобы заниматься исследованиями, учебой, вот и всё. Изменилось время.

Ф. П. Тарасенко. Нехватка времени – вот тебе критерий.

А. В. Медведев. Но почему появилась эта нехватка времени? Я бы хотел напомнить: История, Время оставили нам имена выдающихся ученых, педагогов, ученых-учителей. Их оценивали не по показателям, а по результатам содеянного. И они навсегда вошли в Историю.

Тем не менее сегодня оценивать как-то нужно. Один из возможных путей – доклад на профессорском собрании, которое, наверное, есть в каждом университете, и оценка преподавателя научным сообществом. Определение процедуры, технологии оценивания целесообразно отдать на «откуп» профессорскому собранию каждого университета. Конечно же, опыт прошлого, и российский, и зарубежный, а также сегодняшние условия следует учитывать.

Голос из зала. Александр Васильевич, вы видите выход?

А. В. Медведев. Я убежден, что из сегодняшней ситуации, впрочем, как и из любой другой, выход есть. Было бы легкомысленно с моей стороны так сразу давать рекомендации. Это дело всего научно-образовательного сообщества. Хочу только сказать, что все равно придется ответить на вечные российские вопросы, означенные в заголовке дискуссии.

Ну, а теперь общий итог:

– с 90-х гг. прошлого века уровень образования начал резко падать, одновременно упала востребованность в научно-педагогических кадрах;

– при всём этом, в каком бы состоянии ни находилось образование, всегда были и будут хорошие студенты (только их становится всё меньше и меньше);

– существенная черта современного процесса образования – это слияние образовательной и научно-исследовательской деятельности;

– обычная ситуация – студент хорошо заканчивает бакалавриат, уже имеет публикации, поступает в магистратуру, в дальнейшем – в аспирантуру, потом находит хорошо оплачиваемое место работы, и защита диссертации существенно затягивается;

– критерии, по-видимому, следует сохранить, но принимать окончательное решение о рейтинге преподавателя следовало бы не только по критериям, а с учётом их;

– было бы целесообразным устанавливать рейтинг преподавателя (с учётом его показателей) по результатам его доклада на «профессорском собрании», которые существуют, наверное, во всех университетах;

– первой ступенью при оценке рейтинга преподавателя считать принятые показатели и соответствующие им весовые коэффициенты. Второй – результаты голосования по докладу, сделанному на профессорском собрании;

– изменить показатели, предъявляемые преподавателям при оценке их рейтинга, таким образом, чтобы не стимулировать бесконечного числа разного рода публикаций, методических пособий и т. п;

– следовало бы во главу угла при определении рейтинга преподавателя поставить научный и научно-педагогический результат, который также оценивать на профессорском собрании;

– оценка рейтинга университетов осуществляется по показателям и процедуре, выработанными Мин-

обрнауки. Конечно же, здесь целесообразно учесть мнение и научного сообщества;

– ВАК следовало бы создать при правительстве России, как это было ранее; в частности, это было отмечено в газете «Поиск» (№ 32–33, 2010 г.);

– было бы целесообразным исключить из показателей работы диссертационных советов количество защищенных диссертаций.

Материалы дискуссии
подготовлены к публикации
профессором А. В. Медведевым

P. S. Когда материал дискуссии уже был подготовлен к опубликованию, от профессора В. А. Филимонова было получено электронное письмо по результатам 4-й конференции «Рефлексивный театр ситуационного центра», проходившей в г. Омске. Приведем некоторые выдержки из письма.

«Большинство высказываний содержат критику существующих концепций, подходов, методов, политики и т. п., а также рецепты „правильных“ концепций, подходов, методов, политики и т. п. Практически не обсуждаются сценарии, когда ничего из „правильного“ набора реализовано не будет, и прогнозов трансформации образования вплоть до его ликвидации нет».

«Мой пессимистический прогноз состоит как раз в ликвидации образования по аналогии с ликвидацией коммунистического варианта цивилизации в СССР».

«Логично заняться сохранением генофонда образования: образцов его реализации Мастерами. Во время войны в Ленинграде сохранили селекционный фонд. Надо же: хотел найти ссылку на историю, а нашел свежее подтверждение своему тезису и не радуюсь…»

Библиографические ссылки

1. Марчук Г. И. Молодым о науке. М. : Мол. гвардия, 1980.

2. Юнг Р. Ярче тысячи солнц. М. : Госатомиздат, 1961.

3. Имя Россия. Исторический выбор 2008. М. : АСТ : Астрель, 2008.

Раздел
2

МАТЕМАТИКА, МЕХАНИКА, ИНФОРМАТИКА

УДК:629.7.018.4:620.178.3

В. А. Бернс

ОЦЕНКА ТОЧНОСТИ ОПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК СОБСТВЕННЫХ ТОНОВ ПРИ НАЛИЧИИ СЛУЧАЙНЫХ ОШИБОК В ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Сделаны оценки точности определения собственных частот, обобщенных масс и обобщенных коэффициентов демпфирования собственных тонов колебаний конструкций по результатам резонансных испытаний. Полагалось, что в величинах действительной и мнимой составляющих перемещений точек конструкции присутствуют случайные ошибки измерений. Исследования проводились методом статистического моделирования.

Ключевые слова: резонансные испытания, ошибки измерений, статистическое моделирование.

Исходными данными для нахождения собственных частот, обобщенных масс и обобщенных коэффициентов демпфирования собственных тонов являются экспериментально определенные величины действительной U и мнимой V составляющих перемещений точек конструкции для ряда частот вынужденных колебаний при известных силах возбуждения. Характеристики вынужденных колебаний измеряются в испытаниях с некоторой погрешностью, поэтому представляет практический интерес анализ чувствительности способов определения обобщенных динамических характеристик к ошибкам в экспериментальных данных. Решению этой задачи посвящены, например, работы [1; 2].

Для определения обобщенной массы (a_i) и обобщенного коэффициента демпфирования (h_i) i-го собственного тона колебаний использовались следующие формулы [3]:

$$a_i = \frac{V_i^{\mathrm{T}}(\lambda_i E_i - F_i)}{(1+\lambda_i^2)(p_i^2-\omega^2)V_i^{*2}},$$

$$h_i = \frac{V_i^{\mathrm{T}}(E_i + \lambda_i F_i)}{(1+\lambda_i^2)V_i^{*2}},\ i = 1,2,...,L,$$

где E_i, F_i – векторы синфазной и квадратурной составляющих сил возбуждения, обеспечивающие выделение i-го тона колебаний на частоте ω (если собственный тон можно выделить без введения квадратурной составляющей возбуждения, то такой случай отмечается условием $F = 0$); λ_i – котангенс сдвига фаз между перемещениями и синфазной составляющей возбуждения; p_i – собственная частота тона, определяемая по переходу λ_i через 0; V_i* – мнимая составляющая перемещения точки нормирования формы тона. Если положить $\lambda_i = 0$ при $F \neq 0$ (условие «фиктивного» фазового резонанса), то получим известную формулу определения обобщенных масс введением квадратурной составляющей возбуждения (см., например: [1]).

Рассматривается влияние только погрешностей измерений составляющих вынужденных колебаний на точность определения характеристик собственных тонов, так как экспериментальное оборудование позволяет поддерживать частоту вынужденных колебаний с большой точностью. Отклонения же в задаваемых величинах сил возбуждения можно (в рассматриваемом случае линейных систем) учесть в точности определения перемещений.

Оценка влияния случайных ошибок измерений производилась методом статистического моделирования. В качестве математической модели реальной конструкции использовалось уравнение вынужденных крутильных колебаний стержня постоянного сечения под действием системы сосредоточенных возбуждающих моментов. Такое уравнение имеет точное аналитическое решение.

Будем считать, что ошибки измерений действительной и мнимой составляющих вынужденных колебаний распределяются по нормальному закону с нулевым математическим ожиданием. Для моделирования ошибок использовался датчик случайных чисел. Создание ошибок требуемого уровня осуществлялось соответствующим изменением дисперсии с использованием свойства нормального распределения – случайная величина x с вероятностью 0,997 удовлетворяет соотношению

$$|x| \le 3\sigma, \qquad (1)$$

где σ – дисперсия. Вычисление дисперсии по формуле $\sigma = \varepsilon/300$, где ε – величина ошибки измерений в процентах, означает, что с вероятностью 0,997 погрешности измерений не превышают ε. Исследование влияния случайных ошибок измерения действительной и мнимой составляющих вынужденных колебаний на точность определения обобщенных динамических характеристик проводилось по следующему алгоритму:

– вычислялись точные значения составляющих вынужденных колебаний в точках «регистрации» для ряда частот исследуемого диапазона;

– вводились случайные отклонения заданного уровня в величины составляющих колебаний;

– производился подбор возбуждения [3] по искаженному ошибками отклику;

– определялись собственные частоты, обобщенные массы и обобщенные коэффициенты демпфирования;

– анализировалось изменение первых двух моментов распределения результатов расчета обобщенных характеристик с увеличением числа независимых «испытаний» на каждой частоте с целью получения достоверных оценок математического ожидания и дисперсии;

– по результатам расчетов строились гистограммы и проверялись гипотезы о законе распределения этих результатов.

Расчеты проводились при разном уровне ошибок измерений составляющих перемещений, числе сил возбуждения, демпфировании в системе, а также для различных способов настройки чувствительности измерительного оборудования.

Рассматривалось два вида зависимости абсолютной погрешности измерений действительной ΔU и мнимой ΔV составляющих перемещений конструкции от величин этих составляющих: величины ΔU и ΔV в каждой точке регистрации колебаний на любой частоте возбуждения пропорциональны измеряемым величинам составляющих перемещений:

$$\Delta U = \varepsilon U/100\ \%\ ,\ \Delta V = \varepsilon V/100\ \%; \qquad (2)$$

абсолютные погрешности измерений ΔU и ΔV в каждой точке регистрации колебаний на любой частоте возбуждения пропорциональны наибольшей из измеряемых величин:

$$\Delta U = \Delta V = \tilde{\varepsilon}\ - z\ /\ 100\ \%, \qquad (3)$$

где z есть наибольшая из $|U|$ и $|V|$ в данной точке, а «~» отличает погрешности вида (2) от погрешностей вида (3). Абсолютные погрешности вида (2) реализуются таким подбором чувствительности измерительного оборудования, что каждое из измеряемых значений составляющих перемещений соответствует пределу измерения. Для погрешностей вида (3) пределу измерений соответствует величина z.

Распределения случайных величин собственных частот, обобщенных масс и обобщенных коэффициентов демпфирования полагались также нормальными. Для оценки достоверности этого положения использовался критерий χ^2 [4]. Установлено, что с вероятностью 0,58…0,76 отклонения распределений обобщенных характеристик от нормального носят случайный характер (ограниченность выборки, погрешности расчетов при проведении оценки и т. п.). Здесь указаны минимальное и максимальное значения вероятностей, полученных во всех расчетах. Эти результаты подтверждают справедливость сделанного предположения о нормальности распределения обобщенных характеристик [4] и позволяют использовать соотношение (1) для нахождения абсолютной погрешности их определения.

Для выделения влияния только случайных ошибок в составляющих вынужденных колебаний и отсеивания ошибок, вызванных, например, несоответствием числа степеней свободы модели и исходной системы, относительная погрешность определения обобщенных характеристик вычислялась делением абсолютной погрешности на соответствующее математическое ожидание. Для контроля смещения оценки обобщенных характеристик найденное математическое ожидание сравнивалось с величиной собственной частоты, обобщенной массы или обобщенного коэффициента демпфирования, полученной в обычном расчете при точных исходных данных.

В расчетах относительной погрешности обобщенной массы ε_a, обобщенного коэффициента демпфирования ε_h и собственной частоты ε_p (рис. 1–5) символом δ обозначен логарифмический декремент колебаний, а N – число сил возбуждения.

Отметим, что оценки величин ε_a и ε_h проводились, в основном, вблизи собственной частоты тона, так как в этой области частот вынужденных колебаний минимальны погрешности определения обобщенных характеристик собственных тонов за счет несоответствия числа степеней свободы математической модели числу степеней свободы реальной системы, ограниченности числа сил возбуждения, взаимного влияния собственных тонов и т. д.

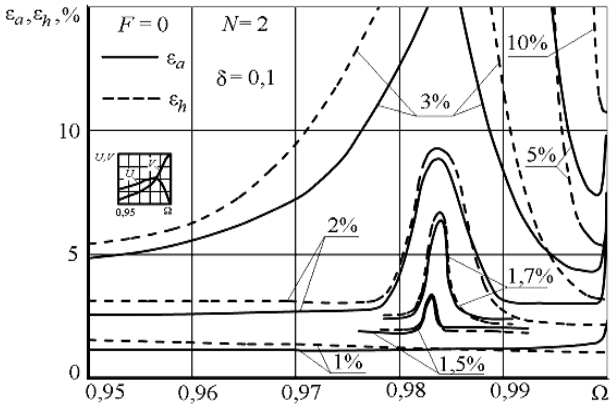


Рис. 1. Погрешности определения обобщенных характеристик при различных ошибках в перемещениях

Рассмотрим зависимости величин ε_a и ε_h от относительной частоты монофазного возбуждения ($F = 0$) для различного уровня ошибок измерений (рис. 1). Анализ результатов показывает, что ε_a и ε_h увеличиваются с приближением относительной частоты вынужденных колебаний к некоторому значению, которое обозначим за Ω^*, а затем погрешности ε_h падают при $\Omega \to 1$. Аналогичное падение ε_a сменяется резким ростом вблизи $\Omega = 1$.

Рост ε_a и ε_h в окрестности $\Omega = \Omega^*$ нелинейно зависит от погрешностей измерения перемещений ε (этого роста не происходит при $\varepsilon \le 1{,}2$ %). Если сопоставить кривые ε_a и ε_h с зависимостями составляющих перемещений в точках регистрации от частоты вынужденных колебаний (слева на рис. 1 показаны зависимости $U(\Omega)$ и $V(\Omega)$ в точке нормирования тона), то оказывается, что рост ε_a и ε_h сопровождает подъем действительной составляющей. При этом частота Ω^* соответствует максимуму абсолютного значения этой составляющей.

Для объяснения характера зависимостей $\varepsilon_a\ (\Omega)$ и $\varepsilon_h\ (\Omega)$ в окрестности $\Omega = \Omega^*$ получим выражения для ε_a и ε_h для системы с одной степенью свободы. Если не учитывать погрешности экспериментального определения собственной частоты тона, то

$$\varepsilon_a \approx [(1+\alpha)(1+\lambda^2)(1+\Delta E/E)/(1+\lambda^2(1+\alpha)^2)/ \\ /(1+\tilde{\Delta}\,V/V) - 1] \cdot 100\ \%\ , \qquad (4)$$

$$\varepsilon_h \approx [(1+\lambda^2)(1+\Delta E/E)/(1+\lambda^2(1+\alpha)^2)/ /(1+\tilde{\Delta} V/V)-1]\cdot 100\ \%, \quad (5)$$

где $\alpha = 2\varepsilon/(1+\varepsilon)$; ΔE – погрешности подбора возбуждения, вызванные неточностью измерений действительной и мнимой составляющих перемещений; $\tilde{\Delta} V$ – погрешности определения мнимой составляющей из-за неточности измерения и подбора возбуждения. Из формул (4) и (5) следует, что опережающий рост действительной составляющей колебаний по сравнению с мнимой при $\Omega \to \Omega^*$ и сопутствующее ему увеличение погрешности ΔU, приводящие к погрешностям в определении мнимой составляющей $\Delta - V > \Delta V$, увеличивают величины ε_a и ε_h, так как здесь $\Delta V/V > \varepsilon$.

Рост ε_a вблизи собственной частоты объясняется тем, что фактически обобщенная масса находится по разности величин $\beta = (\lambda/(1+\lambda^2))V^{\mathrm{T}}E$ на двух частотах колебаний (одна из них соответствует $\lambda = 0$), т. е. определяется разностью близких величин, каждая из которых измерена с некоторой погрешностью.

С увеличением числа сил возбуждения N происходит более резкий рост ε_a и ε_h при $\Omega \to \Omega^*$, но с приближением Ω к 1 величины ε_a и ε_h могут уменьшаться с ростом N (рис. 2).

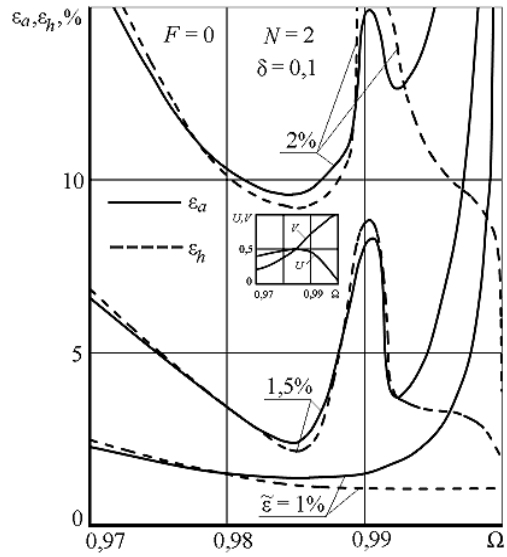


Рис. 2. Погрешности определения обобщенных характеристик при различных ошибках вида $\tilde{\varepsilon}$ в перемещениях

Определение обобщенной массы через разность величин β на двух частотах колебаний удобно использовать для объяснения результатов расчета ε_a при различном демпфировании. Так, сдвиг влево частоты колебаний, на которой начинается рост ε_a, с увеличением декремента является следствием того, что с ростом демпфирования резонансный пик становится менее острым, а зависимость $U(\Omega)$ – более пологой в окрестности собственной частоты. Поэтому разность β, сопоставимая с точностью этой величины, возникает при большем отличии частоты вынужденных колебаний от собственной.

С изменением способа настройки чувствительности оборудования с (2) на (3) меняется характер зависимостей $\varepsilon_a(\Omega)$ и $\varepsilon_h(\Omega)$ (см. рис. 2). Кривые ε_a и ε_h имеют локальный минимум на частоте колебаний, соответствующей равенству действительной и мнимой составляющих, т. е. равенству относительных ошибок измерения ε_u и ε_v этих составляющих.

Большие величины ε_a и ε_h при меньших частотах являются следствием роста ε_v с увеличением различия между U и V. Появление локального максимума с увеличением частоты колебаний объясняется так же, как и рост ε_a и ε_h в окрестности Ω^* (см. рис. 1), влиянием погрешностей в подборе возбуждения. При $\Omega \to 1$ величина ε_h падает с уменьшением λ и снижением влияния погрешностей определения этого параметра (в формуле для h стоит величина $(1+\lambda^2)$). В то же время ε_a увеличивается с ростом погрешностей в λ, так как в вычислениях обобщенной массы участвует величина $\lambda/(1+\lambda^2)$. То, что этот рост ε_a начинается при большей разности частот $|p-\omega|$ (ср. рис. 1 и 2), является следствием больших погрешностей $\varepsilon_u = \Delta V/U$.

Отметим, что при использовании монофазного возбуждения математические ожидания случайных величин обобщенной массы и обобщенного коэффициента демпфирования практически совпадают с соответствующими точными значениями этих обобщенных характеристик.

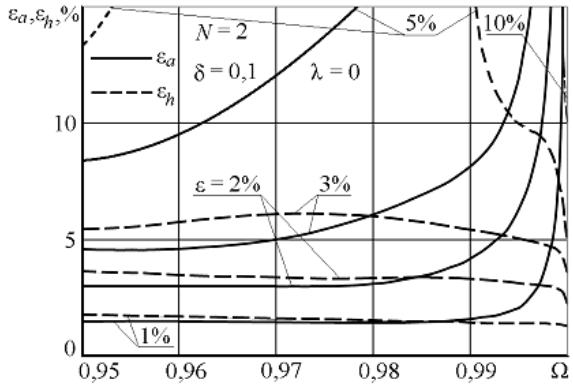


Рис. 3. Погрешности определения обобщенных характеристик при различных ошибках в перемещениях (немонофазное возбуждение)

Аналогичные расчеты (рис. 3) проводились при определении обобщенных характеристик с использованием немонофазного возбуждения, причем введением мнимой составляющей возбуждения создавался режим фиктивного фазового резонанса.

Основной особенностью использования немонофазного возбуждения является то, что величина параметра монофазных колебаний λ задается заранее.

Таким образом, заранее задается вектор действительной составляющей перемещений после подбора сил. Поэтому рост U при $\Omega \to \Omega^*$, увеличение числа сил возбуждения и изменение способа настройки измерительного оборудования не оказывает существенного влияния на характер кривых ε_a и ε_h.

Определяющее влияние на точность определения обобщенных характеристик с использованием немо-

нофазного возбуждения оказывают погрешности подбора сил. Здесь подбираются две составляющие возбуждения – действительная E и мнимая F, поэтому рост ε_a начинается при большем удалении частоты вынужденных колебаний от собственной, чем при использовании только монофазного возбуждения (см. рис. 1). Рост погрешности определения обобщенной массы с увеличением числа сил немонофазного возбуждения отмечается также в работе [2].

Важным результатом исследования влияния погрешностей измерения составляющих перемещений системы на точность определения обобщенной массы введением немонофазного возбуждения является то, что обнаружено различие между математическим ожиданием M_a случайных величин обобщенной массы и точным значением массы a (рис. 4). Кроме того, это различие увеличивается с приближением частоты фиктивного фазового резонанса к собственной частоте и ростом погрешности измерений.

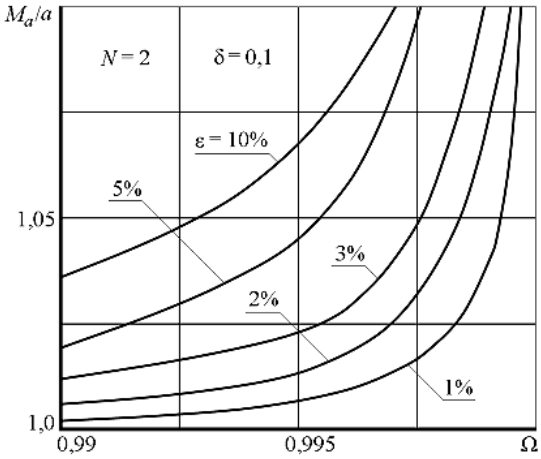


Рис. 4. Отношение математического ожидания к точному значению обобщенной массы

Рассмотрим результаты расчетов относительной погрешности определения собственной частоты p по переходу через нуль параметра монофазных колебаний λ с использованием линейной аппроксимации зависимости $\lambda(\omega)$ в окрестности $\lambda = 0$ (рис. 5). Анализ результатов показывает, что при $\delta = 0{,}1$ и числе сил возбуждения $N = 2$ погрешности определения собственной частоты не превышают 1 % при $\varepsilon \le 10$ %. Получено также, что погрешности определения собственной частоты растут с увеличением числа сил возбуждения и изменением способа настройки оборудования с (2) на (3).

В целом по результатам оценки влияния случайных ошибок измерения действительной и мнимой составляющей перемещений на точность определения обобщенных характеристик можно сделать следующие выводы:

– величина обобщенного коэффициента демпфирования определяется наиболее точно на собственной частоте тона и погрешность его определения соизмерима с точностью измерения составляющих вынужденных колебаний;

– чем ближе частоты собственных тонов, тем с меньшими погрешностями нужно измерять перемещения для сохранения неизменной точности определения обобщенной массы;

– использование немонофазного возбуждения для определения обобщенных масс тонов с близкими собственными частотами может привести к значительным погрешностям, так как рост ε_a при $\Omega \to 1$ начинается при значительном удалении частоты вынужденных колебаний от собственной частоты тона. При этом математическое ожидание случайных величин обобщенной массы является смещенной оценкой ее точного значения, и это смещение усиливается при $\Omega \to 1$;

– при определении обобщенной массы с использованием монофазного возбуждения большое значение имеет способ настройки чувствительности измерительного оборудования. Настройка чувствительности по наибольшей из составляющих колебаний приводит к недопустимым погрешностям в оценке обобщенной массы вблизи собственной частоты тона;

– наименьшая относительная погрешность расчета обобщенной массы при монофазном возбуждении примерно в 1,5 раза превосходит погрешности измерения составляющих вынужденных колебаний. Положение диапазона частот, где реализуется такая точность, по отношению к собственной частоте тона зависит от величины ошибок измерений, числа сил возбуждения, уровня демпфирования в системе и не может быть указано заранее;

– при выборе числа сил возбуждения необходимо учитывать уровень возможных ошибок измерений в процессе резонансных испытаний, так как рост погрешностей определения обобщенных характеристик с увеличением N может существенно превосходить повышение точности их расчета за счет лучшего выделения исследуемого тона. Если учесть, что современное оборудование практически не позволяет производить регистрацию перемещений (скоростей, ускорений) с точностью выше 2…3 %, то нецелесообразно одновременное использование более трех-четырех силовозбудителей;

– с целью повышения точности определения обобщенной массы и коэффициента демпфирования необходимо, чтобы в точке нормировки тона были наибольшие перемещения.

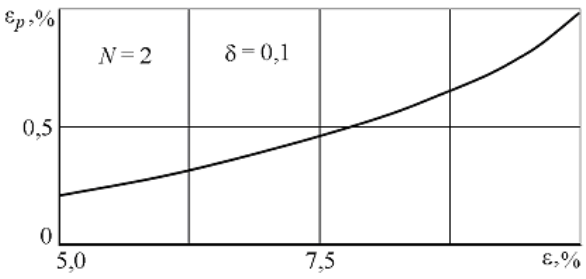


Рис. 5. Погрешности определения собственной частоты

Итак, как отмечено выше, наибольшие погрешности возникают при определении обобщенных масс.

Поэтому необходима разработка мероприятий, направленных на снижение этих погрешностей. К таким мероприятиям можно отнести, например, сглаживание экспериментальных зависимостей и расчет обобщенных масс без использования измерений действительной составляющей перемещений вблизи собственных частот конструкции.

Библиографические ссылки

1. Смыслов В. И. Некоторые вопросы методики многоточечного возбуждения при экспериментальном исследовании колебаний упругих конструкций // Учен. записки ЦАГИ им. Н. Е. Жуковского. 1972. Т. 3. № 5. С. 110–118.

2. Жаров Е. А., Смыслов В. И. Точность определения колебательных характеристик упругой конструкции при резонансных испытаниях с многоточечным возбуждением // Учен. записки ЦАГИ им. Н. Е. Жуковского. 1976. Т. 7. № 5. С. 88–97.

3. Бернс В. А. Особенности определения характеристик собственных тонов при многоканальном возбуждении и измерении колебаний // Аэродинамика и прочность конструкций летат. аппаратов: тр. Всерос. науч.-техн. конф., посвящ. 60-летию отделений аэродинам. летат. аппаратов и прочности авиац. конструкций (15–17 июня 2004 г.). Новосибирск, 2005. С. 295–297.

4. Вентцель Е. С. Теория вероятностей. М. : Наука, 1969.

V. A. Berns

ASSESSMENT OF DETERMINATION ACCURACY OF EIGENTONES CHARACTERISTICS IN THE PRESENCE OF RANDOM ERRORS IN THE EXPERIMENTAL DATA

Assessment of determination accuracy of natural frequencies, generalized masses and damping coefficients of eigentones of structures based on resonant tests were made. It was assumed that there are random errors of measurement in the values of real and imaginary components of the structure points' displacement. Studies were carried out by the method of statistical modeling.

Keywords: resonant tests, measurement errors, statistical modeling.

© Бернс В. А., 2010

УДК 004.932

М. В. Дамов, А. Г. Зотин

ПОСТРОЕНИЕ ИЗОБРАЖЕНИЯ СЦЕНЫ СОВМЕЩЕНИЕМ ПОСЛЕДОВАТЕЛЬНЫХ КАДРОВ

Представлена концепция построения изображения сцены при помощи совмещения кадров видеопоследовательности или серии фотоснимков. Построение сцены выполняется на основе карты смещений, отображающей векторы перемещений блоков.

Ключевые слова: реконструкция видеопоследовательности, совмещение изображений, построение сцены, карта смещений.

Уровень развития современной вычислительной техники позволяет решать задачи большой вычислительной сложности, к которым относится обработка видеопоследовательностей. Построение изображения сцены видеопоследовательности относится к классу задач, которые необходимы для выполнения реконструкции видеопоследовательности. Реконструкция видеопоследовательностей является актуальным направлением в связи с возрастающей потребностью потенциальных заказчиков в таких специалистах, как специалисты по работе с видеоархивами, специалисты в области производства и реставрации кино, специалисты по подготовке телепрограмм к эфиру, а также специалисты по анализу визуальных данных, полученных различными методами, а именно аэрофотосъемкой, спутниковой фотографией, лазерной локацией и другими системами датчиков. Но в общем случае под сценой видеопоследовательности понимается часть кинофильма или последовательность изображений, снятых с одного ракурса в течение некоторого времени. В работе [1] приведен один из алгоритмов разделения видеопоследовательности на сцены.

Картой смещений (*disparity map*) называется двумерный массив, элементами которого являются двумерные векторы. Каждый вектор задает сдвиг от точки на первом изображении до соответствующей точки на втором изображении. Эта информация используется для построения изображения сцены на основе нескольких соседних кадров. Таким образом, алгоритм построения относится к пространственно-временному типу. Существуют три основных подхода к определению параметров глобального движения: подход с использованием аппарата особых точек; подход с использованием векторов движения блоков; глобальный поиск. В данной статье рассматривается подход с использованием векторов движения блоков или окрест-

ностей. Преимуществом предлагаемого подхода является использование пирамиды детальности. Сначала карта смещений ищется для сильно уменьшенных копий изображений. Найденные значения являются начальными для карт смещений для более детальных копий и т. д. Таким образом, на каждом уровне детальности требуется лишь обновить карту смещений, что значительно сокращает время вычисления и вероятность нахождения ложных значений. Вместе с тем алгоритм предполагает наличие на сцене достаточно крупных объектов, т. е. кусочно-гладкую карту смещений. Приведем алгоритм поиска карты смещений.

Имеется пара изображений, для которой требуется построить карту смещений. Строятся две пирамиды детальности – для каждого изображения. Под пирамидой в области обработки изображений подразумевается представление исходного изображения в множестве изображений меньшего разрешения, дополнительно обработанных одним из фильтров сглаживания. Формирование пирамиды происходит путем сглаживания изображения на предыдущем уровне и выбором точек с шагом более одного пикселя при помощи билинейной интерполяции. Полуширина σ функции Гаусса связана с отношением k ($k > 1$) размеров изображений пирамиды на соседних уровнях:

$$k = \sigma\pi/2, \qquad (1)$$

где под полушириной функции Гаусса понимается расстояние между двумя крайними значениями независимой переменной, для которых значение функции равно половине ее максимального значения.

Такой выбор, с одной стороны, оставляет в сглаженном изображении только те частоты, которые будет содержать уменьшенное изображение, а с другой стороны, не приводит к потере деталей. Также происходит формирование пирамиды детальности для карты смещений. Процесс поиска карты смещений происходит постепенно, начиная с вершины пирамиды детальности. Обработка изображений вблизи вершины пирамиды и вблизи основания различна. Для изображений вблизи вершин пирамид (сильно уменьшенные изображения) ищется геометрическое преобразование (аффинное, проективное) изображения в первой пирамиде детальности, совмещающее его в целом с изображением на том же уровне второй пирамиды детальности. Поиск преобразования представляет собой модификацию алгоритма, приведенного в работе [2], и описан ниже. Найденное преобразование между изображениями на заданном уровне пирамиды детальности позволяет вычислить карту смещений на том же уровне. При переходе к следующему уровню детального разрешения, изображение в первой пирамиде преобразуется согласно карте смещений для текущего уровня, таким образом, на следующем уровне ищется только уточнение карты смещений.

Начиная с некоторого уровня, в пирамиде детальности более нельзя найти простое геометрическое преобразование, совмещающее изображения. Изображение разбивается на квадраты предопределенного размера (со стороной от 8 до 16 пикселей), и для каждого ищется геометрическое преобразование (обычно сдвиг) и карта смещений тем же способом, который применялся для всего изображения. Очевидно, что поиск преобразования уместен только в том случае, если один квадрат не захватывает объекты на разном расстоянии от камеры, в противном случае разные части квадрата должны испытывать разный сдвиг. С целью уменьшения скачков смещений на границах квадратов необходимо проводить усреднение смещений. Чем больше степень перекрытия квадратов, тем более сглаженной получается карта смещений, но вместе с тем возрастает время работы.

В качестве входных данных рассматривается пара изображений $I_1(\vec{x})$, $I_2(\vec{x})$ и карта смещений $\vec{d}(\vec{x},\vec{\omega})$. Вектор $\vec{\omega}$ есть вектор параметров, задающий модель карты смещений. Ниже рассматриваются две модели карты смещений: сдвиг и преобразование подобия.

Сдвиг:

$$\vec{\omega} = (t_x, t_y), \qquad (2)$$

$$\vec{d}(\vec{x},\vec{\omega}) = l\vec{t}, \qquad (3)$$

где $\vec{x}$ – некоторые известные координаты пикселя в изображении; t_x, t_y – смещение пикселя между изображениями по координатам.

Преобразование подобия:

$$\vec{\omega} = (t_x, t_y, \lambda, \varphi), \qquad (4)$$

$$\vec{d}(\vec{x},\vec{\omega}) = (1+\lambda)R_\varphi\vec{x} + l\vec{t} - \vec{x}, \qquad (5)$$

$$R_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \qquad (6)$$

где λ – коэффициент масштабирования изображения; φ – угол поворота изображения.

Параметр l вводится для обезразмеривания компонент $\vec{\omega}$ и приведения их значений к одному порядку величины так, что каждая компонента не превосходит нескольких десятых. Его численное значение полагается равным размеру диагонали изображения:

$$l = \sqrt{w^2 + h^2}, \qquad (7)$$

где w и h – ширина и высота изображения соответственно. В обоих случаях

$$\vec{d}(\vec{x},0) = 0. \qquad (8)$$

Поиск карты смещений сводится к поиску $\vec{\omega}$. Пусть $\varepsilon(\vec{\omega})$ обозначает интеграл от квадрата разности второго и преобразованного первого изображений:

$$\varepsilon(\vec{\omega}) = \int_\Omega \left(I_1\left(\vec{x} + \vec{d}(\vec{x},\vec{\omega})\right) - I_2\left(\vec{x}\right)\right)^2 d\vec{x},$$

$$\varepsilon(\vec{\omega}) = \int_\Omega \left(I_1\left(\vec{x} + \vec{d}(\vec{x},\vec{\omega})\right) - I_2\left(\vec{x}\right)\right)^2 d\vec{x}, \qquad (9)$$

где Ω – размеры изображения.

Ниже используется обозначение

$$\vec{y} = \vec{x} + \vec{d}(\vec{x},\vec{\omega}). \qquad (10)$$

Требуется найти $\vec{\omega}$, минимизирующий $\varepsilon(\vec{\omega})$:

$$\frac{\partial \varepsilon}{\partial \vec{\omega}} = 2\int_{\Omega}\left(I_1(\vec{y}) - I_2(\vec{x})\right)\frac{\partial I_1(\vec{y})}{\partial \vec{y}}\frac{\partial \vec{y}}{\partial \vec{\omega}}d\vec{x} = 0. \quad (11)$$

Ниже используются обозначения

$$\vec{g}^T \equiv \frac{\partial I_1(\vec{y})}{\partial \vec{y}}, \quad (12)$$

$$Y \equiv \frac{\partial \vec{y}}{\partial \vec{\omega}} = \frac{\partial \vec{d}(\vec{x},\vec{\omega})}{\partial \vec{\omega}}. \quad (13)$$

В предположении сглаженности $I_1(\vec{x})$, малого отличия $\vec{x}$ от $\vec{y}$ и с учетом (8) с точностью до членов первого порядка получается:

$$\vec{g}(\vec{y}) = \vec{g}(\vec{x}), \quad (14)$$

$$I_1(\vec{y}) = I_1(\vec{x}) + (\vec{g}, \vec{y} - \vec{x}), \quad (15)$$

$$\vec{y} - \vec{x} = Y\vec{\omega}. \quad (16)$$

Используя (15), (16), уравнение (11) переписывается в виде

$$\int_{\Omega}\left(I_1(\vec{x}) - I_2(\vec{x}) + \vec{g}^T Y\vec{\omega}\right)\vec{g}^T Y d\vec{x} = 0. \quad (17)$$

Из (17) получается линейная система относительно $\vec{\omega}$:

$$A\vec{\omega} = \vec{b}, \quad (18)$$

где

$$A = \int_{\Omega} Y^T \vec{g}\vec{g}^T Y^T d\vec{x}, \quad (19)$$

$$\vec{b} = \int_{\Omega}\left(I_2(\vec{x}) - I_1(\vec{x})\right)Y^T \vec{g}\, d\vec{x}. \quad (20)$$

Поиск решения системы (18) требует аккуратности, так как изображение (или его фрагмент) может не обладать достаточно выраженной текстурой для достоверного определения $\vec{\omega}$. Идея состоит в том, чтобы оценить погрешность матрицы **A** по априорному значению шума n на изображении и радиусу сглаживания σ исходных изображений. Далее система (18) решается с применением SVD-разложения и сингулярные числа меньшего порогового обнуляются [2]. Для оценки порогового сингулярного числа матрицы **A** рассмотрим случайное изображение с дисперсией интенсивности пикселей $<n^2>$ и нулевым средним значением: $<n> = 0$. В случае отсутствия границ в локальной области для $\sigma \to 1$ выполняется оценка для дисперсии производных рассматриваемого изображения [2]:

$$\mu^2 \equiv \frac{<|\vec{g}|^2>}{2} = \frac{<n^2>}{8\pi\sigma^4}. \quad (21)$$

Оценка (21) может быть легко получена путем перехода в спектральную область. Рассматриваемое изображение можно представить в виде спектра однородной и случайной функции, дисперсия которой может быть вычислена из теоремы о равенстве интеграла от квадрата модуля сигнала интегралу от квадрата его спектра. Дифференцирование изображения сводится к умножению его спектра на соответствующую производную функции Гаусса, что позволяет оценить μ^2. Зная модель искомого преобразования и оценку (21), можно оценить минимальное сингулярное число матрицы **A**:

$$S_{\min} = \mu^2 l^2 wh. \quad (22)$$

Сингулярные числа матрицы **A**, меньшие $S_{\min}$, должны быть обнулены при вычислении псевдообратной матрицы. В случае оценки минимального сингулярного значения при работе с фрагментом переменными w, h, l в выражении (22) обозначены размеры фрагмента.

Приведем конкретные модели карт смещений и алгоритм совмещения изображений, полученные на основе аналитических выкладок.

Сдвиг:

$$Y = \begin{pmatrix} l & 0 \\ 0 & l \end{pmatrix}. \quad (23)$$

Преобразование подобия:

$$Y = \begin{pmatrix} l & 0 & x & -y \\ 0 & l & y & -x \end{pmatrix}. \quad (24)$$

Алгоритм совмещения изображений:

Вход: первое изображение I_1, второе изображение I_2.

Выход: преобразование, переводящее первое изображение во второе изображение.

Шаг 1. Вычислить производные первого изображения путем свертки с функцией Гаусса. Полуширина функции Гаусса должна быть порядка ожидаемых значений в карте смещений.

Шаг 2. Составить матрицу **A** системы согласно выражению (19).

Шаг 3. Составить правую часть $\vec{b}$ в соответствии с выражением (20) и решить систему уравнений (18).

Шаг 4. Преобразовать второе изображение согласно обращенному преобразованию, построенному из $\vec{\omega}$. Это исключает необходимость пересчета **A** на каждой итерации.

Шаг 5. Если модуль вектора $\vec{\omega}$ не превышает заданной величины, то перейти к шагу 7.

Шаг 6. Перейти к шагу 3.

Шаг 7. Конец алгоритма.

Как можно видеть из тестов (см. рисунок), для кусочно-гладких поверхностей алгоритм работает удовлетворительно. Неверные значения получаются вблизи границ объектов. Время работы алгоритма пропорционально числу пикселей на изображениях.

Рассмотренный алгоритм имеет два слабых места: во-первых, на границах объектов находятся неправильные значения смещений, во-вторых, предлагаемый подход в принципе неприменим в случае наличия мелких, расположенных близко к камере объектов. Это связано с тем, что на «низких» разрешениях объект пропадает, а при высоком разрешении сдвиг объекта получается слишком большой, чтобы предлагаемый алгоритм мог его зафиксировать. Однако анализ разностного кадра может выявить подобные проблемные объекты, и для них возможно применение других методов в качестве пост-обработки.

Изображение сцены тестовой видеопоследовательности после работы алгоритма (*слева*);
один из кадров тестовой видеопоследовательности до работы алгоритма (*справа*)

Следует отметить также, что в результате работы алгоритма получены взаимно-однозначные соответствия между парами пикселей первого и второго изображения, верные для значительного количества точек изображения (исключая области затенения, occlusions). Это создает возможность использовать устойчивые методы для нахождения фундаментальной матрицы в эпиполярной геометрии. Таким образом, возможен второй проход алгоритма, выполняемый при наложенных эпиполярных ограничениях.

Необходимо отметить, что в качестве исходных данных для построения карты смещений также можно использовать данные, полученные с помощью аппарата особых точек или с использованием аппарата глобального поиска. Недостаток алгоритмов, использующих как основу особые точки, – низкая производительность, что обусловлено высокой вычислительной сложностью операций по поиску (*Feature Selection*) и сопровождению (*Feature Tracking*) особых точек [3]. Преимущество подобных алгоритмов заключается в высокой надежности определения локального движения. Алгоритмы глобального поиска определяют оптимальное преобразование без использования информации о локальном движении [4]. Вычислительная сложность этих алгоритмов чрезвычайно высока, что практически исключает возможность их применения для решения реальных задач.

Таким образом, на этом этапе имеются два частично перекрывающихся изображения I_1 и I_2; карта смещений, представляющая собой массив векторов смещений (Δx, Δy); а также преобразование, переводящее первое изображение во второе (преобразование подобия или сдвига).

Массив векторов, найденный алгоритмом определения движения, имеет некоторые характерные свойства, а именно: вектор движения с большой вероятностью найден неверно, если: 1) ошибка определения вектора движения велика (поскольку это означает, что сопоставленные блоки мало похожи); 2) вектор движения значительно отличается от соседних векторов (следует из гладкости поля векторов движения); дисперсия окрестности мала (как правило, в этом случае блок мало отличается от соседних окрестностей). Исходя из этого, следует ввести оценку качества или доверия, учитывающую особенности поля векторов.

В работе [5] для определения надежности векторов движения предлагается ввести функцию доверия следующего вида:

$$fb_{x,y} = \left(a \cdot e_{x,y} + \frac{b}{{d_{x,y}}^2} + c \cdot \sigma_{x,y} \right)^{-1}, \qquad (25)$$

где a, b, c – параметры; $e_{x,y}$ – ошибка определения вектора движения (сумма абсолютных разностей яркостей пикселей окрестностей вектора в позиции (x, y) из текущего кадра и сопоставленной окрестности из предыдущего кадра; $d_{x,y}$ – дисперсия окрестности в позиции (x, y) из текущего кадра; $\sigma_{x,y}$ – среднеквадратичное отклонение вектора движения в позиции (x, y) от соседних векторов (26):

$$\sigma_{x,y} = \frac{1}{4} \cdot \sum_{i=-1}^{1} \sum_{j=-1}^{1} \left(\left(p^X_{x,y} - p^X_{x+i,y+j} \right)^2 + \left(p^Y_{x,y} - p^Y_{x+i,y+j} \right)^2 \right), \qquad (26)$$

где $p^X_{x,y}$, $p^Y_{x,y}$ – проекция вектора движения в позиции (x, y) на ось абсцисс и ординат соответственно. В работе [5] приводятся следующие значения параметров функции доверия: $a = 0{,}25$; $b = 32$; $c = 1$, полученные экспериментальным путем. Вектора движения с малым значением функции доверия отбрасываются, остальные вектора составляют множество надежных векторов.

При формировании изображении сцены I_s следует принимать во внимание особенности карты смещения. Карта смещения отображает только вектора для сходных (подобных) точек, находящихся на перекрывающихся изображениях I_1 и I_2. В связи с этим для точек, вновь появившихся или исчезнувших объектов, вектора смещений будут отсутствовать. На карте смещения отображение таких точек будет иметь вид вектора нулевого размера.

Для генерации сцены на основе карты смещения возможно использование двух подходов. Суть первого подхода заключается в том, что осуществляется достройка начального кадра, в ходе которой происходит расширение границ и заполнение новых блоков информацией по данным с вновь считанного кадра и претерпевшим определенные преобразования. Второй подход заключается в том, что для каждого нового кадра изменение размера происходит путем добавления так называемого шлейфа, полученного в резуль-

тате трансформации граничной информации с предыдущих кадров, и сопоставлением.

Поскольку при построении полного изображения сцены для нас наибольший интерес представляет область на границах кадра, то для повышения быстродействия целесообразно использовать граничные данные для анализируемого кадра. Принимая во внимание возможность неравномерного движения на границах кадра, нужно проводить расчет параметров смещения для анализируемых регионов, однако внутри самого региона может также существовать неравномерность. Исходя из вышесказанного, в качестве базиса для формирования изображения сцены решено применять информацию о векторах локального движения для сегментов и данные о преобразованиях подобия и сдвига.

Общий алгоритм для формирования изображения сцены на основе карты смещения будет состоять из следующих шагов.

Шаг 1. Формирование предварительных граничных сегментов. На данном шаге в зависимости от размера анализируемой области формируются граничные сегменты с шириной и высотой соответственно $W_\omega \times W_h$, вычисляемыми из размерности кадра путем деления на 8, и происходит генерация карты сегментов M_{segm}, учитывающих их пространственное положение в соответствии с начальным делением.

Шаг 2. Расчет векторов локального движения из кадра I_1 в кадр I_2. В ходе этого шага происходит расчет среднего, максимального и минимального векторов движения для каждого сегмента. Для расчета среднего вектора движения можно выбрать следующие пути получения значения VSs_i: как среднего арифметического или медианного значения из соответствующих элементов массива смещений.

Шаг 3. Анализ векторов локального движения и формирование векторов смещения. На этом шаге для каждого сегмента анализируются локальные вектора движения из кадра I_1 в кадр I_2 и вычисляются расстояния от сопоставимого сегмента для кадра I_2 с границами кадра I_2.

Шаг 4. Анализ полученных векторов смещения для всех сегментов и формирование карты расширения M_{scale}. Данная карта формируется как прямоугольная область, учитывающая максимально возможное смещение для каждого сегмента. По завершении генерации карты расширения производится разметка неиспользуемых областей, т. е. тех областей, в которых движение однозначно не определено.

Шаг 5. Заполнение изображения сцены на основе карты расширения и векторов смещения с учетом возможного преобразования. Для каждого вектора смещения (Δx, Δy) окрестности (x, y), не равного нулю, перенести эту окрестность в изображение сцены. Для каждого вектора смещения (Δx, Δy) окрестности (x, y), равного нулю, проанализировать соседние векторы. Используя данные соседних векторов, вычислить местоположение такой окрестности в изображении сцены. Затем поместить окрестность в изображение сцены, используя вычисленное местоположение.

После работы алгоритма формирования сцены имеется изображение сцены некоторой высоты и ширины в виде массива пикселей. Достоверность этого изображения, построенного на основе нескольких соседних кадров одной сцены видеопоследовательности, можно оценить визуально с помощью экспертной оценки. На рисунке приведен результат работы алгоритма после обработки сцены тестовой видеопоследовательности, а также найденное поле векторов движения.

Таким образом, результатом исследовательской работы является алгоритм формирования изображения сцены на основе соседних кадров сцены видеопоследовательности. Предлагаемый алгоритм применим для решения многих актуальных задач, например, для построения панорамного изображения, стабилизации и сжатия видеопоследовательности, а также реконструкции видеопоследовательности в системах видеоредактирования и электронных видеоархивах.

Библиографические ссылки

1. Дамов М. В. Восстановление фона в областях кадра с объектами малого размера в видеопоследовательности // Вестник СибГАУ. 2010. Вып. 1(27). С. 52–56.

2. Моулер К., Форсайт Дж., Малькольм М. Машинные методы математических вычислений : пер. с англ. М. : Мир, 1980.

3. Making good features to track better / T. Tommasini, A. Fusiello, E. Trucco, V. Roberto // Proc. of IEEE Computer Society Conf. on Computer Vision Pattern Recognition. 1998. P. 145–149.

4. Litvin A., Konrad J., Karl W. C. Probabilistic Video Stabilization Using Kalman Filtering and Mosaicking // Proc. of SPIE Conf. on Electronic Imaging, 2003. P. 663–674.

5. Солдатов С. А., Стрельников К. Н., Ватолин Д. С. Быстрое и надежное определение глобального движения в видеопоследовательностях // Материалы 16-й Междунар. конф. по компьютер. графике и ее приложениям. 2006. С. 430–437.

M. V. Damov, A. G. Zotin

SCENE IMAGE CONSTRUCTION BY WAY OF IMAGE SEQUNCE SUPERPOSITION

In paper it is presented concept of scene image construction by way of video sequence frames or photo shot series superposition. Frame superposition based on construction of disparity map imaged vectors of shift blocks

Keywords: video sequence reconstruction, image superposition, scene construction, disparity map.

© Дамов М. В., Зотин А. Г., 2010

УДК 621.391.82

А. В. Кузовников, В. Г. Сомов, В. И. Лавров, А. Л. Дерябин, В. А. Анжина

СПОСОБ ФОРМИРОВАНИЯ ПОМЕХОЗАЩИЩЕННЫХ СИГНАЛОВ

Рассмотрен способ формирования широкополосных сигналов (ШПС) с непосредственным расширением спектра, при котором псевдослучайная последовательность (ПСП) модулируется не гармонической несущей, а биортогональной вейвлет-функцией (W). Показано, что данный способ модуляции ПСП приводит к значительному расширению ширины полосы спектра полученного сигнала. Проведен сравнительный анализ относительной помехозащищенности сигналов, модулированных биортогональной вейвлет-функцией, и ФМ-2 (BPSK).

Ключевые слова: псевдослучайная последовательность, широкополосный сигнал, методы формирования и модуляции сигнала, помехозащищенность.

Интенсивное развитие систем связи различного назначения в условиях сильной загруженности частотного диапазона приводит к необходимости разработки новых типов сигналов. По сообщениям [1; 2], участились случаи несанкционированного использования ресурсов систем связи. Для борьбы с этими негативными факторами постоянно ведется доработка имеющихся и формирование новых типов сигналов. Согласно требованиям, предъявляемым к новым сигналам, они должны обладать следующими качествами:

– способностью противостоять организованной помехе (т. е. обеспечивать связь в условиях сложной радиоэлектронной обстановки);

– широкой полосой сигнала, обеспечивающей сложность обнаружения («скрытность») сигнала;

– электромагнитной совместимостью с другими системами связи за счет формирования сигнала, работающего «ниже» уровня шума [3].

Под помехозащищенностью понимают способность системы противостоять воздействию мощных помех. Помехозащищенность включает в себя скрытность системы связи и ее помехоустойчивость [4].

В настоящее время наиболее распространены следующие способы формирования помехоустойчивых сигналов.

1. Формирование сложных сигналов с помощью частотно-временного кодирования [3].

2. Псевдослучайное формирование сложных дискретных частотно-манипулированных сигналов [3; 4].

3. Псевдослучайная перестройка рабочей частоты [5].

4. Модуляция несущей частоты бинарными псевдослучайными последовательностями (ПСП) [6].

Анализ известных источников литературы [3–6] показал, что наиболее перспективными являются методы, основанные на модуляции несущей частоты бинарными псевдослучайными последовательностями (ПСП) [6]. Псевдослучайная бинарная последовательность заданной длительности – это последовательность, сформированная по определенным правилам из дискретных элементов 0 и 1 так, чтобы ее корреляционные свойства были близки к соответствующим свойствам шумовой реализации той же длительности. Такие последовательности обладают следующими свойствами:

– М-последовательность является периодической, с периодом, состоящим из N символов;

– боковые пики периодической автокорреляционной функции сигналов, образованных М-последовательностью, равны $(1/N)$.

Сформированная данным способом последовательность используется для формирования высокочастотного широкополосного сигнала. Если полученную последовательность сложных информационных символов подать на модулятор и выполнить манипуляцию одного из параметров несущего колебания, то в результате образуется сложный сигнал, свойства и структура которого полностью определяются свойствами поданной последовательности и видом выбранной модуляции.

В статье проведен анализ использования в качестве модулирующей функции негармонической биортогональной вейвлет-функции.

В общем случае биортогональные (рис. 1) вейвлет-функции [7] используют два дуальных вейвлет-базиса $\psi_{m,k}(t)$ и $\psi^*_{m,k}(t)$, которые удовлетворяют требованию биортогональности скалярного произведения этих вейвлетов:

$$\left(\psi^*_{m,k}(t), \psi^*_{m,k}(t)\right) = \delta_{m,k;m',k'}. \qquad (1)$$

Для оценки помехозащищенности сигналов, модулированных биортогональными вейвлет-функциями, было проведено моделирование [8], которое включает сравнительный анализ помехозащищенности сигналов, модулированных биортогональными вейвлет-функциями, и широкополосных сигналов ФМ-2.

В качестве расширяющей последовательности использовалась ПСП Голда длиной 31. Скорость передачи (V) информационного символа при этом составляла 1 кбит/с .

Согласно [9], полиномы для псевдослучайной последовательности Голда длиной 31, имеют следующий вид (рис. 2):

– для верхнего плеча схемы: $g_1(p) = p^5 + p^2 + 1$;

– для нижнего плеча схемы: $g_2(p) = p^5 + p^4 + p^2 + p + 1$.

Сформированы следующие сигналы: полученный при модуляции данной ПСП сигнал ФМ-2 (BPSK) (рис. 3); сигнал с той же длиной ПСП, но модулированный биортогональной вейвлет-функцией (рис. 4).

Анализ полученных спектров (рис. 5, 6) показал увеличение ширины полосы сигнала, модулированного биортогональной вейвлет-функцией, в 10 раз, по сравнению с шириной полосы сигнала, модулированного ФМ-2 (BPSK).

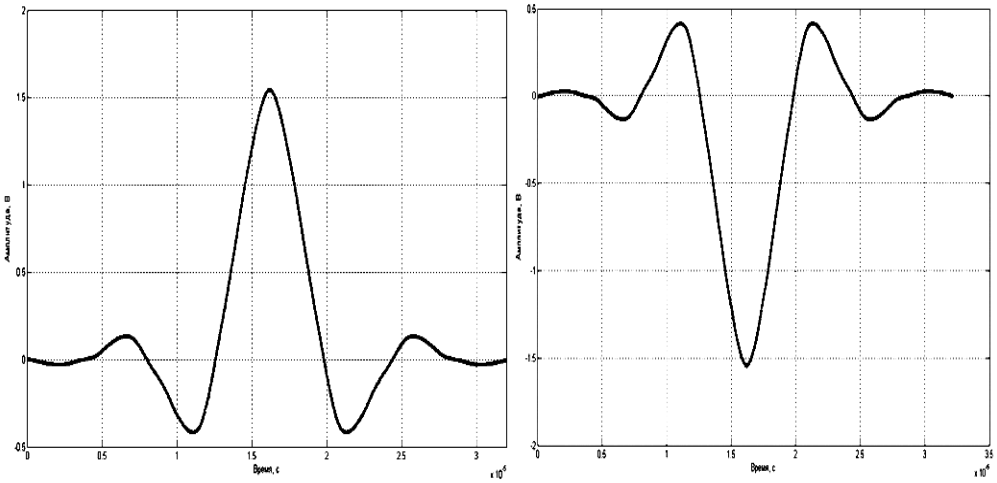


Рис. 1. Общий вид биортогонального вейвлета

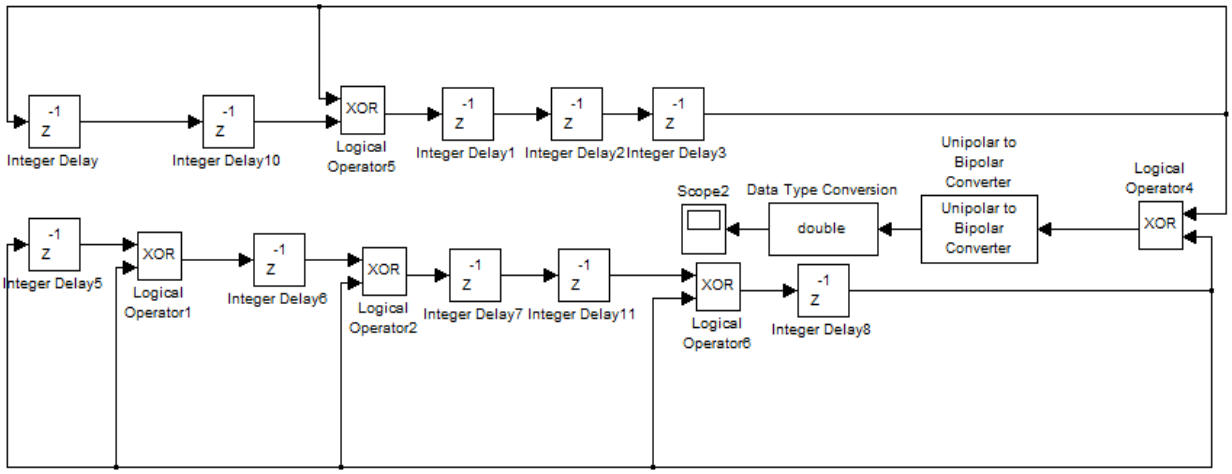


Рис. 2. Схема реализации последовательности Голда в Simulink MatLab

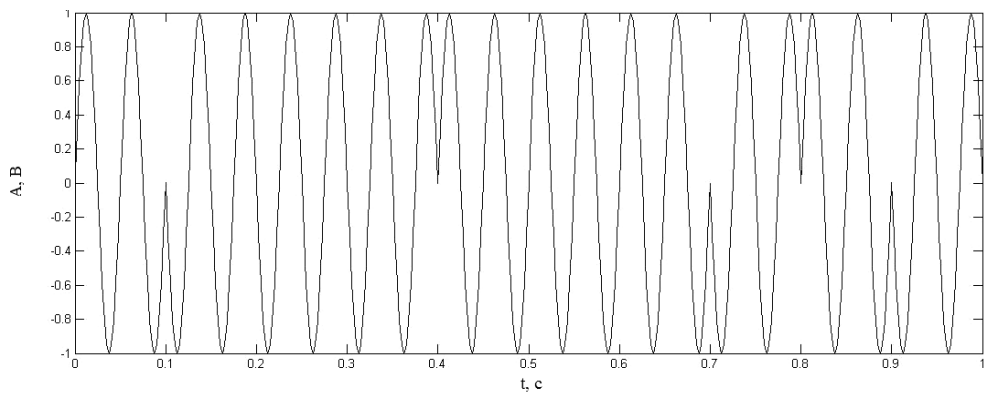


Рис. 3. Сигнал, модулированный BPSK

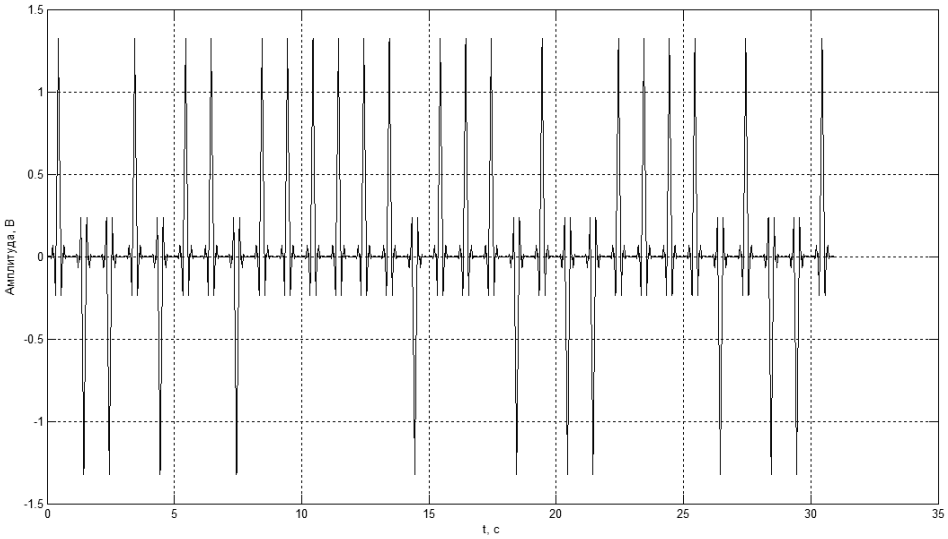


Рис. 4. Сигнал, модулированный биортогональной вейвлет-функцией

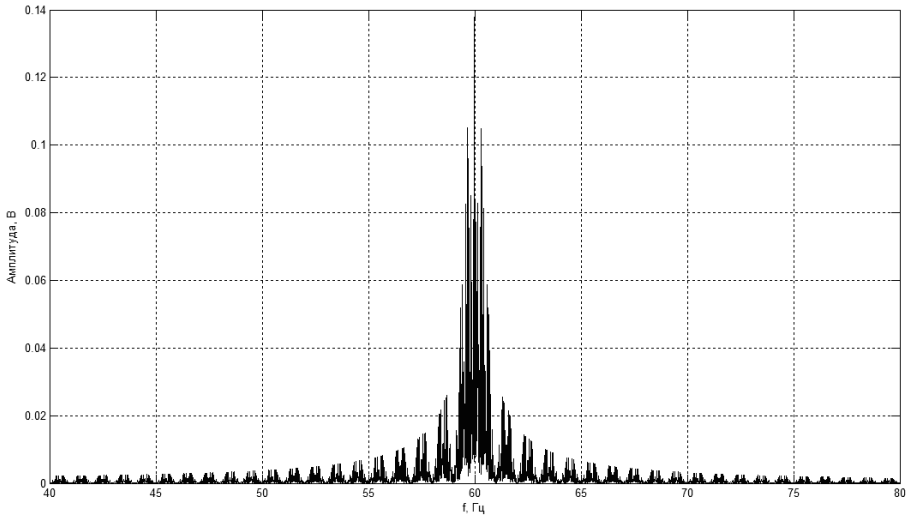


Рис. 5. Спектр сигнала, модулированного BPSK

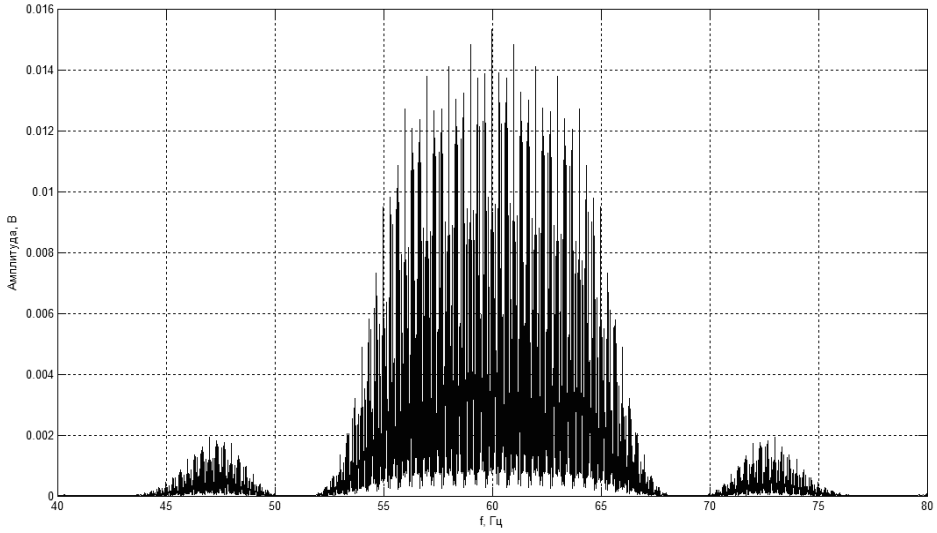


Рис. 6. Спектр сигнала, модулированного биортогональной вейвлет-функцией

При этом увеличение помехозащищенности, согласно [5; 6], пропорционально увеличению ширины полосы сигнала. Для количественной оценки величины относительной помехозащищенности рассмотрена система связи со следующими параметрами:

– общая ширина полосы системы связи $\Delta f = 70$ МГц ;

– тактовая частота информационного сигнала $f_{\text{ТИ}} = 9,6$ кГц; ;

– скорость передачи символов $V_{\text{сим}} = 9,6$ кбит/с .

Расчет относительной помехозащищенности при использовании биортогональной вейвлет-функции, проведен согласно [6]:

$$П_{0(\text{W})\text{ШПС}} = \frac{M \cdot f_{\text{ТШ}}}{\eta_{\text{С}} f_{\text{ТИ}} h_0^2}, \qquad (2)$$

где $f_{\text{ТШ}}$ – тактовая частота ПСП; $f_{\text{ТИ}}$ – тактовая частота информационного сигнала; M – величина, показывающая, во сколько раз ширина спектра сигнала, модулированного вейвлет-функцией, больше сигнала, модулированного BPSK.

Максимально возможное число каналов в рассматриваемой системе связи рассчитывалось по формуле

$$K_{(\text{W})\text{ШПС}} = \frac{\Delta F}{M \cdot f_{\text{ТИ}} \cdot N_{\text{ПСП}}}. \qquad (3)$$

Значения относительной помехозащищенности Π_0, тактовой частоты ПСП $f_{\text{ТШ}}$, числа каналов K для сигналов, модулированных ФМ-2 (BPSK) и биортогональной вейвлет-функцией при различной длине кодирующей ПСП $N_{\text{ПСП}}$, приведены в таблице.

Сравнение графиков относительной помехозащищенности сигналов, полученных обоими способами, от тактовой частоты ПСП (рис. 7) и числа каналов (рис. 8), показало, что при одинаковой длине ПСП сигналы, модулированные биортогональной вейвлет-функцией, обладают большей относительной помехозащищенностью по сравнению с сигналами, модулированными ФМ-2 (BPSK). Выигрыш относительной помехозащищенности сигналов, модулированных биортогональной вейвлет-функцией, составляет порядка 10 дБ. Однако число каналов связи, которые можно реализовать в данной системе при использовании такого широкополосного сигнала, сокращается в 10 раз.

Итак, результаты проведенных исследований показали, что рассматриваемый способ формирования широкополосных сигналов позволяет повысить относительную помехозащищенность. Результаты математического моделирования показали, что выигрыш помехозащищенности обусловлен увеличением ширины полосы сигнала, модулированного биортогональной вейвлет-функцией. Предложенный способ формирования широкополосных сигналов является наиболее эффективным для систем связи или передачи информации, когда требуется высокая скрытность и помехозащищенность.

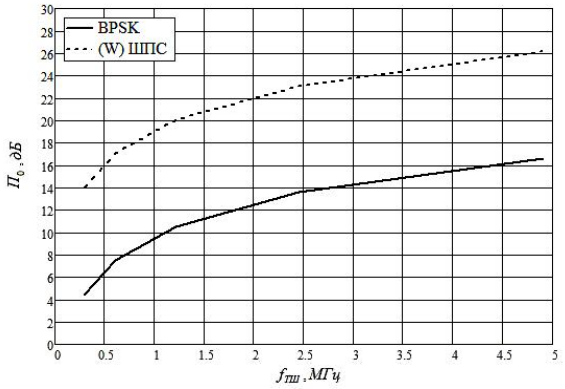


Рис. 7. График зависимости относительной помехозащищенности от тактовой частоты ПСП для разных типов сигналов

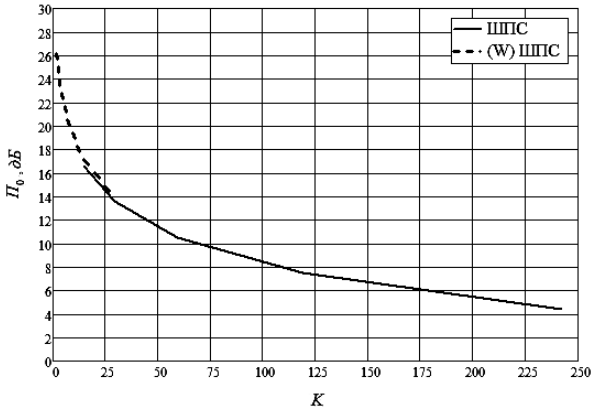


Рис. 8. График зависимости относительной помехозащищенности от числа каналов для разных типов сигналов

Библиографические ссылки

1. Колюбакин В. Конференция в Дубне // ТЕЛЕ-Спутник. Май. 1999. С. 40.

2. Панько. С. П., Сухотин В. В. Несанкционированный доступ в системы спутниковых коммуникаций // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. 2002. № 4. С. 15–27.

3. Семенов И. И. Широкополосные системы радиосвязи : учеб. пособие. Омск, 2000.

4. Варакин Л. Е. Системы связи с шумоподобными сигналами. М. : Радио и связь, 1985.

5. Помехозащищенность систем радиосвязи с расширением спектра сигналов методом псевдослучайной перестройки рабочей частоты / В. И. Борисов [и др.]. М. : Радио и связь, 2000.

6. Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью / В. И. Борисов [и др.]. М. : Радио и связь, 2003.

7. Смоленцев Н. К. Основы теории вейвлетов. Вейвлеты в MATLAB. М. : ДМК Пресс, 2008.

8. Дьяконов В. П. MATLAB 6.5 SP1/7 + Simulink 5/6. Обработка сигналов и проектирование фильтров. М. : Солон-Пресс, 2005.

9. Прокис Дж. Цифровая связь: пер. с англ. / под ред. Д. Д. Кловского. М. : Радио и связь. 2000.

Сравнение относительной помехозащищенности разных типов сигналов

$N_{ПСП}$	31	63	127	255	511
$f_{ТШ}$, МГц	0,298	0,605	1,219	2,448	4,906
K_{BPSK}	241	119	59	29	14
$K_{(W)ШПС}$	26	13	6	3	1
$П_{0BPSK}$, дБ	4,41	7,49	10,54	13,57	16,58
$П_{0(W)ШПС}$, дБ	13,96	17,04	20,08	23,11	26,13

A. V. Kuzovnikov, V. G. Somov, V. I. Lavrov, A. L. Deryabin, V. A. Anzhina

METHOD OF FORMING OF NOISE-STOP SIGNALS

The article considers way of generation of broad-band signals (BBS) with direct expansion of a spectrum at which the pseudo-random sequence (PRS) is modulated not with harmonious bearing but with biorthogonal wavelet-function. It is shown, that this way of modulation PNS leads to considerable dilation of a bandwidth of a spectrum of the received signal. Comparative analysis of relative jam-protection of PSK BBS and wavelet (W) BBS signals is carried out.

Keywords: pseudorandom sequence, broad-band signal, methods of forming and modulation of signals, jam-protection.

© Кузовников А. В., Сомов В. Г., Лавров В. И., Дерябин А. Л., Анжина В. А., 2010

УДК 539

А. В. Лопатин, Р. А. Удальцов

СИММЕТРИЧНЫЕ КОЛЕБАНИЯ ТРЕХСЛОЙНОЙ ПЛАСТИНЫ*

Решена задача определения частоты симметричных колебаний трехслойной пластины с одинаковыми композитными несущими слоями и ортотропным заполнителем. Основное дифференциальное уравнение четвертого порядка получено с помощью принципа Гамильтона. Приведена формула для частоты симметричных колебаний трехслойной пластины с шарнирно-закрепленными несущими слоями.

Ключевые слова: трехслойная пластина, симметричные колебания.

Колебания трехслойных пластин отличаются большим разнообразием форм движения несущих слоев и заполнителя. Наиболее изученными из них являются поперечные изгибные колебания, при которых оба несущих слоя и заполнитель движутся в одну сторону. Вместе с тем трехслойные пластины могут совершать колебания, формы которых отличаются от форм поперечных колебаний. К ним относятся симметричные колебания трехслойных пластин с одинаковыми несущими слоями. При таких колебаниях срединная плоскость пластины не движется, а несущие слои и части заполнителя, лежащие по разные стороны срединной плоскости, движутся в противоположных направлениях. Очевидно, что для моделирования симметричных колебаний трехслойных пластин необходимо учитывать податливость заполнителя. Отметим, что число исследований, в которых рассматривается влияние податливости заполнителя на динамическое поведение трехслойных пластин, невелико. Одним из первых было исследование, выполненное Фростингом и Томсоном [1]. Анализ решенных к настоящему времени вибрационных задач, в которых учитывается влияние податливости заполнителя на динамическое поведение трехслойных пластин, позволяет сделать вывод, что эти исследования далеки от своего завершения и могут быть дополнены новыми результатами.

Уравнения движения. Рассмотрим трехслойную пластину, состоящую из двух одинаковых композитных слоев и ортотропного заполнителя. Введем декартову систему координат x, y, z, связанную со срединной плоскостью трехслойной пластины. Пусть a, b – размеры пластины по осям x и y, а t, h – толщины несущего слоя и заполнителя.

*Работа выполнена при поддержке федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг.

Для получения дифференциального уравнения симметричных колебаний трехслойной пластины воспользуемся вариационным принципом Гамильтона [2]. Отметим два положения движущейся пластины – в момент времени τ_1 и в момент времени τ_2. При этом $\tau_2 > \tau_1$. Рассмотрим далее интеграл действия Гамильтона

$$S = \int_{\tau_1}^{\tau_2} L\,dt\,, \tag{1}$$

где L – функция Лагранжа, определяемая следующим образом:

$$L = T - U, \tag{2}$$

где T – кинетическая и U – потенциальная энергии трехслойной пластины, совершающей симметричные колебания.

Кинетическая энергия пластины определяется формулой

$$T = T_t + T_h, \tag{3}$$

где T_t – кинетическая энергия несущего слоя; T_h – кинетическая энергия заполнителя.

Для потенциальной энергии трехслойной пластины будем иметь

$$U = U_t + U_h, \tag{4}$$

где U_t – потенциальная энергия изгиба несущего слоя; U_h – потенциальная энергия деформации заполнителя.

Определим значения кинетической T_t и потенциальной U_t энергий несущего слоя, который совершает изгибные колебания. Выражения для T_t и U_t будут иметь следующий вид:

$$T_t = \frac{1}{2} B_{\rho t} \int_0^a \int_0^b \left(\partial w / \partial \tau\right)^2 dxdy; \tag{5}$$

$$U_t = \frac{1}{2} \int_0^a \int_0^b \left[D_{11} \left(\frac{\partial^2 w}{\partial x^2} \right)^2 + 2D_{12} \frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2} + D_{22} \left(\frac{\partial^2 w}{\partial y^2} \right)^2 + 4D_{33} \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 \right] dxdy, \tag{6}$$

где τ – время; $w = w(x, y, \tau)$ – прогиб несущего слоя; D_{11}, D_{22}, D_{12}, D_{33} – изгибные жесткости ортотропного несущего слоя; $B_{\rho t}$ – инерциальный параметр несущего слоя [3].

Кинетическая и потенциальная энергии заполнителя как трехмерного ортотропного тела определяются следующими выражениями:

$$T_h = \frac{1}{2} \int_0^a \int_0^b \int_0^{h/2} \rho_h \left[\left(\frac{\partial u_x}{\partial \tau} \right)^2 + \left(\frac{\partial u_y}{\partial \tau} \right)^2 + \left(\frac{\partial u_z}{\partial \tau} \right)^2 \right] dxdydz; \tag{7}$$

$$U_h = \frac{1}{2} \int_0^a \int_0^b \left(\sigma_x e_x + \sigma_y e_y + \sigma_z e_z + \tau_{xy} e_{xy} + \tau_{xz} e_{xz} + \tau_{yz} e_{yz} \right) dxdydz, \tag{8}$$

где ρ_h – плотность материала заполнителя; u_x, u_y, u_z – перемещения вдоль осей x, y, z; σ_x, σ_y, σ_z – нормальные напряжения; τ_{xy}, τ_{xz}, τ_{yz}, – касательные напряжения; e_x, e_y, e_z – продольные деформации; e_{xy}, e_{xz}, e_{yz} – деформации сдвига.

Компоненты напряжений и деформаций в ортотропном заполнителе связаны между собой законом Гука:

$$e_x = \sigma_x / E_x - \nu_{xy}\sigma_y / E_y - \nu_{xz}\sigma_z / E_z,$$
$$e_y = \sigma_y / E_y - \nu_{yx}\sigma_x / E_x - \nu_{yz}\sigma_z / E_z, \tag{9}$$

$$e_z = \sigma_z / E_z - \nu_{zx}\sigma_x / E_x - \nu_{zy}\sigma_y / E_{yz}, \tag{10}$$

$$e_{xy} = \tau_{xy} / G_{xy}, \tag{11}$$

$$e_{xz} = \tau_{xz} / G_{xz}, \quad e_{yz} = \tau_{yz} / G_{yz}, \tag{12}$$

где E_x, E_y, E_z – модули упругости материала; G_{xy}, G_{xz}, G_{yz} – модули сдвига; ν_{xy}, ν_{xz}, ν_{yx}, ν_{yz}, ν_{zx}, ν_{zy} – коэффициенты Пуассона.

Деформации и перемещения в заполнителе связаны между собой геометрическими соотношениями

$$e_x = \partial u_x / \partial x, \quad e_y = \partial u_y / \partial y, \quad e_z = \partial u_z / \partial z,$$
$$e_{xy} = \partial u_x / \partial y + \partial u_y / \partial x, \quad e_{xz} = \partial u_x / \partial z + \partial u_z / \partial x, \tag{13}$$
$$e_{yx} = \partial u_y / \partial z + \partial u_z / \partial y.$$

Напряжения, деформации и перемещения, возникающие в заполнителе, являются в общем случае функциями трех координат и времени. Это обстоятельство существенно затрудняет получение как аналитического, так и численного решений рассматриваемой динамической задачи. Однако симметричные колебания обладают особенностью, которая позволит значительно упростить исходные уравнения для заполнителя. Дело в том, что при таких колебаниях перемещение u_z значительно больше перемещений u_x и u_y, появление которых обусловлено только эффектом Пуассона. Поэтому в рассматриваемой задаче можно принять, что в заполнителе тангенциальные перемещения отсутствуют, т. е.

$$u_x = 0, \quad u_y = 0. \tag{14}$$

Отметим, что при симметричных колебаниях допущение (14) будет заметно нарушаться только в небольшой области около краев заполнителя $x = 0$, a и $y = 0$, b.

Подставляя равенства (14) в геометрические соотношения (13), будем иметь

$$e_x = 0, \qquad e_y = 0, \tag{15}$$

$$e_{xy} = 0, \tag{16}$$

$$e_z = \partial u_z / \partial z, \quad e_{xz} = \partial u_z / \partial x, \quad e_{yz} = \partial u_z / \partial y. \tag{17}$$

С учетом уравнений (15) и (16) выражение (8), определяющее потенциальную энергию деформации заполнителя, примет вид

$$U_h = \frac{1}{2} \int_0^a \int_0^b \int_0^{h/2} \left(\sigma_z e_z + \tau_{xz} e_{xz} + \tau_{xz} e_{xz} + \tau_{yz} e_{yz} \right) dxdydz. \tag{18}$$

Выразим напряжения, входящие в формулу (18), через деформации. Для определения σ_z воспользуемся соотношениями (9) и (10). Учитывая равенства (15), из уравнений (9) выразим напряжения σ_x и σ_y через напряжение σ_z. Подставляя полученный результат в формулу (10), будем иметь

$$\sigma_z = \overline{E_z} e_z, \tag{19}$$

где $\overline{E_z}$ – приведенный модуль упругости материала заполнителя в направлении оси z:

$$\overline{E_z} = E_z \left[1 - \nu_{zx} \left(\nu_{xz} + \nu_{yz} \nu_{xy} \right) + \right. \\ \left. + \nu_{zy} \left(\nu_{yz} + \nu_{yx} \nu_{xz} \right) / \left(1 - \nu_{yx} \nu_{xy} \right) \right]^{-1}. \tag{20}$$

Трансверсальные напряжения τ_{xz} и τ_{yz} определим из физических соотношений (12):

$$\tau_{xz} = G_{xz} e_{xz}, \quad \tau_{yz} = G_{yz} e_{yz}. \tag{21}$$

Подставим равенства (19) и (21) в формулу (18) и заменим деформации, используя геометрические соотношения (17). Тогда для потенциальной энергии деформации заполнителя будем иметь

$$U_h = \frac{1}{2} \int_0^a \int_0^b \int_0^{h/2} \left[\overline{E_z} \left(\partial u_z / \partial z \right)^2 + G_{xz} \left(\partial u_z / \partial x \right)^2 + \right. \\ \left. + G_{yz} \left(\partial u_z / \partial y \right)^2 \right] dxdydz. \tag{22}$$

Подставляя равенства (14) в формулу (7), получим

$$T_h = \frac{1}{2} \int_0^a \int_0^b \int_0^{h/2} \rho_h \left(\partial u_z / \partial \tau \right)^2 dxdydz. \tag{23}$$

Таким образом, потенциальная и кинетическая энергии заполнителя зависят только от перемещения u_z. При симметричных колебаниях трехслойной пластины функция u_z должна удовлетворять следующим условиям:

$$u_z \left(x, y, z, \tau \right) = \begin{cases} w\left(x, y, \tau \right), & z = h/2, \\ 0, & z = 0. \end{cases} \tag{24}$$

Учитывая эти условия, представим перемещение u_z в следующем виде

$$u_z \left(x, y, z, \tau \right) = f\left(z \right) w\left(x, y, \tau \right). \tag{25}$$

Здесь функция $f(z)$ задает характер распределения перемещений u_z по толщине заполнителя. Из уравнений (24) и (25) следует, что $f(h/2) = 1$ и $f(0) = 0$. При этом зависимость $f(z)$ должна обладать возможностью регулировать закон затухания перемещения по толщине заполнителя. Поэтому зададим $f(z)$ в следующем виде:

$$f\left(z \right) = \left(z / \left(h/2 \right) \right)^{\xi}, \tag{26}$$

где ξ – неизвестный, подлежащий определению параметр, который задает характер затухания нормального перемещения по толщине заполнителя. Изменение параметра ξ приводит к изменению характера затухания перемещения u_z от величины прогиба несущего слоя w до 0. При $\xi = 1$ функция $f(z)$ определяет линейный закон затухания u_z по толщине заполнителя. При $\xi > 1$ основное затухание u_z происходит в сужающейся зоне, которая приближается к несущему слою пластины. При $\xi < 1$ зона затухания u_z увеличивается и смещается к срединной плоскости пластины. Таким образом, параметр ξ позволяет изменять размеры зоны заполнителя, в которой происходит затухание перемещения u_z. Величина ξ будет вычисляться в процессе определения частоты симметричных колебаний трехслойной пластины.

Подставляя равенства (25) и (26) в выражения (22) и (23) и выполняя интегрирование, получим

$$U_h = \frac{1}{2} \int_0^a \int_0^b \left[R_z w^2 + K_x \left(\frac{\partial w}{\partial x} \right)^2 + K_y \left(\frac{\partial w}{\partial y} \right)^2 \right] dxdy; \tag{27}$$

$$T_h = \frac{1}{2} B_{\rho h} \int_0^a \int_0^b \left(\frac{\partial w}{\partial \tau} \right)^2 dxdy. \tag{28}$$

Здесь

$$R_z = \overline{E_z} \left(2/h \right) \varphi\left(\xi \right), \quad K_x = G_{xz} \left(h/2 \right) \theta\left(\xi \right), \\ K_y = G_{yz} \left(h/2 \right) \theta\left(\xi \right), \tag{29}$$

$$B_{\rho h} = \rho_h \left(h/2 \right) \theta\left(\xi \right). \tag{30}$$

Функции $\varphi(\xi)$, $\theta(\xi)$ определяются следующим образом:

$$\varphi\left(\xi \right) = \xi^2 / \left(2\xi - 1 \right), \quad \theta\left(\xi \right) = 1 / \left(2\xi + 1 \right). \tag{31}$$

В равенствах (29) и (30) R_z, K_x, K_y – жесткостные параметры заполнителя; $B_{\rho t}$ – инерциальный параметр заполнителя. Выражения (27) и (29) определяют потенциальную энергию деформации и кинетическую энергию заполнителя в соответствии с допущением об отсутствии тангенциальных перемещений при симметричных колебаниях пластины.

При известных значениях потенциальной и кинетической энергий несущего слоя и заполнителя определим функцию Лагранжа для всей пластины. Подставляя уравнения (3), (4) в равенство (2), получим

$$L = T_t + T_h - (U_t + U_h). \tag{32}$$

Учитывая формулы (5), (6), (27) и (28), представим функцию Лагранжа (32) в следующем виде:

$$L = \int_0^a \int_0^b \Phi \left(\frac{\partial^2 w}{\partial x^2}, \frac{\partial^2 w}{\partial y^2}, \frac{\partial^2 w}{\partial x \partial y}, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial w}{\partial \tau}, w \right) dxdy. \tag{33}$$

Здесь

$$\Phi = \frac{1}{2} \left\{ \left(B_{\rho t} + B_{\rho h} \right) \left(\frac{\partial w}{\partial \tau} \right)^2 - \left[D_{11} \left(\frac{\partial^2 w}{\partial x^2} \right)^2 + \right. \right. \\ + 2 D_{12} \frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2} + D_{22} \left(\frac{\partial^2 w}{\partial y^2} \right)^2 + 4 D_{33} \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 + \\ \left. \left. + R_z w^2 + K_x \left(\frac{\partial w}{\partial x} \right)^2 + K_y \left(\frac{\partial w}{\partial y} \right)^2 \right] \right\}. \tag{34}$$

Подставляя (33) в (1), будем иметь

$$S = \int_{\tau_1}^{\tau_2} \int_0^a \int_0^b \Phi \left(\frac{\partial^2 w}{\partial x^2}, \frac{\partial^2 w}{\partial y^2}, \frac{\partial^2 w}{\partial x \partial y}, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial w}{\partial \tau}, w \right) dxdy. \tag{35}$$

В соответствии с принципом Гамильтона, интеграл действия (35) в промежутке времени $\tau_2 - \tau_1$ для действительного движения трехслойной пластины

имеет стационарное значение. Тогда вариация функционала (35) равна нулю, т. е.

$$\delta S = 0. \tag{36}$$

Уравнение, описывающее симметричные колебания трехслойной пластины, является уравнением Эйлера для функционала (35). Это уравнение имеет следующий вид:

$$\frac{\partial^2}{\partial x^2}\left[\frac{\partial \Phi}{\partial\left(\partial^2 w/\partial x^2\right)}\right]+\frac{\partial^2}{\partial x\partial y}\left[\frac{\partial \Phi}{\partial\left(\partial^2 w/\partial x\partial y\right)}\right]+$$
$$+\frac{\partial^2}{\partial y^2}\left[\frac{\partial \Phi}{\partial\left(\partial^2 w/\partial y^2\right)}\right]-\frac{\partial}{\partial x}\left[\frac{\partial \Phi}{\partial\left(\partial w/\partial x\right)}\right]- \tag{37}$$
$$-\frac{\partial}{\partial y}\left[\frac{\partial \Phi}{\partial\left(\partial w/\partial y\right)}\right]-\frac{\partial}{\partial \tau}\left[\frac{\partial \Phi}{\partial\left(\partial w/\partial \tau\right)}\right]+\frac{\partial \Phi}{\partial w}=0.$$

Подставляя (34) в (37), будем иметь

$$D_{11}\partial^4 w/\partial x^4+2\left(D_{12}+2D_{33}\right)\partial^4 w/\partial x^2\partial y^2+$$
$$+D_{22}\partial^4 w/\partial y^4-K_x\partial^2 w/\partial x^2-K_y\partial^2 w/\partial y^2+ \tag{38}$$
$$+R_z w+\left(B_{\rho t}+B_{\rho h}\right)\partial^2 w/\partial \tau^2=0.$$

Уравнение (38) представляет собой основное дифференциальное уравнение симметричных колебаний трехслойной пластины.

Рассматривая свободные колебания трехслойной пластины, представим прогиб несущего слоя в следующем виде:

$$w(x,y,\tau)=w(x,y)e^{i\omega\tau}, \tag{39}$$

где i – мнимая единица; ω – круговая частота колебаний; $w(x, y)$ – функция формы, описывающая моды колебаний несущего слоя.

Получим уравнение для определения формы несущего слоя при симметричных колебаниях. Подставляя (39) в (38), будем иметь

$$D_{11}\partial^4 w/\partial x^4+2\left(D_{12}+2D_{33}\right)\partial^4 w/\partial x^2\partial y^2+$$
$$+D_{22}\partial^4 w/\partial y^4-K_x\partial^2 w/\partial x^2-K_y\partial^2 w/\partial y^2+ \tag{40}$$
$$+R_z w-\left(B_{\rho t}+B_{\rho h}\right)\omega^2 w=0.$$

Здесь и далее $w = w(x, y)$.

Трехслойная пластина с шарнирно-закрепленными несущими слоями. Определим частоту симметричных колебаний трехслойной пластины, у которой края несущих слоев шарнирно закреплены. Представим решение уравнения (40) в следующем виде:

$$w=\sum_{m=1}\sum_{n=1} w_{mn}\sin\lambda_m x\sin\lambda_n y, \tag{41}$$

где m, n – число полуволн вдоль осей x и y соответственно; $\lambda_m = m\pi/a$, $\lambda_n = n\pi/a$, w_{mn} – неизвестные числа.

Подставляя (41) в (40), получим

$$D_{11}\lambda_m^4+2\left(D_{12}+2D_{33}\right)\lambda_m^2\lambda_n^2+D_{22}\lambda_n^4-$$
$$-K_x\lambda_m^2-K_y\lambda_n^2+R_z=\left(B_{\rho t}+B_{\rho h}\right)\omega_{mn}^2, \tag{42}$$

где ω_{mn} – частота колебаний, соответствующая номерам mn.

Приведем уравнение (42) к безразмерному виду. Для этого умножим его на a^2b^2 и разделим на $(D_{11}D_{22})^{1/2}$. В результате преобразований получим

$$\eta_{mn}=\zeta_{mn}\gamma_{mn}. \tag{43}$$

Здесь

$$\zeta_{mn}=\pi^4\left(\alpha m^4+2\beta m^2n^2+n^4/\alpha\right), \tag{44}$$

$$\gamma_{mn}=1/1+r_\rho\left(1+\pi^2 r_x\alpha m^2+r_y n^2/a/\zeta_{mn}+t_z/\zeta_{mn}\right), \tag{45}$$

$$\alpha=\sqrt{D_{11}/D_{22}}\,b^2/a^2,\quad \beta=\left(D_{12}+2D_{33}\right)/\sqrt{D_{11}D_{22}}, \tag{46}$$

$$r_x=K_x a^2/D_{11},\quad r_y=K_y b^2/D_{22},$$
$$t_z=R_z a^2b^2/\sqrt{D_{11}D_{22}},\quad r_\rho=B_{\rho h}/B_{\rho t}, \tag{47}$$

$$\eta_{mn}=\omega_{mn}^2 B_{\rho t}a^2b^2/\sqrt{D_{11}D_{22}}. \tag{48}$$

Равенство (48) определяет безразмерный частотный параметр η_{mn} несущего слоя как отдельной пластины, не связанной с заполнителем. В формуле (43) величина γ_{mn} характеризует степень влияния заполнителя на частотный параметр несущего слоя. Величина ζ_{mn} при этом определяет частотный параметр только пластины несущего слоя. На самом деле, если заполнитель отсутствует, то из формул (29), (30) и (47) следует, что $r_x = 0$, $r_y = 0$, $t_z = 0$, $r_\rho = 0$. В этом случае из равенства (45) будем иметь $\gamma_{mn} = 1$. Тогда $\eta_{mn} = \zeta_{mn}$. Используя формулы (29) и (30), представим равенства (47) в следующем виде:

$$r_x=p_x\theta(\xi),\quad r_y=p_y\theta(\xi),$$
$$t_z=s_z\varphi(\xi),\quad r_\rho=p_\rho\theta(\xi) \tag{49}$$

Здесь

$$p_x=G_{xz}ha^2/(2D_{11}),\quad p_y=G_{yz}hb^2/(2D_{22}),$$
$$s_z=2\tilde{E}a^2b^2/(h(D_{11}D_{22})^{1/2}),$$
$$p_\rho=p_h h/\left(2B_{\rho t}\right). \tag{50}$$

Подставляя (49) в (45), будем иметь

$$\gamma_{mn}(\xi)=\left[1+\theta(\xi)\pi^2\left(p_x\alpha m^2+p_y n^2/\alpha\right)/\right.$$
$$\left./\zeta_{mn}+\varphi(\xi)s_z/\zeta_{mn}\right]/\left(1+p_\rho\theta(\xi)\right). \tag{51}$$

Учитывая равенство (51), запишем выражение (43) в следующем виде:

$$\eta_{mn}=\zeta_{mn}\gamma_{mn}(\xi). \tag{52}$$

Таким образом, частотный параметр зависит от параметра ξ, который задает характер затухания нормального перемещения по толщине заполнителя.

Определим частоту симметричных колебаний трехслойной пластины. Из равенства (48) будем иметь

$$\omega_{mn}=\frac{\sqrt{\eta_{mn}}}{ab}\sqrt{\frac{\sqrt{D_{11}D_{22}}}{B_{\rho t}}}. \tag{53}$$

Подставляя (52) в (53), получим

$$\omega_{mn}=\psi_{mn}(\xi)\omega_{mnt}, \tag{54}$$

где $\psi_{mn}(\xi) = (\gamma_{mn}(\xi))^{1/2}$. Величина ω_{mnt}, представляющая собой частоту колебаний только пластины несущего слоя, определяется следующим выражением:

$$\omega_{mnt} = \frac{\lambda_{mn}}{ab}\sqrt{\frac{\sqrt{D_{11}D_{22}}}{B_{\rho t}}}, \qquad (55)$$

где $\lambda_{mn} = (\zeta_{mn})^{1/2}$.

Из формулы (54) видно, что частота симметричных колебаний трехслойной пластины зависит от величины ξ. Истинное значение частоты колебаний для каждой комбинации m и n может быть получено с помощью процедуры минимизации по параметру ξ. В этом случае

$$\omega_{mn} = \varphi_{mn}\omega_{mnt}, \qquad (56)$$

где

$$\varphi_{mn} = \min_{\xi_l \le \xi \le \xi_r}\left[\psi_{mn}(\xi)\right], \qquad (57)$$

где ξ_r – ξ_l – диапазон изменения параметра ξ.

Формула (56) определяет частоту симметричных колебаний трехслойной пластины. Как видно, частота колебаний ω_{mn} представляет собой произведение частоты колебаний несущего слоя ω_{mnt} и частотного коэффициента φ_{mn}, который учитывает динамическое поведение заполнителя. Отметим, что формула (56) оказывается особенно удобной для параметрического анализа, в котором исследуется влияние геометрических, упругих и инерциальных параметров трехслойной пластины на частоты колебаний.

Получим расчетные формулы для трехслойной пластины с однородными ортотропными несущими слоями. Изгибные жесткости такого несущего слоя определяются следующим образом:

$$\begin{aligned} D_{11} &= \overline{E_{xt}}t^3/12, \quad D_{12} = \overline{E_{xt}}\nu_{xyt}t^3/12, \\ D_{22} &= \overline{E_{yt}}t^3/12, \quad D_{33} = G_{xyt}t^3/12, \end{aligned} \qquad (58)$$

$$\begin{aligned} \overline{E_{xt}} &= E_{xt}/\left(1-\nu_{xyt}\nu_{yxt}\right), \\ \overline{E_{yt}} &= E_{yt}/\left(1-\nu_{xyt}\nu_{yxt}\right), \end{aligned} \qquad (59)$$

где E_{xt}, E_{yt} – модули упругости; G_{xyt} – модуль сдвига; ν_{xyt}, ν_{yxt} – коэффициенты Пуассона. Для инерциального параметра однородного несущего слоя будем иметь

$$B_{\rho t} = \rho_t t, \qquad (60)$$

где ρ_t – плотность материала несущего слоя.

Подставляя (58) и (59) в (46), получим

$$\begin{aligned} \alpha &= \sqrt{E_{xt}/E_{yt}}\,b^2/a^2, \\ \beta &= \left(\overline{E}_{xt}\nu_{xyt} + 2G_{xyt}\right)\sqrt{E_{xt}/E_{yt}}. \end{aligned} \qquad (61)$$

Определим далее безразмерные параметры p_x, p_y, s_z, p_ρ. Учитывая формулы (58) и (60), из равенств (50) будем иметь

$$\begin{aligned} p_x &= 6\frac{G_{xz}}{\overline{E}_{xt}}\frac{h}{t}\frac{a^2}{t^2}, \qquad p_y = 6\frac{G_{yz}}{\overline{E}_{yt}}\frac{h}{t}\frac{b^2}{t^2}, \\ s_z &= 24\frac{\overline{E}_z}{\sqrt{\overline{E}_{xt}\overline{E}_{yt}}}\frac{a^2}{t^2}\frac{b^2}{t^2}\left(\frac{h}{t}\right)^{-1}, \end{aligned} \qquad (62)$$

$$p_\rho = \frac{1}{2}\frac{\rho_h}{\rho_t}\frac{h}{t}. \qquad (63)$$

Подставляя D_{11} и D_{22} из (58) и $B_{\rho t}$ из (60) в выражение (55), получим

$$\omega_{mnt} = \lambda_{mn}\frac{t}{ab}\sqrt{\frac{\sqrt{\overline{E}_{xt}\overline{E}_{yt}}}{12\rho_t}}. \qquad (64)$$

Формула (64) определяет частоту колебаний однородного ортотропного несущего слоя.

Если несущие слои выполнены из изотропного материала, то для модулей упругости будем иметь

$$E_{xt} = E_{yt} = E_t, \qquad (65)$$

$$\overline{E}_{xt} = \overline{E}_{yt} = \overline{E}_t, \qquad (66)$$

$$\overline{E}_{xt}\nu_{xyt} = \overline{E}_t\nu_t, \quad G_{xyt} = 0{,}5\left(1-\nu_t\right)/\overline{E}_t, \qquad (67)$$

где $\overline{E}_t = E_t(1-\nu_t^2)$; E_t – модуль упругости; ν_t – коэффициент Пуассона.

Подставляя (65), (66) и (67) в равенства (61), получим

$$\alpha = b^2/a^2, \quad \beta = 1. \qquad (68)$$

С учетом соотношений (66), выражения (62), определяющие безразмерные параметры p_x, p_y и s_z, примут следующий вид:

$$\begin{aligned} p_x &= 6\frac{G_{xz}}{\overline{E}_t}\frac{h}{t}\frac{a^2}{t^2}, \quad p_y = 6\frac{G_{yz}}{\overline{E}_t}\frac{h}{t}\frac{b^2}{t^2}, \\ s_z &= 24\frac{\overline{E}_z}{\overline{E}_t}\frac{a^2}{t^2}\frac{b^2}{t^2}\left(\frac{h}{t}\right)^{-1}. \end{aligned} \qquad (69)$$

Подставляя (66) в (64) для частоты колебаний изотропного несущего слоя, будем иметь

$$\omega_{mnt} = \lambda_{mn}\frac{t}{ab}\sqrt{\frac{\overline{E}_t}{12\rho_t}}. \qquad (70)$$

Структуры трехслойных пластин с композиционными несущими слоями и ортотропным заполнителем обладают большим разнообразием геометрических, упругих и инерциальных параметров. Каждый из этих параметров тем или иным образом может влиять на частоты симметричных колебаний. Исследование такого влияния в общем случае является достаточно громоздкой задачей. Для того чтобы сделать параметрический анализ обозримым и в то же время продемонстрировать возможности разработанной динамической модели, рассмотрим далее квадратную трехслойную пластину с изотропными несущими слоями и изотропным заполнителем. Получим расчетные формулы для такой пластины. Пусть в заполнителе

$$E_x = E_y = E_z = E_h, \qquad (71)$$

$$G_{xy} = G_{xz} = G_{yz} = G_h, \qquad (72)$$

$$\nu_{xy} = \nu_{xz} = \nu_{yx} = \nu_{yz} = \nu_{zx} = \nu_{zy} = \nu_h, \qquad (73)$$

где E_h, G_h, – модули упругости; ν_h – коэффициент Пуассона изотропного материала заполнителя.

С учетом равенства (71) и (73) выражение (20), определяющее приведенный модуль упругости материала заполнителя, примет следующий вид:

$$\overline{E}_h = E_h\left[1-2\nu_h^2/\left(1-\nu_h\right)\right]^{-1}. \qquad (74)$$

Подставляя (72) в (69) и учитывая, что $a = b$ и $\overline{E}_z = \overline{E}_h$, получим

$$p_x = p_y = p = 6\frac{G_h}{\overline{E}_t}\frac{h}{t}\frac{a^2}{t^2},$$
$$s_z = s = 24\frac{\overline{E}_h}{\overline{E}_t}\frac{a^4}{t^4}\left(\frac{h}{t}\right)^{-1}. \quad (75)$$

Определим далее для рассматриваемой пластины параметры ζ_{mn}, λ_{mn} и $\gamma_{mn}(\xi)$. Из первого равенства (68) при $a = b$ будем иметь $\alpha = 1$. Подставляя $\alpha = 1$ и $\beta = 1$ в уравнение (44), получим

$$\zeta_{mn} = \pi^4\left(m^2 + n^2\right)^2. \quad (76)$$

Тогда

$$\lambda_{mn} = \sqrt{\zeta_{mn}} = \pi^2\left(m^2 + n^2\right). \quad (77)$$

Подставляя $\alpha = 1$, $p_x = p_y = p$, $s_z = s$ в уравнение (51) и учитывая равенство (76), будем иметь

$$\gamma_{mn}(\xi) = \left[1 + p\theta(\xi)/\lambda_{mn} + s\varphi(\xi)/\zeta_{mn}\right]/ \left/\left(1 + p_\rho\theta(\xi)\right)\right. . \quad (78)$$

Частота колебаний квадратного изотропного несущего слоя может быть получена из формулы (70), если в ней принять $a = b$. Тогда

$$\omega_{mnt} = \lambda_{mn}\frac{t}{a^2}\sqrt{\frac{\overline{E}_t}{12\rho_t}}. \quad (79)$$

Частота колебаний трехслойной пластины определяется по-прежнему с помощью формулы (56).

Сделаем еще одно упрощение при решении рассматриваемой задачи. Будем в дальнейшем отыскивать частоты колебаний только для случая, когда $m = n$. Получим необходимые расчетные формулы, заменяя при этом получающийся двойной индекс mn на одинарный индекс m. Из уравнений (76), (77), (78) и (79), полагая в них $m = n$, получим

$$\zeta_m = 4\pi^4 m^4, \quad \lambda_m = 2\pi^2 m^2, \quad (80)$$

$$\gamma_m(\xi) = \left[1 + p\theta(\xi)/\lambda_m + s\varphi(\xi)/\zeta_m\right]/\left(1 + p_\rho\theta(\xi)\right), \quad (81)$$

$$\omega_{mt} = \lambda_m\frac{t}{a^2}\sqrt{\frac{\overline{E}_t}{12\rho_t}}. \quad (82)$$

Частота симметричных колебаний рассматриваемой трехслойной пластины может быть найдена из следующих формул (см. уравнения (56), (57):

$$\omega_m = \varphi_m\omega_{mt}, \quad (83)$$

$$\varphi_m = \min_{\xi_l \le \xi \le \xi_r}\psi_m(\xi), \quad (84)$$

где

$$\psi_m(\xi) = \sqrt{\gamma_m(\xi)}. \quad (85)$$

Рассмотрим порядок определения частоты колебаний для произвольного номера m. Пусть известны размеры a, h, t, упругие характеристики E_t, ν_t, E_h, ν_h (или G_h) и плотности ρ_t, ρ_h. С помощью формул (63), (75) определим безразмерные параметры p_ρ, p, s. Затем используя равенства (31), (81), (84), выполним процедуру минимизации функции $\psi_m(\xi)$ по параметру ξ и найдем коэффициент φ_m согласно формуле (84). Завершая вычисления, определим частоту ω_{mt} (82) и частоту симметричных колебаний трехслойной пластины ω_m (83).

Выполним верификацию разработанной модели симметричных колебаний трехслойной пластины. Для этого сравним частоты колебаний, в одном случае полученные с помощью этой модели, а в другом – с помощью метода конечных элементов. Пусть рассматриваемая трехслойная пластина имеет $a = 0,2$ м; $h = 0,04$ м; $t = 0,000\,5$ м; $E_t = 69$ ГПа; $\nu_t = 0,33$; $E_h = 165$ МПа; $\nu_p = 0,031\,25$; $\rho_t = 2\,700$ кг/м^3; $\rho_h = 30$ кг/м^3. Определим в примере частоту колебаний $f_m = \omega_m/2\pi$ для $m = 1, 3, 5$. Результаты расчетов, полученные с помощью разработанной модели, имеют следующие значения: $f_1 = 11\,800$ Гц, $f_3 = 13\,184$ Гц, $f_5 = 15\,469$ Гц.

Для определения частот колебаний с помощью метода конечных элементов был использован пакет COSMOS/M. Пластины несущих слоев моделировались элементами SHELL4, а трехмерный заполнитель моделировался элементами SOLID.

Результаты вычислений частот с использованием конечно-элементного моделирования обозначим $f_m^{\mathrm{FEM}}(m = 1, 3, 5)$. Были получены следующие значения частот колебаний: $f_1^{\mathrm{FEM}} = 11\,803$ Гц, $f_3^{\mathrm{FEM}} = 13\,191$ Гц, $f_5^{\mathrm{FEM}} = 15\,483$ Гц.

Сравнение частот колебаний f_m и $f_m^{\mathrm{FEM}}(m = 1, 3, 5)$ показывает, что относительная разница между ними составляет 0,025; 0,054; 0,090 % соответственно. Это дает основание утверждать, что разработанная модель позволяет с высокой точностью определять частоты симметричных колебаний трехслойной пластины. Частотный анализ, выполненный по приведенным выше формулам, требует минимальных вычислительных ресурсов, не сравнимых с вычислительными ресурсами, необходимыми для решения этой задачи методом конечных элементов. Преимущества разработанной модели будут особенно заметны в высокочастотном вибрационном анализе. При высокочастотных колебаниях длина полуволны становится заметно меньше толщины заполнителя. Если такую задачу решать методом конечных элементов, то для реалистичного описания движения трехмерного заполнителя надо будет использовать элементы очень малого размера. Это потребует значительных вычислительных ресурсов и увеличит время решения.

Таким образом, решена задача о симметричных колебаниях трехслойной пластины, структура которой состоит из двух одинаковых композитных несущих слоев и ортотропного заполнителя. С использованием вариационного принципа Гамильтона получено дифференциальное уравнение, описывающее движение трехслойной пластины. При выводе этого уравнения были использованы два предположения, касающиеся деформирования заполнителя. Согласно

первому предположению, в заполнителе отсутствуют тангенциальные перемещения. В соответствии со вторым предположением нормальные перемещения в заполнителе изменяются от прогиба несущего слоя до нуля по нелинейному закону. Были получены формулы, определяющие частоту симметричных колебаний трехслойной пластины, у которой края несущих слоев шарнирно закреплены. Показано, что частота колебаний трехслойной пластины может быть представлена в виде произведения частоты колебаний несущего слоя и частотного коэффициента, который учитывает динамическое поведение заполнителя. Приведены формулы, определяющие величину частотного коэффициента для трехслойных пластин с ортотропными и изотропными несущими слоями. С помощью метода конечных элементов была выполнена верификация разработанной модели симметричных колебаний трехслойной пластины. Из сравнения результатов вычислений следует, что полученные в работе формулы позволяют с высокой достоверностью определять частоты симметричных колебаний трехслойных пластин. Разработанная модель дает возможность находить частоты с точностью, сопоставимой с точностью аналогичных расчетов, выполненных в пакетах COSMOS/M, ANSYS. При этом частотный анализ, выполненный на основе представленной модели, требует минимальных вычислительных ресурсов.

Библиографические ссылки

1. Frosting Y., Thomson O. T. High-Order Free Vibration of Sandwich Panels with a Flexible Core // Intern. J. of Solids and Structures. 2004. № 41(5–6). P. 1697–1724.

2. Langhaar H. L. Energy Methods in Applied Mechanics. N. Y. : John Wiley & Sons, 1962.

3. Васильев В. В. Механика конструкций из композиционных материалов. М. : Машиностроение, 1988.

A. V. Lopatin, R. A. Udaltsov

WRINKLING VIBRATIONS OF SANDWICH PLATE

The task of frequency test of sandwich plate wrinkling vibrations, when plate is composed of two similar composite bearing layers and orthotropic filler, is solved in this paper. Hamilton's principle was applied to derive a basic differential equation of forth order. Frequency formula for symmetrical vibrations of sandwich plate with pinned bearing layers is given.

Keywords: sandwich plate, wrinkling vibrations.

© Лопатин А. В., Удальцов Р. А., 2010

УДК 621.393.3

В. Б. Малинкин, Е. В. Малинкин, Е. Ф. Кураш, О. В. Соболева

ИНВАРИАНТНАЯ ВОЛОКОННО-ОПТИЧЕСКАЯ СИСТЕМА ПЕРЕДАЧИ И ЕЁ ХАРАКТЕРИСТИКИ

Предложен метод борьбы с искажениями, вносимыми волоконно-оптической линией связи. Метод основан на использовании инвариантного равенства. Определены основные технические характеристики.

Ключевые слова: инвариантный, волоконно-оптическая система передачи, фотоприемник.

В подавляющем большинстве случаев в волоконно-оптических системах передачи (ВОСП) для передачи информационного сигнала используется классическая амплитудная модуляция.

Вероятность ошибочного приема регенераторов составляет 10^{-10} в соответствии с рекомендациями Международного союза электросвязи (МСЭ) (ITU-T G.707. Network node interface for the Synchronous Digital Hierarchy 2004). В более поздних рекомендациях МСЭ (ITU-T G.975. Forward error correction for submarine systems. 1996) предлагается использовать устройства защиты от ошибок (УЗО), работа которых основана на специальном кодировании сигнала передачи с помощью циклических кодов. При скорости передачи 10 Гбит/с и выше создать устройства защиты от ошибок, работающие в реальном масштабе времени, сложно.

Между тем уменьшение вероятности ошибки можно достичь другими способами. Один из них предлагается ниже.

Постановка задачи. Имеем ВОСП (рис. 1). В качестве передатчика используется лазер. В качестве приемника используется фотоприемное устройство. Для передачи информационного сигнала используется второе окно прозрачности.

Необходимо синтезировать алгоритм передачи информационного сигнала, основанный на инвариантном способе обработки информации.

Решение поставленной задачи. Структура ВОСП включает в себя передающие и приемные устройства и ВОЛП.

Следует заметить, что сквозной тракт ВОСП во втором окне прозрачности [1] является линейным при условии, что мощность на выходе передатчика не превышает допустимую величину 1 мВт. *Z*-преобразование сигналов приема $Y(Z)$ на выходе ФПУ (на *i*-м блоке обработки) определится по формуле:

$$Y_i(Z) = [G(Z)\cdot H_0(Z))\cdot H_1(Z))\cdot H_2(Z)]_i, \qquad (1)$$

где $G_i(Z)$ – *Z*-преобразование сигнала передачи на *i*-м блоке обработки; $H_{0i}(Z)$ – передаточная характеристика формирователя сигналов на *i*-м блоке; $H_{1i}(Z)$ – передаточная характеристика волоконно-оптической линии передачи на *i*-м блоке; $H_{2i}(Z)$ – передаточная характеристика приемника (ФПУ) на *i*-м блоке.

В работе [2] приведено инвариантное равенство, справедливое для любого линейного четырехполюсника:

$$\frac{G_i(Z)}{G_{i-1}(Z)} = \frac{Y_i(Z)}{Y_{i-1}(Z)}. \qquad (2)$$

Равенство (2) справедливо для физически реализуемых систем, когда знаменатели не равны нулю.

Любая ВОСП является консервативной системой, характеристики которой разделены на интервалы стационарности.

Таким образом,

$$H_{\Sigma(i-1)}(Z) \approx H_{\Sigma i}(Z) \approx H_{\Sigma(i+1)}(Z) \approx \ldots, \qquad (3)$$

где $H_{\Sigma i}(Z) = H_{0i}(Z)\cdot H_{1i}(Z)\cdot H_{2i}(Z)$ – передаточная характеристика сквозного тракта ВОСП на *i*-м блоке.

При подстановке выражения (3) в (2) с учетом (1), равенство (2) переходит в тождество.

При переходе от *Z*-изображений к амплитудно-фазовым спектрам имеем:

$$\frac{G_i(jk\omega_1)}{G_{i-1}(jk\omega_1)} = \frac{Y_i(jk\omega_1)}{Y_{i-1}(jk\omega_1)}. \qquad (4)$$

Равенство (4), в свою очередь, распадается на равенство отношений амплитудных спектров и равенство разности цифровых спектров:

$$\left.\begin{aligned} &\frac{G_i(k\omega_1)}{G_{i-1}(k\omega_1)} = \frac{Y_i(k\omega_1)}{Y_{i-1}(k\omega_1)}, \\ &\varphi_i(k\omega_1) - \varphi_{i-1}(k\omega_1) = \\ &= \psi_i(k\omega_1) - \psi_{i-1}(k\omega_1), \end{aligned}\right\} \qquad (5)$$

где $G_i(k\omega_1)$ и $G_{i-1}(k\omega_1)$ – амплитудные спектры на входе формирователя сигналов на *i*-м и (*i*–1)-блоках; $Y_i(k\omega_1)$ и $Y_{i-1}(k\omega_1)$ – амплитудные спектры на выходе ФПУ сигналов на *i*-м и (*i*–1)-блоках; $\varphi_i(k\omega_1)$ и $\varphi_{i-1}(k\omega_1)$ – фазовые спектры сигналов на входе формирователя сигналов на *i*-м и (*i*–1)-блоках; $\psi_i(k\omega_1)$ и $\psi_{i-1}(k\omega_1)$ – фазовые спектры на выходе ФПУ сигналов на *i*-м и (*i*–1)-блоках.

Первое равенство (5) повторяет принцип относительной амплитудной модуляции (ОАМ), а второе – принцип относительной фазовой модуляции (ОФМ). Таким образом, для достижения минимальной вероятности ошибки в ВОСП необходимо модулирующий параметр «вложить» в отношение *Z*-изображений сигнала передачи на соседних блоках обработки, а на приемной стороне модулирующий параметр извлечь, путем сравнения соседних блоков.

Формирование информационных сигналов в такой системе осуществляется на входе формирователя сигналов, а демодуляция – на выходе ФПУ. Будем далее называть такую систему «инвариантной волоконно-оптической системой передачи» (ИВОСП).

Процесс формирования сигналов будет выглядеть следующим образом:

$$\frac{G_1(Z)}{G_0(Z)} = S_{\text{мод}1}(Z) \to G_1(Z) = G_0(Z)\cdot S_{\text{мод}1}(Z);$$

$$\frac{G_2(Z)}{G_1(Z)} = S_{\text{мод}2}(Z) \to G_2(Z) =$$

$$= G_1(Z)\cdot S_{\text{мод}2}(Z) = G_0(Z)\cdot S_{\text{мод}1}(Z)\cdot S_{\text{мод}2}(Z);$$

$$G_N(Z) = G_0(Z)\cdot\prod_{i=1}^{N} S_{\text{мод}i}(Z), \qquad (6)$$

где $G_0(Z)$ – *Z*-изображение информационного сигнала на начальном блоке (сигнал обучения).

Однако реализовать алгоритм модуляции согласно выражению (6) нельзя, так как при длинных сеансах связи $N \to \infty$ и нерекурсивный фильтр будет физически нереализуемым. Структура формирователя сигналов при $N = 4$ (рис. 2) содержит четыре блока задержки, ключ, умножитель БПФ и ОБПФ. Количество отводов может быть любым.

Модулирующий параметр $S_{\text{мод}}(nT)$ в блоке БПФ преобразуется в $S_{\text{мод}}(Z)$.

Процесс формирования сигналов передачи в каждом блоке содержит 2 этапа. На первом этапе ключ К1 замкнут. В точке «б» сигнал определится формулой

$$G_i(Z) = G_0(Z)\cdot\prod_{k=0}^{5} S_{\text{мод}(i-k)}(Z). \qquad (7)$$

На втором этапе ключ К1 разомкнут. В точке «б» сигнал (рис. 3) определится как

$$G'_i(Z) = G_0(Z)\cdot\prod_{k=0}^{4} S_{\text{мод}(i-k)}(Z). \qquad (8)$$

В соответствии с законами цифровой фильтрации на приемной стороне каждый блок умножается на передаточную характеристику сквозного тракта. Представим *Z*-изображения сигналов приема в виде

$$\left.\begin{aligned} &Y_{i-1}(Z) = G_{i-1}(Z)\cdot H_{\Sigma i-1}(Z), \\ &Y'_{i-1}(Z) = G'_{i-1}(Z)\cdot H_{\Sigma i-1}(Z), \\ &Y_i(Z) = G_i(Z)\cdot H_{\Sigma i}(Z), \\ &Y'_i(Z) = G'_i(Z)\cdot H_{\Sigma i}(Z). \end{aligned}\right\} \qquad (9)$$

Процесс демодуляции заключается в делении первой части $G_i(Z)$ на $G_i'(Z)$. Тогда

$$S'_{\text{моді}}(Z) = \frac{G_0(Z) \cdot \prod_{k=0}^{5} S_{\text{мод}(i-k)}(Z) \cdot H_{\sum i}(Z)}{G_0(Z) \cdot \prod_{k=0}^{4} S_{\text{мод}(i-k)}(Z) \cdot H_{\sum i}(Z)}. \quad (10)$$

Справедливость выражения (10) основана на свойствах относительности среды распространения сквозного тракта ИВОСП и справедливости выражения (3).

Структура приемной части ИВОСП приведена на рис. 4.

Следует заметить, что в данном алгоритме производится компенсация АЧИ и ФЧИ сквозного тракта ИВОСП. Это в свою очередь приводит к компенсации дисперсионных свойств ВОЛП, к увеличению отношения сигнал/шум и уменьшению вероятности ошибки.

Технические характеристики метода. Необходимо указать достоинства и недостатки разработанного метода.

К неоспоримым достоинствам можно отнести компенсацию АЧИ и ФЧИ среды распространения. Это позволяет увеличить длину регенерационного участка при сохранении вероятности ошибочного приема либо существенно уменьшить вероятность ошибки при заданной длине участка регенерации.

К недостаткам можно отнести увеличение скорости передачи информационного сигнала. По-существу, в сигнал передачи введена избыточность, что и позволило улучшить качественные показатели.

Однако наряду с компенсацией АЧИ и ФЧИ наблюдается увеличение аддитивных шумов.

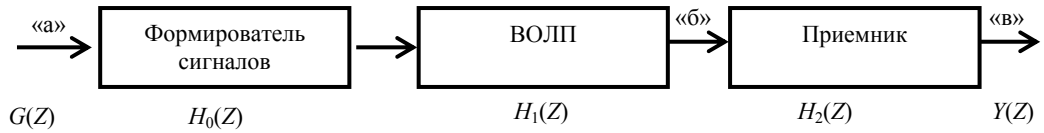


Рис. 1. Структура ВОСП

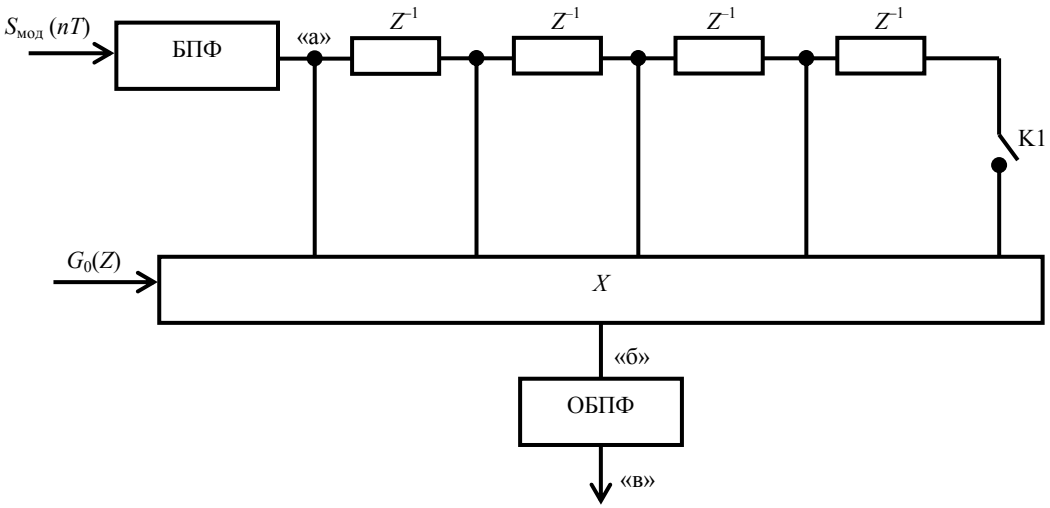


Рис. 2. Структура модулятора ИВОСП

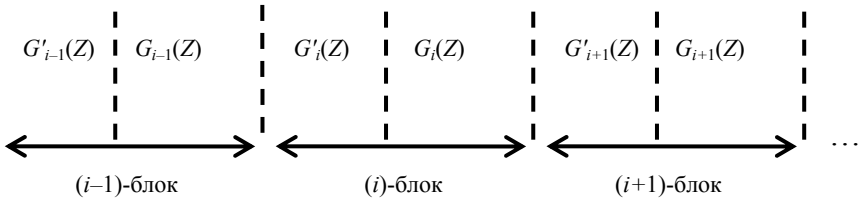


Рис. 3. Представление сигнала передачи

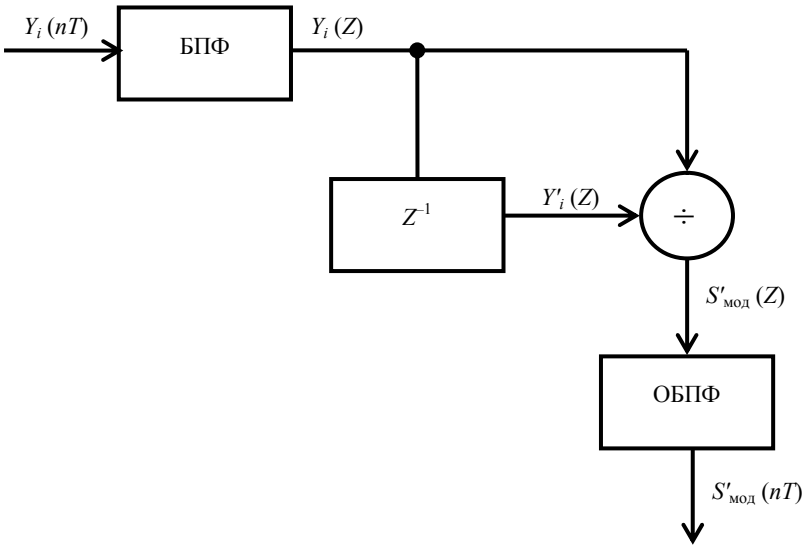


Рис. 4. Структура демодулятора ИВОСП

Оценим величину собственных шумов, используя известное соотношение [3]:

$$\sigma^2 = \frac{\Delta_0^2}{12}\sum_{n=0}^{\infty} h^2(nT) + \frac{\Delta^2}{12}\sum_{j=1}^{N}\sum_{n=0}^{\infty} h_j^2(nT) \ , \qquad (11)$$

где Δ_0 – шаг квантования входного слова; $h(nT)$ – импульсная реакция цифрового фильтра; $h_j(nT)$ – импульсная реакция усеченного цифрового фильтра от j-го источника шума; Δ – шаг квантования обработки сигнала в ЦФ; N – число отводов ЦФ.

Обычно в расчетах принимают $h(nT) = h_j(nT)$ и $\Delta_0 = \Delta$. Тогда выражение (11) упрощается.

Величина дополнительного собственного шума на передаче для $N = 4$ определится формулой

$$\sigma^2_{\text{собств.ПРД}} = \frac{\Delta^2}{12}\cdot 5 \cdot \sum_{n=0}^{\infty} h^2(nT) = \frac{5\Delta^2}{3}, \qquad (12)$$

а величина дополнительного собственного шума на приеме – формулой

$$\sigma^2_{\text{собств.ПРМ}} = \frac{4\Delta^2}{12} = \frac{\Delta^2}{3}. \qquad (13)$$

Выражения для общей величины дополнительного шума запишется как

$$\sigma^2_{\sum\text{собств}} = \sigma^2_{\sum\text{собств.ПРД}} + \sigma^2_{\sum\text{собств.ПРМ}} = 2\Delta^2. \qquad (14)$$

При поступлении шума канала связи (фотонный шум) величина его на выходе блока ОБПФ будет [2]:

$$\sigma^2_{\text{доп.КС}} = \sigma^2_{\text{КС}} \sum_{n=1}^{\infty} h^2(nT) = 2\sigma^2_{\text{КС}}, \qquad (15)$$

где $\sigma^2_{\text{КС}}$ – мощность шума канала связи.

Таким образом, разработана структура ИВОСП, позволяющая компенсировать АЧИ и ФЧИ среды распространения. Определены технические характеристики. Разработанный метод может найти широкое применение в волоконно-оптических системах передачи.

Библиографические ссылки

1. Заславский К. Е. Оптические волокна для систем связи : учеб. пособие / Сиб. гос. ун-т телекоммуникаций и информатики. Новосибирск, 2008.

2. Малинкин В. Б. Повышение помехоустойчивости модифицированных фильтров Калмана в относительных компенсационных методах : дис. … д-ра техн. наук. Омск, 2003.

3. Гольденберг Л. М., Матюшкин Б. В., Поляк М. Н. Цифровая обработка сигналов. М. : Радио и связь, 1990.

V. B. Malinkin, E. V. Malinkin, H. F. Kurash, O. V. Soboleva,

INVARIANT METHOD OF INFORMATION TRANSMISSION IN FIBER-OPTIC TRANSMISSION SYSTEMS

Synthesized method of controlling the distortions introduced by fiber-optic communications line. The method is based on the use of invariant equality. The main technical characteristics were determined.

Keywords: invariant fiber-optic transmission system.

© Малинкин В. Б., Малинкин Е. В., Кураш Е. Ф., Соболева О. В., 2010

УДК 519.7

М. В. Наумов

ГЕНЕРАЦИЯ ЧАСТИЧНО БЕНТ-ФУНКЦИЙ

Для булевых функций в криптографии важны следующие характеристики: сбалансированность, нелинейность, критерий распространения, корреляционная иммунность, степень и отсутствие ненулевых линейных структур. Частично бент-функции – это класс булевых функций, которые представляют интерес, поскольку могут обладать ценными криптографическими свойствами.

Предлагаются два алгоритма генерации частично бент-функций. Эти алгоритмы были реализованы и исследованы; второй из них позволяет улучшать криптографические свойства генерируемых функций.

Ключевые слова: булевы функции, частично бент-функции, корреляционная иммунность.

В работе использованы следующие обозначения.

$Z_2 = \{ 0, 1 \}$; $n \in N$, $G = \{ 0, 1 \}^n$ – множество всех двоичных векторов длины n;

$\oplus$ – сумма по модулю 2, $u \oplus v$ – вектор, каждый бит которого равен сумме по модулю 2 соответствующих бит векторов u и v;

$(u, v) = u_1 v_1 \oplus ... \oplus u_n v_n$ – скалярное произведение векторов u и v из G;

$f: G \to \{0, 1\}$ – булева функция от n переменных;

$w(f) = |\{ s \in G : f(s) \neq 0 \}|$ – вес функции f;

$P_2(n)$ – множество булевых функций от n переменных;

$W_f(s) = \sum_{x \in G} (-1)^{f(x)+(x,s)}$ – преобразование Уолша–Адамара функции f, $W_f: G \to Z$;

NW_f – количество ненулевых значений функции $W_f(s)$;

$\Delta_f(s) = \sum_{x \in G} (-1)^{f(x)+f(x \oplus s)}$ – функция автокорреляции, $\Delta_f : G \to Z$;

$N\Delta_f$ – количество ненулевых значений функции автокорреляции;

$d(f, g) = |\{ s \in G : f(s) \neq g(s) \}| = w(f \oplus g)$ – расстояние по Хэммингу между f и g из $P_2(n)$;

$d(f, T) = \min_{g \in T} d(f, g)$ – расстояние между функцией $f \in P_2(n)$ и множеством $T \subset P_2(n)$;

$x^0 = 1, x^1 = x$, $f(x_1, ..., x_n) = \bigoplus_{(i_1, ..., i_n) \in G} a_{i_1, ..., i_n} x_1^{i_1} \cdot ... \cdot x_n^{i_n}$ – представление функции $f \in P_2(n)$ в виде алгебраической нормальной формы (АНФ), $a_{i_1, ..., i_n} \in \{0, 1\}$;

$\deg (x_1^{i_1} \cdot ... \cdot x_n^{i_n}) = \sum_{k=1}^{n} i_k$ – степень монома $x_1^{i_1} \cdot ... \cdot x_n^{i_n}$, число входящих в него переменных;

$\deg f$ – степень функции f, наибольшая из степеней мономов ее АНФ;

$A(n) = \{ (a, x) \oplus a_0 : a \in G, a_0 \in \{0, 1\} \}$ – класс аффинных функций, $A(n) \subset P_2(n)$;

$N_f = d(f, A(n))$ – нелинейность функции f;

КИ(m) – корреляционная иммунность порядка m;

КР(m) – критерий распространения порядка m.

Основные понятия. Для генерации частично бент-функций применяются обычные бент-функции.

Бент-функцией называется функция $f \in P_2(n)$, у которой $W_f(s) = \pm 2^{n/2}$ для всех $s \in G$.

Утверждение [1]. f – бент-функция, если и только если для всех $s \in G$, $s \neq 0$ выполняется условие $\Delta_f(s) = 0$.

Бент-функции существуют только для четного n, так как W_f – целочисленная функция.

Бент-функции можно генерировать с помощью следующей теоремы.

Теорема (*конструкция Мейорана–МакФарланда*) [2]. Пусть h – любая перестановка на $Z_2^{n/2}$, g – произвольная булева функция от $n/2$ переменных. Тогда функция $f(u', u'') = (u', h(u'')) \oplus g(u'')$ является бент-функцией от n переменных.

Функция $f \in P_2(n)$ называется корреляционно-иммунной порядка m (КИ(m)), если любая ее подфункция от $(n - m)$ переменных, полученная фиксацией остальных m переменных, имеет вес $\frac{w(f)}{2^m}$. Очевидно, что если функция КИ(m), то она КИ(k) для любого $k < m$.

Известно, что $f(x)$ – КИ(m), если и только если $W_f(s) = 0$ для всех $s \in G$ таких, что $1 \le w(s) \le m$.

Корреляционная иммунность позволяет противостоять корреляционной атаке.

К сожалению, бент-функции не могут быть корреляционно-иммунными и сбалансированными, поэтому изучаются различные расширения бент-функций.

Говорят, что функция f удовлетворяет критерию распространения (КР) по направлению a, если $\Delta_f(a) = 0$, т. е. производная $D_a f = f(x) \oplus f(x \oplus a)$ сбалансирована.

Выполнение критерия распространения по направлению a приводит к тому, что $f(x) = f(x \oplus a)$ с вероятностью 0,5, т. е. знание значения $f(x \oplus a)$ не поможет нам узнать $f(x)$.

Говорят, что функция f удовлетворяет критерию распространения порядка m (КР(m)), если $\Delta_f(s) = 0$ для всех $s \in G$ таких, что $1 \le w(s) \le m$.

Критерий распространения – это свойство, которое характеризует поведение функции в момент, когда некоторые ее координаты инвертированы. Это свойство булевой функции подобно свойству диффузии для криптосистемы. Функции, применяемые в блочных шифрах, должны иметь высокий порядок КР.

Частично бент-функции. Функция $f \in P_2(n)$ называется частично бент-функцией, если $NW_f \cdot N\Delta_f = 2^n$.

Для любой бент-функции по определению $NW_f = 2^n$, а по критерию Ротхауза $N\Delta_f = 1$, поэтому класс бент-функций содержится в классе частично бент-функций.

Частично бент-функции представляют интерес в криптографии, потому что для них произведение $NW_f \cdot N\Delta_f$ имеет минимальное значение, что дает основание надеяться на то, что число ненулевых значений преобразования Уолша–Адамара и функции автокорреляции также мало, однако это совсем не гарантирует КИ и КР, ведь может оказаться $W_f(s) \neq 0$ или $\Delta_f(s) \neq 0$ на единственном векторе s веса 1. Очевидно, частично бент-функции могут удовлетворять критерию распространения по многим направлениям (в зависимости от $N\Delta_f$), важному свойству безопасности.

Обычные бент-функции используются в стандарте связи CDMA, в блоках замены (S-boxes), в некоторых шифрах, хэш-функциях [3]. Возможно, частично бент-функции можно использовать в тех же самых областях.

Генератор частично бент-функций 1. Генерировать частично бент-функции можно с помощью следующей теоремы.

Теорема 1 [4]. Функция $f \in P_2(n)$ является частично бент-функцией тогда и только тогда, когда существует невырожденная матрица A с элементами из Z_2 и вектор $\beta \in Z_2^n$ такие, что $f(xA \oplus \beta) = g(x_0) \oplus (x_1, t)$, где $x = (x_0\, x_1),\ x \in Z_2^n,\ x_0 \in Z_2^{2h},\ x_1 \in Z_2^{n-2h},\ t \in Z_2^{n-2h}$; g – бент-функция.

Опишем этот генератор и вспомогательный генератор невырожденной матрицы. Бент-функции будем строить с помощью конструкции Мейорана–МакФарланда.

Генератор невырожденной матрицы

Вход: $n \in N$.

Выход: невырожденная матрица A размерности $n \times n$.

1. Будем представлять матрицу A как множество строк. Возьмем в качестве ее первой строки произвольный ненулевой вектор e_1 длины n, $A = \{e_1\}$. Пусть $M = \{0, e_1\}$. Под M будем понимать линейную оболочку строк матрицы A.

2. Если в A есть n строк, т. е. $M = G$, то возвращаем A.

3. Берем произвольный вектор $e \in G \setminus M$.

4. $A = A \cup e,\ M = M \cup \{m \oplus e : m \in M\}$.

5. Идем на шаг 2.

Генератор частично бент-функций 1

Вход: $n \in N$, $0 \le h \le \lfloor n/2 \rfloor$.

Выход: $r \in P_2(n)$ – частично бент-функция или ответ «не существует».

1. Если $n = 2h$, то генерируем бент-функцию $r \in P_2(n)$ и возвращаем ее.

2. Генерируем бент-функцию $g(x_0)$ на Z_2^{2h} и произвольный вектор $t \in Z_2^{n-2h}$.

3. Для всех $x = (x_0\, x_1) \in G$ полагаем $f(x) = g(x_0) \oplus (x_1, t)$.

4. Генерируем невырожденную матрицу A и вектор $\beta \in Z_2^n$.

5. Для всех $x \in G$ полагаем $r(x) = f(xA \oplus \beta)$. Возвращаем r.

Замечание. Нелинейность функции $r(x)$ равна $N_r = 2^{n-1} - 2^{n-h-1}$. Такая же нелинейность у функций, которые выдает генератор 2. Это следует из теорем, которые приводятся в [5].

Определенные на подпространстве бент-функции. Следующие три определения взяты из [6].

Пусть $S \subset G$. Функция $f : S \to \{0,1\}$ называется частично определенной булевой функцией. Обозначать такие функции мы будем $f|_S$, а называть просто – функциями, определенными на S.

Неполным преобразованием Уолша–Адамара для f называется функция $W_f^S(\upsilon) = \sum_{u \in S} (-1)^{f(u)+(u,\upsilon)}$, определенная на G.

Функция f – частично определенная бент-функция, если для всех $\upsilon \in G$ выполняется $W_f^S(\upsilon) = \pm\sqrt{|S|}$. Мы будем называть f бент-функцией, определенной на S.

Пусть E' – подпространство G и f – функция, определенная на E'. Введем функцию $\Delta_f^{E'}(\upsilon) = \sum_{u \in E'} (-1)^{f(u)+f(u \oplus \upsilon)}$, определенную на E', и назовем ее функцией автокорреляции для f на подпространстве E' (определение из [7]).

Приведенное ниже утверждение 2 будет использовано в генераторе 2 частично бент-функций.

Утверждение 1 [7] (характеризация определённых на подпространстве бент-функций). Пусть E' – подпространство G с базисом $\{e_1, \ldots, e_{2h}\}$ и f – частично определенная функция на нем.

f – частично определенная бент-функция на E', если и только если для всех $s' \in E',\ s' \neq 0$ выполняется $\Delta_f^{E'}(s') = 0$.

Сформулируем и докажем следующее утверждение.

Утверждение 2 (о бент-функции, определённой на подпространстве). Пусть E' – подпространство G

с базисом $\{e_1, \ldots, e_{2h}\}$ и f – функция, определенная на E'.

Определим булеву функцию g от $2h$ переменных так:

$$\forall x = x_1 \ldots x_{2h} \in Z_2^{2h}, \quad g(x) = f(x_1 e_1 \oplus \ldots \oplus x_{2h} e_{2h}).$$

Тогда

$\forall s' \in E', s' \neq 0, \quad \Delta_f^{E'}(s') = 0 \quad \Leftrightarrow \quad g$ – бент-функция.

$$\triangleright \quad x' = x_1 e_1 \oplus \ldots \oplus x_{2h} e_{2h}, \; s' = s_1 e_1 \oplus \ldots \oplus s_{2h} e_{2h},$$

$$\Delta_f^{E'}(s') = \sum_{x' \in E'} (-1)^{f(x') + f(x' \oplus s')} = \sum_{x \in Z_2^{2h}} (-1)^{g(x) + g(x \oplus s)} = \Delta_g(s),$$

$$\forall s' \in E', \; s' \neq 0;$$

$$\Delta_f^{E'}(s') = 0 \quad \Leftrightarrow \quad \forall s \in Z_2^{2h}, \; s \neq 0;$$

$$\Delta_g(s) = 0. \quad \triangleleft$$

Как видим, утверждение 2 дает нам способ построения определенной на подпространстве бент-функции с помощью обычной бент-функции.

Генератор частично бент-функций 2. В [5] было впервые введено понятие частично бент-функции и сформулирована приведенная ниже теорема.

Теорема 2. Любая функция $f \in P_2(n)$ удовлетворяет неравенству $NW_f \cdot N\Delta_f \geq 2^n$.

Следующие утверждения эквивалентны:

а) функция f – частично бент-функция;

б) $\exists z \in G, \quad \forall s \in G, \; \Delta_f(s) = 0 \vee \Delta_f(s) = (-1)^{(s,z)} 2^n$;

в) G разлагается в прямую сумму подпространств $E = \{x \in G : \; \Delta_f(x) \neq 0\}$ *и* E' (E' четной размерности) так, что $f|_{E'}$ – бент-функция и $\forall x \in E, \quad \forall y \in E', \quad f(x \oplus y) = f(y) \oplus (x, z)$. Здесь z – любое из тех, которые могут быть выбраны в пункте (*б*).

Поскольку вектор z встречается и далее в теореме 3, назовем его базовым вектором аффинной части для частично бент-функции f.

Следующий генератор основан на теореме 2 и утверждении 2:

Вход: $n \in N$, $0 \leq h \leq \lfloor n/2 \rfloor$.

Выход: $f \in P_2(n)$, $N_f = 2^{n-1} - 2^{n-h-1}$ или ответ – «не существует».

1. Если $n = 2h$, то генерируем бент-функцию $f \in P_2(n)$ и возвращаем ее.

2. Генерируем бент-функцию $g \in P_2(2h)$ и произвольный вектор $z \in G$.

3. Генерируем базисы пространств E' и E, в прямую сумму которых раскладывается G. Это n линейно независимых булевых векторов длины n, $\{e_1, \ldots, e_{2h}\}$ – базис E', $\{e_{2h+1}, \ldots, e_n\}$ – базис E. Базисы получаем так: генерируем невырожденную матрицу A размерности $n \times n$, ее первые $2h$ строчек – базис E', остальные – базис E.

4. Для всех $x = (x_1, \ldots, x_n) \in Z_2^n$ полагаем $f(\overset{n}{\underset{i=1}{\oplus}} x_i e_i) = g(x_1, \ldots, x_{2h}) \oplus \overset{n}{\underset{i=2h+1}{\oplus}} x_i (e_i, z)$.

Анализ работы генераторов частично бент-функций. Время работы генераторов примерно одинаково, поскольку они похожи по построению. В генераторе 2 $f(\overset{n}{\underset{i=1}{\oplus}} x_i e_i) = f(xA)$, т. е. генерация невырожденной матрицы используется в обоих генераторах частично бент-функций и занимает 50…60 % всего времени их работы.

В дальнейшем имеет смысл использовать генератор 2, поскольку он позволяет явно задавать пространство E и вектор z, с помощью определенного выбора которых можно попытаться улучшить КИ и КР, задать сбалансированность. Для этого используется следующая теорема.

Теорема 3 [5]. Пусть f – частично бент-функция, z – вектор из пункта (*ii*) теоремы 2, $E = \{x \in G : \; \Delta_f(x) \neq 0\}$. Функция f является:

а) сбалансированной $\Leftrightarrow f|_E \neq \text{const}$;

б) несбалансированной $\Leftrightarrow f|_E = \text{const}$ и $w(f) = 2^{n-1} \pm 2^{n-h-1}$, где $\dim E = n - 2h$, $0 \leq h \leq \lfloor n/2 \rfloor$;

в) корреляционно-иммунной порядка $k \Leftrightarrow$ смежный класс $z \oplus E^{\perp}$ содержит только векторы веса больше k или нулевого;

г) сбалансированной корреляционно-иммунной порядка $k \Leftrightarrow$ класс $z \oplus E^{\perp}$ содержит только векторы веса больше k;

д) удовлетворяет критерию распространения РС(k) $\Leftrightarrow E$ не содержит векторов веса w, $1 \leq w \leq k$.

В этой теореме $E^{\perp} = \{x \in G : \forall e \in E \quad (x, e) = 0\}$. Используя теорему 2, можно пытаться улучшать порядок КИ и КР частично бент-функций.

Улучшение порядка КИ и КР. О работе генератора 2 без улучшений можно судить по следующей таблице (табл. 1).

В генераторе 2 есть возможность напрямую задавать для частично бент-функции базовый вектор аффинной части z и пространство линейных структур E. Это означает, что можно попытаться улучшить порядок КИ и КР для генерируемой функции, так как из теоремы 2 известно, что порядок КИ определяется смежным классом $z \oplus E^{\perp}$, а порядок КР – пространством E.

Начнем с критерия распространения. Выдаваемая функция удовлетворяет КР(k) тогда и только тогда, когда все ненулевые векторы пространства E имеют вес больше k. Нужно сконструировать E таким образом, чтобы число k было как можно больше.

Если $\dim E = n - 2h = 1$, т. е. базис пространства E состоит из единственного вектора, то мы берем в качестве этого вектора вектор из всех единиц. Он обеспечивает максимально возможный порядок КР, равный $2h$. Например, улучшенный генератор 2 выдает в этом случае (n = 15, h = 7) функцию с порядком КИ, равным единице, и порядком КР, равным 14.

Множество {1, 2, …, n} можно разбить на dim E непересекающихся множеств K_i с числом элементов, превосходящим или равным $\lfloor n/\dim E \rfloor$. С каждым множеством K_i сопоставим вектор длины n, в котором единицы стоят только в позициях из K_i. Очевидно, эти dim E векторов линейно независимы. Таким образом, всегда можно добиться порядка КР, равного $\lfloor n/\dim E \rfloor - 1$.

Пусть p – порядок КР. Начиная с p = 1 (или $\lfloor n/\dim E \rfloor - 1$, если это не нуль) и до $p = 2h$ пробуем строить пространство E, обеспечивающее КР(p), с помощью описанного ниже алгоритма; если построить E удается – увеличиваем p, иначе выходим из цикла и E остается тем, что было построено для предыдущего p.

Алгоритм построения пространства E

Вход: n, dim E, p – требуемый порядок КР.

Выход: базис E или ответ – «не найден».

1. Если dim E = 1, то возвращаем базис E – вектор из всех единиц.

2. Берем в качестве e_1 случайный вектор. Если он не подошел, т. е. его вес $\le p$, то генерируем случайное число t от p + 1 до n и пусть e_1 – это вектор, у которого первые t позиций заполнены единицами. Полагаем $M = \{0, e_1\}$, $K = G \setminus M$, i = 2. Здесь M – линейная оболочка построенных векторов базиса E, K – множество кандидатов на место e_i.

3. Строим e_i. Сначала пробуем на место e_i случайный вектор. Если он подходит, идем на шаг 4, иначе вычеркиваем этот вектор из K. Перебираем векторы из K с начала или с конца до тех пор, пока не найдем подходящий вектор; неподходящие векторы вычеркиваем. Если нашли подходящий вектор (его вес больше p, и он не приводит к появлению в M ненулевых векторов веса $\le p$), то идем на шаг 4, иначе ответ – «не найден».

4. $M = M \cup \{m \oplus e_i : m \in M\}$,

$K = K \setminus \{m \oplus e_i : m \in M\}$.

5. Если базис E размерности dim E еще не построен, то увеличиваем i на единицу и идем на шаг 3, иначе возвращаем базис E.

Функция с n = 15, h = 5 генерируется с использованием алгоритма построения E примерно за три секунды. Для n = 21, h = 7 время работы этого алгоритма неизвестно; оно слишком велико. Ниже приводятся характеристики функций, построенных с применением этого алгоритма (табл. 2).

Как видим, порядок КР действительно улучшился. Функции, для которых он был меньше четырех, перестали выдаваться вообще.

После построения пространства E находим для него произвольное прямое дополнение, это будет пространство E'. Функцию вычисления прямого дополнения нетрудно реализовать на основе предложенного ранее генератора невырожденной матрицы.

Вычисление прямого дополнения E′ пространства E

Вход: E, dim E.

Выход: базис E'.

1. $M = E$, i = 1.

2. Если $i > n - \dim E$, то возвращаем $\{e_1, \ldots, e_{n-\dim E}\}$ – базис E'.

3. Берем в качестве e_i произвольный вектор из $G \setminus M$.

4. Расширяем M: $M = M \cup \{m \oplus e_i : m \in M\}$, $i = i +$ 1. Идем на шаг 2.

Этот алгоритм, так же как и генератор невырожденной матрицы, можно немного улучшить: если мы нашли последний вектор, то линейную оболочку остальных векторов M расширять не нужно, поскольку она больше не понадобится.

Таблица 1

1 000 функций, выданных генератором 2: n = 15, h = 5; сбалансированных – 966

	не РС(1)	РС(1)	РС(2)	РС(3)	РС(4)	РС(5)
не КИ(1)	6	27 (3)	107 (8)	170 (6)	42	3
КИ(1)	–	45	213 (5)	253 (9)	60 (3)	–
КИ(2)	3	11	14	9	–	–

Примечание. Здесь и далее в таблицах: без скобок – число сбалансированных функций, в скобках – несбалансированных.

Таблица 2

1 000 функций, выданных генератором 2 с улучшенным КР: n = 15, h = 5; сбалансированных – 966

	не КР(1)	КР(1)	КР(2)	КР(3)	КР(4)	КР(5)	КР(6)
не КИ(1)	–	–	–	–	–	338 (2)	109
КИ(1)					4 (1)	391 (21)	102
КИ(2)					–	16 (2)	6
КИ(3)					–	–	(8)

Из базисов E и E' составляется матрица A (см. генератор 2).

Теперь попробуем улучшить порядок КИ. На пространство $E^{\perp}$ повлиять не можем, так как уже построили E; $E^{\perp}$ вычисляем как множество векторов из G, каждый из которых ортогонален всем векторам базиса E. Осталось перебрать все значения z и выбрать среди них то, которое обеспечивает самый высокий порядок КИ. Пишем вспомогательную функцию, которая по z и $E^{\perp}$ определяет, какой порядок КИ обеспечивает смежный класс $z \oplus E^{\perp}$. Для этого проверяем, принадлежат ли $z \oplus E^{\perp}$ векторы веса 1, 2 и т. д.

Можно перебирать все значения z из G, однако это необязательно, поскольку $z_1 \oplus E^{\perp} = z_2 \oplus E^{\perp}$, если и только если $z_2 \oplus z_1 \in E^{\perp}$. Очевидно, достаточно перебрать все значения z из пространства E_1 – прямого дополнения $E^{\perp}$; $\dim E_1 = n - \dim E^{\perp} = \dim E = n - 2h$. Как находить прямое дополнение, нам известно. Базис $E^{\perp}$ находится примерно так же, как базис G в генераторе невырожденной матрицы. Для тех значений n и h, которые мы использовали ранее, перебор z не занимает существенного времени – функция генерируется примерно за то же время. О влиянии улучшения z на характеристики генерируемой функции можно судить по приведенным ниже таблицам (табл. 3, 4). Большое число несбалансированных функций в них объясняется тем, что перебор значений z начинается с нуля, а остальные значения z улучшения КИ не дают.

Также была сделана попытка одновременного улучшения КИ и КР.

Выясним, как можно улучшить порядок КИ функции с помощью аффинной добавки.

Влияние аффинной добавки на значения преобразования Уолша–Адамара такое: если $\tilde{f}(x) = f(x) \oplus (x,t) \oplus t_0$, где $t \in G, t_0 \in \{0,1\}$, то $W_{\tilde{f}}(s) = W_f(s \oplus t)$, т. е. аффинная добавка приводит к сдвигу значений преобразования Уолша–Адамара. Пусть f является частично бент-функцией, тогда

$$W_{\tilde{f}}(s) \neq 0 \Leftrightarrow W_f(s \oplus t) \neq 0 \Leftrightarrow$$

$$\Leftrightarrow s \oplus t \in z \oplus E^{\perp} \Leftrightarrow s \in z \oplus t \oplus E^{\perp},$$

т. е. аффинная добавка к частично бент-функции приводит лишь к изменению z. Улучшить КИ частично бент-функции с помощью аффинной добавки – это значит перебрать все значения вектора z. Как это можно сделать, мы уже знаем.

Предложенный ниже алгоритм может применяться для улучшения КИ любой булевой функции. На практике он не исследовался.

Алгоритм улучшения порядка КИ функции с помощью аффинной добавки

Вход: n, $f \in P_2(n)$.

Выход: $t \in G$ или ответ – «нельзя улучшить».

1. Находим вес функции f. Если f является константой или ее вес нечетный, то ответ – «нельзя улучшить».

2. Находим все значения преобразования Уолша–Адамара функции f, вычисляем множество $E = \{s \in G : W_f(s) \neq 0\}$ и k – порядок КИ функции f.

3. Если функция несбалансированная, то вычисляем максимальное m такое, что для всех $s \in G$ выполняется $2^{m+1} \mid W_f(s)$, иначе вычисляем максимальное m такое, что для всех $s \in G$ выполняется $2^{m+2} \mid W_f(s)$. По теореме о необходимом условии КИ [1] мы не сможем получить порядок КИ выше m. Если $k = m$, то ответ – «нельзя улучшить».

4. $i = 1$. $T = G$. T – это множество возможных значений вектора t.

5. Нам известно, что для функции $\tilde{f}(x) = f(x) \oplus (x,t) \oplus t_0, t_0 \in \{0,1\}$ множество $t \oplus E = \{s \in G : W_{\tilde{f}}(s) \neq 0\}$. Для каждого $t \in T$ и каждого $j \in \{j \in G : w(j) = i\}$ проверяем: если $t \oplus j \in E$, то $T = T \setminus t$. В итоге в T остаются только те значения t, которые обеспечивают КИ(i) для $\tilde{f}$. Если T не пусто, то берем любое $t \in T$, в противном случае поступаем так: если $i = 1$ или $i - 1 \leq k$, то ответ – «нельзя улучшить», иначе возвращаем t.

6. Если $i = m$, то возвращаем t. $i = i + 1$. Идем на шаг 4.

Таблица 3

1 000 функций, выданных генератором 2 с улучшенным КИ: $n = 15$, $h = 5$; сбалансированных – 328

	не КР(1)	КР(1)	КР(2)	КР(3)	КР(4)	КР(5)	КР(6)
не КИ(1)	–						
КИ(1)	–	–	(142)	48 (291)	44 (190)	–	–
КИ(2)	–	–	189 (49)	47	–	–	–

Таблица 4

1 000 функций, выданных генератором 2 с улучшенным КИ и КР: $n = 15$, $h = 5$; сбалансированных – 277

	не КР(1)	КР(1)	КР(2)	КР(3)	КР(4)	КР(5)	КР(6)
не КИ(1)	–	–	–	–	–	–	–
КИ(1)					(1)	8 (435)	–
КИ(2)					1	268 (51)	–
КИ(3)					–	–	(236)

Если этот алгоритм выдает t, то функция $\tilde{f}$ будет иметь более высокий порядок КИ, нежели f. Известно, что аффинная добавка никак не влияет на порядок КР.

Итак, изложены результаты реализации и исследования двух генераторов частично бент-функций, описаны возможности по улучшению порядков КИ и КР таких функций. Работоспособность алгоритмов проверена на практике, изучены свойства генерируемых функций.

Автор выражает благодарность И. А. Панкратовой и К. В. Сафонову за постоянное внимание к работе и ценные замечания.

Библиографические ссылки

1. Carlet C. Boolean Functions for Cryptography and Error Correcting Codes [Electronic resource] // INRIA. 2010. URL: http://www-rocq.inria.fr/codes/Claude.Carlet/chap-fcts-Bool-corr.pdf. (дата обращения: 08.06.2010).

2. Токарева Н. Н. Бент-функции: результаты и приложения. Обзор работ // Прикл. дискретная математика. 2009. № 1 (3). С. 15–36.

3. Bent-function [Electronic resource] // Wikipedia. 2010. URL: http://en.wikipedia.org/wiki/Bent_function (дата обращения: 08.06.2010).

4. Zheng Y., Zhang X. M. Plateaued functions // ICICS'99. Lecture Notes in Computer Science. 1999. Vol. 1726. P. 284–300.

5. Carlet C. Partially-bent functions // Design, Codes and Cryptography. 1993. Vol. 3. №. 2. P. 135–145.

6. Токарева Н. Н. Обобщения бент-функций. Обзор работ // Дискрет. анализ и исследование операций. 2010. Т. 17. № 1. С. 34–64.

7. Агафонова И. В. Криптографические свойства нелинейных булевых функций [Electronic resource] // Семинар по дискретному гармоническому анализу и геометрическому моделированию DHA & CAGD. 2007. URL: http://dha.spb.ru/PDF/cryptoBOOLEAN.pdf (дата обращения: 08.06.2010).

M. V. Naumov

THE GENERATION OF PARTIALLY-BENT FUNCTIONS

Most important characteristics of cryptographic functions are balancedness, nonlinearity, propagation criterion, correlation immunity, degree and non-existence of nonzero linear structure. Partially-bent functions form super-class of the class of bent functions. These functions may achieve desirable characteristics.

Two algorithms for generation of partially-bent functions were supposed and studied. Second algorithm may improve cryptographic characteristics of generated functions.

Keywords: Boolean functions, partially-bent functions, correlation immunity.

© Наумов М. В., 2010

УДК 519.62

В. А. Нестеров

ПОПЕРЕЧНЫЕ КОЛЕБАНИЯ ПЛАСТИНЫ, ПОДАТЛИВОЙ ПРИ ТРАНСВЕРСАЛЬНОМ СДВИГЕ

Рассматривается конечно-элементный модальный расчет пластины с низкой трансверсальной сдвиговой жесткостью. В каждом из четырех узлов прямоугольного конечного элемента пластины в качестве основных кинематических параметров присутствуют углы трансверсального сдвига. На примере анализа собственных колебаний композитной пластины показана актуальность разработанной конечно-элементной модели. Представлены результаты численного исследования влияния граничных условий неклассического вида на величины частот собственных колебаний.

Ключевые слова: пластина, метод конечных элементов, трансверсальный сдвиг, модальный анализ.

Композитные конструкции, обладающие высокой степенью весового совершенства, часто используются в производстве космической техники. Композитные пластины отличаются низкой сдвиговой жесткостью по отношению к трансверсальным напряжениям. Учет указанной особенности при реализации численного расчета приводит к повышению порядка разрешающих уравнений за счет введения в рассмотрение углов трансверсального сдвига.

Разрешающие уравнения теории метода конечных элементов (МКЭ) для задачи о собственных колебаниях пластины с низкой трансверсальной сдвиговой жесткостью получим вариационным способом. Для этого запишем выражение полной энергии колеблющейся пластины

$$\tilde{E} = \tilde{U} + \tilde{T}, \qquad (1)$$

где $\tilde{U}$ – потенциальная энергия деформации; $\tilde{T}$ – кинематическая энергия движения.

Выражение для потенциальной энергии деформации пластины как трехмерного тела имеет вид

$$\tilde{U} = \frac{1}{2}\int\limits_{(V)}\int\int (\tilde{\sigma}_x\tilde{e}_x + \tilde{\sigma}_y\tilde{e}_y + \tilde{\sigma}_z\tilde{e}_z + \\ + \tilde{\tau}_{xy}\tilde{e}_{xy} + \tilde{\tau}_{xz}\tilde{e}_{xz} + \tilde{\tau}_{yz}\tilde{e}_{yz})\, dxdydz, \qquad (2)$$

где напряжения и деформации являются функциями координат и времени

$$\tilde{\sigma} = \tilde{\sigma}(x, y, z, t) \;; \qquad \tilde{e} = \tilde{e}(x, y, z, t), \qquad (3)$$

где $\tilde{\sigma}$ – любой параметр из $\tilde{\sigma}_x, \tilde{\sigma}_y, \tilde{\sigma}_z, \tilde{\tau}_{xy}, \tilde{\tau}_{yz}, \tilde{\tau}_{xz}$; $\tilde{e}$ – любой параметр из $\tilde{e}_x, \tilde{e}_y, \tilde{e}_z, \tilde{e}_{xy}, \tilde{e}_{yz}, \tilde{e}_{xz}$.

Геометрические соотношения имеют вид

$$\tilde{e}_x = \frac{\partial\tilde{u}_x}{\partial x};\;\; \tilde{e}_y = \frac{\partial\tilde{u}_y}{\partial y};\;\; \tilde{e}_z = \frac{\partial\tilde{u}_z}{\partial z};\;\; \tilde{e}_{xy} = \frac{\partial\tilde{u}_x}{\partial y} + \frac{\partial\tilde{u}_y}{\partial x};$$
$$\tilde{e}_{yz} = \frac{\partial\tilde{u}_y}{\partial z} + \frac{\partial\tilde{u}_z}{\partial y};\;\; \tilde{e}_{xz} = \frac{\partial\tilde{u}_x}{\partial z} + \frac{\partial\tilde{u}_z}{\partial x}, \qquad (4)$$

где $\tilde{u}_x, \tilde{u}_y, \tilde{u}_z$ – проекции перемещения произвольной точки на соответствующие оси координат. Они также являются функциями координат и времени:

$$\tilde{u}_x = \tilde{u}_x(x, y, z, t); \qquad \tilde{u}_y = \tilde{u}_y(x, y, z, t);$$
$$\tilde{u}_z = \tilde{u}_z(x, y, z, t). \qquad (5)$$

При гармонических колебаниях закон изменения перемещений по времени можно представить в следующем виде:

$$\tilde{u}_x = u_x \sin\omega t; \qquad \tilde{u}_y = u_y \sin\omega t; \qquad \tilde{u}_z = u_z \sin\omega t, \qquad (6)$$

где ω – круговая частота колебаний, а перемещения u_x, u_y, u_z зависят только от координат:

$$u_x = u_x(x, y, z); \;\; u_y = u_y(x, y, z); \;\; u_z = u_z(x, y, z). \qquad (7)$$

Подставляя (5) в (4), получим:

$$\tilde{e} = e\sin\omega t, \qquad (8)$$

где e – любой параметр из e_x, e_y, e_z, e_{xy}, e_{yz}, e_{xz}, для которых справедливы геометрические соотношения

$$e_x = \frac{\partial u_x}{\partial x};\;\; e_y = \frac{\partial u_y}{\partial y};\;\; e_z = \frac{\partial u_z}{\partial z};\;\; e_{xy} = \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x};$$
$$e_{yz} = \frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y};\;\; e_{xz} = \frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}. \qquad (9)$$

В силу линейности физических соотношений можно записать

$$\tilde{\sigma} = \sigma\sin\omega t, \qquad (10)$$

где σ – любой из параметров σ_x, σ_y, σ_z, τ_{xy}, τ_{yz}, τ_{xz}, для которых справедливы физические соотношения

$$e_x = \frac{\sigma_x}{E_x} - \mu_{xy}\frac{\sigma_y}{E_y} - \mu_{xz}\frac{\sigma_z}{E_z};\;\; e_y = \frac{\sigma_y}{E_y} - \mu_{yx}\frac{\sigma_x}{E_x} - \mu_{yz}\frac{\sigma_z}{E_z};$$
$$e_z = \frac{\sigma_z}{E_z} - \mu_{zx}\frac{\sigma_x}{E_x} - \mu_{zy}\frac{\sigma_y}{E_y};\;\; e_{xy} = \frac{\tau_{xy}}{G_{xy}}; \qquad (11)$$
$$e_{yz} = \frac{\tau_{yz}}{G_{yz}}; \qquad e_{xz} = \frac{\tau_{xz}}{G_{xz}},$$

где $E_{x(y,z)}$ – модуль упругости соответствующего направления; $G_{xy(yz,xz)}$ – модуль сдвига в соответствующей плоскости; μ_{xy}, μ_{yx}, μ_{yz}, μ_{zy}, μ_{xz}, μ_{zx} – коэффициенты Пуассона.

Имеет место свойство симметрии упругих постоянных:

$$E_x\mu_{xy} = E_y\mu_{yx}; \quad E_y\mu_{yz} = E_z\mu_{zy}; \quad E_x\mu_{xz} = E_z\mu_{zx}. \qquad (12)$$

Используя принятые в [1] допущения, выражение потенциальной энергии деформации (2) можно привести к виду

$$\tilde{U} = U\sin^2\omega t, \qquad (13)$$

где U определяется выражением

$$U = \frac{1}{2}\int\limits_{(V)}\int\int \left\{\sigma_x\left(\frac{\partial u}{\partial x} + z\frac{\partial\theta_x}{\partial x}\right) + \sigma_y\left(\frac{\partial v}{\partial y} + z\frac{\partial\theta_y}{\partial y}\right) + \right.$$
$$+\tau_{xy}\left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} + z\left(\frac{\partial\theta_x}{\partial y} + \frac{\partial\theta_y}{\partial x}\right)\right] + \qquad (14)$$
$$\left. + \tau_{xz}\psi_x + \tau_{yz}\psi_y\right\} dxdydz.$$

Выражение кинетической энергии колебания пластины имеет вид

$$\tilde{T} = \frac{1}{2}\int\limits_{(V)}\int\int \left(\tilde{v}_x^2 + \tilde{v}_y^2 + \tilde{v}_z^2\right)\rho\, dxdydz, \qquad (15)$$

где ρ – плотность материала (в общем случае является функцией координат); $\tilde{v}_x$, $\tilde{v}_y$, $\tilde{v}_z$ – проекции вектора скорости на соответствующие оси координат, связанные с перемещениями дифференциальными зависимостями

$$\tilde{v}_x = \frac{\partial\tilde{u}_x}{\partial t}; \qquad \tilde{v}_y = \frac{\partial\tilde{u}_y}{\partial t}; \qquad \tilde{v}_z = \frac{\partial\tilde{u}_z}{\partial t}. \qquad (16)$$

Подставляя (5) в (16), а затем результат – в (15), получим

$$\tilde{T} = T\cos^2\omega t, \qquad (17)$$

где

$$T = \frac{1}{2}\int\limits_{(V)}\int\int \left(u_x^2 + u_y^2 + u_z^2\right)\rho\,\omega^2\, dxdydz. \qquad (18)$$

Приравнивая максимальные значения потенциальной и кинетической энергии, которые определяются выражениями (14) и (18), получим

$$\frac{1}{2}\int\limits_{(V)}\int\int \left\{\sigma_x\left(\frac{\partial u}{\partial x} + z\frac{\partial\theta_x}{\partial x}\right) + \sigma_x\left(\frac{\partial v}{\partial y} + z\frac{\partial\theta_y}{\partial y}\right) + \right.$$
$$\left. + \tau_{xy}\left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} + z\left(\frac{\partial\theta_x}{\partial y} + \frac{\partial\theta_y}{\partial x}\right)\right] + \tau_{xz}\psi_x + \tau_{yz}\psi_y\right\} \times$$

$$\times dxdydz - \frac{1}{2}\int\limits_{(V)}\int\int \rho\,\omega^2\left[\left(u+z\theta_x\right)^2 + \left(v+z\theta_y\right)^2 + w^2\right]dxdydz = 0. \quad (19)$$

При выводе (19) учтено, что

$$u_x = u + z\theta_x\ ;\ u_y = v + z\theta_y\ ;\ u_z = w, \quad (20)$$

где u и v – перемещения точек начальной плоскости вдоль осей X и Y соответственно; w – прогиб точек начальной плоскости; θ_x, θ_y – углы наклона сечения, определяемые соотношениями

$$\theta_x = \psi_x - \frac{\partial w}{\partial x};\quad \theta_y = \psi_y - \frac{\partial w}{\partial y}. \quad (21)$$

Проинтегрируем (19) по толщине пластины, т. е. по координате z в пределах от $-s$ до $h - s$. В результате преобразований получим

$$\frac{1}{2}\int_0^a\int_0^b\left(N_x\varepsilon_x + N_y\varepsilon_y + N_{xy}\varepsilon_{xy} + M_x\chi_x + M_y\chi_y + M_{xy}\chi_{xy} + Q_x\psi_x + Q_y\psi_y\right)dxdy - \frac{1}{2}\omega^2\int_0^a\int_0^b\left\{B_\rho\left(u^2+v^2+w^2\right) + 2C_\rho\left[u\left(\psi_x - \frac{\partial w}{\partial x}\right) + v\left(\psi_y - \frac{\partial w}{\partial y}\right)\right] + D_\rho\left[\left(\psi_x - \frac{\partial w}{\partial x}\right)^2 + \left(\psi_y - \frac{\partial w}{\partial y}\right)^2\right]\right\}dxdy = 0. \quad (22)$$

Здесь N, M, Q – внутренние погонные усилия, определяемые следующими соотношениями:

$$\begin{aligned} N_x &= B_{11}\varepsilon_x + B_{12}\varepsilon_y + C_{11}\chi_x + C_{12}\chi_y\ ;\\ N_y &= B_{21}\varepsilon_x + B_{22}\varepsilon_y + C_{21}\chi_x + C_{22}\chi_y\ ;\\ N_{xy} &= B_{33}\varepsilon_{xy} + C_{33}\chi_{xy}\ ;\ Q_x = K_x\,\psi_x\ ;\ Q_y = K_y\,\psi_y;\\ M_{xy} &= C_{33}\varepsilon_{xy} + D_{33}\chi_{xy}\ ;\\ M_x &= C_{11}\varepsilon_x + C_{12}\varepsilon_y + D_{11}\chi_x + D_{12}\chi_y\ ;\\ M_y &= C_{21}\varepsilon_x + C_{22}\varepsilon_y + D_{21}\chi_x + D_{22}\chi_y\ , \end{aligned} \quad (23)$$

где B, C, D и K – параметры мембранной, смешанной, изгибной и трансверсальной жесткости соответственно; B_ρ, C_ρ, D_ρ – параметры инерции пластины, вычисляемые по формулам

$$B_\rho = \int_{-s}^{h-s}\rho\,dz\ ;\ C_\rho = \int_{-s}^{h-s}\rho\,z\,dz\ ;\quad D_\rho = \int_{-s}^{h-s}\rho\,z^2\,dz. \quad (24)$$

Сделаем последние преобразования в функционале (22). Приведем его к виду, удобному для реализации КЭ процедуры. Для этого с помощью физических соотношений (23) выразим усилия через деформации, которые также выразим через перемещения и углы трансверсального сдвига с помощью геометрических соотношений

$$\varepsilon_x = \frac{\partial u}{\partial x}\ ;\ \varepsilon_y = \frac{\partial v}{\partial y}\ ;\ \varepsilon_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\ ;\ \chi_x = \frac{\partial \theta_x}{\partial x}\ ;$$

$$\chi_y = \frac{\partial \theta_y}{\partial y}\ ;\ \chi_{xy} = \frac{\partial \theta_x}{\partial y} + \frac{\partial \theta_y}{\partial x}. \quad (25)$$

В итоге с учетом (21) получим

$$\begin{aligned} &\int_0^a\int_0^b\left\{B_{11}\left(\frac{\partial u}{\partial x}\right)^2 + 2C_{11}\frac{\partial u}{\partial x}\left(\frac{\partial \psi_x}{\partial x} - \frac{\partial^2 w}{\partial x^2}\right) + D_{11}\left(\frac{\partial \psi_x}{\partial x} - \frac{\partial^2 w}{\partial x^2}\right)^2 + \right.\\ &+2B_{12}\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} + 2C_{12}\frac{\partial u}{\partial x}\left(\frac{\partial \psi_y}{\partial y} - \frac{\partial^2 w}{\partial y^2}\right) + 2C_{12}\frac{\partial v}{\partial y}\left(\frac{\partial \psi_x}{\partial x} - \frac{\partial^2 w}{\partial x^2}\right) +\\ &+2D_{12}\left(\frac{\partial \psi_x}{\partial x} - \frac{\partial^2 w}{\partial x^2}\right)\left(\frac{\partial \psi_y}{\partial y} - \frac{\partial^2 w}{\partial y^2}\right) + B_{22}\left(\frac{\partial v}{\partial y}\right)^2 +\\ &+2C_{22}\frac{\partial v}{\partial y}\left(\frac{\partial \psi_y}{\partial y} - \frac{\partial^2 w}{\partial y^2}\right) + D_{22}\left(\frac{\partial \psi_y}{\partial y} - \frac{\partial^2 w}{\partial y^2}\right)^2 +\\ &+B_{33}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 + 2C_{33}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\left(\frac{\partial \psi_x}{\partial x} + \frac{\partial \psi_y}{\partial y} - 2\frac{\partial^2 w}{\partial x \partial y}\right) +\\ &\left.+D_{33}\left(\frac{\partial \psi_x}{\partial x} + \frac{\partial \psi_y}{\partial y} - 2\frac{\partial^2 w}{\partial x \partial y}\right)^2 + K_x\psi_x^2 + K_y\psi_y^2\right\}dxdy -\\ &-\omega^2\int_0^a\int_0^b\left\{B_\rho\left(u^2+v^2+w^2\right) + 2C_\rho\left[u\left(\psi_x - \frac{\partial w}{\partial x}\right) + v\left(\psi_y - \frac{\partial w}{\partial y}\right)\right] +\right.\\ &\left.+D_\rho\left[\left(\psi_x - \frac{\partial w}{\partial x}\right)^2 + \left(\psi_y - \frac{\partial w}{\partial y}\right)^2\right]\right\}dxdy = 0. \end{aligned} \quad (26)$$

Функционал (26) позволяет получить разрешающие уравнения для задачи о свободных колебаниях пластины.

Выполним модальный расчет пластины с помощью МКЭ. Будем рассматривать четырехузловой конечный элемент пластины, для которого вектор узловых кинематических параметров имеет вид

$$\boldsymbol{\delta}_e = \{\,\boldsymbol{\delta_1}\ \boldsymbol{\delta_2}\ \boldsymbol{\delta_3}\ \boldsymbol{\delta_4}\}^T, \quad (27)$$

где $\boldsymbol{\delta_i}$ (I = 1, 2, 3, 4) – вектор неизвестных i-го узла

$$\boldsymbol{\delta_i} = \left\{w_i \quad \left(\frac{\partial w}{\partial x}\right)_i \quad \left(\frac{\partial w}{\partial y}\right)_i \quad \psi_{xi} \quad \psi_{yi} \quad u_i \quad v_i\right\}^T. \quad (28)$$

Кинематические переменные внутри элемента представим следующим вектором

$$\boldsymbol{\delta} = \left\{w \quad \frac{\partial w}{\partial x} \quad \frac{\partial w}{\partial y} \quad \psi_x \quad \psi_y \quad u \quad v\right\}^T. \quad (29)$$

Его компоненты определяются через узловые значения (28) с помощью соотношения

$$\boldsymbol{\delta} = \boldsymbol{P}\,\boldsymbol{\delta}_e, \quad (30)$$

где $\boldsymbol{P}$ – матрица размерностью 7 × 28. Это так называемая матрица функций формы.

Подставим выражения для компонент вектора $\boldsymbol{\delta}$ (29) в функционал (26), выполним интегрирование по площади элемента, в результате получим функцию полной энергии колебания конечного элемента пластины. Эта функция зависит от компонент вектора узловых кинематических параметров (28), минимизация по которым приводит к системе однородных алгебраических уравнений:

$$\boldsymbol{K_e}\,\boldsymbol{\delta_e} - \omega^2\,\boldsymbol{S_e}\,\boldsymbol{\delta_e} = \mathbf{0}, \qquad (31)$$

где $\boldsymbol{K_e}$ – матрица жесткости; $\boldsymbol{S_e}$ – матрица инерциальных параметров конечного элемента пластины (обе – симметричные, размером 28×28).

В выражениях для компонентов матрицы $\boldsymbol{K_e}$ фигурируют параметры жесткости, а в выражениях для компонентов матрицы $\boldsymbol{S_e}$ – параметры инерции. Например:

$$k_{11} = \frac{2}{5}\,\frac{14D_{33}a^2b^2 + 5D_{12}b^2a^2 + 10D_{11}b^4 + 10D_{22}a^4}{a^3b^3};$$

$$s_{11} = \frac{1\,727B_\rho b^2a^2 + 5\,520D_\rho b^2 + 5\,520D_\rho a^2}{12\,600\,a\,b}.$$

Если структура пакета пластины симметрична относительно срединной плоскости и система координат связана со срединной плоскостью, то в этом случае смешанные жесткости пластины C и инерционный параметр C_ρ становятся равным нулю, и структура матрицы инерции упрощается за счет появления в ней множества нулевых компонентов.

Представим матрицу жесткости и матрицу инерции конечного элемента пластины в блочном виде:

$$\boldsymbol{K_e} = \begin{bmatrix} \boldsymbol{K_{I-I}} & \boldsymbol{K_{I-II}} & \boldsymbol{K_{I-III}} & \boldsymbol{K_{I-IV}} \\ \boldsymbol{K_{II-I}} & \boldsymbol{K_{II-II}} & \boldsymbol{K_{II-III}} & \boldsymbol{K_{II-IV}} \\ \boldsymbol{K_{III-I}} & \boldsymbol{K_{III-II}} & \boldsymbol{K_{III-III}} & \boldsymbol{K_{II-IV}} \\ \boldsymbol{K_{IV-I}} & \boldsymbol{K_{IV-II}} & \boldsymbol{K_{IV-III}} & \boldsymbol{K_{IV-IV}} \end{bmatrix};$$

$$\boldsymbol{S_e} = \begin{bmatrix} \boldsymbol{S_{I-I}} & \boldsymbol{S_{I-II}} & \boldsymbol{S_{I-III}} & \boldsymbol{S_{I-IV}} \\ \boldsymbol{S_{II-I}} & \boldsymbol{S_{II-II}} & \boldsymbol{S_{II-III}} & \boldsymbol{S_{II-IV}} \\ \boldsymbol{S_{III-I}} & \boldsymbol{S_{III-II}} & \boldsymbol{S_{III-III}} & \boldsymbol{S_{III-IV}} \\ \boldsymbol{S_{IV-I}} & \boldsymbol{S_{IV-II}} & \boldsymbol{S_{IV-III}} & \boldsymbol{S_{IV-IV}} \end{bmatrix}; \qquad (32)$$

где $\boldsymbol{K_{i-j}}$ и $\boldsymbol{S_{i-j}}$ ($\boldsymbol{i,j}$ = **I, II, III, IV**) – подматрицы размером 7 × 7. Ненулевые компоненты этих матриц определяются теми же выражениями, что и соответствующие компоненты матриц жесткости и инерции, фигурирующих в уравнениях системы (31).

Глобальная система разрешающих уравнений в задаче модального расчета пластины имеет вид

$$\boldsymbol{K_\Sigma}\,\boldsymbol{\Delta} - \omega^2\,\boldsymbol{S_\Sigma}\,\boldsymbol{\Delta} = \mathbf{0}, \qquad (33)$$

где $\boldsymbol{\Delta}$ – глобальный вектор узловых неизвестных,

$$\boldsymbol{\Delta} = \{\boldsymbol{\delta_1}\ \boldsymbol{\delta_2}\ \ldots\ \boldsymbol{\delta_i}\ \ldots\ \boldsymbol{\delta_N}\}^T, \qquad (34)$$

где N – общее число узлов в системе; $\boldsymbol{K_\Sigma}$ – глобальная матрица жесткости; $\boldsymbol{S_\Sigma}$ – глобальная матрица инерции.

Система (33) в математическом смысле является обобщенной задачей на собственные значения, решая которую (после учета в ней граничных условий), определим частоты и формы собственных колебаний пластины.

Протестируем полученный алгоритм решения на примере модального расчета изотропной квадратной пластины, жестко защемленной по контуру. Зададим следующие размеры пластины и механические свойства материала: длина стороны пластины $a = b = 1$ м, толщина пластины $h = 3$ мм, модуль упругости $E = 210$ ГПа, коэффициент Пуассона $\mu = 0{,}3$, плотность $\rho = 7\,800$ кг/м^3.

В результате расчета, выполненного по нашей модели с двадцатью элементами вдоль каждой из сторон пластины, определены первые пять значений собственных частот (табл. 1) и соответствующие им формы колебаний. Здесь в качестве формы колебаний будем рассматривать распределение функции прогибов. Для оценки полученных результатов выполнен расчет в пакете COSMOS/M, где для моделирования использованы элементы тонкой оболочки SHELL4.

Таблица 1

Значения частот собственных колебаний

Номер моды	Частоты собственных колебаний, Гц	
	Тестируемое решение	Решение COSMOS/M
1	26,908	26,976
2	54,831	55,017
3	54,831	55,017
4	80,483	81,112
5	98,287	98,630

Результаты модального расчета, выполненного с помощью нашей модели (рис. 1), имеют хорошее совпадение с решением в пакете COSMOS/M (рис. 2). Это касается и частот собственных колебаний (см. табл. 1) и форм мод (рис. 3–6).

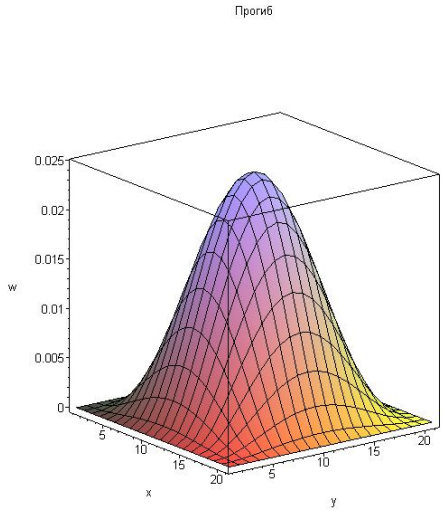


Рис. 1. Форма колебаний, соответствующая первой собственной частоте

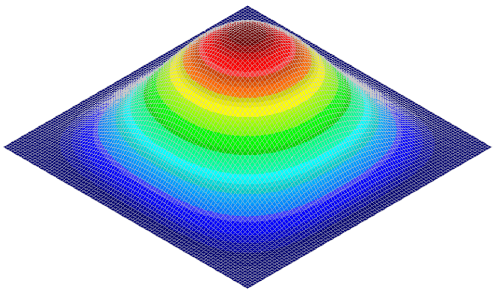


Рис. 2. Форма колебаний, соответствующая первой собственной частоте (COSMOS/M)

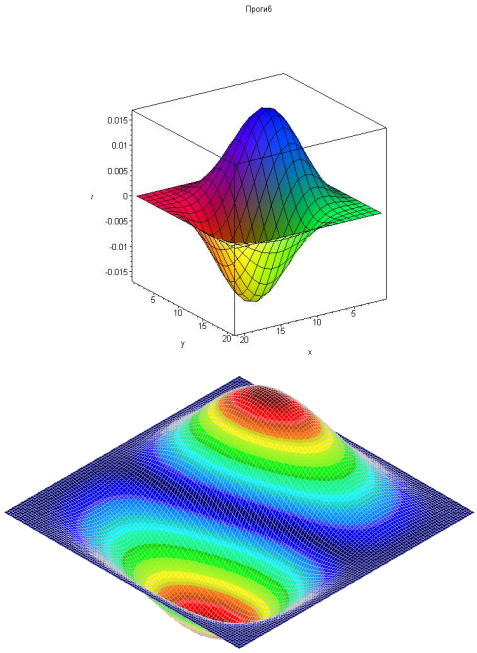


Рис. 3. Форма колебаний, соответствующая второй собственной частоте

Выполним численное исследование влияния вида граничных условий на величины частот собственных колебаний квадратной композитной пластины. Будем сравнивать два вида защемления: классическое (I) и со свободным сдвигом (II). Примем длины сторон пластины $a = b = 1$ м, а ее толщину $h = 10$ мм.

Будем полагать пластину изготовленной из однонаправленного композита с углами укладки $\varphi = \pm 45^{\circ}$. Зададим следующие механические свойства (углепластик):

– модуль упругости вдоль волокон $E_1 = 180$ ГПа;
– модуль упругости поперек волокон $E_2 = 6{,}2$ ГПа;
– модуль сдвига $G_{12} = 5$ ГПа;
– коэффициент Пуассона $\mu_{12} = 0{,}007$;
– плотность $\rho = 1\,500$ кг/м^3.

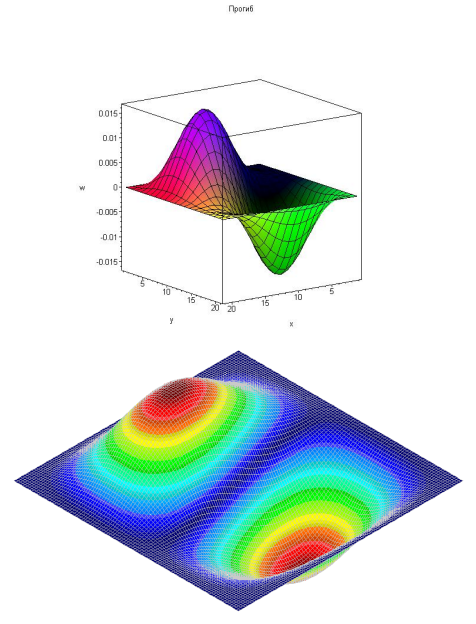


Рис. 4. Форма колебаний, соответствующая третьей собственной частоте

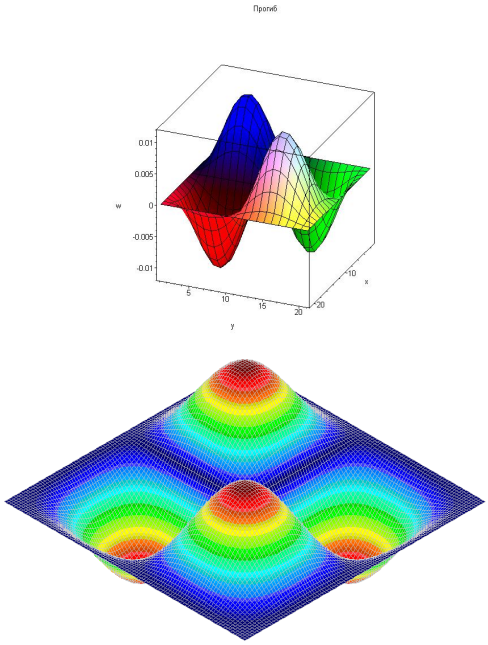


Рис. 5. Форма колебаний, соответствующая четвертой собственной частоте

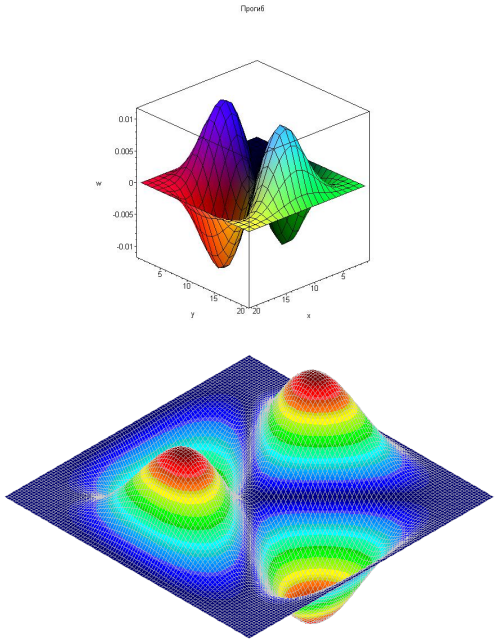


Рис. 6. Форма колебаний, соответствующая пятой собственной частоте

Упругие параметры пластины, которую будем считать однослойной и условно однородной, вычислим по формулам

$$A_{11} = \overline{E_1}\cos^4\varphi + \overline{E_2}\sin^4\varphi + 2\left(\overline{E_1}\mu_{12} + 2G_{12}\right)\sin^2\varphi\cos^2\varphi;$$

$$A_{22} = \overline{E_1}\sin^4\varphi + \overline{E_2}\cos^4\varphi + 2\left(\overline{E_1}\mu_{12} + 2G_{12}\right)\sin^2\varphi\cos^2\varphi;$$

$$A_{12} = A_{21} = \overline{E_1}\mu_{12} + \left(\overline{E_1} + \overline{E_2} - 2\left(\overline{E_1}\mu_{12} + 2G_{12}\right)\right)\sin^2\varphi\cos^2\varphi;$$

$$A_{33} = \left(\overline{E_1} + \overline{E_2} - 2\,\overline{E_1}\mu_{12}\right)\sin^2\varphi\cos^2\varphi + G_{12}\cos^2 2\varphi,$$

где $\overline{E_1}$ и $\overline{E_2}$ – приведенные модули упругости, вычисляемые по формуле

$$\overline{E}_{1(2)} = \frac{E_{1(2)}}{1-\mu_{12}\mu_{21}},$$

а коэффициент Пуассона μ_{21} определяется из условия симметрии упругих постоянных:

$$\mu_{21} = \frac{\mu_{12}E_1}{\overline{E}_2}.$$

Модули трансверсального сдвига примем равными G_{12}.

Коэффициенты жесткости пластины вычислим по формулам

$$B_{11} = A_{11}h;\ B_{22} = A_{22}h;\ B_{21} = A_{21}h;$$

$$B_{33} = A_{33}h;\ D_{11} = \frac{A_{11}h^3}{12};\ D_{22} = \frac{A_{22}h^3}{12};$$

$$D_{12} = \frac{A_{12}h^3}{12};\ D_{21} = \frac{A_{21}h^3}{12};\ D_{33} = \frac{A_{33}h^3}{12};$$

$$K_x = K_y = G_{12}h;$$

$$C_{ij} = 0 \quad (ij = 11, 22, 12, 21, 33),$$

а параметры инерции – по формулам

$$B_\rho = \rho h;\ C_\rho = 0;\ D_\rho = \rho h^3/12$$

Результаты вычислений в виде значений первых пяти частот собственных колебаний поместим в таблицу (табл. 2). Здесь же приведены частоты для пластин с толщинами 30 и 50 мм.

Формы колебаний соответствуют тем, что приведены на рис. 2–6.

Анализ представленных результатов говорит о том, что в задаче модального расчета результаты, полученные при защемлении со свободным сдвигом краев, отличаются от тех, что определены при классическом защемлении. Это различие нарастает по мере увеличения толщины пластины, достигая максимума (по первой частоте) в рассматриваемой модели в 10,6 %.

Проведенная работа позволяет сформулировать следующие выводы.

1. Получен энергетический функционал для решения задачи о собственных колебаниях пластин с низкой трансверсальной жесткостью.

2. Разработана конечно-элементная модель податливой при сдвиге пластины, вектор узловых неизвестных которой включает углы трансверсального сдвига. Сформированы матрица жесткости и матрица инерции для соответствующего элемента пластины.

3. В результате проведенного численного исследования на примере расчета композитной пластины показана актуальность разработанной модели при проведении модального анализа конструкций с низкой трансверсальной сдвиговой жесткостью, а также при учете граничных условий неклассического вида.

Библиографическая ссылка

1. Васильев В. В. Механика конструкций из композиционных материалов. М. : Машиностроение, 1988.

Таблица 2

Значения частот собственных колебаний композитной пластины

Номер моды	Значения собственных частот, Гц					
	h = 10 мм		h = 30 мм		h = 50 мм	
	Полное защемление	Защемление со свободным сдвигом	Полное защемление	Защемление со свободным сдвигом	Полное защемление	Защемление со свободным сдвигом
1	111,24	110,42	323	306	509	455
2	222,09	220,72	634	608	975	900
3	222,09	220,72	634	608	975	900
4	349,33	347,40	977	943	1 457	1 370
5	380,66	378,74	1 066	1 033	1 592	1 512

V. A. Nesterov

VIBRATION OF A PLATE WITH LOW TRANSVERSE SHEAR STIFFNESS

Modal finite element analysis of a plate with low transverse shear stiffness is considered. There are two transverse shear strains at each of four nodes. In modal analysis of a composite plate the urgency of the developed finite element model is shown. Results of numerical research of influence of nonclassic boundary conditions are presented.

Keywords: a plate, finite element method, transverse shear strains, modal analysis.

© Нестеров В. А., 2010

УДК 519.62

В. А. Нестеров, А. В. Лопатин

ВЫВОД ВЫРАЖЕНИЯ ЭНЕРГИИ ДЕФОРМАЦИИ ОБОЛОЧКИ В ГЕОМЕТРИЧЕСКИ НЕЛИНЕЙНОЙ ПОСТАНОВКЕ

В геометрически нелинейной постановке рассматривается вариационная задача о деформировании оболочки, податливой при трансверсальном сдвиге. Представлен вывод выражения потенциальной энергии деформации, которое может быть использовано для получения уравнений устойчивости для оболочек с низкой трансверсальной жесткостью.

Ключевые слова: оболочка, трансверсальный сдвиг, вариационная задача.

В последнее время в производстве ракетно-космической техники все чаще используются композиционные материалы, которые, обладая большой удельной прочностью и жесткостью, позволяют создавать конструкции высокой степени весового совершенства. Однако композиты, в отличие от традиционных конструкционных материалов, отличаются рядом особенностей, которые усложняют классические расчетные модели. Одной из таких особенностей является низкая сдвиговая жесткость по отношению к трансверсальным напряжениям.

Метод конечных элементов – самый распространенный среди современных численных методов расчета конструкций. В общем случае вывод разрешающих уравнений МКЭ предполагает формирование энергетического функционала полной потенциальной энергии объекта. При решении задачи об устойчивости оболочек необходимо оперировать выражением потенциальной энергии деформации в геометрически нелинейной постановке.

Общее выражение потенциальной энергии деформации запишем на основе уравнения, представленного в [1]:

$$U = \frac{1}{2}\iiint\limits_{(V)} (\sigma^*_\alpha \Delta_\alpha + \sigma^*_\beta \Delta_\beta + \sigma^*_\gamma \Delta_\gamma + \tau^*_{\alpha\beta}\Delta_{\alpha\beta} + \\ + \tau^*_{\alpha\lambda}\Delta_{\alpha\gamma} + \tau^*_{\beta\gamma}\Delta_{\beta\gamma}) H_1 H_2 \, d\alpha\, d\beta\, d\gamma, \quad (1)$$

где коэффициент 1/2 перед интегралом обусловлен линейной зависимостью между напряжениями и деформациями (для линейно-упругого материала); $\sigma^*_\alpha, \sigma^*_\beta, \sigma^*_\gamma$, $\tau^*_{\alpha\beta}, \tau^*_{\alpha\gamma}$ и $\tau^*_{\beta\gamma}$ – обобщенные компоненты напряжения. Их выражения записываются следующим образом:

$$\sigma^*_\alpha = \frac{S^*_1}{S_1}\frac{\sigma_\alpha}{1+\xi_\alpha};\quad \tau^*_{\alpha\beta} = \frac{S^*_1}{S_1}\frac{\tau_{\alpha\beta}}{1+\xi_\beta};\quad \tau^*_{\alpha\gamma} = \frac{S^*_1}{S_1}\frac{\tau_{\alpha\gamma}}{1+\xi_\gamma};$$
$$\tau^*_{\beta\alpha} = \frac{S^*_2}{S_2}\frac{\tau_{\beta\alpha}}{1+\xi_\alpha};\quad \sigma^*_\beta = \frac{S^*_2}{S_2}\frac{\sigma_\beta}{1+\xi_\beta};\quad \tau^*_{\beta\gamma} = \frac{S^*_2}{S_2}\frac{\tau_{\beta\gamma}}{1+\xi_\gamma};$$
$$\tau^*_{\gamma\alpha} = \frac{S^*_3}{S_3}\frac{\tau_{\gamma\alpha}}{1+\xi_\alpha};\quad \tau^*_{\gamma\beta} = \frac{S^*_3}{S_3}\frac{\tau_{\gamma\beta}}{1+\xi_\beta};\quad \sigma^*_\gamma = \frac{S^*_3}{S_3}\frac{\sigma_\gamma}{1+\xi_\gamma}, \quad (2)$$

где ξ_α, ξ_β, ξ_γ – относительные удлинения волокон, первоначально направленных по k_1, k_2, k_3 (орты системы координат α, β, γ). Они вычисляются по формулам

$$\xi_\alpha = \sqrt{1+2\Delta_\alpha} - 1;\quad \xi_\beta = \sqrt{1+2\Delta_\beta} - 1;$$
$$\xi_\gamma = \sqrt{1+2\Delta_\gamma} - 1 \quad (3)$$

В выражениях (2) $-\frac{S^*_1}{S_1};\ \frac{S^*_2}{S_2};\ \frac{S^*_3}{S_3}$ – коэффициенты изменения величины площадок, первоначально перпендикулярных k_1, k_2, k_3 соответственно. Они определяются через компоненты деформации Δ:

$$\frac{S^*_1}{S_1} = \sqrt{(1+2\Delta_\beta)(1+2\Delta_\gamma) - \Delta^2_{\beta\gamma}};$$
$$\frac{S^*_2}{S_2} = \sqrt{(1+2\Delta_\alpha)(1+2\Delta_\gamma) - \Delta^2_{\alpha\gamma}}; \quad (4)$$
$$\frac{S^*_3}{S_3} = \sqrt{(1+2\Delta_\alpha)(1+2\Delta_\beta) - \Delta^2_{\alpha\beta}}.$$

Для компонентов напряжения справедливо соотношение

$$\tau^*_{ij} = \tau^*_{ji}. \quad (5)$$

Введем допущение

$$\sigma^*_\alpha = \sigma_\alpha;\ \ \sigma^*_\beta = \sigma_\beta;\ \ \sigma^*_\gamma = \sigma_\gamma;$$
$$\tau^*_{\alpha\beta} = \tau_{\alpha\beta};\ \ \tau^*_{\alpha\gamma} = \tau_{\alpha\gamma};\ \ \tau^*_{\beta\gamma} = \tau_{\beta\gamma}. \quad (6)$$

Оно справедливо при малых величинах Δ, когда комплексы

$$\frac{S^*_i}{S_i}\frac{1}{1+\xi_j}$$

можно принять равными единице.

Общие выражения для компонентов деформации Δ запишем в виде

$$\Delta_\alpha = e_\alpha + \frac{1}{2}\left[e^2_\alpha + \left(\frac{1}{2}e_{\alpha\beta} + \Omega_\gamma\right)^2 + \left(\frac{1}{2}e_{\alpha\gamma} - \Omega_\beta\right)^2\right];$$
$$\Delta_\beta = e_\beta + \frac{1}{2}\left[e^2_\beta + \left(\frac{1}{2}e_{\alpha\beta} - \Omega_\gamma\right)^2 + \left(\frac{1}{2}e_{\beta\gamma} + \Omega_\alpha\right)^2\right];$$
$$\Delta_\gamma = e_\gamma + \frac{1}{2}\left[e^2_\gamma + \left(\frac{1}{2}e_{\alpha\gamma} + \Omega_\beta\right)^2 + \left(\frac{1}{2}e_{\beta\gamma} - \Omega_\alpha\right)^2\right]; \quad (7)$$
$$\Delta_{\alpha\beta} = e_{\alpha\beta} + e_\alpha\left(\frac{1}{2}e_{\alpha\beta} - \Omega_\gamma\right) + e_\beta\left(\frac{1}{2}e_{\alpha\beta} + \Omega_\gamma\right) + \\ + \left(\frac{1}{2}e_{\alpha\gamma} - \Omega_\beta\right)\left(\frac{1}{2}e_{\beta\gamma} + \Omega_\alpha\right);$$

$$\Delta_{\alpha\gamma} = e_{\alpha\gamma} + e_{\alpha}\left(\frac{1}{2}e_{\alpha\gamma} + \Omega_{\beta}\right) + e_{\gamma}\left(\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta}\right) + \left(\frac{1}{2}e_{\alpha\beta} + \Omega_{\gamma}\right)\left(\frac{1}{2}e_{\beta\gamma} - \Omega_{\alpha}\right);$$

$$\Delta_{\beta\gamma} = e_{\beta\gamma} + e_{\beta}\left(\frac{1}{2}e_{\beta\gamma} - \Omega_{\alpha}\right) + e_{\gamma}\left(\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha}\right) + \left(\frac{1}{2}e_{\alpha\beta} - \Omega_{\gamma}\right)\left(\frac{1}{2}e_{\alpha\gamma} + \Omega_{\beta}\right).$$

Введем упрощения:

– пренебрежем квадратами и произведениями деформаций e_{α}, e_{β}, $e_{\alpha\beta}$;

– примем угол поворота Ω_{γ} равным нулю;

– в выражениях трансверсальных деформаций оставим только линейные члены.

В результате вместо (7) получим

$$\Delta_{\alpha} = e_{\alpha} + \frac{1}{2}\left(\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta}\right)^2;\ \Delta_{\beta} = e_{\beta} + \frac{1}{2}\left(\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha}\right)^2;$$

$$\Delta_{\alpha\beta} = e_{\alpha\beta} + \left(\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta}\right)\left(\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha}\right); \qquad (8)$$

$$\Delta_{\gamma} = e_{\gamma};\quad \Delta_{\alpha\gamma} = e_{\alpha\gamma};\quad \Delta_{\beta\gamma} = e_{\beta\gamma}.$$

Линеаризованные компоненты деформации e_{α}, e_{β}, e_{γ}, $e_{\alpha\beta}$, $e_{\alpha\gamma}$, $e_{\beta\gamma}$ и проекции угла поворота Ω_{α}, Ω_{β}, Ω_{γ} можно выразить через проекции перемещения U, V, W:

$$e_{\alpha} = \frac{1}{H_1}\frac{\partial U}{\partial\alpha} + \frac{1}{H_1H_2}\frac{\partial H_1}{\partial\beta}V + \frac{1}{H_1H_3}\frac{\partial H_1}{\partial\gamma}W,$$

$$e_{\beta} = \frac{1}{H_2}\frac{\partial V}{\partial\beta} + \frac{1}{H_2H_3}\frac{\partial H_2}{\partial\gamma}W + \frac{1}{H_2H_1}\frac{\partial H_2}{\partial\alpha}U,$$

$$e_{\gamma} = \frac{1}{H_3}\frac{\partial W}{\partial\gamma} + \frac{1}{H_3H_1}\frac{\partial H_3}{\partial\alpha}U + \frac{1}{H_3H_2}\frac{\partial H_3}{\partial\beta}V, \qquad (9)$$

$$e_{\alpha\beta} = e_{\beta\alpha} = \frac{H_2}{H_1}\frac{\partial}{\partial\alpha}\left(\frac{V}{H_2}\right) + \frac{H_1}{H_2}\frac{\partial}{\partial\beta}\left(\frac{U}{H_1}\right),$$

$$e_{\alpha\gamma} = e_{\gamma\alpha} = \frac{H_3}{H_1}\frac{\partial}{\partial\alpha}\left(\frac{W}{H_3}\right) + \frac{H_1}{H_3}\frac{\partial}{\partial\gamma}\left(\frac{U}{H_1}\right),$$

$$e_{\beta\gamma} = e_{\gamma\beta} = \frac{H_2}{H_3}\frac{\partial}{\partial\gamma}\left(\frac{V}{H_2}\right) + \frac{H_3}{H_2}\frac{\partial}{\partial\beta}\left(\frac{W}{H_3}\right);$$

$$2\Omega_{\alpha} = \frac{1}{H_2H_3}\left[\frac{\partial}{\partial\beta}(H_3W) - \frac{\partial}{\partial\gamma}(H_2V)\right],$$

$$2\Omega_{\beta} = \frac{1}{H_1H_3}\left[\frac{\partial}{\partial\gamma}(H_1U) - \frac{\partial}{\partial\alpha}(H_3W)\right], \qquad (10)$$

$$2\Omega_{\gamma} = \frac{1}{H_1H_2}\left[\frac{\partial}{\partial\alpha}(H_2V) - \frac{\partial}{\partial\beta}(H_1U)\right].$$

С учетом (9) и (10) запишем выражения комплексов, фигурирующих в (8):

$$\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta} = \frac{1}{2}\left[\frac{1}{H_1}\frac{\partial W}{\partial\alpha} + \frac{1}{H_1}\left(H_1\frac{\partial U}{\partial\gamma} - U\frac{\partial H_1}{\partial\gamma}\right) - \frac{1}{H_1}\left(U\frac{\partial H_1}{\partial\gamma} + H_1\frac{\partial U}{\partial\gamma}\right) + \frac{1}{H_1}\frac{\partial W}{\partial\alpha}\right] = \frac{1}{H_1}\left(\frac{\partial W}{\partial\alpha} - U\frac{\partial H_1}{\partial\gamma}\right);$$

$$\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha} = \frac{1}{2}\left[\frac{1}{H_2}\left(H_2\frac{\partial V}{\partial\gamma} - V\frac{\partial H_2}{\partial\gamma}\right) + \frac{1}{H_2}\frac{\partial W}{\partial\beta} + \frac{1}{H_2}\frac{\partial W}{\partial\beta} - \frac{1}{H_2}\left(H_2\frac{\partial V}{\partial\gamma} - V\frac{\partial H_2}{\partial\gamma}\right)\right] = \frac{1}{H_2}\left(\frac{\partial W}{\partial\beta} - V\frac{\partial H_2}{\partial\gamma}\right).$$

Здесь учтено, что для оболочек $H_3 = 1$.

Для дальнейших преобразований понадобятся выражения для коэффициентов Ламе:

$$H_1 = A_1\left(1 + \frac{\gamma}{R_1}\right); \qquad H_2 = A_2\left(1 + \frac{\gamma}{R_2}\right) \qquad (11)$$

и следующие кинематические соотношения [2]:

$$U = u + \gamma\theta_{\alpha};\ V = v + \gamma\theta_{\beta};\ W = w; \qquad (12)$$

$$\theta_{\alpha} = \psi_{\alpha} + \frac{u}{R_1} - \frac{1}{A_1}\frac{\partial w}{\partial\alpha};\quad \theta_{\beta} = \psi_{\beta} + \frac{v}{R_2} - \frac{1}{A_2}\frac{\partial w}{\partial\beta}. \qquad (13)$$

С учетом (11), (12) и (13) получим:

$$\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta} = \frac{1}{H_1}\left(\frac{\partial\omega}{\partial\alpha} - \left(u + \gamma\theta_{\alpha}\right)\frac{A_1}{R_1}\right) =$$

$$= \frac{A_1}{H_1}\left(\frac{1}{A_1}\frac{\partial\omega}{\partial\alpha} - \frac{u}{R_1}\right) - \frac{A_1}{H_1}\frac{\gamma}{R_1}\theta_{\alpha} = \frac{A_1}{H_1}\left(\psi_{\alpha} - \theta_{\alpha}\right) - \qquad (14)$$

$$-\frac{A_1}{H_1}\frac{\gamma}{R_1}\theta_{\alpha} = \frac{A_1}{H_1}\psi_{\alpha} - \theta_{\alpha}\frac{A_1}{H_1}\left(1 + \frac{\gamma}{R_1}\right) = \frac{A_1}{H_1}\psi_{\alpha} - \theta_{\alpha};$$

$$\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha} = \frac{1}{H_2}\left(\frac{\partial\omega}{\partial\beta} - \left(\upsilon + \gamma\theta_{\beta}\right)\frac{A_2}{R_2}\right) = \frac{A_2}{H_2}\left(\frac{1}{A_2}\frac{\partial\omega}{\partial\beta} - \frac{\upsilon}{R_2}\right) -$$

$$-\frac{A_2}{H_2}\frac{\gamma}{R_2}\theta_{\beta} = \frac{A_2}{H_2}\left(\psi_{\beta} - \theta_{\beta}\right) - \frac{A_2}{H_2}\frac{\gamma}{R_2}\theta_{\beta} = \frac{A_2}{H_2}\psi_{\beta} -$$

$$-\theta_{\beta}\frac{A_2}{H_2}\left(1 + \frac{\gamma}{R_2}\right) = \frac{A_2}{H_2}\psi_{\beta} - \theta_{\beta}.$$

Соотношения для углов поворота (13) можно записать в следующем виде:

$$\theta_{\alpha} = \psi_{\alpha} + \omega_{\alpha};\qquad \theta_{\beta} = \psi_{\beta} + \omega_{\beta}, \qquad (15)$$

где

$$\omega_{\alpha} = \frac{u}{R_1} - \frac{1}{A_1}\frac{\partial w}{\partial\alpha};\qquad \omega_{\beta} = \frac{v}{R_2} - \frac{1}{A_2}\frac{\partial w}{\partial\beta}. \qquad (16)$$

С учетом (15) выражения для комплексов (14) можно переписать в виде

$$\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta} = \psi_{\alpha}\left(\frac{A_1}{H_1} - 1\right) - \omega_{\alpha};$$

$$\frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha} = \psi_{\beta}\left(\frac{A_2}{H_2} - 1\right) - \omega_{\beta}. \qquad (17)$$

Введем еще одно упрощение: в нелинейных членах компонентов деформации положим $\psi_{\alpha} = 0$ и $\psi_{\beta} = 0$, тогда вместо (17) получим

$$\frac{1}{2}e_{\alpha\gamma} - \Omega_{\beta} = -\omega_{\alpha};\qquad \frac{1}{2}e_{\beta\gamma} + \Omega_{\alpha} = -\omega_{\beta}. \qquad (18)$$

Подставим (18) в (8) и запишем

$$\Delta_{\alpha} = e_{\alpha} + \frac{1}{2}\omega_{\alpha}^2;\ \Delta_{\beta} = e_{\beta} + \frac{1}{2}\omega_{\beta}^2;$$

$$\Delta_{\alpha\beta} = e_{\alpha\beta} + \omega_\alpha\omega_\beta\ ;\ \ \Delta_\gamma = e_\gamma\ ;$$
$$\Delta_{\alpha\gamma} = e_{\alpha\gamma}\ ;\ \ \Delta_{\beta\gamma} = e_{\beta\gamma}. \qquad (19)$$

Подставим законы распределения перемещений (12) в первое, второе и четвертое соотношения системы (9). В результате получим

$$e_\alpha = \frac{A_1}{H_1}\left(\varepsilon_\alpha + \gamma\chi_\alpha\right);$$
$$e_\beta = \frac{A_2}{H_2}\left(\varepsilon_\beta + \gamma\chi_\beta\right); \qquad (20)$$
$$e_{\alpha\beta} = \frac{A_1}{H_1}\left(\varepsilon_{\alpha\beta} + \gamma\chi_{\alpha\beta}\right) + \frac{A_2}{H_2}\left(\varepsilon_{\beta\alpha} + \gamma\chi_{\beta\alpha}\right),$$

где

$$\varepsilon_\alpha = \frac{1}{A_1}\frac{\partial u}{\partial\alpha} + \frac{v}{A_1A_2}\frac{\partial A_1}{\partial\beta} + \frac{w}{R_1};$$
$$\varepsilon_\beta = \frac{1}{A_2}\frac{\partial v}{\partial\beta} + \frac{u}{A_1A_2}\frac{\partial A_2}{\partial\alpha} + \frac{w}{R_2};$$
$$\varepsilon_{\alpha\beta} = \frac{1}{A_1}\frac{\partial v}{\partial\alpha} - \frac{u}{A_1A_2}\frac{\partial A_1}{\partial\beta};\ \ \varepsilon_{\beta\alpha} = \frac{1}{A_2}\frac{\partial u}{\partial\beta} - \frac{v}{A_1A_2}\frac{\partial A_2}{\partial\alpha};$$
$$\chi_\alpha = \frac{1}{A_1}\frac{\partial\theta_\alpha}{\partial\alpha} + \frac{\theta_\beta}{A_1A_2}\frac{\partial A_1}{\partial\beta}; \qquad (21)$$
$$\chi_\beta = \frac{1}{A_2}\frac{\partial\theta_\beta}{\partial\beta} + \frac{\theta_\alpha}{A_1A_2}\frac{\partial A_2}{\partial\alpha};\ \ \chi_{\alpha\beta} = \frac{1}{A_1}\frac{\partial\theta_\beta}{\partial\alpha} - \frac{\theta_\alpha}{A_1A_2}\frac{\partial A_1}{\partial\beta};$$
$$\chi_{\beta\alpha} = \frac{1}{A_2}\frac{\partial\theta_\alpha}{\partial\beta} - \frac{\theta_\beta}{A_1A_2}\frac{\partial A_2}{\partial\alpha}.$$

Запишем общее выражение полной энергии деформации (1), заменив в нем компоненты напряжения в соответствии с (6), а компоненты деформации – выражениями (19):

$$U = \frac{1}{2}\int\int\int\left\{\sigma_\alpha\left(e_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \sigma_\beta\left(e_\beta + \frac{1}{2}\omega_\beta^2\right) + \sigma_\gamma e_\gamma + \right.$$
$$\left. + \tau_{\alpha\beta}\left(e_{\alpha\beta} + \omega_\alpha\omega_\beta\right) + \tau_{\alpha\gamma}e_{\alpha\gamma} + \tau_{\beta\gamma}e_{\beta\gamma}\right\} H_1H_2\,d\alpha\,d\beta\,d\gamma. \qquad (22)$$

Физические соотношения в нелинейной теории имеют вид

$$\sigma_\alpha = A_{11}\Delta_\alpha + A_{12}\Delta_\beta\ ;\ \ \sigma_\beta = A_{21}\Delta_\alpha + A_{22}\Delta_\beta\ ;$$
$$\tau_{\alpha\beta} = A_{33}\Delta_{\alpha\beta}, \qquad (23)$$

где

$$\Delta_\alpha = e_\alpha + \frac{1}{2}\omega_\alpha^2\ ;\ \ \Delta_\beta = e_\beta + \frac{1}{2}\omega_\beta^2\ ;$$
$$\Delta_{\alpha\beta} = e_{\alpha\beta} + \omega_\alpha\omega_\beta.$$

Для трансверсальных напряжений примем линейные соотношения

$$\tau_{\alpha\gamma} = G_{\alpha\gamma}e_{\alpha\gamma}\ ;\ \tau_{\beta\gamma} = G_{\beta\gamma}e_{\beta\gamma}. \qquad (24)$$

Подставим физические соотношения (23), (24) в (22) и получим

$$U = \frac{1}{2}\int\limits_{(V)}\int\int\left(A_{11}\Delta_\alpha^2 + 2A_{12}\Delta_\alpha\Delta_\beta + A_{22}\Delta_\beta^2 + \right.$$
$$\left. + A_{33}\Delta_{\alpha\beta}^2 + G_{\alpha\gamma}e_{\alpha\gamma}^2 + G_{\beta\gamma}e_{\beta\gamma}^2\right) H_1H_2\,d\alpha\,d\beta\,d\gamma \qquad (25)$$

или

$$U = \frac{1}{2}\int\limits_{(V)}\int\int\left\{A_{11}\left(e_\alpha + \frac{1}{2}\omega_\alpha^2\right)^2 + 2A_{12}\left(e_\alpha + \frac{1}{2}\omega_\alpha^2\right)\times\right.$$
$$\times\left(e_\beta + \frac{1}{2}\omega_\beta^2\right) + A_{22}\left(e_\beta + \frac{1}{2}\omega_\beta^2\right)^2 + A_{33}\left(e_{\alpha\beta} + \omega_\alpha\omega_\beta\right)^2 + \qquad (26)$$
$$\left. + G_{\alpha\gamma}e_{\alpha\gamma}^2 + G_{\beta\gamma}e_{\beta\gamma}^2\right\} H_1H_2\,d\alpha\,d\beta\,d\gamma.$$

Подставим в (26) выражения для относительных деформаций (20):

$$U = \frac{1}{2}\int\int\int\left\{A_{11}\left[\frac{A_1}{H_1}\left(\varepsilon_\alpha + \gamma\chi_\alpha\right) + \frac{1}{2}\omega_\alpha^2\right]^2 + \right.$$
$$+ A_{22}\left[\frac{A_2}{H_2}\left(\varepsilon_\beta + \gamma\chi_\beta\right) + \frac{1}{2}\omega_\beta^2\right]^2 +$$
$$+ 2A_{12}\left[\frac{A_1}{H_1}\left(\varepsilon_\alpha + \gamma\chi_\alpha\right) + \frac{1}{2}\omega_\alpha^2\right]\times \qquad (27)$$
$$\times\left[\frac{A_2}{H_2}\left(\varepsilon_\beta + \gamma\chi_\beta\right) + \frac{1}{2}\omega_\beta^2\right] +$$
$$+ A_{33}\left[\frac{A_1}{H_1}\left(\varepsilon_{\alpha\beta} - \gamma\chi_{\alpha\beta}\right) + \frac{A_2}{H_2}\left(\varepsilon_{\beta\alpha} - \gamma\chi_{\beta\alpha}\right) + \omega_\alpha\omega_\beta\right]^2 +$$
$$\left. + G_{\alpha\gamma}e_{\alpha\gamma}^2 + G_{\beta\gamma}e_{\beta\gamma}^2\right\} H_1H_2\,d\alpha\,d\beta\,d\gamma.$$

Выполним в (27) следующие преобразования. Выделим комплексы

$$\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\ ;\qquad \frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\ ;$$
$$\frac{A_1}{H_1}\varepsilon_{\alpha\beta} + \frac{A_2}{H_2}\varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\ .$$

Это позволит записать квадраты и произведения нелинейных деформаций в виде

$$\Delta_\alpha^2 = \left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)^2 + 2\gamma\frac{A_1}{H_1}\chi_\alpha\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\frac{A_1^2}{H_1^2}\chi_\alpha^2\ ;$$
$$\Delta_\beta^2 = \left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right)^2 + 2\gamma\frac{A_2}{H_2}\times$$
$$\times\chi_\beta\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma^2\frac{A_2^2}{H_2^2}\chi_\beta^2\ ; \qquad (28)$$
$$\Delta_\alpha\Delta_\beta = \left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma\frac{A_1}{H_1}\times$$
$$\times\chi_\alpha\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma\frac{A_2}{H_2}\chi_\beta\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\frac{A_1}{H_1}\frac{A_2}{H_2}\chi_\alpha\chi_\beta;$$
$$\Delta_{\alpha\beta}^2 = \left(\frac{A_1}{H_1}\varepsilon_{\alpha\beta} + \frac{A_2}{H_2}\varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right)^2 +$$
$$+2\gamma\left(\frac{A_1}{H_1}\chi_{\alpha\beta} + \frac{A_2}{H_2}\chi_{\beta\alpha}\right)\left(\frac{A_1}{H_1}\varepsilon_{\alpha\beta} + \frac{A_2}{H_2}\varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) +$$
$$+\gamma^2\left(\frac{A_1}{H_1}\chi_{\alpha\beta} + \frac{A_2}{H_2}\chi_{\beta\alpha}\right)^2.$$

Заменим истинные деформации трансверсального сдвига их осредненными значениями

$$e_{\alpha\gamma} = \psi_\alpha \; ; \qquad e_{\beta\gamma} = \psi_\beta . \tag{29}$$

Перепишем (27) с учетом (28), (29):

$$U = \frac{1}{2}\int\limits_{(V)}\int\int\Bigg\{ A_{11}\Bigg[\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)^2 + 2\gamma\frac{A_1}{H_1}\times$$

$$\times\chi_\alpha\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\frac{A_1^2}{H_1^2}\chi_\alpha^2\Bigg] +$$

$$+A_{22}\Bigg[\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right)^2 + 2\gamma\frac{A_2}{H_2}\chi_\beta\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma^2\frac{A_2^2}{H_2^2}\chi_\beta^2\Bigg] +$$

$$+2A_{12}\Bigg[\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma\frac{A_1}{H_1}\chi_\alpha\left(\frac{A_2}{H_2}\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) +$$

$$+\gamma\frac{A_2}{H_2}\chi_\beta\left(\frac{A_1}{H_1}\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\frac{A_1}{H_1}\frac{A_2}{H_2}\chi_\alpha\chi_\beta\Bigg] + \tag{30}$$

$$+A_{33}\Bigg[\left(\frac{A_1}{H_1}\varepsilon_{\alpha\beta} + \frac{A_2}{H_2}\varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right)^2 + 2\gamma\left(\frac{A_1}{H_1}\chi_{\alpha\beta} + \frac{A_2}{H_2}\chi_{\beta\alpha}\right)\times$$

$$\times\left(\frac{A_1}{H_1}\varepsilon_{\alpha\beta} + \frac{A_2}{H_2}\varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) + \gamma^2\left(\frac{A_1}{H_1}\chi_{\alpha\beta} + \frac{A_2}{H_2}\chi_{\beta\alpha}\right)^2\Bigg] +$$

$$+G_{\alpha\gamma}\psi_\alpha^2 + G_{\beta\gamma}\psi_\beta^2\Big\}\, H_1H_2\, d\alpha\, d\beta\, d\gamma \; .$$

Запишем (30) для случая, когда не учитывается изменение метрики по толщине:

$$U = \frac{1}{2}\int\limits_{(V)}\int\int\Bigg\{ A_{11}\Bigg[\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)^2 + 2\gamma\chi_x\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\chi_\alpha^2\Bigg] +$$

$$+A_{22}\Bigg[\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right)^2 + 2\gamma\chi_\beta\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma^2\chi_\beta^2\Bigg] + \tag{31}$$

$$+2A_{12}\Bigg[\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \gamma\chi_\alpha\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) +$$

$$+\gamma\chi_\beta\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + \gamma^2\chi_\alpha\chi_\beta\Bigg] +$$

$$+A_{33}\Big[\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right)^2 + 2\gamma\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right)\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) +$$

$$+\gamma^2\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right)^2\Big] + G_{\alpha\gamma}\psi_\alpha^2 + G_{\beta\gamma}\psi_\beta^2\Big\}\, d\alpha\, d\beta\, d\gamma.$$

Проинтегрируем (31) по толщине оболочки:

$$U = \frac{1}{2}\int\int\Bigg\{ B_{11}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)^2 + 2C_{11}\chi_\alpha\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + D_{11}\chi_\alpha^2 +$$

$$+\; B_{22}\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right)^2 + \; 2C_{22}\chi_\beta\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + D_{22}\chi_\beta^2 +$$

$$+2B_{12}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right)\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + 2C_{12}\chi_\alpha\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + \tag{32}$$

$$+\; 2C_{12}\chi_\beta\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + 2D_{12}\chi_\alpha\chi_\beta + B_{33}\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right)^2 +$$

$$+2C_{33}\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right)\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) +$$

$$+D_{33}\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right)^2 + K_\alpha\psi_\alpha^2 + K_\beta\psi_\beta^2\Big\}\, d\alpha\, d\beta.$$

Рассмотрим физические соотношения (23). Вычислим с их помощью усилия и моменты по формулам

$$N_\alpha = \int\limits_{-s}^{h-s}\sigma_\alpha\, d\gamma \; ; \quad N_{\alpha\beta} = \int\limits_{-s}^{h-s}\tau_{\alpha\beta}\, d\gamma \; ; \; N_\beta = \int\limits_{-s}^{h-s}\sigma_\beta\, d\gamma \; ;$$

$$Q_\alpha = \int\limits_{-s}^{h-s}\tau_{\alpha\gamma}\, d\gamma \; ; \; Q_\beta = \int\limits_{-s}^{h-s}\tau_{\beta\gamma}\, d\gamma \; ;$$

$$M_\alpha = \int\limits_{-s}^{h-s}\gamma\sigma_\alpha\, d\gamma \; ; \; M_{\alpha\beta} = \int\limits_{-s}^{h-s}\gamma\tau_{\alpha\beta}\, d\gamma \; ;$$

$$M_\beta = \int\limits_{-s}^{h-s}\gamma\sigma_\beta\, d\gamma. \tag{33}$$

Подставим в (33) соотношения для напряжений (23), в которых предварительно выразим Δ_α , Δ_β , $\Delta_{\alpha\beta}$ через относительные деформации e_α, e_β, $e_{\alpha\beta}$. Последние, в свою очередь, выразим через обобщенные деформации ε_α, ε_β, $\varepsilon_{\alpha\beta}$, $\varepsilon_{\beta\alpha}$, χ_α, χ_β, $\chi_{\alpha\beta}$, $\chi_{\beta\alpha}$, т. е. используя формулы (4.20), в которых не будем учитывать изменение метрики по толщине:

$$e_\alpha = \varepsilon_\alpha + \gamma\chi_\alpha \; ; \quad e_\beta = \varepsilon_\beta + \gamma\chi_\beta \; ;$$

$$e_{\alpha\beta} = \varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \gamma\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right). \tag{34}$$

Тогда физические соотношения (23) примут вид

$$\sigma_\alpha = A_{11}\left(\varepsilon_\alpha + \gamma\chi_\alpha + \frac{1}{2}\omega_\alpha^2\right) + A_{12}\left(\varepsilon_\beta + \gamma\chi_\beta + \frac{1}{2}\omega_\beta^2\right);$$

$$\sigma_\beta = A_{21}\left(\varepsilon_\alpha + \gamma\chi_\alpha + \frac{1}{2}\omega_\alpha^2\right) + A_{22}\left(\varepsilon_\beta + \gamma\chi_\beta + \frac{1}{2}\omega_\beta^2\right); \tag{35}$$

$$\tau_{\alpha\beta} = A_{33}\left[\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \gamma\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right) + \omega_\alpha\omega_\beta\right].$$

Подставим (35) в (33) и получим

$$N_\alpha = B_{11}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + B_{12}\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + C_{11}\chi_\alpha + C_{12}\chi_\beta \; ;$$

$$N_\beta = B_{21}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + B_{22}\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + C_{21}\chi_\alpha + C_{22}\chi_\beta \; ;$$

$$N_{\alpha\beta} = B_{33}\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) + C_{33}\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right) \; ; \tag{36}$$

$$M_\alpha = C_{11}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + C_{12}\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + D_{11}\chi_\alpha + D_{12}\chi_\beta \; ;$$

$$M_\beta = C_{21}\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + C_{22}\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) + D_{21}\chi_\alpha + D_{22}\chi_\beta \; ;$$

$$M_{\alpha\beta} = C_{33}\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) + D_{33}\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right) \; ;$$

$$Q_\alpha = K_\alpha\psi_\alpha \; ; \; Q_\beta = K_\beta\psi_\beta .$$

С учетом (36) выражение для потенциальной энергии деформации можно записать в следующем виде:

$$U = \frac{1}{2}\int\int\Bigg\{ N_\alpha\left(\varepsilon_\alpha + \frac{1}{2}\omega_\alpha^2\right) + N_\beta\left(\varepsilon_\beta + \frac{1}{2}\omega_\beta^2\right) +$$

$$+N_{\alpha\beta}\left(\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} + \omega_\alpha\omega_\beta\right) + M_\alpha\chi_\alpha + M_\beta\chi_\beta + \tag{37}$$

$$+\; M_{\alpha\beta}\left(\chi_{\alpha\beta} + \chi_{\beta\alpha}\right) + Q_\alpha\psi_\alpha + Q_\beta\psi_\beta\Big\}\, d\alpha\, d\beta.$$

Функционал (37) является выражением потенциальной энергии деформации, которое следует использовать при решении нелинейных задач о деформировании оболочек с низкой трансверсальной сдвиговой жесткостью.

Оно также необходимо для получения разрешающих уравнений устойчивости оболочек, в расчетных моделях которых учитывается трансверсальный сдвиг.

Библиографические ссылки

1. Новожилов В. В. Теория упругости. Л. : Судпромгиз, 1958.

2. Васильев В. В. Механика конструкций из композиционных материалов. М. : Машиностроение, 1988.

V. A. Nesterov, A. V. Lopatin

ENERGY DERIVATION OF DEFORMATION OF THE SHELL IN GEOMETRICAL NONLINEAR ARRANGEMENT

In geometrically nonlinear arrangement a variational problem of deformation of shell with low value of transverse shear stiffness is considered. The conclusion of expression of potential energy of deformation which can be used for generation of the equations of stability for shells with low value of transverse shear stiffness.

Keywords: a shell, transverse shear, variational problem.

© Нестеров В. А., Лопатин А. В., 2010

УДК 535.5; 538

Е. А. Попов

ВЛИЯНИЕ УПРУГОЙ ПОДСИСТЕМЫ МАГНИТНОГО ДИЭЛЕКТРИКА НА ЕГО ОПТИЧЕСКИЙ СПЕКТР

Обсуждаются результаты изменения тонкой структуры оптического спектра антиферромагнитного диэлектрика под действием одноосного давления. По расщеплению и смещению полос поглощения света определены состояния ионов, порождающие серии полос поглощения, оценено изменение компонент кристаллического поля на поглощающем свет ионе.

Ключевые слова: экситон, магнон, поглощение света, антиферромагнетик.

Спектры поглощения света кристаллов, содержащих 3d-, 4f- и другие ионы, формируются из спектров входящих в их состав ионов, изменённых действием кристаллического поля и взаимодействиями между ионами. Положение полос поглощения хорошо описывается в теории поля лигандов [1]. Интенсивность полос поглощения и тонкая структура спектров не объясняется в рамках этой теории и является следствием коллективных возбуждений различных подсистем кристалла (электронной, магнитной, упругой), хотя исходно и порождается прямыми электронными переходами в 3d-, 4f- и других ионах. В антиферромагнитных кристаллах, содержащих $3d^5$-ионы (Mn^{2+}, Fe^{3+}), основное состояние которых ^6S, прямые электронные переходы запрещены по спину и по чётности в электродипольном приближении. Поэтому интенсивные линии в спектрах поглощения света таких кристаллов имеют сложную многочастичную природу. Связать полосы поглощения спектра с конкретными электронными переходами возможно лишь после анализа реакции полос спектра на внешние воздействия. В данной работе рассмотрено изменение спектра оптического поглощения антиферромагнитного кристалла Rb_2MnCl_4 при действии на него одноосного давления с целью выявления прямых источников серий полос поглощения в районе частот ~26 500 $см^{-1}$, образованных с участием перехода $^6A_{1g}(^6S) \rightarrow {}^4T_{2g}(^4D)$ в ионах Mn^{2+}.

Кристаллы Rb_2MnCl_4 при комнатной температуре имеют тетрагональную структуру симметрии D_{4h}^{17}. При температуре ниже $T_N = 57$ К в кристалле устанавливается антиферромагнитный порядок с анизотропией типа лёгкая ось. Магнитные моменты направлены параллельно C_4 оси симметрии кристалла [2]. Элементарная ячейка при переходе из парамагнитного в антиферромагнитное состояние удваивается, а симметрия понижается до ромбической.

Рассмотрим поляризованные спектры поглощения света кристалла Rb_2MnCl_4 в области энергии ~26 000 $см^{-1}$ при температуре $T = 4{,}2$ К при приложенном к нему внешнем одноосном давлении, направленном вдоль оси C_4 (рис. 1). Поглощение света в этой области энергии происходит благодаря переходам $^6A_{1g}(^6S) \rightarrow {}^4T_{2g}(^4D)$ в ионах Mn^{2+}. Поляризация полос поглощения света, связанных с прямыми переходами (экситонными полосами) и парными экситон-магнонными в этой области спектра, должна соответ-

ствовать правилам отбора, приведённым в [3]. Экситоны симметрии Γ_2 («like»), происходящие из состояний электронов $|T_2, 0\rangle$, должны наблюдаться в σ-поляризации, а экситоны симметрии Γ_5 («unlike»), происходящие из $|T_2, -1\rangle$, – в α- и π-поляризации. В обозначении состояний электронов $|\Gamma, M_\Gamma\rangle$ Γ – орбитальная компонента, M_Γ – проекция орбитального момента. При этом экситонные полосы должны быть магнитодипольными и иметь очень маленькую интенсивность. Экситоны Γ_2 должны порождать магнонные полосы-спутники электродипольной природы в π-поляризации, а экситонам Γ_5 должны сопутствовать магнонные полосы-спутники в α- и σ-поляризации.

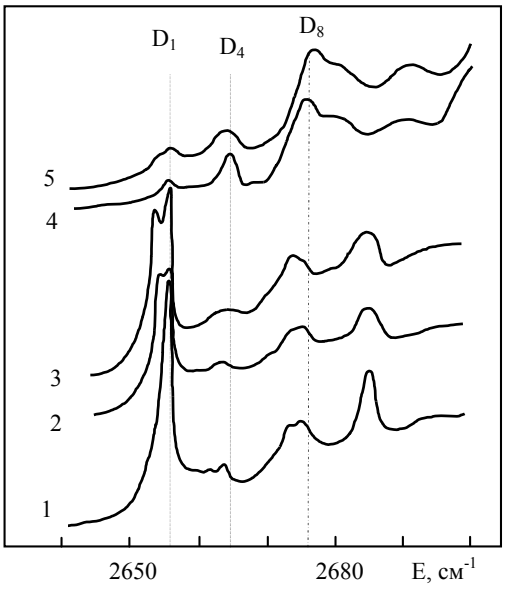


Рис. 1. Изменение спектра поглощения света Rb_2MnCl_4 при температуре $T = 4{,}2$ К под действием одноосного давления $p \perp C_4$(кбар): 1 – 0; 2 – 0,9; 3 – 1,25; 4 – 0; 5 – 1,25. Направление распространения света $k \perp C_4$; направление вектора электрического поля световой волны E: 1, 2 и 3 – $E \parallel C_4$ (σ-поляризация); 4, 5 – $E \perp C_4$ (π-поляризация)

Обнаружить слабые экситонные полосы в спектре не удаётся. Однако косвенно их положение может быть установлено. Для этого можно воспользоваться расчётом изменения положения экситонных полос, происходящих из состояния $^4T_{2g}(^4D)$ иона марганца, при учёте спин-орбитального взаимодействия и тетрагональной компоненты кристаллического поля [4]. При использовании базисных функций $|^4T_{2g}, M_\Gamma, M_S\rangle$ диагонализация гамильтониана $\mathbf{H} = D(M_\Gamma^2 - 2/3) \pm \pm HS_z + H_S$ даёт зависимость энергии и относительной интенсивности прямых электронных переходов, происходящих из $^4T_{2g}(^4D)$-состояний ионов Mn^{2+}. Здесь $H_S = \frac{1}{\xi}\sum_\alpha s_\alpha l_\alpha$; ξ – константа спин-орбитальной связи; s_α и l_α – операторы спинового и орбитального моментов электрона α; D – величина тетрагональной компоненты кристаллического поля; $H = g'\mu_B H_E$ – обменная энергия (g' – g-фактор возбуждённого состояния).

Энергия таких экситонных полос не зависит от величины кубической компоненты кристаллического поля.

Рассмотрим результаты расчёта зависимостей поглощения света для экситонов Γ_2 и Γ_5 от величины тетрагональной компоненты кристаллического поля D при $H = 120$ см$^{-1}$ (рис. 2). Положение «0» на шкале абсцисс выбрано условно. При этом выбрано значение параметра спин-орбитальной связи $\Lambda = 40$ см$^{-1}$, дающее наилучшее согласие поведения экситонных полос при приложении к кристаллу внешнего магнитного поля, переводящего кристалл в спин-флоп фазу ($H > 56$ кЭ) [5]. Здесь $\Lambda = -\left(1/3\sqrt{10}\right)\left\langle {}^4T_2 \left| \sum s \cdot l / \xi \right| {}^4T_2 \right\rangle$.

Если выбрать значение тетрагональной компоненты кристаллического поля $D = 250$ см$^{-1}$, то взаимное расположение полос D_1 , D_4 и D_8 (см. рис. 1) совпадёт со взаимным расположением полос E_1, E_4 и E_8 (см. рис. 2). Тогда естественно их следует считать магнонными спутниками экситонных полос E_1, E_4 и E_8 соответственно. Это полностью согласуется с правилами отбора для магнонных полос спутников экситонов симметрии Γ_2 и Γ_5 [3]. Смешение полос D_1, D_4 и D_8 при увеличении давления соответствует изменению энергии экситонных полос E_1, E_4 и E_8 при увеличении величины тетрагональной компоненты кристаллического поля. Так, энергия полосы E_1 уменьшается, а полос E_4 и E_8 растёт, что наблюдается и для полос D_1, D_4 и D_8.

Форма экситон-магнонных полос D_1, D_4 и D_8 указывает на слабое экситон-магнонное взаимодействие [6]. Поэтому экситонные полосы должны иметь энергию, отличающуюся от энергии экситон-магнонных полос на величину энергии магнона на границе зоны Бриллюэна, т. е. на 80 см$^{-1}$ [7]. Наблюдаемые энергии экситон-магнонных полос и вычисленные энергии экситонных полос приведены ниже:

Полоса	E_1	E_4	E_8	D_1	D_4	D_8
Энергия (см$^{-1}$)	26 472	24 569	26 691	26 552	26 649	26 771

По смещению положения полос D_1, D_4 и D_8, приняв во внимание поведение экситонов в зависимости от величины тетрагональной компоненты кристаллического поля (см. рис. 2), можно оценить изменение кристаллического поля на поглощающем свет ионе. При величине давления 1,25 кбар изменение тетрагональной компоненты кристаллического поля, оцененное по смещению центра тяжести полосы D_1, составляет ~5 см$^{-1}$.

Значительное расщепление полосы D_1 достаточно необычно. В кристалле обменное поле ~10^3 кЭ. Внешним давлением, таким, какое было достигнуто в эксперименте, изменить взаимную ориентацию магнитных моментов подрешёток вряд ли возможно. Поэтому природа расщепления полосы – это бетевское

расщепление, связанное с понижением симметрии в антиферромагнитной фазе до ромбической и появление анизотропии деформаций в базисной плоскости кристалла при действии одноосного давления.

Таким образом, установлены электронные состояния, порождающие тонкую структуру спектра поглощения света в области переходов $^6A_{1g}(^6S) \rightarrow {}^4T_{2g}(^4D)$ в ионах Mn^{2+} кристалла Rb_2MnCl_4 и определены их энергии, а также определено изменение тетрагональной компоненты кристаллического поля на ионе марганца при деформации кристалла под действием одноосного давления.

Библиографические ссылки

1. Берсукер И. Б. Электронное строение и свойства координационных соединений. Л. : Химия, 1976.

2. Magnetic structure and two-dimensional behavior of Rb_2MnCl_4 and Cs_2MnCl_4 / A. Epstain, E. Gurewitz, J. Makovsky, H. Shaked // Phys. Rev. 1970. Vol. B2. № 9. P. 3703–3706.

3. Попов Е. А. Симметрийный анализ элементарных возбуждений различных подсистем кристалла // Вестник НИИСУВПТ. Красноярск, 2009.

4. Optical and Magnetic Circular Dichroism Study of the 2D Antiferromagnet Rb2MnCl4 / A. Vervoitte, J. C. Canit, B. Briat, U. Cambli // Phys. Stat. Sol. (b). 1984. Vol. 124. № 1. P. 87–102.

5. Попов Е. А. Оптические и магнитооптические свойства антиферромагнитных хлоридов марганца : дис. … д-ра физ.-мат. наук. Красноярск, 2004.

6. Попов Е. А., Овчинников С. Г. Магнонные полосы-спутники в оптическом спектре антиферромагнитного Rb_2MnCl_4 // ФТТ. 2003. Т. 45. Вып. 8. С. 1429–1431.

7. Попов Е. А. Тонкая структура оптического спектра и многочастичные возбуждения в Rb_2MnCl_4 // Изв. вузов. Физика. 2003. № 10. С. 10–13.

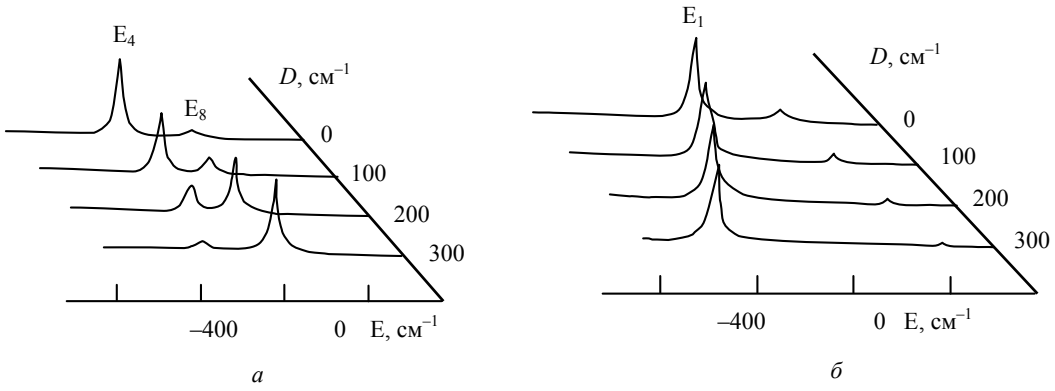


Рис. 2. Зависимость формы и энергии полос поглощения экситонов, происходящих из электронных состояний: $|T_2, -1\rangle$ (*слева*) и $|T_2, 0\rangle$ (*справа*) [4] от тетрагональной компоненты *D* кристаллического поля

E. A. Popov

INFLUENCE OF NONRIGID SUBSYSTEM OF MAGNETIC DIELECTRIC ON ITS OPTICAL SPECTRUM

In the article we discuss results of variations of fine structure of optical spectrum of antiferromagnetic dielectric under uniaxial pressure. Basing on light absorption lines splitting and displacement we determined ion states, initiating light absorption line series and estimated variation of crystal field components at light-absorbing ion.

Keywords: exciton, magnon, light absorption, antiferromagnet.

© Попов Е. А., 2010

УДК 681.786.23; 004.9

В. О. Рутковский, М. А. Рутковская

МЕТОД ПОЛУЧЕНИЯ ТРЕХМЕРНЫХ ЦИФРОВЫХ МОДЕЛЕЙ ТЕХНИЧЕСКИХ ОБЪЕКТОВ, ОСНОВАННЫЙ НА ПРИМЕНЕНИИ ИСКУССТВЕННЫХ ТЕКСТУР

Представлен новый метод пассивного оптического трехмерного сканирования, основанный на предварительной подготовке измеряемого объекта путем нанесения на него искусственной текстуры и обработке результатов с несколькими уровнями детализации.

Ключевые слова: 3D-сканер, оптические измерения, текстура, калибровка.

Трехмерные модели технических объектов могут быть созданы различными методами прямого моделирования, а также методами обратного моделирования – получением трехмерных моделей по существующим физическим объектам или имеющимся компьютерным моделям. Первым этапом обратного моделирования обычно является процесс 3D-сканирования, позволяющий с высокой точностью определить геометрию объекта. Технология 3D-сканирования существует более 14 лет, но только в последнее время возможности программно-аппаратного обеспечения сделали ее доступной для различных областей применения. При выборе метода измерения важны следующие критерии:

– точность – степень соответствия данных измерения реальному физическому объекту;

– уровень шума – хаотическое отклонение точек от измеряемой поверхности;

– разрешение – оно должно быть менее половины размера самой мелкой существенной детали объекта. Прямой связи разрешения с точностью не существует;

– диапазон измерений – минимальное и максимальное расстояние от объекта до сканера. Если размер объекта превышает диапазон, необходимо сделать множество измерений. Близкий критерий: покрытие – возможность измерять труднодоступные участки;

– универсальность – возможность измерять объекты различных размеров, форм и оптических свойств в различных условиях освещения. Позволяет уменьшить количество необходимого оборудования;

– мобильность системы сканирования, включая расчёт времени измерения;

– время измерения – состоит из времени установки (включая время на калибровку сканера, а также время на подготовку объекта) и времени сканирования;

– простота применения – возможность подготовить объект и провести сканирование персоналом, не имеющим специальных знаний и навыков. При этом степень контроля над процессом сканирования и соответственно качество измерения снижается;

– безопасность использования – некоторые типы измерительного оборудования вредны для органов зрения и должны применяться с соблюдением правил безопасности.

3D-сканирование нашло применение в различных областях.

В промышленном производстве с помощью 3D-сканирования осуществляется:

– контроль точности изготовления – устанавливается степень отличия от проектной документации;

– обратная инженерия (Reverse engineering) – раскрытие дизайна, функционирования и способа производства продукции конкурентов;

– цифровое архивирование для воспроизводства устаревших компонентов, изделий с утерянной конструкторской документацией, а также геометрических моделей из устаревших систем проектирования без необходимости повторного проектирования;

– разработка компонентов для существующих товаров и изделий, разработка упаковки по форме физического объекта (переоборудование и вторичный рынок);

– репродукция в масштабе существующих объектов (модели и макеты);

– производство оснастки для ремонта, обновления и замены изношенных деталей.

3D-сканирование широко применяется в дизайне. Это обеспечивает эргономичность – т. е. сочетание проектируемого изделия в CAD-системе с формой тела человека, траекторией движения конечностей и заданным расположением относительно органов чувств; эстетичность – объекты, созданные дизайнером вручную, сканируются для получения цифровой модели (например, глиняные модели, используемые при проектировании корпусов автомобилей). В промышленном дизайне отсканированный объект используется для дизайнерской работы. Отсканированные данные отправляются в 3D CAD-программу, где они становятся базой или ссылкой для проектировочных изменений.

В области искусства и развлечения с помощью 3D-сканирования создаются изображения для трехмерного кино и телевидения, а также 3D-цифровые объекты для игр, анимации, спецэффектов и презентаций.

3D-сканирование применяется также в медицине для репродукции человеческой анатомии для изделий, изготовленных на заказ, или вспомогательных изделий, таких как протезы, слуховые аппараты, мосты/коронки, при создании имплантатов и хирургических инструментов.

3D-сканирование может сыграть важную роль в сохранении культурного наследия, поскольку с его

помощью осуществляется каталогизация, сохранение или восстановление культурных артефактов, скульптур, исторически значимых объектов, зданий, архитектурных орнаментов, а также археологических находок.

Для получения 3D-моделей в настоящее время применяются различные методы измерения формы (рис. 1) [1]. Среди бесконтактных методов измерения наибольшее распространение получили активные методы с применением источников света в виде лазерных сканеров или структурированного света, а также пассивный метод стереоскопического соответствия.

Особенности, область применения, достоинства и недостатки этих методов измерения см. в таблице.

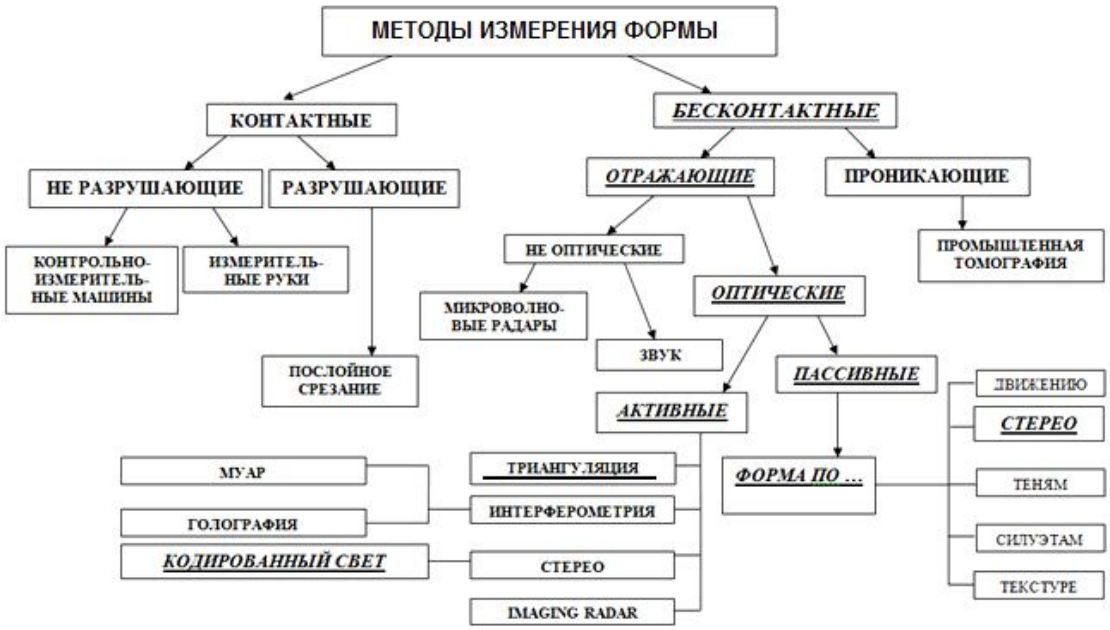


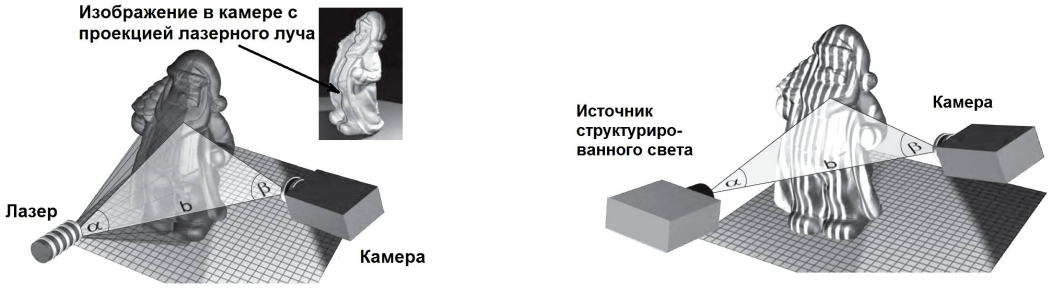
Рис. 1. Основные методы измерения, применяемые для получения трехмерных моделей

Бесконтактные методы измерения формы объекта

Метод	Лазерное сканирование	Структурированный свет	Стереоскопическое соответствие
Принцип работы (рис. 2) [2]	На измеряемый объект проецируется лазерный луч (в виде точек, линии или сетки). Проекция луча снимается камерой и обрабатывается (рис. 2, *а*)	На объект проецируется набор изображений в виде полос с помощью проектора. Изображение снимается камерой и обрабатывается (рис. 2, *б*)	Используется множество камер (или одна камера и разные точки съемки). Координаты точек получаются в результате компьютерной обработки
Основное использование	В системах автоматизированного контроля на производстве	Для измерения лица, частей тела человека, предметов искусства	Для измерения любых, обычно крупных объектов, которые можно снять фотокамерой
Основные недостатки	Большое время сканирования. Сложно измерять подвижные объекты. Опасно для зрения	Невозможна работа вне помещений, желательно затемнение	Трудно автоматизировать. Большие вычислительные ресурсы, множество камер или точек съемки
Основные преимущества	Возможность сканирования очень маленьких объектов и измерений в различных условиях освещения. Простая калибровка	Высокая скорость сканирования – миллионы точек в секунду. Лучший метод измерения подвижных объектов, людей. Безопасно для зрения	Единственный метод измерения, если установка проектора или лазера невозможна. Возможно измерение выбранных точек большого объекта
Точность измерения	Высокая	Высокая	Высокая
Разрешение	Высокое, если время сканирования велико	Высокое, зависит от разрешения проектора	Зависит от разрешения камеры
Универсальность	Требуется различное аппаратное обеспечение для каждой размерной зоны. Качество сильно зависит от оптических свойств объекта	Требуется различное аппаратное обеспечение для каждой размерной зоны. Качество сильно зависит от оптических свойств объекта	Измеряются только точки-маркеры на объекте. Мало зависит от оптических свойств объекта

Окончание таблицы

Метод	Лазерное сканирование	Структурированный свет	Стереоскопическое соответствие
Диапазон измерений	От миллиметров до десятков метров	От 50 мм до 10 м	Не ограничен
Мобильность	Низкая. Существуют мобильные исполнения	Средняя	Высокая. Достаточно камеры и калибров
Время измерения	Большое (от десятков секунд до нескольких минут на измерение)	Очень малое, от миллисекунд до десятка секунд	Время сканирования мало, но время и ресурсы последующей обработки велики
Простота применения	Средняя	Низкая	Высокая
Уровень шумов	Низкий	Средний	Средний
Безопасность	Низкая	Средняя	Высокая



а *б*

Рис. 2. Принцип работы лазерного сканирования (*а*) и сканирования с применением источника структурированного света (*б*)

Современное оборудование объемного сканирования имеет ряд существенных недостатков: высокую цену, низкую универсальность, чувствительность к условиям освещения и оптическим свойствам измеряемых объектов. Проанализировав недостатки трех основных методов, применяемых для 3D-сканирования, авторами статьи предлагается новый метод получения трехмерной геометрии тел сложной формы. Метод основан на усовершенствовании существующего метода стереоскопического соответствия [3; 4].

Суть метода заключается в следующем.

Проводится предварительная подготовка измеряемого объекта путем нанесения на него искусственной текстуры, состоящей из разреженных точек и линий (полученных аэрографией, фломастером и т. п.), дополнительно к опорным меткам, применяемым в исходном методе [4]. Текстура должна быть контрастна объекту и может быть многоцветной. Нанесение текстуры не слишком усложнит подготовку объекта. Обычно при измерении активными методами на объект наносится матовое белое покрытие.

Затем определяются координаты центров опорных меток, далее – геометрия пространственных кривых (рис. 3) и, наконец, строится плотное облако точек по искусственной текстуре.

При использовании данного метода возможно два основных варианта измерений:

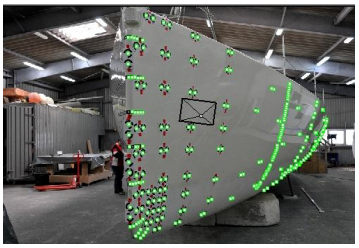
– со свободной камерой. Кроме измеряемого объекта в кадре может находиться одна или несколько калибровочных поверхностей со специально нанесенным рисунком для предварительного грубого определения положения камеры в пространстве и уменьшения объема последующих вычислений;

– с использованием поворотного стола и устройства точного позиционирования камеры. В этом случае размеры объекта будут ограничены размерами устройств позиционирования, но точность существенно вырастет.

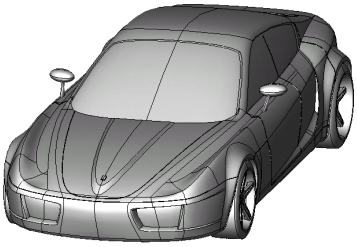
Для получения опорных точек объект обклеивается кодированными метками с размерами 1…2,5 % зоны измерения. Метка представляет собой напечатанные на принтере грубо нарезанные квадратики с разноцветными углами. Центральная часть метки – это окружность с двумя черными и двумя белыми секторами. При обработке снимков центр метки находится с субпиксельной точностью по изображению окружности и точки пересечения секторов. Задача нахождения соответствия кодированных меток решается путем распознавания отдельных меток по цветам их углов с помощью библиотеки OpenCV (URL: http://sourceforge.net/projects/opencvlibrary).

Для вычисления координат точки искусственной текстуры необходимо обработь минимум два снимка, большее количество снимков повышает точность

измерения за счет использования обработки методами математической статистики. Для измерения установим объект в фиксированной системе координат. Положение опорной точки A ($X_AY_AZ_A$) определяется в глобальной системе координат XYZ, связанной с объектом, или a ($x_ay_az_a$) в фотограмметрической системе координат xyz (рис. 4).



а



б

Рис. 3. Процесс получения первичных опорных точек (*а*) и кривых на поверхности объекта (*б*)

Матрица трансформации вращения R может быть определена как

$$R = \begin{bmatrix} \cos\varphi\cos\chi & \sin\omega\sin\varphi\cos\chi + \cos\omega\sin\chi & -\cos\omega\sin\varphi\cos\chi + \sin\omega\sin\chi \\ -\cos\varphi\sin\chi & -\sin\omega\sin\varphi\sin\chi + \cos\omega\cos\chi & \cos\omega\sin\varphi\cos\chi + \sin\omega\cos\chi \\ \sin\varphi & -\sin\omega\cos\varphi & \cos\omega\cos\varphi \end{bmatrix}, \quad (1)$$

где ω, φ, χ – последовательные трансформации вращения первичных осей x', y', z' (оси модели) во вторичные оси x, y, z (оси базовой системы координат) (рис. 4, *а*) [3].

Если переместить начало координат на расстояние T_x, T_y, T_z и применить масштаб K, то получим семь параметров трансформации T_x, T_y, T_z, ω, φ, χ, K [4]. Принцип центральной перспективной проекции показан на рис. 4, *б*. Точка на изображении светочувствительной матрицы камеры a (x_a,y_a, $-f$) с фокусным расстоянием f, центр перспективы 0 ($X_0Y_0Z_0$) и соответствующая точка объекта A($X_AY_AZ_A$) находятся на одной прямой $A0a$. Уравнение колинеарности позволяет найти координаты изображения (x_a,y_a) по координатам объекта ($X_AY_AZ_A$).

$$x_a = -f\frac{(r_{11}(X_A - X_0) + r_{12}(Y_A - Y_0) + r_{13}(Z_A - Z_0)}{(r_{31}(X_A - X_0) + r_{32}(Y_A - Y_0) + r_{33}(Z_A - Z_0)}; \quad (2)$$

$$y_a = -f\frac{(r_{21}(X_A - X_0) + r_{22}(Y_A - Y_0) + r_{23}(Z_A - Z_0)}{(r_{31}(X_A - X_0) + r_{32}(Y_A - Y_0) + r_{33}(Z_A - Z_0)},$$

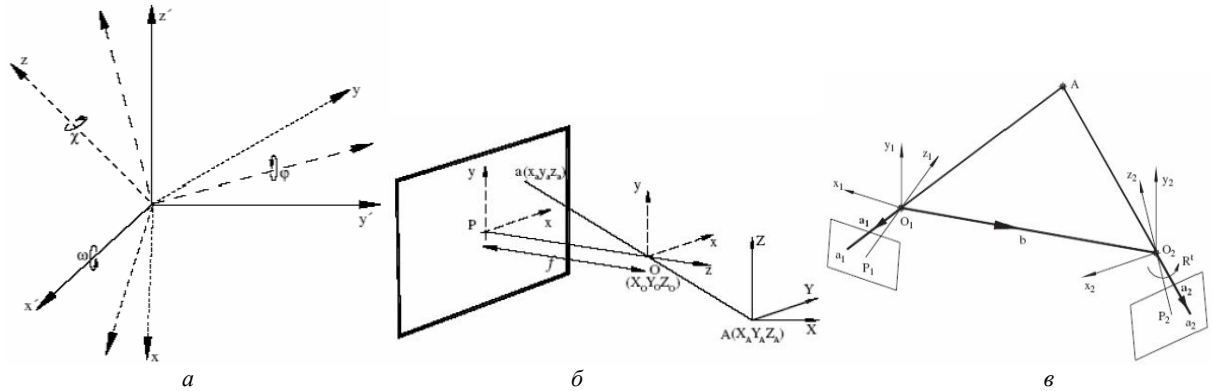
где r_{ij} – элементы матрицы вращения R, для которой необходимо определить углы поворота, т. е. нужно решить уравнение для плоской проекции точки на матрицу камеры. Техника взаимного соответствия изображений дает возможность одновременного нахождения положения камеры относительно объекта и исправления искажений оптики (калибровки). Эти уравнения могут быть записаны в виде

$$F(x, b, a) = 0 \quad (3)$$

или

$$\begin{vmatrix} 1 & \frac{b_y}{b_x} & \frac{b_z}{b_x} \\ x_1 & y_1 & -f \\ x_2 & y_2 & z_2 \end{vmatrix} = 0, \quad (4)$$

где b_x, b_y и b_z – расстояния между центром проекции двух фотографий (рис. 4, *в*) в системе XYZ; x_1, y_1 – координаты точки на проекции первой камеры; f – фокусное расстояние; x_2, y_2, z_2 – координаты точки на проекции второй камеры относительно координатной системы первой.



а *б* *в*

Рис. 4. Матрица трансформации вращения (*а*); центральная перспективная проекция на сенсор камеры (*б*); проекция трехмерной точки A в точки a_1 и a_2 на два различных снимка с различными положениями камеры (*в*)

С учетом матриц вращения и перемещения получим

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = R \begin{bmatrix} x_2' \\ y_2' \\ z_2' \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}. \quad (5)$$

Здесь R определяется из уравнения (1). После определения положения координат камеры по первому снимку положение остальных снимков будет определяться в системе координат первого.

Чтобы определить положение в абсолютной системе координат сцены, нужно описать геометрию опорных точек объекта в глобальной системе координат с помощью ориентации. Положение оси Z определяется как нормальное к плоскости, на которой располагается измеряемый объект. Масштабный фактор определяется по данным обработки множества снимков с элементом геометрии с известным размером. Выравнивание объекта достигается вращением относительно осей X и Z (т. е. $\varphi = 0$ в матрице вращения).

Уравнения трансформации для фотограмметрической системы (xyz) относительно глобальной системы координат (XYZ) примут вид

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = K R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}, \quad (6)$$

где $(X_0Y_0Z_0)$ – координаты камеры в глобальной системе координат измеряемого объекта; K – предварительно вычисленный для данной камеры масштабный коэффициент (при измерениях со свободной камерой вычисляется обработкой изображений калибра с известной геометрией); R – матрица вращения с углами вращения ω и χ, полученными ранее ($\varphi = 0$). Координаты любой опорной точки в системе координат (XYZ) будут вычисляться из уравнения (6).

Для варианта со свободной камерой ее координаты вычисляются по фотограмметрическим меткам, в варианте с позиционирующим механизмом все угловые положения камеры на снимках известны (возможна дополнительная калибровка по объектам простой формы с известной точной геометрией).

На последнем этапе получаем плотные облака точек в системах координат реперных точек и кривых при помощи сопоставления множества изображений с искусственными текстурами, используя модификацию метода Patch-based Multi-view Stereo [5]. Модификация заключается в использовании компенсации искажений оптики методом Tsai [6] и применении разных весов точек в изображениях точек на матрице. Для обеспечения большей точности и разрешения дополнительно производится съемка выбранных зон с большим увеличением и разрешением.

Метод был испытан при измерении двух объектов, сложных для измерения обычными оптическими методами – стальной полированной сферы с диаметром 50 мм и прозрачного акрилового кубика с размером стороны 50 мм. На объекты были нанесены по 24 опорных метки (очень тонкая пленка 3 × 3 мм) и текстуры из капелек черной туши с плотностью около 0,6 мм. При обработке снимков искуственной текстуры было выявлено около 36 800 точек на кубике и 19 400 на сфере. Погрешность измерения опорных точек со свободной камерой после обработки 72 снимков с разрешением 12 Mpix не превысила 0,02 мм (0,04 %). Погрешность измерения искуственных точек текстуры составила не более 0,05 мм (0,1 %).

Новый метод получения трехмерных цифровых моделей технических объектов, основанный на применении искусственных текстур, отличается следующими достоинствами:

– низкой стоимостью оборудования для сканирования – достаточно использовать обычные цифровые фото или видеокамеры с набором распечатанных меток и подготовить объект путем нанесения искусственных текстур;

– возможностью получения качественных результатов на объектах со сложными оптическими свойствами поверхностей – с высокой отражающей способностью, прозрачностью, высокой степенью поглощения света;

– возможностью съемки подвижных и быстро деформируемых или колеблющихся объектов в условиях естественного освещения;

– простотой последующей обработки. После сканирования другими методами требуется трудоемкая обработка, включающая в себя получение частичных облаков точек, объединение облаков точек в единую модель, наложение на модель сетки кривых и обтягивание ее лоскутами поверхностей. При использовании нового метода сразу получаем геометрическую единую модель объекта, возможно, с нанесенной сеткой кривых. Нанесение кривых на этапе подготовки объекта проще, чем на этапе обработки компьютерной модели;

– возможностью сохранения естественного цвета и текстуры объектов за счет нанесения разреженной структуры и неполного окрашивания;

– возможностью проработки очень мелких деталей путем нанесения более густого облака точек в выбранных областях и дополнительной съемки этих мест.

Планируется дальнейшая последующая обработка облака точек с целью получения поверхностных и твердотельных NURBs-моделей.

Библиографические ссылки

1. A low cost 3D scanner based on structured light / C. Rocchini, P. Cignoni, C. Montani et al. // Eurographics. 2001. Vol. 20. № 3. P. 1–10

2. Teutsch C. Model-based Analysis and Evaluation of Point Sets from Optical 3D Laser Scanners : PhD Thesis // Magdeburger Schriften zur Visualisierung. Vol. 1. Shaker Verlag, 2007.

3. Control of structural problems in cultural heritage monuments using close-range photogrammetry and computer methods / P. Arias, J. Herraez, H. Lorenzo et al. // Computers and Structures. 2005. № 83. P. 1754–1766.

4. Zhizhuo W. Principles of photogrammetry. Beijing, China : Publishing House of Surveying and Mapping, 1990.

5. Furukawa Y., Ponce J. Accurate, Dense, and Robust Multi-View Stereopsis // IEEE Trans. on Pattern Analysis and Machine Intelligence. 2009. Vol. 32. № 8. P. 1362–1376.

6. Wilson R. G. Tsai camera calibration software [Electronic resourse] // URL: http://www.cs.cmu.edu/~rgw/TsaiCode.html (date of visit: 12.12.2010).

V. O. Rutkovskiy, M. A. Rutkovskaya

METHOD OF 3D DIGITALIZATION OF TECHNICAL OBJECTS BASED ON ARTIFICIAL TEXTURES

It is represented new method of passive refractive optical 3D scanning which is based on preliminary preprocessing of measured objects by creation on its surfaces of artificial textures and post processing on a different level of details.

Keywords: 3D-scanner, close range photogrammetry, passive optical measurement, texture, calibration.

© Рутковский В. О., Рутковская М. А., 2010

УДК 004.7

С. П. Саханский, М. Е. Якобсон

МОДЕЛЬ УПРАВЛЕНИЯ СКОРОСТЬЮ ВЫТЯГИВАНИЯ МОНОКРИСТАЛЛОВ КРЕМНИЯ

Предложена математическая модель формирования скорости вытягивания на установке выращивания монокристаллов кремния по способу Чохральского, которая позволяет автоматизировать процесс ввода программного задания скорости в микропроцессорную систему управления.

Ключевые слова: скорость вытягивания, монокристалл кремния, осевой градиент.

Управление скоростью вытягивания кристаллов полупроводниковых материалов, выращиваемых из расплава по способу Чохральского, заключается в следующем: во вращающемся вокруг своей оси с угловой скоростью $W_т$ тигле с внутренним диаметром D находится жидкий расплавленный металл (рис. 1); кристалл диаметром d вытягивается из расплава со скоростью вытягивания $V_з$ и угловой скоростью вращения $W_з$; процесс проводится в камере, наполненной инертным газом аргоном. Общий вид микропроцессорной установки вытягивания кристаллов кремния приведен на рис. 2.

Температура расплава формируется за счет управления мощностью нагревателя на основе показаний датчика температуры $T_з$ боковой поверхности нагревателя. Управление скоростью вытягивания кристаллов кремния по программе позволяет сформировать плоский фронт кристаллизации при выращивании прямого конуса, цилиндрической части и обратного конуса кристалла, что необходимо для многих марок кристалла и обеспечивает минимум дислокаций, т. е. нарушений монокристаллической структуры в выращенном монокристалле. Отклонение от цилиндрической формы по длине кристалла повышает вероятность образования дислокаций в местах перегиба формы, за счет повышенных механических напряжений в них, возникающих в процессе охлаждения кристалла.

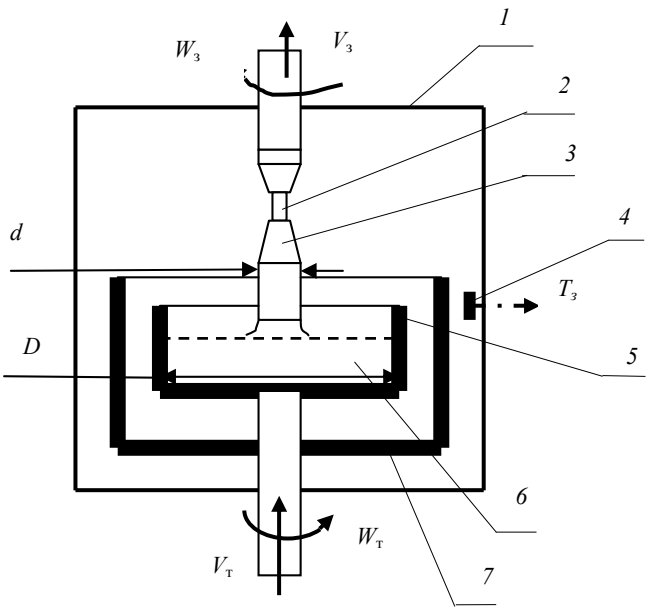


Рис. 1. Схема выращивания кристаллов кремния:
1 – камера; *2* – затравка; *3* – кристалл; *4* – датчик температуры; *5* – тигель; *6* – расплав металла; *7* – нагреватель

По данным многих исследователей [1–4], обеспечение условий получения плоского фронта кристаллизации при выращивании кристаллов кремния и германия (с необходимой величиной осевого градиен-

та температур в выращиваемом кристалле) позволяет получать кристаллы с совершенной структурой (минимумом дислокаций).

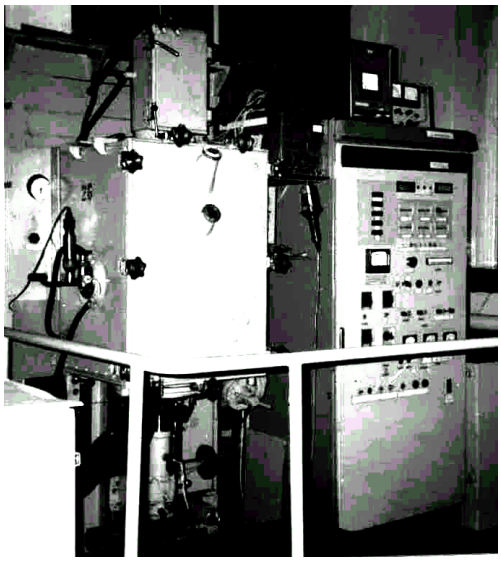


Рис. 2. Общий вид микропроцессорной установки вытягивания кристаллов кремния

Превышение скорости вытягивания кристалла кремния некоторой пороговой величины приводит к прогибу фронта кристалла вверх и отрыву фронта кристаллизации кристалла от расплава, т. е. к срыву процесса монокристаллического роста. Уменьшение скорости ниже другой пороговой величины приводит к опусканию фронта кристаллизации кристалла вниз и касанию твердой фазы фронта кристаллизации дна тигля, что ведет к поломке технологической оснастки и аварийной остановке процесса. Поэтому в практике выращивания монокристаллов кремния актуальной является проблема разработки математического аппарата для задания скорости вытягивания кристалла, исходя из условия плоского фронта кристаллизации. Для получения монокристаллов кремния высокого качества необходимо обеспечить форму фронта кристаллизации, близкую к плоской на участках выращивания прямого конуса, цилиндрической части и обратного конуса кристалла, что дает возможность получить одновременный устойчивый монокристаллический рост. Ниже приведены математические выражения для формирования закона изменения скорости вытягивания $V_з(x)$, которые позволяют автоматизировать процесс ввода этого параметра в систему управления установкой. Модель управления скоростью вытягивания (рис. 3) монокристалла кремния можно представить в виде выражения

$$V(x) = F(\boldsymbol{Z}, \boldsymbol{Y}, \lambda, L(x), x),$$

где λ – величина среднего перегрева расплава; x – координата перемещения вдоль оси кристалла; $L(x)$ – линейный осевой градиент в твердом кристалле; $\boldsymbol{Z}$ – вектор задания геометрии выращиваемого кристалла; $\boldsymbol{Y}$ – вектор задания теплофизических параметров материала.

Математическую зависимость величины среднего перегрева расплава λ относительно температуры кристаллизации материала [3; 4] можно представить виде следующего выражения:

$$\lambda = C_t \cdot \frac{\left(L - C_V \cdot V_з\right)}{d} \qquad (1)$$

где $C_V = \rho_ж \cdot \frac{E}{\lambda_{тв}}$; $C_t = 4 \cdot \frac{\sigma \cdot \lambda_{тв}}{\rho_ж \cdot \lambda_ж \cdot g}$; $\lambda = \left(T - T_к\right)$;

$V_з$ – скорость вытягивания кристалла; $T_к$ – температура кристаллизации материала; T – температура расплава в зоне фронта кристаллизации; L – линейный осевой градиент в твердом кристалле; E – удельная теплота плавления материала; $\lambda_ж$ – коэффициент теплопроводности расплава; $\lambda_{тв}$ – коэффициент теплопроводности кристалла; g – ускорение свободного падения; d – диаметр кристалла; σ – поверхностное натяжение расплава.

В общем случае средний перегрев λ есть функция осевого градиента в кристалле L и скорости вытягивания кристалла $V_з$. Процесс управления основными параметрами выращивания кристалла показан на рис. 4.

Используя линейную аппроксимацию параметров на узловых участках выращивания кристалла, получены выражения для определения задания скорости вытягивания $V_з(x)$ на конусной, цилиндрической части кристалла и при формировании обратного конуса соответственно:

$$V_з(x) = V_0 - \frac{x \cdot \left[V_0 - V_1\right]}{x_1}; \qquad (2)$$

$$V_з(x) = V_1 - \frac{\left(x - x_1\right) \cdot \left[V_1 - V_2\right]}{\left(x_2 - x_1\right)}; \qquad (3)$$

$$V_з(x) = V_2 + \frac{\left(x - x_2\right) \cdot \left[V_3 - V_2\right]}{\left(x_3 - x_2\right)}, \qquad (4)$$

где V_0 – начальная скорость вытягивания кристалла; V_1 – скорость вытягивания кристалла при завершении формирования обратного конуса; V_2 – скорость вытягивания кристалла при завершении формирования цилиндра; V_3 – скорость вытягивания кристалла при завершении формирования обратного конуса; x_1 – координата завершения формирования прямого конуса кристалла; x_2 – координата завершения формирования цилиндрической части кристалла; x_3 – координата завершения формирования обратного конуса кристалла; x – координата перемещения вдоль оси кристалла.

Для автоматического определения значений координат скорости вытягивания в узловых точках (V_0, V_1, V_2, V_3) преобразуем выражение (1) к виду выражения (5), введя технологические поправочные коэффициенты β_i:

$$V_з(d) = \left[L - \frac{\lambda \cdot d_з}{C_t}\right] \cdot \frac{\beta_i}{C_V}, \qquad (5)$$

где $V_з(d)$ – скорость вытягивания кристалла; $d_з$ – задание диаметра кристалла на прямом конусе, цилиндрической части и обратном конусе; β_i – технологический коэффициент снижения скорости (0,9…0,7).

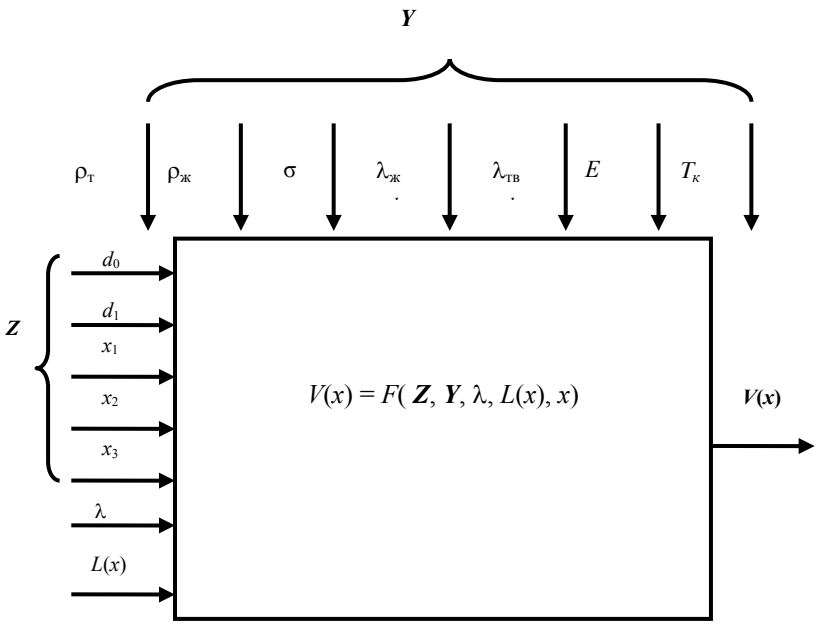


Рис. 3. Модель управления скоростью вытягивания монокристалла кремния

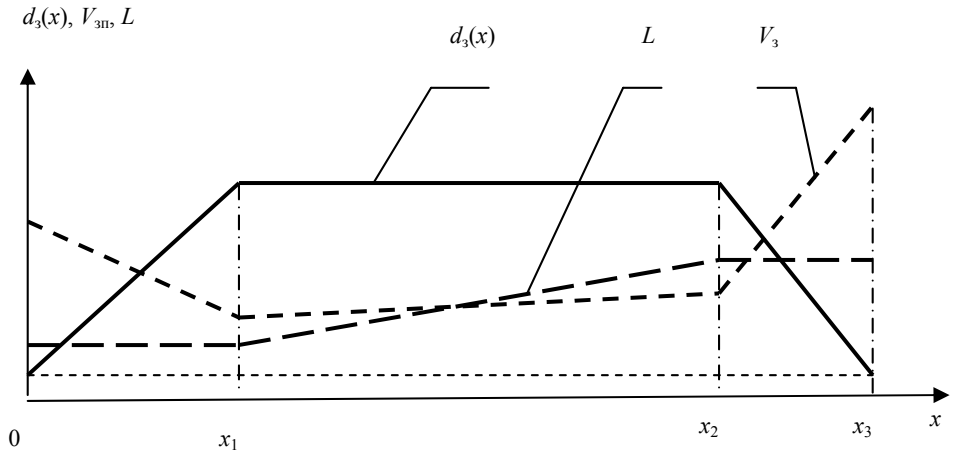


Рис. 4. Задание параметров процесса выращивания кристаллов:
$d_з(x)$ – задание диаметра; $V_з$ – задание скорости вытягивания; L – задание осевого градиента;
x – перемещение вдоль оси кристалла

Из (5) для узловых точек скорости вытягивания получим выражения:

$$V_0 = \left[L_0 - \frac{\lambda \cdot d_0}{C_t} \right] \cdot \frac{\beta_0}{C_V}; \quad (6)$$

$$V_1 = \left[L_0 - \frac{\lambda \cdot d_1}{C_t} \right] \cdot \frac{\beta_1}{C_V}; \quad (7)$$

$$V_2 = \left[L_1 - \frac{\lambda \cdot d_1}{C_t} \right] \cdot \frac{\beta_2}{C_V}; \quad (8)$$

$$V_3 = \left[L_1 - \frac{\lambda \cdot d_0}{C_t} \right] \cdot \frac{\beta_3}{C_V}, \quad (9)$$

где L_0 – осевой градиент в начале цилиндрической части кристалла; L_1 – осевой градиент в конце цилиндрической части кристалла; V_0, V_1, V_2, V_3 – узловые точки скорости вытягивания; d_0 – диаметр вытягиваемой шейки кристалла; d_1 – диаметр цилиндрической части кристалла.

Технологические поправочные коэффициенты β_i введены для корректировки скорости вытягивания, исходя из технологических требований (например, равномерности легирования кристалла по его длине, при допущении некоторой величины прогиба фронта кристаллизации). Осевой градиент в выращиваемом кристалле кремния можно рассчитать на основе непрерывного измерения температуры пирометром в области выращивания на боковой поверхности кри-

сталла (на расстоянии 8…10 мм от расплава) в процессе роста, при этом пирометр устанавливается на смотровое окно камеры установки вытягивания. Другим способом определения осевого градиента в выращиваемом кристалле является метод непрерывного измерения температуры термопарами (введенными в камеру установки), вблизи боковой поверхности выращиваемого кристалла. Применяя микропроцессорные системы управления выращиванием кристаллов кремния, вычисление осевого градиента можно автоматизировать. На основе выражений (1)–(9), с подстановкой в них физических постоянных полупроводникового материала кремния [5], на рис. 5 и 6 приведен пример моделирования задания диаметра кристалла кремния и его скорости вытягивания на всех участках.

Моделирование скоростей вытягивания кристаллов кремния и германия [4] показало, что рабочий диапазон скорости вытягивания германия (0,2…1 мм/мин) и кремния (2…10 мм/мин) отличается на порядок, что согласуется с теплофизическими параметрами данных материалов. Основной трудностью для применения данной модели при выращивании монокристаллов кремния является практическое измерение и вычисление осевых градиентов при выращивании новых марок кристаллов.

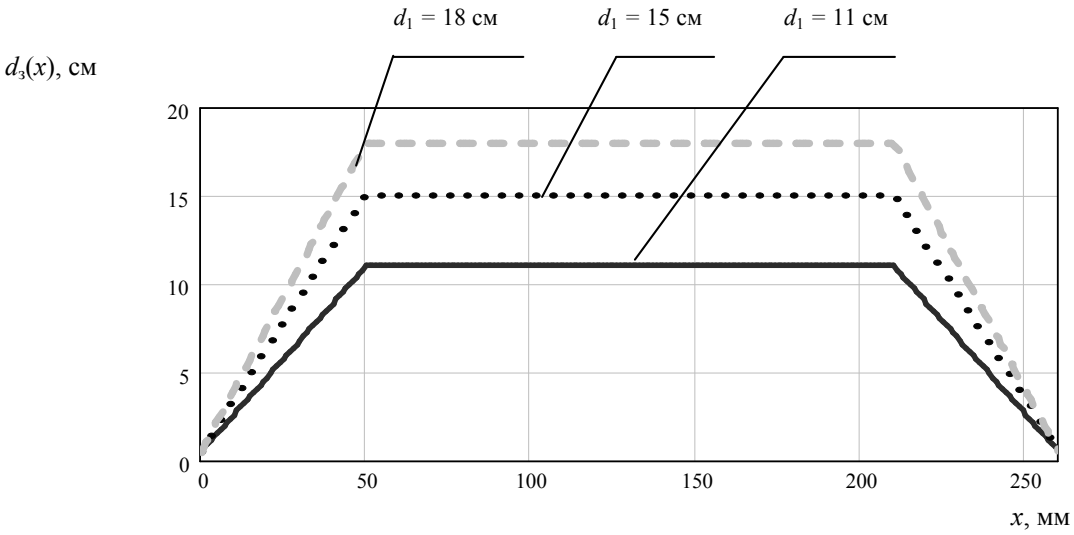


Рис. 5. Задание диаметра кристалла кремния: при x_1 = 50 мм; x_2 = 210 мм; x_3 = 260 мм; d_0 = 0,5 см

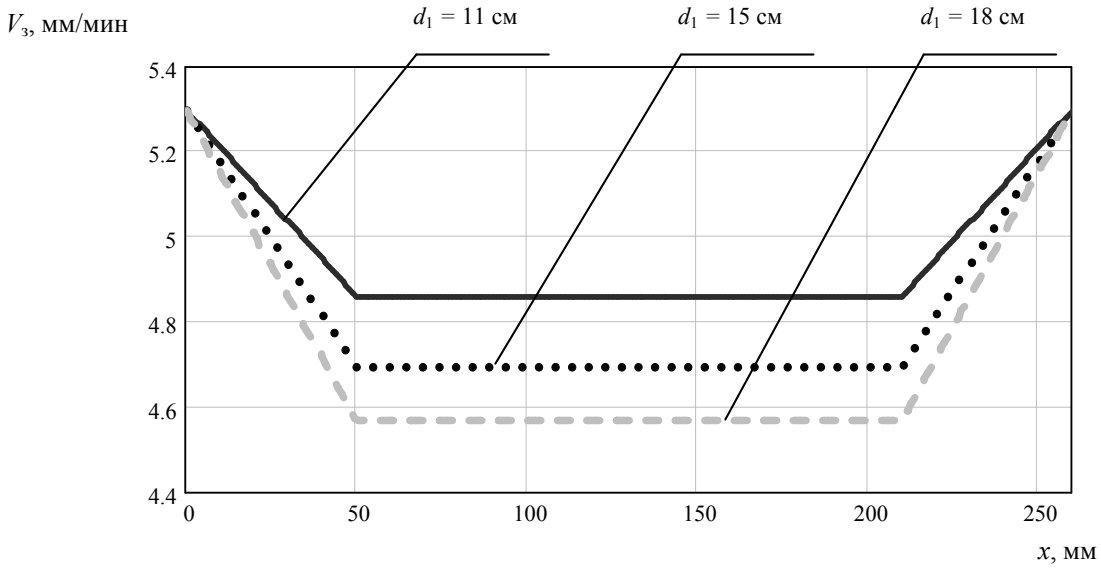


Рис. 6. Задание скорости вытягивания кристалла кремния: при λ = 0,2 °С; L_0 = 15 °С/см; L_1 = 15 °С/см; $\beta_1 = \beta_2 = \beta_3 = \beta_4$ = 0,8; d_0 = 0,5 см; x_1 = 50 мм; x_2 = 210 мм; x_3 = 260 мм; d_0 = 0,5 см

Итак, предложена математическая модель формирования скорости вытягивания монокристаллов кремния, выращиваемых по способу Чохральского. Это позволяет автоматизировать процесс ввода параметра скорости в систему управления и сокращает цикл производства готовой продукции требуемого качества, обеспечивая при выращивании кристалла форму фронта кристаллизации, близкую к плоской.

Библиографические ссылки

1. Лоусон У. Д., Нильсен С., Херл Д. Т. Процессы роста и выращивание монокристаллов : пер. с англ. / под ред. Н. Н. Шефталя. М. : Изд-во иностр. лит., 1963.

2. Kumaragurubaran S., Subramanian C. Growth of paratellurite crystals: effect of axial temperaturte gradient on the quality of the crystals // Crystal Growth. 2000. Vol. 211. P. 276–280.

3. Саханский С. П. Управление скоростью вытягивания на установке выращивания монокристаллов германия // Приборы и системы. Управление, контроль, диагностика. 2008. № 4. С. 7–11.

4. Саханский С. П. Управление процессом выращивания монокристаллов германия : монография / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2008.

5. Бабичев А. П., Бабушкина Н. А., Братковский А. М. Физические величины : справочник / под. ред. И. С. Григорьева, Е. З. Мелихова. М. : Энергоатомиздат, 1991.

S. P. Sakhanskiy, M. E. Yakobson

PULLING RATE ASSIGNMENT OF SILICONIUM SINGLE CRYSTTALS GROWING

Mathematical expression was offered to form pulling set on silicon monocrystal growing system by Czochralski which allows to automate the process of introduction of programmed set of speed in microprocessor control system.

Keywords: speed pulling, silicon single crystal, axial gradient.

© Саханский С. П., Якобсон М. Е., 2010

УДК 621.317.7.023

А. А. Шауэрман, М. С. Жариков, А. В. Борисов

АВТОМАТИЗИРОВАННЫЙ ИЗМЕРИТЕЛЬ КОМПЛЕКСНОГО КОЭФФИЦИЕНТА ОТРАЖЕНИЯ НА ОСНОВЕ ЛОГАРИФМИЧЕСКОГО УСИЛИТЕЛЯ

Описано применение логарифмического усилителя для построения измерителя комплексного коэффициента отражения (ККО) оконечных устройств диапазона СВЧ. Рассмотрены методы измерения ККО датчиком, параметры которого уточняются в процессе калибровки. Проанализированы источники неопределенности измерений.

Ключевые слова: комплексный коэффициент отражения, СВЧ-измерение, логарифмический усилитель, датчик.

Интенсивное развитие систем радиосвязи, в частности спутниковой связи, расширение частотного диапазона, ставит проблему создания точных радиоизмерительных приборов. В настоящее время отечественной промышленностью выпускается явно недостаточное количество таких приборов. Особенно это касается СВЧ-диапазона, который используется в спутниковой связи. От точности согласования антенно-фидерных устройств в диапазоне рабочих частот существенно зависит эффективность передачи информации по каналам связи.

Основным параметром, характеризующим степень согласования высокочастотного тракта, является комплексный коэффициент отражения – комплексная величина, равная отношению амплитуд отражённой и падающей волны. На сегодняшний день для измерения ККО существуют различные приборы, обладающие высокой точностью и широким диапазоном, но стоимость таких приборов не позволяет применять их массово. Другой класс недорогих приборов способен измерять только модуль коэффициента отражения, что не всегда достаточно. Поэтому актуальной задачей является разработка прибора для измерения комплексного коэффициента отражения, с низкой стоимостью.

В самом простейшем случае ККО можно определить с помощью измерителя комплексных отношений, на один вход которого подается отраженная волна, а на второй – сигнал с генератора. Падающая и отраженная волны могут быть выделены с помощью направленных ответвителей. В этом случае ККО определяется по формуле

$$\dot{\Gamma} = B \cdot \frac{\dot{b}_1}{\dot{b}_2},$$

где $\dot{b}_1$ и $\dot{b}_2$ – амплитуды отраженной и падающей волн соответственно; B – коэффициент пропорциональности, определяемый коэффициентом ответвления направленных ответвителей.

Точность такого измерения ограничена технологическими сложностями изготовления прецизионного измерительного преобразователя, особенно сложно выдержать параметры в широкой полосе частот, что неизбежно несет в себе систематическую ошибку измерения. Авторами разработан цифровой измеритель комплексного коэффициента отражения, лишенный указанного недостатка, в разработанной измерительной системе параметры измерительного преобразователя определяются уже после изготовления на этапе калибровки, тем самым устраняется систематическая ошибка, вызванная неидеальностью преобразователя. В качестве измерителя комплексных отношений использован интегральный двуканальный логарифмический усилитель AD8302 (рис. 1).

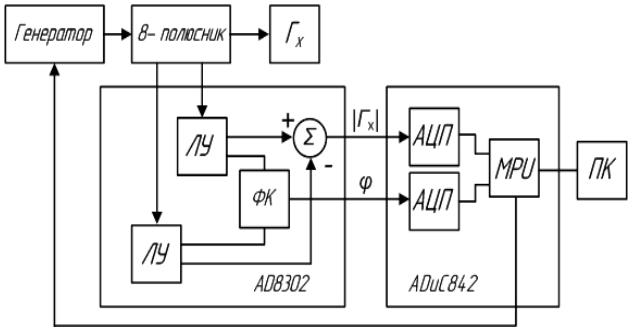


Рис. 1. Структурная схема анализатора сигнала

Измерительная система состоит из управляемого генератора СВЧ, 8-полюсного измерительного преобразователя, к входу которого подключена измеряемая нагрузка (Γ_x), логарифмических усилителей (ЛУ), фазового компаратора (ФК), сумматора, аналого-цифровых преобразователей (АЦП), управляющего микропроцессора (MPU) и персонального компьютера ПК. Микропроцессор и АЦП входят в состав микроконтроллера ADuC842. Логарифмические усилители, фазовый компаратор и сумматор входят в состав интегральной микросхемы AD8302. Определение отношения сигналов выполняется суммированием логарифмов входных сигналов. На выходе сумматора формируются напряжения, пропорциональные отношению модулей двух сигналов, а на выходе ФК – фазовому сдвигу между ними. Эти напряжения переводятся в цифровую форму с помощью АЦП в составе микроконтроллера. Далее результаты измерения передаются в ПК для дальнейшей цифровой обработки и индикации на мониторе. В качестве управляемого генератора использован широкополосный синтезатор частоты ADF4350. Синтезатор позволяет формировать сигнал с частотой от 137,5 до 4 400 МГц. Управление параметрами сигнала производится через 3-проводный интерфейс (ADF4350 data sheet. Analog Devices. Norwood, MA, U.S.A).

В приведенной системе назначение микропроцессора заключается в автоматической регистрации значений напряжений на выходе векторного измерителя и в управлении процессом измерения, калибровки и передачи полученных данных в ПК.

Персональный компьютер в измерительной системе используется не только для индикации и сохранения результата измерения, но и для повышения точности, при этом появляется возможность автоматизации измерительных и вычислительных операций, определения и введения поправок непосредственно в процессе измерения. Одновременно может быть учтено и предельно уменьшено влияние источников случайных погрешностей на результат измерения.

В разработанной системе роль индикатора выполняет микросхема AD8302, представляющая собой полностью интегрированную систему для измерения отношения мощностей двух сигналов и фазового сдвига между ними. На выходе микросхемы формируются напряжения, пропорциональные разности уровней входных сигналов и разности фаз между этими сигналами. Напряжение может быть измерено с помощью внешнего аналого-цифрового преобразователя.

Напряжение на выходе логарифмического усилителя может быть описано выражением

$$V_{OUT} = V_{SLP} \cdot \lg\left(\frac{V_{INA}}{V_{INB}}\right), \quad (1)$$

где V_{OUT} – напряжение на выходе логарифмического усилителя; V_{INA} – амплитуда входного сигнала первого канала; V_{INB} – амплитуда входного сигнала второго канала.

Коэффициент V_{SLP} характеризует наклон передаточной характеристики логарифмического усилителя, имеет размерность В/дБ. Допустимое отношение входных сигналов 60 дБ.

Также на выходе микросхемы формируется напряжение, пропорциональное разности фаз между двумя входными сигналами:

$$V_{PHS} = V_{\Phi} \cdot \left|\Phi\left(V_{INA}\right) - \Phi\left(V_{INB}\right)\right|, \quad (2)$$

где V_{PHS} – напряжение на выходе фазового компаратора; $\Phi\left(V_{INA}\right)$ – фаза сигнала на первом входе; $\Phi\left(V_{INB}\right)$ – фаза сигнала на втором входе; V_{Φ} – некоторый коэффициент пропорциональности, если фазовый сдвиг измеряется в градусах, то размерность будет В/градус.

В качестве измерительного преобразователя может быть использован 8-полюсник. Структурная схема использования 8-полюсника для измерения ККО приведена на рис. 3.

В схеме генератор подключен к первому полюсу, объект измерения ко второму полюсу, а третий и четвертый полюс используются в качестве индикаторных

каналов. Согласно [1; 2], выходящая волна для *j*-го полюса нагруженного многополюсника определяется выражением

$$\dot{b}_j = \frac{\dot{A}_j + \dot{B}_j \cdot \dot{\Gamma}_2}{1 + \dot{C} \cdot \dot{\Gamma}_2} \cdot \dot{a}_\Gamma, \qquad (3)$$

где *A*, *B*, *C* – эквивалентные параметры многополюсника; $j = 3, 4$; $\dot{a}_\Gamma$ – амплитуда волны генератора; $\dot{\Gamma}_2$ – ККО нагрузки, подключенной ко второму порту 8-полюсника.

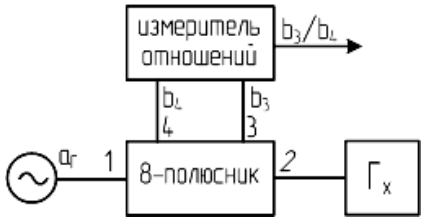


Рис. 2. Структурная схема измерителя ККО с 8-полюсным датчиком

Параметры *A*, *B*, *C* могут быть выражены через матрицу рассеяния (*S*-матрица) нагруженного многополюсника следующим образом:

$$\dot{A}_j = \frac{\det[D]_{(jS1)(22)(11)}}{\det[D]_{(22)}};$$

$$\dot{B}_j = -\frac{\det[D]_{(jS1)(2S2)(11)}}{\det[D]_{(22)}}; \qquad (4)$$

$$\dot{C} = -\frac{\det[D]_{(2S2)}}{\det[D]_{(22)}},$$

где $\det[D]_{(jSk)}$ – определитель матрицы *D*, у которой *j*-й столбец заменен *k*-м столбцом матрицы *D*; $\det[D]_{(jk)}$ – определитель матрицы, полученной из матрицы *D* вычеркиванием *j*-го столбца и *k*-й строки. Матрица *D* может быть получена из матрицы *S* следующим образом [1]:

$$[D] = [1] - [S] \cdot [\Gamma]^{-1},$$

где $[\Gamma]$ – матрица, диагональные элементы которой – комплексные коэффициенты отражений от нагрузок, подключенных к соответствующим портам многополюсника, остальные элементы равны нулю; $[1]$ – единичная матрица.

Если в качестве индикаторного канала использовать измеритель отношений, то для показаний индикатора, используя выражение (4), можно записать

$$\dot{N} = \frac{\dot{b}_3}{\dot{b}_4} = \frac{\dot{A}_3 + \dot{B}_3 \cdot \dot{\Gamma}_x}{\dot{A}_4 + \dot{B}_4 \cdot \dot{\Gamma}_x} = \frac{\dot{A} + \dot{B} \cdot \dot{\Gamma}_x}{1 + \dot{C} \cdot \dot{\Gamma}_x}, \qquad (5)$$

где $\dot{A} = \frac{A_3}{A_4}; \dot{B} = \frac{B_3}{A_4}; \dot{C} = \frac{B_4}{A_4}; \dot{N}$ – показание индикатора, $\dot{\Gamma}_x$ – значение измеряемой нагрузки, подключенной ко второму порту 8-полюсника.

Таким образом, эквивалентные параметры 8-полюсного датчика при использовании в качестве индикатора измерителя отношений описываются тремя параметрами: $\dot{A}, \dot{B}, \dot{C}$, при этом показания индикатора не зависят от амплитуды генератора.

Из выражения (5) по показанию индикатора определяется значение измеряемой нагрузки:

$$\dot{\Gamma}_x = \frac{\dot{N} - \dot{A}_x}{\dot{B} - \dot{N} \cdot \dot{C}}. \qquad (6)$$

Выражение (6) может быть использовано для измерения неизвестного ККО (Γ_x) оконечного устройства СВЧ-диапазона.

Анализируя измерительное уравнение (6), можно сделать вывод, что наибольший динамический диапазон измеряемых значений ККО достигается при условии

$$\dot{A} \Big/ \dot{B} \to 0,$$
$$\dot{C} \to 0. \qquad (7)$$

Сформулируем требование для измерительного датчика для обеспечения максимального динамического диапазона измеряемого ККО. Используя (4), можно записать выражения для нахождения эквивалентных параметров датчика, через параметры нагруженного многополюсника:

$$\dot{A} = \frac{\det[D]_{(3S1)(22)(11)}}{\det[D]_{(4S1)(22)(11)}},$$

$$\dot{B} = \frac{\det[D]_{(3S1)(2S2)(11)}}{\det[D]_{(4S1)(22)(11)}},$$

$$\dot{C} = \frac{\det[D]_{(4S1)(2S2)(11)}}{\det[D]_{(4S1)(22)(11)}}.$$

Если индикаторные каналы идеально согласованы ($\Gamma_3 = \Gamma_4 = 0$), тогда

$$\dot{A} = \frac{\dot{S}_{3,1}}{\dot{S}_{4,1}};$$

$$\dot{B} = -\frac{\dot{S}_{2,1} \cdot \dot{S}_{3,2} - \dot{S}_{2,2} \cdot \dot{S}_{3,1}}{\dot{S}_{4,1}}; \qquad (8)$$

$$\dot{C} = -\frac{\dot{S}_{2,1} \cdot \dot{S}_{4,2} - \dot{S}_{2,2} \cdot \dot{S}_{4,1}}{\dot{S}_{4,1}}.$$

Для выполнения условия (7) параметры многополюсника $\dot{S}_{3,1}$, $\dot{S}_{2,2}$, $\dot{S}_{4,2}$ должны быть наименьшими, а параметры $\dot{S}_{4,1}$, $\dot{S}_{2,1}$ и $\dot{S}_{3,2}$ – наибольшими. Исходя из требований, можно заключить, что в качестве 8-полюсного датчика может быть использован направленный ответвитель на связанных микрополосковых линиях.

Расчет электромагнитной структуры датчика производился в среде автоматизированного проектирования Microwave Office AWR. Топология датчика и по-

лученные *S*-параметры в полосе частот приведены на рис. 3 и 4 соответственно.

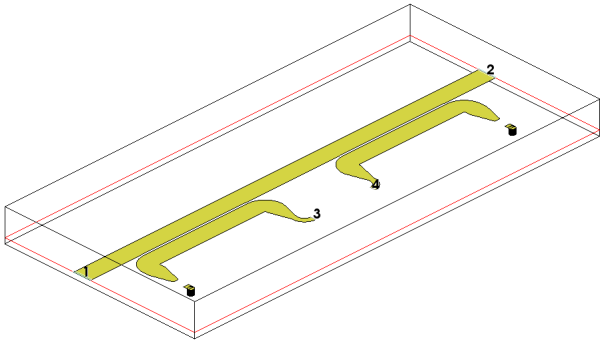


Рис. 3. Топология датчика

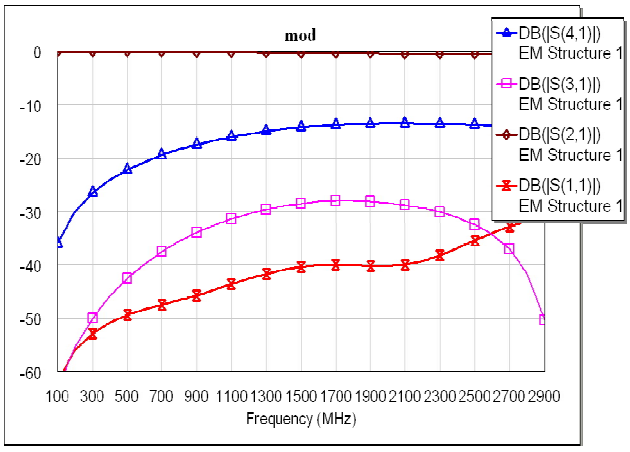


Рис. 4. График зависимости *S*-параметров 8-полюсного датчика от частоты

Несмотря на то что эквивалентные параметры могут быть найдены аналитическим путем на этапе проектирования, при таком подходе датчик должен быть изготовлен с высокой точностью, что влечет за собой ряд технологических проблем. Кроме того, при проектировании датчика довольно сложно учесть ККО коаксиально-полоскового перехода, а также переходников для нагрузок различных сечений. Поэтому оптимальным решением представляется нахождение эквивалентных параметров датчика путем его калибровки. Для этого к измерительному порту датчика подключаются три нагрузки с известными значениями ККО $\left(\dot{\Gamma}_1, \dot{\Gamma}_2, \dot{\Gamma}_3\right)$, на основании (5) составляется система уравнений

$$\begin{aligned} \dot{A} + \dot{B}\cdot\dot{\Gamma}_1 - \dot{N}_1\cdot\dot{\Gamma}_1\cdot\dot{C} &= \dot{N}_1; \\ \dot{A} + \dot{B}\cdot\dot{\Gamma}_2 - \dot{N}_2\cdot\dot{\Gamma}_2\cdot\dot{C} &= \dot{N}_2; \\ \dot{A} + \dot{B}\cdot\dot{\Gamma}_3 - \dot{N}_3\cdot\dot{\Gamma}_3\cdot\dot{C} &= \dot{N}_3. \end{aligned} \quad (9)$$

Решение этой системы позволяет определить эквивалентные параметры $\dot{A}, \dot{B}, \dot{C}$:

$$\begin{aligned} \dot{A} &= \Delta_1\cdot\Delta^{-1}; \\ \dot{B} &= \Delta_2\cdot\Delta^{-1}; \\ \dot{C} &= \Delta_3\cdot\Delta^{-1}, \end{aligned} \quad (10)$$

где Δ – определитель системы уравнений (9); $\Delta_i\,(i=1,2,3)$ – определители, полученные из Δ заменой *i*-го столбца на столбец свободных членов системы (9).

При разработке любого измерительного средства всегда необходимо определить погрешности измерений, свойственные разработанной системе. Так как измеряемая величина комплексная, то оценка погрешности измерения производится отдельно для модуля и для фазы. Для анализа погрешности нахождения ККО необходимо сопоставить измерительное уравнение (6) и выражения для нахождения эквивалентных параметров датчика на этапе калибровки (10):

$$\begin{aligned} \dot{\Gamma}_x = \frac{\dot{N}-\dot{A}_x}{\dot{B}-\dot{N}\cdot\dot{C}} = \frac{\dot{N}-\Delta_1\cdot\Delta^{-1}}{\Delta_2\cdot\Delta^{-1}-\dot{N}\cdot\Delta_3\cdot\Delta^{-1}} = \\ = \frac{\Delta\cdot\dot{N}-\Delta_1}{\Delta_2-\dot{N}\cdot\Delta_3}. \end{aligned} \quad (11)$$

Из формулы (11) следует, что погрешность определения комплексного коэффициента отражения зависит от значения коэффициентов отражения калибровочных устройств $\left(\dot{\Gamma}_1, \dot{\Gamma}_2, \dot{\Gamma}_3\right)$, от погрешности их аттестации и от погрешностей измерения $\dot{N}, \dot{N}_1, \dot{N}_2, \dot{N}_3$. На основании (11) можно записать выражения для модуля и фазы ККО:

$$\begin{aligned} \left|\dot{\Gamma}_x\right| &= \left|\frac{\Delta\cdot\dot{N}-\Delta_1}{\Delta_2-\dot{N}\cdot\Delta_3}\right|; \\ \varphi_{\dot{\Gamma}_x} &= \arg\left(\frac{\Delta\cdot\dot{N}-\Delta_1}{\Delta_2-\dot{N}\cdot\Delta_3}\right). \end{aligned} \quad (13)$$

Проанализируем погрешность измерения модуля и фазы $\dot{N}$ с помощью логарифмического усилителя. Модуль отношения $\dot{N}$ по линейной шкале можно найти из выражения (1):

$$\left|\dot{N}\right| = \frac{V_{INA}}{V_{INB}} = 10^{\frac{V_{OUT}}{V_{SLP}}} = 10^{0{,}05\cdot N_{\text{дБ}}}, \quad (14)$$

где $N_{\text{дБ}}$ – отношение амплитуд сигналов, выраженное в дБ.

Тогда погрешность нахождения $\Delta\left|\dot{N}\right|$ будет определяться как

$$\Delta\left|\dot{N}\right| = \frac{d\left|\dot{N}\right|}{dN_{\text{дБ}}}\cdot\Delta N_{\text{дБ}} = 0{,}05\ln(10)\cdot 10^{0{,}05\cdot N_{\text{дБ}}}\cdot\Delta N_{\text{дБ}}, \quad (15)$$

где $\Delta N_{\text{дБ}}$ – ошибка определения модуля отношений двух волн, выраженная в дБ.

Ошибка $\Delta N_{\text{дБ}}$ определяется по документации на микросхему AD8302 и является функцией от $N_{\text{дБ}}$. Из

документации на микросхему AD8302 (AD8302 data sheet. Analog Devices. Norwood, MA, U.S.A) ясно, что ошибка определения модуля при температуре окружающей среды 25 °С в широком диапазоне от –20 дБ до 20 дБ не превышает 0,2 дБ, но при выходе за указанный диапазон резко возрастает. Ошибка фазы сильно зависит от фазового соотношения анализируемых сигналов, при малой разности фаз и при разности фаз 180° ошибка возрастает. Поэтому для анализа погрешности следует рассматривать ошибку модуля и фазы не как некоторое постоянное число, а как функцию

$$\Delta\varphi_N = f\left(\varphi_N\right),$$
$$\Delta N_{\text{дБ}} = f\left(N_{\text{дБ}}\right),$$

При этом считается, что погрешность аналого-цифрового преобразования пренебрежимо мала. Для расчета погрешностей измерения ККО функции $\Delta\varphi_N\left(\varphi_N\right)$ и $\Delta N_{\text{дБ}}\left(N_{\text{дБ}}\right)$ могут быть заменены интерполяционным полиномом.

Расчет погрешностей произведен численными методами в среде графического программирования LabVIEW. Коэффициенты отражения калибровочных устройств выбирались равными $0{,}166\cdot e^{j10^\circ}$, $0{,}333\cdot e^{j90^\circ}, 1\cdot e^{j160^\circ}$ с погрешностью аттестации $\pm 0{,}01$ по модулю и $\pm 1^\circ$ по фазе. В качестве измерительного датчика использовался датчик с теоретически рассчитанными на этапе проектирования эквивалентными параметрами, расчет производился для измерений на частоте 1 900 МГц.

Рассмотрим графики зависимости погрешности измерения модуля и фазы коэффициента отражения нагрузки от фазы и модуля измеряемой (рис. 6, 7) нагрузки. Погрешности приведены для доверительного интервала 0,95.

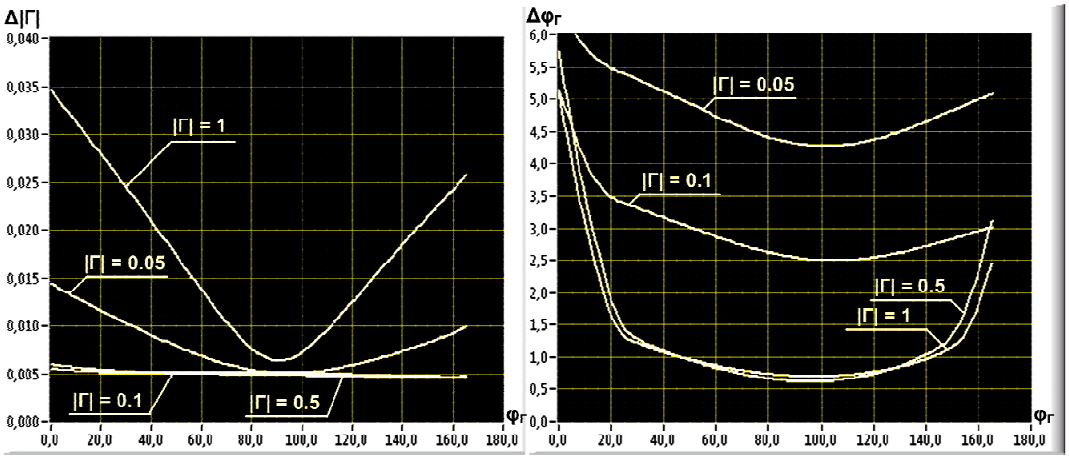


Рис. 6. Зависимость погрешности измерения модуля (*а*)
и фазы ККО (*б*) от фазы измеряемого коэффициента отражения

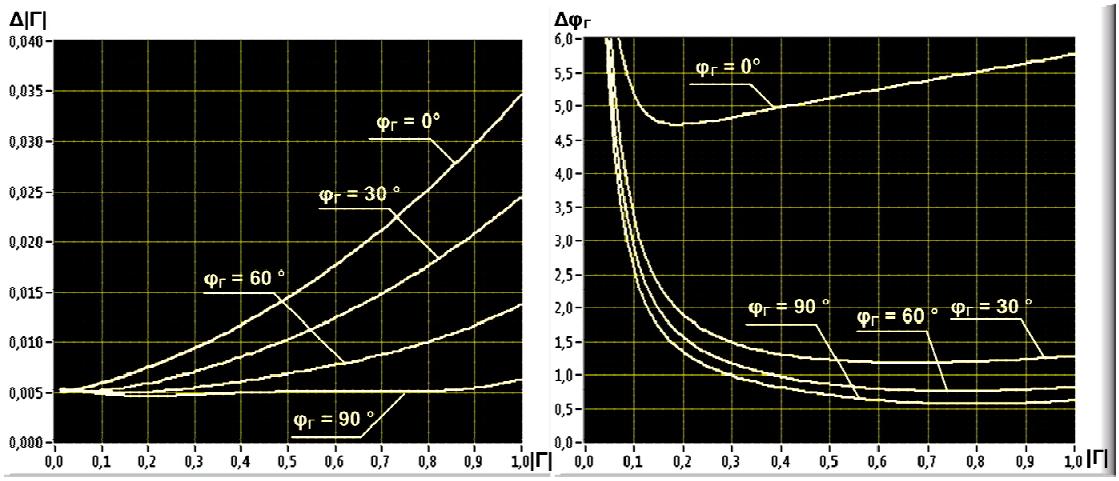


Рис. 7. Зависимость погрешности измерения модуля (*а*)
и фазы ККО (*б*) от модуля измеряемого коэффициента отражения

Анализ графиков позволяет сделать следующие выводы.

1. Погрешность измерения модуля ККО резко возрастает при фазе измеряемой нагрузки, близкой к нулю.

2. Погрешность измерения фазы ККО увеличивается при измерении нагрузки с фазой, близкой к $0°$ или $180°$. Вызвано это низкой точностью измерения разности фаз сигналов измерителем отношений AD8302 в диапазоне значений измеряемых фаз $0...30°$ и $150...180°$.

Разработанное устройство способно измерять ККО в широком диапазоне частот (100 МГц…2,7 ГГц), при этом неопределенность измерения модуля ККО не превышает $\pm 0,035$, а для подавляющего большинства нагрузок значительно меньше; точность измерения фазы в худшем случае составляет $\pm 6°$. Разработанный измерительный прибор может быть использован во время проверки, настройки и разработки различных радиотехнических устройств в условиях промышленного производства и лабораторий, в том числе в составе автоматизированных измерительных систем.

Библиографические ссылки

1. Glenn F. E. The Six-Port Reflectometer: An Alternative Network Analyzer // IEEE Trans. on Microwave Theory and Techniques. 1977. Vol. 25. № 12. P. 1075–1080.

2. Анализ метода измерения коэффициентов отражений скалярным рефлектометром с переменными параметрами / Ю. В. Рясный, А. В. Борисов, А. Н. Лоскутов, М. С. Чашков // Электросвязь. 2007. № 9. С. 42–43.

A. A. Shauerman, M. S. Zharikov, A. V. Borisov

AUTOMATED COMPLEX REFLECTION COEFFICIENT MEASURING DEVICE ON THE BASIS OF LOGARITHMIC AMPLIFIER

Application of logarithmic amplifier to design the microwave range complex reflection coefficient (CRC) measuring device of terminals is described in this article. The CRC measuring methods using the sensor, which has the parameters précising during the measuring, are considered there. The measurement error sources are also analyzed.

Keywords: complex reflection coefficient, microwave frequency range, measurement, logarithmic amplifier, sensor.

© Шауэрман А. А., Жариков М. С., Борисов А. В., 2010

СВЕДЕНИЯ ОБ АВТОРАХ

Агафонов Евгений Дмитриевич – кандидат технических наук, доцент кафедры системного анализа Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Сибирскую аэрокосмическую академию имени академика М. Ф. Решетнева в 1997 г. Область научных интересов – математическое моделирование, проектирование сложных систем, задачи идентификации и управления. E-mail: agafonov@gmx.de.

Анжина Валерия Александровна – инженер ОАО «Информационные спутниковые системы» имени академика М. Ф. Решетнева», аспирант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончила Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2007 г. Область научных интересов – радиотехника, радиоэлектроника, цифровая обработка сигналов. Т. 8902-914-65-07.

Белолипецкий Виктор Михайлович – доктор физико-математических наук, профессор, заведующий отделом вычислительных моделей в гидрофизике Института вычислительного моделирования Сибирского отделения Российской академии наук. Окончил Новосибирский государственный университет в 1968 г. Область научных интересов – гидрофизика, гидродинамика, математическое моделирование, моделирование водных экосистем, экология. E-mail: belolip@icm.krasn.ru.

Бериков Владимир Борисович – доктор технических наук, ведущий научный сотрудник, доцент Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончил Новосибирский государственный университет в 1986 г. Область научных интересов – распознавание образов, кластерный анализ, обработка изображений. E-mail: berikov@math.nsc.ru.

Бернс Владимир Андреевич – кандидат технических наук, доцент, начальник отдела динамической прочности Сибирского научно-исследовательского института авиации имени С. А. Чаплыгина. Окончил Новосибирский электротехнический институт в 1975 г. Область научных интересов – динамика и прочность летательных аппаратов. E-mail: v.bern@yandex.ru.

Бобров Александр Валерьевич – аспирант Алтайского государственного технического университета. Окончил Алтайский государственный технический университет в 2007 г. Область научных интересов – методы анализа данных, теория баз данных. E-mail: 22bav@mail.ru.

Бодягин Игорь Александрович – аспирант факультета прикладной математики и информатики Белорусского государственного университета. Окончил Белорусский государственный университет в 2009 г. Область научных интересов – статистический анализ случайных последовательностей. E-mail: bodiagin@cosmostv.by.

Бондаренко Анатолий Николаевич – доктор физико-математических наук, доцент, ведущий научный сотрудник Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончил Новосибирский государственный университет в 1975 г. Область научных интересов – обратные задачи, обработка информации. E-mail: bondarenkoan1953@mail.ru.

Бондаренко Михаил Анатольевич – магистр Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончил Новосибирский государственный технический университет в 2010 г. Область научных интересов – математическое моделирование. E-mail: bondarenkoma@mail.ru.

Борисов Александр Васильевич – кандидат технических наук, доцент Сибирского государственного университета телекоммуникаций и информатики. Окончил Новосибирский электротехнический институт связи в 1979 г. Область научных интересов – микропроцессоры и ПЛИС, измерения в области СВЧ. E-mail: bav@sibsutis.ru.

Борисова Ирина Артемовна – кандидат технических наук, старший научный сотрудник Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. E-mail: biamia@mail.ru

Вальтер Стеффен – научный сотрудник, исследователь в области медицинской психологии университета г. Ульм (Германия). Окончил университет г. Констанц в 2005 г. Область научных интересов – исследования эмоциональной сферы, нейропсихология, психотерапия, нейронные связи. Т. +4973150061914.

Викентьев Александр Александрович – кандидат физико-математических наук, доцент кафедры дискретной математики и информатики, алгебры и логики, старший научный сотрудник лаборатории анализа данных Новосибирского государственного университета. Окончил Карагандинский государственный университет в 1977 г. Область научных интересов – теория статистических решений, искусственного интеллекта, распознание, теория моделей, стабильность и конструктивные модели. E-mail: vikent@math.nsc.ru.

Воробейчиков Сергей Эрикович – доктор физико-математических наук, доцент, профессор кафедры высшей математики и математического моделирования Томского государственного университета. Окончил Томский государственный университет в 1979 г. Область научных интересов – статистика случайных процессов, финансовая математика. Т. 8(3822)52-95-99.

Генова Светлана Николаевна – кандидат технических наук, старший научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук. Окончила Красноярский государственный университет в 1976 г. Область

научных интересов – вычислительная гидродинамика, информатика. E-mail: sv@icm.krasn.ru.

Густов Павел Валерьевич – инженер по оборудованию отдела телекоммуникационного и сервисного оборудования Красноярского института железнодорожного транспорта – филиала Иркутского государственного университета путей сообщения. Окончил Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2010 г. Область научных интересов – системный анализ, оптимизация технических систем, искусственный интеллект. E-mail: gustovpavel@gmail.com.

Дамов Михаил Витальевич – аспирант кафедры информатики и вычислительной техники Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Сибирский государственный университет имени академика М. Ф. Решетнева в 2005 г. Область научных интересов – методы и алгоритмы обработки и восстановления видеопоследовательностей. E-mail: info@lightsoft.org.

Дедок Василий Александрович – кандидат физико-математических наук, доцент Новосибирского государственного университета. Окончил Новосибирский государственный университет в 2005 г. Область научных интересов – задачи рассеяния, случайное блуждание, самоподобные структуры. E-mail: vasiliy.a.dedok@gmail.com.

Демиденко Николай Данилович – ведущий научный сотрудник СКТБ «Наука» Российской академии наук. Окончил Томский государственный университет в 1963 г. Область научных интересов – исследование и оптимизация систем с распределенными параметрами. Т. 8983-207-62-05.

Демченко Яна Игоревна – бакалавр техники и технологий, магистрант кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – математическое моделирование. E-mail: yademchenko86@mail.ru.

Дерябин Антон Леонидович – инженер ОАО «Информационные спутниковые системы» имени академика М. Ф. Решетнева», аспирант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Томский государственный университет систем управления и радиоэлектроники в 2009 г. Область научных интересов – радиотехника, радиоэлектроника, цифровая обработка сигналов. E-mail: deanle@mail.ru.

Дюбанов Владимир Владимирович – ведущий инженер Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. E-mail: Vladimir.dyubanov@mail.com.

Жариков Михаил Сергеевич – аспирант Сибирского государственного университета телекоммуникаций и информатики. Окончил Сибирский государственный университет телекоммуникаций и информатики в 2009 г. Область научных интересов – автоматизация, измерения в области СВЧ. E-mail: moirfred@mail.ru.

Жучков Константин Викторович – аспирант Института вычислительного моделирования Сибирского отделения Российской академии наук. Окончил Сибирский федеральный университет в 2001 г. Область научных интересов – информационные технологии, автоматизированное построение пользовательского интерфейса. Т. 8923-281-84-69.

Заблоцкая Ксения Владимировна – студент университета г. Ульм (Германия). Окончила Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2007 г. Область научных интересов – математическое моделирование. E-mail: kseniya zablotskaya@uni-ulm.de.

Заблоцкий Сергей Геннадьевич – студент университета г. Ульм (Германия). Окончил Сибирский государственный университет имени академика М. Ф. Решетнева в 2007 г. Область научных интересов – математическое моделирование и оптимизация. E-mail: sergey.zablotskiy@uni-ulm.de.

Загоруйко Николай Григорьевич – доктор технических наук, профессор, заведующий лабораторией анализа данных Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончил Ленинградский институт киноинженеров в 1953 г. Область научных интересов – анализ данных распознание образов. Т. 8913-987-10-02.

Звонков Владимир Борисович – студент кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева, призер молодежных, краевых, всероссийских конференций, лауреат премии Президента РФ. Область научных интересов – моделирование и оптимизация сложных систем, эволюционные алгоритмы. Т. 8(391)291-91-41.

Зотин Александр Геннадьевич – кандидат технических наук, преподаватель кафедры информатики и вычислительной техники Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2005 г. Область научных интересов – методы и алгоритмы обработки видеопоследовательностей, локализация текста в видеопоследовательностях. E-mail: zotin@sibsau.ru.

Игнатьев Дмитрий Александрович – магистрант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – стохастическое моделирование, теория многосвязных систем, непараметрическая теория принятия решений. Т. 8(391)291-91-41.

Ильин Евгений Сергеевич – кандидат технических наук, заместитель директора по информационным технологиям Красноярского института железнодорожного транспорта – филиала Иркутского государственного университета путей сообщения. Окончил Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2002 г. Область научных интересов – системный анализ, оптимизация технических систем, искусственный интеллект. E-mail: gustovpavel@gmail.com.

Кирик Екатерина Сергеевна – кандидат физико-математических наук, научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук. Окончила Красноярский государственный университет в 1999 г. Область научных интересов – математическое моделирование движения людей, статистическое моделирование, непараметрическая статистика. E-mail: kirik@icm.krasn.ru.

Круглов Дмитрий Владимирович – системный администратор Института леса имени В. Н. Сукачева Сибирского отделения Российской академии наук. Окончил Красноярский государственный университет в 1998 г. Область научных интересов – программирование. E-mail: volgurk@mail.ru.

Козин Роман Евгеньевич – бакалавр техники и технологий, магистрант кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – математическое моделирование. E-mail: romka24region@gmail.ru.

Козинкин Леонид Алексеевич – студент механико-математического факультета Новосибирского государственного университета. Окончил инженерный лицей при Новосибирском государственном техническом университете в 2009 г. Область научных интересов – математическое моделирование. E-mail: bondarenkoma@mail.ru.

Кротова Елена Ивановна – кандидат технических наук, доцент Ярославского государственного университета имени П. Г. Демидова. Окончила Ярославский государственный университет имени П. Г. Демидова в 1980 г. Область научных интересов – контроль качества работы систем передачи и обработки информации. E-mail: ken@uniyar.ac.ru.

Кузнецов Алексей Владимирович – кандидат технических наук, доцент кафедры информатики института космических и информационных технологий Сибирского федерального университета. Окончил Красноярский государственный технический университет в 1999 г. Область научных интересов – глобальная оптимизация. E-mail: kuaw26@mail.ru.

Кузовников Александр Витальевич – начальник группы ОАО «Информационные спутниковые системы» имени академика М. Ф. Решетнева», аспирант кафедры электронной техники и телекоммуникаций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2007 г. Область научных интересов – цифровая обработка сигналов, радиолокация. E-mail: uiub@list.ru.

Кураш Елена Федоровна – старший преподаватель кафедры многоканальных электрических связей и оптических систем Сибирского государственного университета телекоммуникаций и информатики. Окончила Сибирский государственный университет телекоммуникаций и информатики в 2001 г. Область научных интересов – адаптивная фильтрация. Т. 8(383)269-82-52.

Кутненко Ольга Андреевна – кандидат технических наук, старший научный сотрудник Института математики имени С. Л. Соболева Сибирского отделения наук Российской академии наук. E-mail: olga@math.nsc.ru.

Лавров Виктор Иванович – главный конструктор общего проектирования космических аппаратов ОАО «Информационные спутниковые системы» имени академика М. Ф. Решетнева», аспирант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Красноярский государственный университет в 1975 г. Область научных интересов – радиотехника, радиолокация, цифровая обработка сигналов. Т. 8(39197)6-49-20.

Лбов Геннадий Сергеевич – доктор технических наук, профессор. Окончил Иркутский государственный университет в 1959 г. Область научных интересов – информатика, многоэкстремальная оптимизация, математическое моделирование, анализ данных, распознание образов.

Лопатин Александр Витальевич – доктор технических наук, профессор, заведующий кафедрой компьютерного моделирования Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил завод-втуз – филиал Красноярского политехнического института в 1979 г. Область научных интересов – механика твердого деформируемого тела. E-mail: lopatin@mail.ru.

Маер Алексей Владимирович – аспирант Курганского государственного университета. Окончил Курганский государственный университет в 2008 г. Область научных интересов – моделирование, непараметрическая статистика. E-mail: alex_povt@mail.ru.

Малинкин Виталий Борисович – доктор технических наук, профессор кафедры многоканальной электрической связи и оптических систем Сибирского государственного университета телекоммуникаций и информатики. Окончил Новосибирский электротехнический институт связи в 1975 г. Область научных интересов – адаптивная фильтрация. E-mail: mvb@neic.nsk.su.

Малинкин Евгений Витальевич – инженер-исследователь Сибирского государственного университета телекоммуникаций и информатики. Окончил Сибирский государственный университет телекоммуникаций и информатики в 2005 г. Область научных интересов – адаптивная фильтрация. Т. 8905-827-30-73.

Мальцева Татьяна Валерьевна – студент кафедры системного анализа Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – идентификация и моделирование, теория непараметрического оценивания. E-mail: nonparametric@mail.ru.

Медведев Александр Васильевич – доктор технических наук, профессор кафедры системного анализа и исследования операций, заведующий отделением кибернетики и системного анализа Института

вычислительного моделирования Сибирского отделения Российской академии наук. Окончил Фрунзенский политехнический институт в 1968 г. Область научных интересов – кибернетика, теория стохастических систем управления. Т. 8(391)290-50-11.

Медведев Алексей Викторович – кандидат физико-математических наук, доцент кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Кемеровский государственный университет в 1984 г. Область научных интересов – экономико-математическое моделирование, инвестиционный анализ, динамическая оптимизация. E-mail: alexm_62@mail.ru.

Минкер Вольфган – профессор инженерных наук, заместитель директора института информационных технологий университета г. Ульм (Германия). Окончил университет г. Карлсруэ в 1993 г. Область научных интересов – устойчивые и гибкие технологии, адаптивное диалоговое управление, интеллектуальные интерфейсы. Т. +497315026254.

Молокова Наталья Викторовна – доцент кафедры моделирования оптимизации систем Сибирского федерального университета. Окончила Красноярский государственный технический университет в 1996 г. Область научных интересов – математическое моделирование, информационные технологии. E-mail: nat_molokova@mail.ru.

Мюллер Томас – научный сотрудник исследовательского центра Даймлера г. Ульм (Германия). Окончил технический университет Мюнхена в 1996 г. Область научных интересов – аналоговые и цифровые аудио- и ТВ-приемники, мультистационарные приемники, тестирование оборудования. Т. +497315052248.

Наумов Максим Викторович – аспирант кафедры прикладной математики Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Томский государственный университет в 2010 г. Область научных интересов – дискретная математика, булевы функции, программирование. E-mail: mnx@yandex.ru.

Нестеров Владимир Анатольевич – кандидат технических наук, доцент кафедры компьютерного моделирования Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил завод-втуз – филиал Красноярского политехнического института в 1987 г. Область научных интересов – механика деформируемого твердого тела. E-mail: misternester@gmail.com.

Новоселов Аркадий Арсеньевич – кандидат физико-математических наук, докторант кафедры высшей и прикладной математики Красноярского государственного торгово-экономического института. Окончил Новосибирский государственный университет в 1976 г. Область научных интересов – общая теория риска, моделирование финансовых рисков и индивидуальных предпочтений, непараметрическое оценивание. E-mail: arcady@novosyolov.com.

Орлова Анна Сергеевна – инженер факультета повышения квалификации преподавателей Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончила Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2005 г. Область научных интересов – непараметрическая статистика. E-mail: annaorlova@sibsau.ru.

Первушин Владимир Федорович – студент кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – непараметрическое моделирование. E-mail: pervushin_vt@ihbox.ru.

Перепелкин Евгений Александрович – доктор технических наук, профессор кафедры прикладной математики Алтайского государственного технического университета. Окончил Алтайский политехнический институт в 1978 г. Область научных интересов – теория автоматизации управления, методы идентификации систем.

Пестунов Игорь Алексеевич – кандидат физико-математических наук, доцент, заведующий лабораторией Института вычислительных технологий Сибирского отделения Российской академии наук. Окончил Новосибирский государственный университет в 1977 г. Область научных интересов – непараметрические методы распознавания образов и кластеризации, обработка многоспектральных изображений, ГИС-технологии. E-mail: pestunov@ict.nsc.ru.

Победаш Павел Николаевич – кандидат физико-математических наук, докторант кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Кемеровский государственный университет в 1991 г. Область научных интересов – математическая экономика, теория оптимизации. E-mail: pobed_pnp@mail.ru.

Полякова Галина Леонидовна – старший преподаватель Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончила Новосибирский государственный университет в 1976 г. Область научных интересов – распознавание образов, информатика, математическое и имитационное моделирование, анализ данных. E-mail: karusel2007@ya.ru.

Попов Евгений Александрович – доктор технических наук, доцент, профессор кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Красноярский государственный университет в 1973 г. Область научных интересов – спектроскопия магнитных диэлектриков, обработка данных. E-mail: saor_popov@sibsau.ru.

Птицын Алексей Владимирович – кандидат технических наук, доцент кафедры безопасных информационных технологий Национального исследовательского университета Санкт-Петербургского государственного политехнического университета информационных технологий, механики и оптики. Окончил Институт точной механики и оптики в 1997 г. Область

научных интересов – информационная безопасность. E-mail: pticin@inbox.ru.

Птицына Лариса Константиновна – доктор технических наук, профессор кафедры системного анализа и управления Санкт-Петербургского государственного политехнического университета информационных технологий, механики и оптики. Окончила Ленинградский индустриальный институт имени М. И. Калинина в 1976 г. Область научных интересов – информационная безопасность. E-mail: ptitsina_lk@inbox.ru.

Рогалев Алексей Николаевич – кандидат физико-математических наук, старший научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук. Окончил Красноярский государственный университет в 1972 г. Область научных интересов – вычислительная математика, численные методы в задачах управления, символьные вычисления, оценки множеств достижимости. E-mail: rogalyov@icm.krasn.ru.

Рубан Анатолий Иванович – доктор технических наук, профессор, заведующий кафедрой информатики института космических и информационных технологий Сибирского федерального университета. Окончил Томский государственный университет в 1965 г. Область научных интересов – идентификация, оптимизация, адаптивное управление, теория чувствительности. E-mail: ruban@mail.ru.

Рутковская Марина Александровна – старший преподаватель кафедры компьютерного моделирования Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончила Сибирскую аэрокосмическую академию в 2000 г. Область научных интересов – геометрическое моделирование, обработка изображений, теория измерений. E-mail: marina_a_b@mail.ru.

Рутковсий Валерий Олегович – директор научно-образовательного центра интегрированных компьютерных технологий Сибирского федерального университета. Окончил завод-втуз – филиал Красноярского политехнического института в 1988 г. Область научных интересов – геометрическое моделирование, обработка изображений, теория измерений. E-mail: ibiplane@mail.ru.

Саханский Сергей Павлович – доктор технических наук, доцент кафедры информационно-управляющих систем Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил завод-втуз – филиал Красноярского государственного политехнического института в 1977 г. Область научных интересов – комплексное решение задач автоматизации процессов выращивания монокристаллов германия, автоматизации процессов управления очисткой сточных вод промышленных предприятий. E-mail: sahanskiv@yandex.ru.

Семенкин Евгений Станиславович – доктор технических наук. профессор кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Кемеровский государственный университет в 1982 г. Область научных интересов – моделирование и оптимизация сложных систем, интеллектуальный анализ даны, эволюционные алгоритмы. E-mail: saor_semenkin@sibsau.ru.

Сергеев Анатолий Николаевич – аспирант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Красноярский государственный университет в 1995 г. Область научных интересов – теория организационных систем, моделирование, управление и принятие решений в условиях неопределенности. Т. 8(391)291-91-41.

Сергеева Наталья Александровна – кандидат технических наук, доцент кафедры информатики Сибирского федерального университета. Окончила Красноярский государственный университет в 1995 г. Область научных интересов – стохастическое управление, непараметрическая идентификация, моделирование систем. E-mail: eterenteva@mail.ru.

Симахин Валерий Ананьевич – кандидат физико-математических наук, профессор Курганского государственного университета. Окончил Томский государственный университет в 1967 г. Область научных интересов – непараметрическая и робастная статистика. E-mail: sva_full@mail.ru.

Синявский Юрий Николаевич – младший научный сотрудник Института вычислительных технологий Сибирского отделения Российской академии наук. Окончил Новосибирский государственный университет в 2000 г. Область научных интересов – непараметрические методы распознавания образов и кластеризации, обработка многоспектральных изображений, геоинформационные системы, веб-технологии. E-mail: fox83@ngs.ru.

Смирнов Павел Олегович – доцент Санкт-Петербургского государственного политехнического университета. Окончил Санкт-Петербургский государственный политехнический университет в 2002 г. Область научных интересов – робастная статистика. E-mail: s.paul@mail.ru.

Соболева Ольга Валерьевна – доцент кафедры систем автоматизированного проектирования Сибирского государственного университета телекоммуникаций и информатики. Окончила Сибирский государственный университет телекоммуникаций и информатики в 2001 г. Область научных интересов – адаптивная фильтрация. Т. 8(383)269-82-52.

Сомов Виктор Григорьевич – профессор кафедры электронной техники и телекоммуникаций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончил Военную инженерную радиотехническую академию ПВО в 1984 г. Область научных интересов – радиотехника, радиолокация. Т. 8913-557-53-19.

Стрельников Аркадий Владимирович – студент кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – непараметрическое моделирование. E-mail: arkstrelvlad@mail.ru.

Тарасенко Феликс Петрович – доктор технических наук, профессор, декан радиофизического факультета Национального исследовательского Томского государственного университета. Окончил Томский государственный университет в 1955 г. Область научных интересов – системный анализ и математическая статистика. E-mail: ftara@ich.isu.ru.

Терентьева Елена Сергеевна – магистрант Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – непараметрическая идентификация систем. E-mail: eterenteva@mail.ru.

Терещенко Юлия Анатольевна – доцент кафедры высшей математики Сибирского федерального университета. Окончила Красноярский государственный университет в 2000 г. Область научных интересов – моделирование и оптимизация систем с распределенными параметрами. Т. 8983-207-62-05.

Терсков Виталий Анатольевич – доктор технических наук, профессор кафедры эксплуатации железных дорог Красноярского института железнодорожного транспорта – филиала Иркутского государственного университета путей сообщения. Окончил Красноярское высшее командное училище радиоэлектроники противоздушной обороны в 1982 г. Область научных интересов – системный анализ, оптимизация технических систем, искусственный интеллект. E-mail: terskovva@mail.ru.

Удальцов Ростислав Андреевич – ведущий инженер-программист Красноярского завода холодильников «Бирюса». Окончил Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева в 2007 г. Область научных интересов – информационные технологии и технологии. E-mail: awam@nekto.com.

Фаустов Артем Викторович – бакалавр техники и технологий, магистрант кафедры системного анализа и исследования операций Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – непараметрические методы идентификации, моделирование систем. E-mail: a.faustov@mail.ru.

Филимонов Вячеслав Аркадьевич – доктор технических наук, профессор, начальник экспериментального учебно-исследовательского центра Омского филиала Института математики имени С. Л. Соболева Сибирского отделения Российской академии наук. Окончил Томский государственный университет в 1970 г. Область научных интересов – учебно-исследовательские ситуационные центры. E-mail: filimono@ofim.oscsbras.ru.

Харин Юрий Семенович – доктор физико-математических наук, профессор Белорусского государственного университета, заслуженный деятель науки Республики Беларусь, академик Международной академии наук высшей школы. Окончил Томский государственный университет в 1971 г. Область научных интересов – прикладная математика и информатика. E-mail: kharin@bsu.by.

Шауэрман Александр Александрович – аспирант Сибирского государственного университета телекоммуникаций и информатики. Окончил Сибирский государственный университет телекоммуникаций и информатики в 2008 г. Область научных интересов – микропроцессорные измерительные системы, измерения в области СВЧ. E-mail: shamrel@mail.ru.

Шевляков Георгий Леонидович – доктор физико-математических наук, профессор Санкт-Петербургского государственного политехнического университета. Окончил Ленинградский государственный университет в 1973 г. Область научных интересов – робастная статистика. E-mail: Georgy.shevlyakov@gmail.ru.

Шестернева Олеся Викторовна – кандидат технических наук, доцент кафедры системного анализа Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Окончила Красноярский государственный технический университет в 1999 г. Область научных интересов – идентификация и моделирование, теория непараметрического оценивания, методы теории автоматизированного управления. E-mail: saor_shesterneva@sibsau.ru.

Щелканов Николай Николаевич – кандидат физико-математических наук, старший научный сотрудник Института оптики атмосферы имени В. Е. Зуева Сибирского отделения Российской академии наук. Окончил Томский государственный университет в 1977 г. Область научных интересов – разработка новых математических методов для построения оптических моделей атмосферы. E-mail: snn@iao.ru.

Юргельян Татьяна Брониславовна – аспирант института математики Сибирского федерального университета. Окончила Сибирский федеральный университет в 2007 г. Область научных интересов – математическое моделирование, моделирование движения людей. E-mail: tata_yurgel@mail.ru.

Якобсон Марк Евгеньевич – студент кафедры информационно-управляющих систем Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. Область научных интересов – автоматизация технологических процессов и производств. E-mail: sahanskiv@yandex.ru.

СОДЕРЖАНИЕ

Вступительное слово ... 3

РАЗДЕЛ 1. КИБЕРНЕТИКА, СИСТЕМНЫЙ АНАЛИЗ, ПРИЛОЖЕНИЯ ... 5

Медведев А. В. К теории непараметрических систем ... 6
Тарасенко Ф. П. Культура субъекта и моделирование ... 14
Загоруйко Н. Г., Борисова И. А., Дюбанов В. В., Кутненко О. А. Функции конкурентного сходства в алгоритмах распознавания комбинированного типа ... 19
Бодягин И. А., Харин Ю. С. Прогнозирование авторегрессионных временных рядов при наличии цензурирования ... 22
Агафонов Е. Д. Непараметрический алгоритм управления нелинейными динамическими системами с использованием скользящих режимов ... 26
Заблоцкая К., Вальтер С., Заблоцкий С., Минкер В. Определение значения переменных с использованием оценки частной производной первого порядка ... 30
Заблоцкий С., Мюллер Т., Минкер В. Оценка деградации качества радиосигнала с помощью нейронных сетей и непараметрической регрессионной модели ... 33
Кузнецов А. В., Рубан А. И. Алгоритмы метода усреднения координат при поиске главных минимумов многоэкстремальных функций ... 36
Лбов Г. С., Полякова Г. Л. Метод прогнозирования в классе логических решающих функций ... 42
Маер А. В., Симахин В. А. Непараметрические датчики для случайных стационарных процессов ... 46
Новоселов А. А. Построение многомерных дискретных распределений с заданной корреляционной структурой ... 49
Новоселов А. А. Параметризация моделей управляемых систем ... 52
Пестунов И. А., Бериков В. Б., Синявский Ю. Н. Сегментация многоспектральных изображений на основе ансамбля непараметрических алгоритмов кластеризации ... 56
Сергеев А. Н. О непараметрических алгоритмах принятия решений ... 64
Симахин В. А. Робастные непараметрические оценки линейных функционалов ... 72
Сергеева Н. А., Терентьева Е. С. О непараметрических оценках функции регрессии и ее производных при наличии пропусков данных ... 75
Смирнов П. О., Шевляков Г. Л. Приближение оценки Q_N параметра масштаба с помощью быстрых *M*-оценок ... 83
Первушин В. Ф., Сергеева Н. А., Стрельников А. В. Прецизионный генератор случайных чисел ... 86
Бериков В. Б. Алгоритм адаптивного планирования ансамбля таксономических деревьев решений ... 91
Бобров А. В., Перепелкин Е. А. Восстановление изображений на основе спектральной матричной нормы ... 94
Бондаренко А. Н., Дедок В. А. Моделирование неадамаровского квантового блуждания для решеток высших размерностей ... 97
Бондаренко М. А. Методы хаотической динамики в задаче анализа временных рядов Вольфа ... 102
Викентьев А. А. О введении метрик на высказываниях экспертов с вероятностями ... 104
Звонков В. Б., Семенкин Е. С. О решении сложных задач многокритериальной оптимизации модифицированными эвристическими стохастическими процедурами прямого поиска ... 108
Козинкин Л. А. Метод прямого статистического моделирования в задаче Томсона ... 113
Кротова Е. И. Метод оценки влияния аддитивных помех на входе приемника сигналов с помощью идентификации видов законов распределения ... 116
Щелканов Н. Н. Робастный метод построения линейной регрессии между двумя физическими величинами с учетом их случайных погрешностей ... 121
Белолипецкий В. М., Генова С. Н. Численное исследование гидроледотермического режима реки Ангары ... 124
Демиденко Н. Д., Терещенко Ю. А. Оптимальное управление процессами с рециркуляцией взаимодействующих потоков ... 127
Демиденко Н. Д., Терещенко Ю. А. Оптимизация процессов разделения многокомпонентных смесей ... 132
Медведев А. В., Победаш П. Н., Семенкин Е. С. Математическая модель глобального социально-экономического развития ... 137
Молокова Н. В. Математическое моделирование процессов нефтезагрязнения пористой среды ... 142

Рогалев А. Н. Гарантированные оценки и построение множеств достижимости для нелинейных управляемых систем ... 148
Птицына Л. К., Птицын А. В. Преодоление неопределенности относительно динамических профилей комплексных систем защиты информации ... 154
Филимонов В. А. Учебно-исследовательский ситуационный центр – полигон для команды системных аналитиков ... 156
Сугоняк В. В., Филимонов В. А. Управление человеческими ресурсами для коллективного исследования и обучения ... 159
Юргельян Т. Б., Кирик Е. С., Круглов Д. В. О валидации модели движения людей SIgMA.CA по данным фундаментальных диаграмм ... 162
Демченко Я. И., Орлова А. С. Об исследовании некоторых непараметрических оценок функции регрессии по наблюдениям ... 166
Игнатьев Д. А., Сергеев А. Н. Непараметрическое моделирование организационных систем ... 171
Козин Р. Е., Медведев А. В. О непараметрическом управлении линейной динамической системой ... 175
Медведев А. В., Фаустов А. В. О непараметрической оценке нестационарной функции регрессии по наблюдениям ... 182
Шестернева О. В., Мальцева Т. В. О выборе параметрической модели в задаче непараметрической идентификации ... 187
Густов П. В., Ильин Е. С., Терсков В. А. Оценка оптимального быстродействия элементов управления сложными техническими системами ... 193

Что делать? Кто виноват? Дискуссия об образовании ... 197

РАЗДЕЛ 2. МАТЕМАТИКА, МЕХАНИКА, ИНФОРМАТИКА ... 207

Бернс В. А. Оценка точности определения характеристик собственных тонов при наличии случайных ошибок в экспериментальных данных ... 208
Дамов М. В., Зотин А. Г. Построение изображения сцены совмещением последовательных кадров ... 212
Кузовников А. В., Сомов В. Г., Лавров В. И., Дерябин А. Л., Анжина В. А. Способ формирования помехозащищенных сигналов ... 217
Лопатин А. В., Удальцов Р. А. Симметричные колебания трехслойной пластины ... 221
Малинкин В. Б., Малинкин Е. В., Кураш Е. Ф., Соболева О. В. Инвариантная волоконно-оптическая система передачи и ее характеристики ... 227
Наумов М. В. Генерация частично бент-функций ... 231
Нестеров В. А. Поперечные колебания пластины, податливой при трансверсальном сдвиге ... 236
Нестеров В. А., Лопатин А. В. Вывод выражения энергии деформации оболочки в геометрически нелинейной постановке ... 242
Попов Е. А. Влияние упругой подсистемы магнитного диэлектрика на его оптический спектр ... 246
Рутковский В. О., Рутковская М. А. Метод получения трехмерных цифровых моделей технических объектов, основанный на применении искусственных текстур ... 243
Саханский С. П., Якобсон М. Е. Модель управления скоростью вытягивания монокристаллов кремния ... 254
Шауэрман А. А., Жариков М. С., Борисов А. В. Автоматизированный измеритель комплексного коэффициента отражения на основе логарифмического усилителя ... 258

Сведения об авторах ... 264

CONTENTS

PART 1. CYBERNETICS, SYSTEMS ANALYSIS, APPLICATIONS 5

Medvedev A. V. To the theoty of nonparametric systems 6
Tarasenko F. P. Culture of an individual and modeling 14
Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. Functions of rival similarity in algorithms of recognition of combined type 19
Badziahin I. A., Kharin Yu. S. Forecasting of autoregressive time series under censoring 22
Agafonov E. D. Nonparametric control algorithm for nonlinear dynamic systems using sliding modes 26
Zablotskaya K., Walter S., Zablotskiy S., Minker W. Determination of variables significance using estimations of the first-order partial derivative 30
Zablotskiy S., Müller T., Minker W. Estimation of radio signal quality degradation by means of neural network and non-parametric regression model 33
Kuznetsov A. V., Rouban A. I. Algorithms of coordinates averaging method for searching of the principal minimum of multiextreme functions 36
Lbov G. S., Polyakova G. L. Forecasting method in class class of logical decision functions 42
Mayer A. V., Simakhin V. A. Nonparametric sensors for stohastic stationary process 46
Novosyolov A. A. Construction of multidimensional discrete distributions with preassigned correlation structure 49
Novosyolov A. A. Parametrization of models of controlled systems 52
Pestunov I. A., Berikov V. B., Sinyavskiy Yu. N. Algorithm for multispectral image segmentation based on ensemble of nonparametric clustering algorithms 56
Sergeev A. N. On nonparametric algorythms of decision-making 64
Simakhin V. A. Robust nonparametric estimation of linear functionals 72
Sergeeva N. A., Terentyeva E. S. Nonparametric estimations of regression function and its derivatives in the presence of data admissions 75
Smirnov P. O., Shevlyakov G. L. Approximation of the Q_N estimate of scale with the help of fast M-estimates 83
Pervushin V. F., Sergeeva N. A., Strelnikov A. V. The random selection precision generator 86
Berikov V. B. Algorithm for adaptive planning of an ensemble of taxonomic decisions trees 91
Bobrov A. V., Perepelkin E. A. Image reconstruction based on spectral matrix norm 94
Bondarenko A. N., Dedok V. A. Modelling of Neadamarov quantum wandering for grill grids of advanced dimensions 97
Bondarenko M. A. Methods for chaotic dynamics in the problem of analysis of Wolfer's time series 102
Vikentiev A. A. Inletting metrics for expert statements with probabilities 104
Zvonkov V. B., Semenkin E. S. Solving complex multiobjective optimization problems with the modified heuristic stochastic algorithms of direct search 108
Kozinkin L. A. Direct statistical simulation method in Thomson's problem 113
Krotova E. I. Method of estimation of influence of additive disturbance on input of a receiver of signals with the help of identification of kinds of distribution lows 116
Shchelkanov N. N. Robust method for construction of linear regression between two physical parameters in view of their random errors 121
Belolipetskii V. M., Genova S. N. Numerical investigation of hydrothermal and ice regimes of the Angara river 124
Demidenko N. D., Tereschenko Ju. A. Optimal control of processes with recirculation of interacting flows 127
Demidenko N. D., Tereschenko Ju. A. Optimization of processes of multicomponent mixtures separation 132
Medvedev A. V., Pobedash P. N., Semenkin E. S. Mathematical model of the worldwide socioeconomic development 137
Molokova N. V. Mathematical modeling of processes of oil pollution of porous environment 142
Rogalyov A. N. Guaranteed estimations and construction of sets of attainability for nonlinear controllable systems 148
Ptitsyna L. K., Ptitsyn A. V. Overcoming of uncertainty of dynamic profiles of complex systems of protection of the information 154
Filimonov V. A. Training and research situational center as a range and a proving ground for a team of system analysts 156

Sugonyak V. V., Filimonov V. A. Human resources management for collective research and training 159
Yurgel'yan T. B., Kirik E. S., Krouglov D. V. On validation of SIgMA.CA model of pedestrian dynamics according to data of fundamental diagrams 162
Demchenko Ya. I., Orlova A. S. About researche of some nonparametric estimations of regression function according observations 166
Ignatyev D. A., Sergeev A. N. Nonparametric modeling of organizational processes 171
Kozin R. E., Medvedev A. V. About nonparametric control of a linear dynamic system 175
Medvedev A. V., Faustov A. V. About nonparametric estimation of nonstationary regression function according observations 182
Shesterneva O. V., Maltseva T. V. About choice of parametric model in nonparametric identification task 187
Gustov P. V., Iluin E. S., Terskov V. A. Evaluation of optimum execution speed of complex technical systems control elements 193

PART 2. MATHEMATICS, MECHANICS, COMPUTER SCIENCE 207

Berns V. A. Assessment of determination accuracy of eigentones characteristics in the presence of random errors in the experimental data 208
Damov M. V., Zotin A. G. Scene image construction by way of image sequnce superposition 212
Kuzovnikov A. V., Somov V. G., Lavrov V. I., Deryabin A. L., Anzhina V. A. Method of forming of noise-stop signals 217
Lopatin A. V., Udaltsov R. A. Wrinkling vibrations of sandwich plate 221
Malinkin V. B., Malinkin E. V., Kurash H. F., Soboleva O. V. Invariant method of information transmission in fiber-optic transmission systems 227
Naumov M. V. The generation of partially-bent functions 231
Nesterov V. A. Vibration of a plate with low transverse shear stiffness 236
Nesterov V. A., Lopatin A. V. Energy derivation of deformation of the shell in geometrical nonlinear arrangement 242
Popov E. A. Influence of nonrigid subsystem of magnetic dielectric on its optical spectrum 246
Rutkovskiy V. O., Rutkovskaya M. A. Method of 3d digitalization of technical objects based on artificial textures 249
Sakhanski S. P., Yakobson M. E. Pulling rate assignment of siliconium single crysttals growing 254
Shauerman A. A., Zharikov. M. S., Borisov A. V. Automated complex reflection coefficient measuring device on the basis of logarithmic amplifier 258

Приложение № 2
к Положению № 63/74 от 15.03.06
о журнале «Вестник СибГАУ»

ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ ДЛЯ ОПУБЛИКОВАНИЯ В ЖУРНАЛЕ «ВЕСТНИК СИБГАУ»

Общие требования. Тексты статей представляются в редакцию журнала в электронном и печатном виде на русском и английском языках. Русскоязычная версия должна сопровождаться английским переводом фамилии автора (авторов); названия статьи; аннотации; ключевых слов (см. пример оформления статьи).

Файл со статьей представляется на любом электронном носителе.

На последней странице ставится подпись автора (авторов) статьи. Количество авторов одной статьи – не более пяти. Автор имеет право публиковаться в выпуске один раз, второй раз – в соавторстве.

Индекс УДК предшествует названию статьи, соответствует заявленной теме и проставляется в верхнем левом углу листа.

Оригинал должен быть чистым, не помятым, без правок, вписанных от руки. Страницы нумеруются карандашом на обороте.

Электронная копия. Статья набирается в программе Microsoft Word.

Содержание. В статье необходимо сформулировать проблемы, отразить объект исследования, достигнутый уровень процесса исследования, новизну результатов, область их применения.

Статья должна заканчиваться выводом. Текст вывода набирается отдельным абзацем (абзацами), в котором акцентируется новизна результатов, эффективность их использования и др.

Объем статьи: 7–8 страниц (включая рисунки, таблицы и библиографические ссылки).

Параметры страницы. Формат А4 (210×297 мм). Поля: правое и левое – 2 см, верхнее и нижнее – 2,5 см.

Текст. Шрифт – Times New Roman, размер 12 пт.

По центру помещаются инициалы и фамилия автора (авторов).

Не допускается (!) в тексте статьи набирать слова прописными (заглавными) буквами, жирным шрифтом или курсивом, а также размещать все указанные элементы в рамках и имитировать оформление набора, выполняемого в журнале.

Ниже по центру шрифтом 12 пт печатается название статьи и через строку курсивом – аннотация, ключевые слова (не более 6).

Основной текст статьи размещается через пробел от аннотации. Межстрочный интервал – одинарный, межбуквенный и междусловный интервал – нормальный, перенос слов не допускается. Заголовки глав должны быть центрированы.

Абзацный отступ равен 0,5 см.

Ссылки на литературные или иные источники оформляются числами, заключенными в квадратные скобки, например [1]. Ссылки должны быть последовательно пронумерованы.

При наличии гранта ссылка на грант помещается внизу полосы под строками основного текста (подстрочное примечание).

Примечания. Смысловые пояснения основного текста или дополнения к нему оформляются в виде внутритекстовых примечаний среди строк основного текста специальной рубрикой, выделенной светлым курсивом: *Примечание.* (одно примечание), *Примечания*: (несколько примечаний). Отделяются от текста *точкой* (если стоят в единственном числе в подбор к тексту примечания). Примечания должны быть последовательно пронумерованы.

Формулы. Простые внутристрочные и однострочные формулы должны быть набраны без использования специальных редакторов – символами (шрифт Symbol). Специальные сложные символы, а также многострочные формулы, которые не могут быть набраны обычным образом, должны быть набраны в редакторе формул Microsoft Equation 3.0. Набор математических формул в пределах всего текста должен быть единообразен:

– русские и греческие символы – прямым шрифтом;
– латинские – курсивом;
– размер обычного символа – 12 пт;
– крупный индекс – 10 пт;
– мелкий индекс – 9 пт;
– крупный символ – 11 пт;
– мелкий символ – 10 пт.

Формулы, набранные отдельными строками, располагают по центру.

Не допускается (!) набор в основном тексте статьи простых латинских, греческих или специальных символов в редакторе формул.

Таблицы должны быть последовательно пронумерованы. Слово «таблица» набирается светлым курсивом с выравниванием вправо, шрифтом 11 пт, например *Таблица 1*, ниже – заглавие таблицы (набирается жирным шрифтом по центру). Если таблица имеет большой объем, то она может быть помещена на отдельной странице, а в случае, когда она имеет значительную ширину, – на странице с альбомной ориентацией.

Иллюстрации. Оформляются отдельным файлом с расширением tiff. Последовательно пронумеровываются обычным шрифтом без кавычек с выравниванием по центру, например, Рис. 1. Могут содержать подрисуночную подпись, шрифтом 11 пт. Иллюстрации могут быть сканированы с оригинала (в градациях серого с разрешением 150 dpi) или выполнены средствами компьютерной графики. Не принимаются цветные иллюстрации и иллюстрации с разрешением 300 dpi и более.

Библиографические ссылки составляются в соответствии с действующими требованиями к библиографическому описанию и помещаются после основного текста (ГОСТ 7.0.5–2008).

К печатному варианту статьи необходимо приложить:

1) экспертное заключение о возможности открытой публикации;

2) внешнюю рецензию (рецензент, желательно, не ниже ученой степени доктора наук);

3) сведения об авторе:

– фамилия, имя, отчество (полностью);

– ученая степень, звание, должность;

– вуз, год его окончания;

– область научных интересов;

– место работы, номер телефона;

– адрес для переписки (для иногородних);

– e-mail.

Внимание! Материалы статьи авторы сдают лично секретарю журнала или высылают почтой с указанием обратного адреса. По электронной почте статьи не принимаются.

Статьи рецензируются. Отклоненные материалы не возвращаются.

Плата с аспирантов за публикацию рукописей не взимается.

ПРИМЕР ОФОРМЛЕНИЯ СТАТЬИ

УДК 621.396.96.001(07)

И. А. Иванов

ДЕЦЕНТРАЛИЗОВАННЫЕ АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ В ДВУХКАНАЛЬНЫХ ИЗМЕРИТЕЛЬНЫХ СИСТЕМАХ*

Рассмотрена децентрализованная обработка информации в двухканальных измерительных системах при косвенном измерении для различных алгоритмов фильтрации оценки вектора состояния в измерительных пунктах и пункте обработки информации. Проведен сравнительный анализ результатов имитационного моделирования синтезированных алгоритмов.

Ключевые слова (не более 6):

Задача обеспечения высокой точности оценивания координат и параметров траектории движения объекта может быть решена за счет применения многоканальных измерительных систем с оптимальной централизованной обработкой.

(Продолжение текста публикуемого материала).

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 00-01-00912).

Библиографические ссылки

1. Гришин Б. П., Казаринов Ю. М. Динамические системы, устойчивые к отказам. М. : Радио и связь, 1985.

2. Медведев А. В. О моделировании организационных процессов // Вестн. Сиб. гос. аэрокосмич. ун-та : сб. науч. тр. Вып. 1. Красноярск, 2000. С. 173–191.

I. A. Ivanov

DECENTRALIZED ALGORITHMS OF INFORMATION PROCESSING IN TWO-CHANNEL MEASURE SYSTEMS

It is covered a decentralized algorithms of information processing in two-channel measure systems in case of an indirect measuring for different filtration algorithms of a condition vector estimation at the reception measure station and the station of information processing. Comparative analysis is carried out with a help of imitation modeling of synthesized algorithms.

Keywords: