

Информатика и её применения

Том 8 Выпуск 1 Год 2014

СОДЕРЖАНИЕ

Анализ и моделирование распределений в эредитарных стохастических системах И. Н. Синицын	2
Анализ задержек при диспетчеризации однородных заданий в условиях неопределенности Ю. Е. Малашенко, И. А. Назарова	12
Оценка надежности сложных систем с восстановлением по результатам испытаний элементов И. В. Павлов	21
Использование принципа равновесия для управления маршрутизацией в транспортных сетях Н. С. Васильев	28
Асимптотические свойства оценки риска при пороговой обработке вейвлет-коэффициентов в модели с коррелированным шумом А. А. Ерошенко, О. В. Шестаков	36
Базис реализации супер-ЭВМ эксафлопсного класса И. А. Соколов, Ю. А. Степченков, С. Г. Бобков, В. Н. Захаров, Ю. Г. Дьяченко, Ю. В. Рождественский, А. В. Сурков	45
Информационная модель технологии представления натурального объекта и изменения его пространственного положения О. П. Архипов, Ю. А. Маньяков, Д. О. Сиротинин	71
Динамические контексты базы данных реляционного типа С. В. Зыкин	77
Интегральное моделирование языковых структур в лингвистических процессорах систем обработки знаний и машинного перевода Е. Б. Козеренко	89
Разработка модели управления процессом обучения с использованием когнитивных технологий В. А. Маренко, О. Н. Лучко, О. С. Лупенцов	99
Общие оценки устойчивости для нестационарных марковских цепей с непрерывным временем А. И. Зейфман, В. Ю. Королев, А. В. Коротышева, С. Я. Шоргин	106
Об аппроксимации и сходимости одномерных параболических интегродифференциальных многочленов и сплайнов В. И. Киреев, М. М. Гершкович, Т. К. Бирюкова	118
Stability analysis of an optical system with random delay lines lengths Е. Morozov, L. Potakhina, and K. De Turck	127
Об авторах	135

АНАЛИЗ И МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ В ЭРЕДИТАРНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ*

И. Н. Сеницын¹

Аннотация: Рассматриваются методы и алгоритмы анализа и моделирования (аналитического, статистического) одно- и многомерных распределений в эредитарных стохастических системах (ЭСтС) с винеровскими и пуассоновскими шумами. Приводятся нелинейные стохастические интегродифференциальные уравнения. Для затухающих физически возможных эредитарных ядер рассматриваются два способа их аппроксимации (на основе линейных операторных уравнений и вырожденных ядер). Устанавливаются алгоритмы приведения ЭСтС к дифференциальным стохастическим системам (ДСтС). Приводится подробный анализ подходов к аналитическому и статистическому моделированию распределений в ЭСтС, приводимых к ДСтС. В основу подходов положены как методы прямого численного интегрирования уравнений ДСтС, так и численного интегрирования для параметров ортогонального разложения плотностей (моментов, коэффициентов ортогонального разложения и др.). Подробно рассматриваются алгоритмы аналитического и статистического моделирования, основанные на методе статистической линеаризации (МСЛ) и методе нормальной аппроксимации (МНА). Получены условия устойчивости алгоритмов на основе МСЛ и МНА. Для задач МСЛ рассматриваются прямые одношаговые сильные методы и алгоритмы численного интегрирования (различной точности) для гладких и разрывных правых частей уравнений ЭСтС. Разработан комплекс тестовых примеров для разрабатываемого в ИПИ РАН инструментального программного обеспечения “IDStS” в среде MATLAB. Подробно рассмотрены задачи анализа и моделирования колебаний осциллятора Дуффинга и релейного осциллятора в эредитарной стохастической среде.

Ключевые слова: аналитическое моделирование; вырожденное (сингулярное) эредитарное ядро; дифференциальная система; интегродифференциальная система; параметризация распределений; система, приводимая к дифференциальной; стохастическая система; эредитарная система

DOI: 10.14357/19922264140101

1 Введение

Как известно [1–3], интегродифференциальные стохастические системы (СтС) являются подходящими математическими моделями так называемых эредитарных СтС. Процессы в ЭСтС, как правило, являются немарковскими.

В прикладных задачах для затухающей эредитарной памяти путем аппроксимации ядер, определяющих память линейными обыкновенными дифференциальными уравнениями или вырожденными ядрами, удается привести интегродифференциальные уравнения к дифференциальным. В результате процессы в таких ДСтС становятся марковскими. Поэтому, как показано в [2–8], оказывается возможным использовать богатый современный арсенал методов и средств аналитического и статистического моде-

лирования. Вопросам моделирования распределений в ДСтС (см. библиографические замечания в [4–8]) посвящена обширная литература.

Статья посвящена вопросам анализа и моделирования одно- и многомерных распределений в негауссовских ЭСтС, описываемых интегродифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами в конечномерных пространствах.

2 Уравнения эредитарных стохастических систем

Рассмотрим ЭСтС, описываемую интегродифференциальным уравнением Ито следующего вида:

* Работа выполнена при финансовой поддержке Программы ОНИТ РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7).

¹ Институт проблем информатики Российской академии наук, sinitsin@dol.ru

$$dX = \left[a(X, t) + \int_{t_0}^t a_1(X(\tau), \tau, t) d\tau \right] dt + \left[b(X, t) + \int_{t_0}^t b_1(X(\tau), \tau, t) d\tau \right] dW_0 + \int_{R_0^q} \left[c(X, t, v) + \int_{t_0}^t c_1(X(\tau), \tau, t, v) \right] dP^0(t, dv) \quad (1)$$

$$\left. \begin{aligned} & \int_{-\infty}^{\infty} |A_{ij}(t, \tau)| d\tau < \infty; \\ & \int_{-\infty}^{\infty} |B_{il}(t, \tau)| d\tau < \infty; \\ & \int_{-\infty}^{\infty} |C_{ij}(t, \tau)| d\tau < \infty. \end{aligned} \right\} \quad (4)$$

с начальным условием $X(t_0) = X_0$.

В (1) приняты следующие обозначения и допущения:

- $X = X(t)$ – p -мерный вектор состояния;
- W_0 – r -мерный винеровский процесс интенсивности $\nu_0 = \nu_0(t)$;
- $\int_{\Delta_t} dP^0(t, A)$ – центрированная пуассоновская мера, удовлетворяющая условию

$$\int_{\Delta_t} dP^0(t, A) = \int_{\Delta_t} dP(t, A) - \int_{\Delta_t} \nu_P(t, A) dt,$$

где $\int_{\Delta_t} dP(t, A)$ – число скачков пуассоновского процесса в интервале Δ_t ;

- $\nu_P(t, A)$ – интенсивность пуассоновского процесса $P(t, A)$;
- v – q -мерный вспомогательный параметр в пространстве R_0^q с выколотым началом;
- A – некоторое борелевское множество пространства R_0^q .

Функции $a = a(X, t)$, $a_1 = a_1(X(\tau), \tau, t)$, $b = b(X, t)$, $b_1 = b_1(X(\tau), \tau, t)$, $c = c(X, t, v)$ и $c_1 = c_1(X(\tau), \tau, t, v)$ имеют размерности $p \times 1$, $p \times 1$, $p \times r$, $p \times r$, $p \times 1$ и $p \times 1$ и допускают представления вида:

$$\left. \begin{aligned} a_1 &= A(t, \tau)\varphi(X(\tau), \tau); \\ b_1 &= B(t, \tau)\psi(X(\tau), \tau); \\ c_1 &= C(t, \tau)\chi(X(\tau), \tau, v). \end{aligned} \right\} \quad (2)$$

Здесь эредитарные ядра $A(t, \tau) = [A_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$), $B(t, \tau) = [B_{il}(t, \tau)]$ ($i = \overline{1, p}$, $l = \overline{1, r}$) и $C(t, \tau) = [C_{ij}(t, \tau)]$ ($i, j = \overline{1, p}$) имеют соответственно размерности $p \times p$, $p \times r$ и $p \times p$ и удовлетворяют следующим условиям физической реализуемости и асимптотического затухания:

$$\left. \begin{aligned} A_{ij}(t, \tau) &= 0, \quad B_{il}(t, \tau) = 0; \\ C_{ij}(t, \tau) &= 0 \quad \forall \tau > t; \end{aligned} \right\} \quad (3)$$

Нелинейные в общем случае функции $\varphi = \varphi(X(\tau), \tau)$, $\psi = \psi(X(\tau), \tau)$ и $\chi = \chi(X(\tau), \tau, v)$ отражают нелинейные свойства ЭСтС, зависят от $X(\tau)$, τ и имеют размерности $p \times 1$, $p \times p$ и $p \times 1$ соответственно.

В случае, если эредитарные ядра удовлетворяют условиям

$$\left. \begin{aligned} A_{ij}(t, \tau) &= \tilde{A}_{ij}(u); \quad B_{il}(t, \tau) = \tilde{B}_{il}(u); \\ C_{ij}(t, \tau) &= \tilde{C}_{ij}(u) \quad (u = t - \tau), \end{aligned} \right\} \quad (5)$$

то говорят об ЭСтС со стационарным затуханием памяти.

Важный класс ядер представляют собой вырожденные (сингулярные) ядра, когда имеют место представления

$$\left. \begin{aligned} A_{ij}(t, \tau) &= A_{ij}^+(t)A_{ij}^-(\tau); \\ B_{il}(t, \tau) &= B_{il}^+(t)B_{il}^-(\tau); \\ C_{ij}(t, \tau) &= C_{ij}^+(t)C_{ij}^-(\tau) \quad (i, l = \overline{1, p}, j = \overline{1, r}) \end{aligned} \right\} \quad (6)$$

Замечание 2.1. В случае, когда подынтегральные функции $c(X, t, v)$ и $c_1(X(\tau), \tau, v)$ в (1) допускают представления

$$\begin{aligned} c(X, t, v) &= b(X, t)c'(v); \\ c_1(X(\tau), \tau, v) &= b(X(\tau), \tau)c'(v), \end{aligned}$$

ЭСтС (1) приводится к виду:

$$\dot{X} = a(X, t) + \int_{t_0}^t a_1(X(\tau), \tau, t) d\tau + \left[b(X, t) + \int_{t_0}^t b_1(X(\tau), \tau, t) d\tau \right] V, \quad (7)$$

если принять

$$V = \dot{W}; \quad W(t) = W_0(t) + \int_{R_0^q} c'(v)P^0(t, dv).$$

Замечание 2.2. Путем введения блочных матриц рассматриваются случаи, обобщающие соответственно представления (2) и (6)

$$\begin{aligned} a_1 &= \sum_{k=1}^N w'_k(t, \tau) \varphi'_k(X(\tau), \tau); \\ b_1 &= \sum_{k=1}^N w''_k(t, \tau) \psi'_k(X(\tau), \tau); \\ c_1 &= \sum_{k=1}^N w'''_k(t, \tau) \chi'_k(X(\tau), \tau, v) \end{aligned}$$

и

$$\begin{aligned} a_1 &= \sum_{k=1}^N \xi_k^l(t) \varphi'_k(X(\tau), \tau); \\ b_1 &= \sum_{k=1}^N \xi_k''(t) \psi'_k(X(\tau), \tau); \\ c_1 &= \sum_{k=1}^N \xi_k'''(t, \tau) \chi'_k(X(\tau), \tau, v) \end{aligned}$$

Замечание 2.3. Очевидно, что случай ЭСтС, описываемых по одной части переменных состояния системы стохастическим интегральным уравнением, а второй части — стохастическим дифференциальным уравнением, является частным случаем (1) и (7).

Решим сначала вспомогательную задачу приведения ЭСтС (1) к ДСтС, предполагая выполненными условия (2)–(4) и (2), (3) и (5) соответственно. Для ЭСтС (7) задача рассматривалась в [2–6].

3 Аппроксимация эредитарных ядер

Рассмотрим ЭСтС (1) при условиях (2)–(4). Будем считать, что эредитарные ядра $A(t, \tau)$, $B(t, \tau)$ и $C(t, \tau)$ удовлетворяют следующим линейным операторным уравнениям:

$$\begin{aligned} F^{At} A(t, \tau) &= H^{At} \delta(t - \tau); \\ F^{Bt} B(t, \tau) &= H^{Bt} \delta(t - \tau); \\ F^{Ct} C(t, \tau) &= H^{Ct} \delta(t - \tau); \\ A(t, \tau) &= A'(t, \tau)^T (H^{A*\tau})^T; \\ A'(t, \tau)^T (F^{A*\tau})^T &= I_h^A \delta(t - \tau), \\ B(t, \tau) &= B'(t, \tau)^T (H^{B*\tau})^T, \\ B'(t, \tau)^T (F^{B*\tau})^T &= I_h^B \delta(t - \tau); \\ C(t, \tau) &= C'(t, \tau)^T (H^{C*\tau})^T; \\ C'(t, \tau)^T (F^{C*\tau})^T &= I_h^C \delta(t - \tau). \end{aligned}$$

Здесь

$$\left. \begin{aligned} F^A &= F^A(t, D) = \sum_{l=0}^{n_A} \alpha_l^A(t) D^l; \\ H^A &= H^A(t, D) = \sum_{l=0}^{m_A} \beta_l^A(t) D^l; \\ F^B &= F^B(t, D) = \sum_{l=0}^{n_B} \alpha_l^B(t) D^l; \\ H^B &= H^B(t, D) = \sum_{l=0}^{m_B} \beta_l^B(t) D^l; \\ F^C &= F^C(t, D) = \sum_{l=0}^{n_C} \alpha_l^C(t) D^l; \\ H^C &= H^C(t, D) = \sum_{l=0}^{m_C} \beta_l^C(t) D^l \end{aligned} \right\} \quad (8)$$

являются известными матричными дифференциальными операторами размерности $h_A \times h_A$, $h_B \times h_B$ и $h_C \times h_C$ порядков $n_A, m_A, n_B, m_B, n_C, m_C$ ($n_A > m_A, n_B > m_B, n_C > m_C$) соответственно; индекс t у операторов означает, что оператор действует на функцию от t при фиксированном τ ; звездочкой обозначен символ сопряжения оператора; I_h^A, I_h^B и I_h^C — единичные $h \times h$ матрицы.

Введем h^A -, h^B - и h^C -мерные векторы посредством соотношений

$$\begin{aligned} U' &= \int_{t_0}^t A(t, \tau) \varphi(X(\tau), \tau) d\tau; \\ U'' &= \int_{t_0}^t B(t, \tau) \psi(X(\tau), \tau) d\tau; \\ U''' &= \int_{t_0}^t C(t, \tau) \chi(X(\tau), \tau, v) d\tau. \end{aligned}$$

Тогда, как известно из теории линейных дифференциальных систем [4–6], переменные U' , U'' и U''' будут удовлетворять следующим линейным дифференциальным уравнениям:

$$\left. \begin{aligned} F^A(t, D)U' &= H^A(t, D)\varphi(X, t); \\ F^B(t, D)U'' &= H^B(t, D)\psi(X, t); \\ F^C(t, D)U''' &= H^C(t, D)\chi(X, t, v). \end{aligned} \right\} \quad (9)$$

Применяя стандартную технику приведения уравнений (9) к форме Коши [4–6], приходим к искомой ДСтС для расширенного вектора состояния $Z = [X^T Z_1^T Z_1''^T Z_1'''^T]^T$ ($Z_1' = U'$, $Z_1'' = U''$, $Z_1''' = U'''$):

$$dZ = a_1^z(Z, t) dt + b_1^z(Z, t) dW_0 + \int_{R_0^z} c_1^z(Z, t, v) dP^0(t, dv). \quad (10)$$

Для случая $h_A = h_B = h_C = h, n_A = n_B = n_C = n$ и $m_A = m_B = m_C = m$ в подробной записи функции $a_1^z(Z, t), b_1^z(Z, t)$ и $c_1^z(Z, t, v)$ имеют следующий вид:

$$a_1^z(Z, t) = \begin{bmatrix} a(X, t) + Z_1' \\ a'(t)Z_1' \\ a''(t)Z_1'' \\ a'''(t)Z_1''' \end{bmatrix}; \quad (11)$$

$$b_1^z(Z, t) = \begin{bmatrix} b(X, t) + Z_1'' \\ b''(t)Z_1'' \\ 0 \\ 0 \end{bmatrix}; \quad (12)$$

$$c_1^z(Z, t, v) = \begin{bmatrix} c(X, t, v) + Z_1''' \\ c'''(t)Z_1''' \\ 0 \\ 0 \end{bmatrix}. \quad (13)$$

При условии существования обратных матриц $(\alpha_n^A)^{-1}, (\alpha_n^B)^{-1}$ и $(\alpha_n^C)^{-1}$ входящие в (11)–(13) переменные и коэффициенты допускают следующую запись:

$$\left. \begin{aligned} Z_{j+1}' &= \dot{Z}_j' - q_j' \varphi(X, t); \\ Z_{j+1}'' &= \dot{Z}_j'' - q_j'' \psi(X, t); \\ Z_{j+1}''' &= \dot{Z}_j''' - q_j''' \chi(X, t, v) \quad (j = \overline{1, (n-1)}); \\ a'(t) &= \end{aligned} \right\} \quad (14)$$

$$= \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \vdots & 0 \\ \vdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & I_h \\ -(\alpha_n^A)^{-1}\alpha_0^A & -(\alpha_n^A)^{-1}\alpha_1^A & \cdots & -(\alpha_n^A)^{-1}\alpha_{n-1}^A \end{bmatrix}; \quad (15)$$

$$a''(t) = \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \vdots & 0 \\ \vdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & I_h \\ -(\alpha_n^B)^{-1}\alpha_0^B & -(\alpha_n^B)^{-1}\alpha_1^B & \cdots & -(\alpha_n^B)^{-1}\alpha_{n-1}^B \end{bmatrix}; \quad (16)$$

$$a'''(t) = \begin{bmatrix} I_h & 0 & \cdots & 0 \\ 0 & I_h & \vdots & 0 \\ \vdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & I_h \\ -(\alpha_n^C)^{-1}\alpha_0^C & -(\alpha_n^C)^{-1}\alpha_1^C & \cdots & -(\alpha_n^C)^{-1}\alpha_{n-1}^C \end{bmatrix}; \quad (17)$$

$$q_j' = (\alpha_n^A)^{-1} \left[\beta_{n-j}^A - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^A q_i^{(l)'} \right]; \quad (18)$$

$$q_n' = (\alpha_n^A)^{-1} \left[\beta_0^A - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^A q_i^{(l)'} \right];$$

$$q_j'' = (\alpha_n^B)^{-1} \left[\beta_{n-j}^B - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^B q_i^{(l)''} \right]; \quad (19)$$

$$q_n'' = (\alpha_n^B)^{-1} \left[\beta_0^B - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^B q_i^{(l)''} \right];$$

$$q_j''' = (\alpha_n^C)^{-1} \left[\beta_{n-j}^C - \sum_{i=0}^{j-1} \sum_{l=0}^{j-i} C_{n-j-l}^{n-j} \alpha_{n-j+i+l}^C q_i^{(l)'''} \right]; \quad (20)$$

$$q_n''' = (\alpha_n^C)^{-1} \left[\beta_0^C - \sum_{i=0}^{n-1} \sum_{l=0}^{n-i} \alpha_{i+l}^C q_i^{(l)'''} \right].$$

Здесь $C_m^n = n!/(m!(n-m)!)$; индекс l означает, что суммирование проводится по всем индексам, исключая l .

Таким образом, справедливо следующие утверждение.

Теорема 3.1. Пусть эредитарные ядра $A(t, \tau), B(t, \tau)$ и $C(t, \tau)$ в ЭСтС (1) удовлетворяют условиям (2)–(4) или (9), причем матрицы α_n^A, α_n^B и α_n^C в (8) обратимы, а функции φ, ψ и χ дифференцируемы по переменным расширенного вектора состояния достаточное число раз. Тогда ЭСтС (1) приводится к ДСтС (10) на основе (11)–(20).

Замечание 3.1. Рассмотрим ЭСтС со стационарным затуханием, когда эредитарные ядра удовлетворяют условиям (5). Тогда вместо выполнения условия (4) достаточно потребовать, чтобы преобразования Лапласа ядер были бы рациональными функциями скалярной переменной s :

$$\int_0^\infty \tilde{A}(u) e^{-su} du = (F^A(s))^{-1} H^A(s);$$

$$\int_0^\infty \tilde{B}(u) e^{-su} du = (F^B(s))^{-1} H^B(s);$$

$$\int_0^\infty \tilde{C}(u) e^{-su} du = (F^C(s))^{-1} H^C(s).$$

Здесь порядок матричных полиномов $H^i(s)$ равен m^i , а порядок полиномов $F^i(s)$ равен n^i , причем $n^i \geq m^i$ ($i = A, B, C$).

Замечание 3.2. В практических задачах элементами эредитарных ядер $A(t, \tau)$, $B(t, \tau)$ и $C(t, \tau)$ являются следующие типовые функции: $e^{-\alpha|u|}$ и $e^{-\alpha|u|}(\cos \omega u + \gamma \sin \omega|u|)$, удовлетворяющие обыкновенным дифференциальным уравнениям первого и второго порядка (см. разд. 6).

Наконец, в том случае, когда выполнены условия (2)–(4), а функции φ , ψ и χ не дифференцируемы по переменным расширенного вектора состояния, целесообразна аппроксимация вырожденными ядрами. Тогда имеют место следующие соотношения:

$$\left. \begin{aligned} \int_{t_0}^t A(t, \tau)\varphi(X(\tau), \tau)d\tau &= A^+Y'; \\ \int_{t_0}^t B(t, \tau)\psi(X(\tau), \tau)d\tau &= B^+Y''; \\ \int_{t_0}^t C(t, \tau)\chi(X(\tau), \tau, v)d\tau &= C^+Y''' \end{aligned} \right\} \quad (21)$$

$$\left. \begin{aligned} \dot{Y}' &= A^-\varphi; \quad Y'(t_0) = 0; \\ \dot{Y}'' &= B^-\psi; \quad Y''(t_0) = 0, \\ \dot{Y}''' &= C^-\chi; \quad Y'''(t_0) = 0; \end{aligned} \right\} \quad (22)$$

$$\begin{aligned} Z &= [X^T Y'^T Y''^T Y'''^T]^T; \quad (23) \\ dZ &= a_2^z(Z, t) dt + b_2^z(Z, t, t) dW_0 + \\ &+ \int_{R_0^q} c_2^z(Z, t, t, v) dP^0(t, dv); \quad (24) \end{aligned}$$

$$\begin{aligned} a_2^z(Z, t) &= \begin{bmatrix} a(X, t) + A^+\varphi \\ A^-\varphi \\ B^-\psi \\ C^-\chi \end{bmatrix}; \\ b_2^z(Z, t) &= \begin{bmatrix} b(X, t) + B^+\psi \\ 0 \\ 0 \\ 0 \end{bmatrix}; \\ c_2^z(Z, t, v) &= \begin{bmatrix} c(X, t, v) + C^+\chi \\ 0 \\ 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Таким образом, имеем следующий результат.

Теорема 3.2. Пусть эредитарные ядра $A(t, \tau)$, $B(t, \tau)$ и $C(t, \tau)$ в ЭСтС (1) удовлетворяют условиям (3),

(4) и (6), а функции φ, ψ, χ не дифференцируемы по переменным расширенного вектора состояния. Тогда ЭСтС (1) приводится к ДСтС (24) на основе (21) и (22).

Замечание 3.3. Аналогичная теорема устанавливается для ЭСтС (7).

4 Обзор подходов к моделированию распределений в эредитарных стохастических системах

В стохастической информатике для вычисления вероятностей событий, связанных со случайными функциями, достаточно знания многомерных распределений. Различают три подхода к анализу и моделированию таких распределений в ЭСтС.

Первый, общий, основан на прямом численном решении уравнений (1) или (7), а для ЭСтС, приводимых к ДСтС интегрированием, — (10) или (24), с последующей статистической обработкой результатов [5–7].

Второй подход справедлив для ЭСтС, приводимых к ДСтС, поскольку он основан на теории марковских процессов и предполагает аналитическое моделирование, т. е. решение детерминированных уравнений в функциональных пространствах (Фоккера–Планка–Колмогорова, Пугачева и др.) для одно- и многомерных распределений [5, 6].

В практических задачах для ЭСтС, приводимых к ДСтС, часто можно рекомендовать комбинированный метод параметрического статистического и аналитического моделирования. При этом будем предполагать, что, во-первых, существуют одно- и многомерные плотности процессов, во-вторых, плотности одно- и многомерных распределений можно параметризовать с помощью вероятностных моментов, квазимоментов, семиинвариантов, коэффициентов ортогональных разложений плотностей, канонических разложений и др. и, в-третьих, по обобщенной формуле Ито составить для Ξ_n стохастические дифференциальные уравнения. При использовании такого подхода под расширенным вектором состояния следует рассматривать вектор $\bar{Z} = [Z^T \Xi_n^T]^T$.

Наконец, отметим, что всегда приведенные к ДСтС нелинейные уравнения ЭСтС линейны относительно интегральных переменных Z', Z'', Z''' или Y', Y'', Y''' .

Как известно [4–8], в задачах корреляционно-аналитического моделирования процессов в СтС с аддитивными шумами широкое распространение

получил МСЛ. Для СтС с мультипликативными шумами развит МНА.

Обобщением МНА-распределений являются различные приближенные методы, основанные на параметризации распределений. Аппроксимация одномерной характеристической функции $g_1(\lambda; t)$ и соответствующей плотности $f_1(z, t)$ известными функциями $g_1^*(\lambda; \Xi_1)$ и $f_1^*(z; \Xi_1)$, зависящими от конечномерного векторного параметра Ξ_1 , сводит задачу приближенного определения одномерного распределения к выводу из уравнения для характеристических функций обыкновенных дифференциальных уравнений, определяющих Ξ_1 как функцию времени. Это относится и к остальным многомерным распределениям. При аппроксимации многомерных распределений целесообразно выбирать последовательности функций $\{g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)\}$ и $\{f_n^*(z_1, \dots, z_n; \Xi_n)\}$, каждая пара которых находилась бы в такой зависимости от векторного параметра Ξ_n , чтобы при любом n множество параметров, образующих вектор Ξ_n , включало в качестве подмножества множество параметров, образующих вектор Ξ_{n-1} . Тогда при аппроксимации n -мерного распределения придется определять только те координаты вектора Ξ_n , которые не были определены ранее при аппроксимации функций $g_1, f_1, \dots, g_{n-1}, f_{n-1}$.

В зависимости от того, что представляют собой параметры, от которых зависят функции $f_n^*(z_1, \dots, z_n; \Xi_n)$ и $g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)$, аппроксимирующие неизвестные многомерные плотности $f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ и характеристические функции $g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$, используются различные приближенные методы решения уравнений, определяющих многомерные распределения вектора состояния системы Y_t , в частности методы моментов, семиинвариантов, ортогональных разложений и др.

5 Алгоритмы аналитического моделирования эредитарных стохастических систем, основанные на методах нормальной аппроксимации и статистической линеаризации

Для ЭСтС (7) при условии (2) с аддитивным шумом V , когда $b = b(t)$ и $b_1 = b_1(\tau, t)$, можно рекомендовать МСЛ для эквивалентной линеаризации нелинейных функций $a(X, t)$ и $a_1(X(\tau), \tau)$. В результате получим нелинейное интегральное урав-

нение для математических ожиданий и линейное уравнение относительно центрированных переменных. Применяя к последнему уравнению корреляционную теорию линейных систем [4, 5, 8], получим уравнения для ковариационной матрицы $K^X(t)$ и матрицы ковариационных функций $K^X(t_1, t_2)$.

Для ЭСтС, приводимых к (10) или (24), уравнения МНА имеют следующий вид:

$$f_1(z; t) = [(2\pi)^{n_z} |K|]^{-1/2} \times \exp \left[-\frac{1}{2} (z^T - m^T) K^{-1} (z - m) \right];$$

$$f_n(z_1, \dots, z_n; t_1, \dots, t_n) = [(2\pi)^{n n_z} |\bar{K}|]^{-1/2} \times \exp \left[\frac{1}{2} (\bar{z}_n^T - \bar{m}_n^T) \bar{K}_n^{-1} (\bar{z}_n - \bar{m}_n) \right]; \quad (25)$$

$$\dot{m} = \Phi_1(m, K, t); \quad (26)$$

$$\begin{aligned} \dot{K} &= \Phi_2(m, K, t) = \\ &= \Phi_{21}(m, K, t) + \Phi_{21}(m, K, t)^T + \Phi_{22}(m, K, t); \quad (27) \end{aligned}$$

$$\begin{aligned} \frac{\partial K(t_1, t_2)}{\partial t_2} &= \\ &= K(t_1, t_2) K(t_2)^{-1} \Phi_{21}(m(t_2), K(t_2), t_2)^T \end{aligned}$$

при $t_1 < t_2$ и начальном условии $K(t_1, t_2) = K(t_1)$. Здесь введены следующие обозначения:

$$\left. \begin{aligned} \Phi_1(m, K, t) &= M_N a^z(Z, t); \\ \Phi_{21}(m, K, t) &= M_N a^z(Z, t) (Z - m)^T; \\ \Phi_{22}(m, K, t) &= M_N b^z(Z, t) \nu_0 b^z(Z, t)^T + \\ &+ M_N \int_{R_0^z} c^z(Z, t, v) c^z(Z, t, v)^T \nu_P(t, dv), \end{aligned} \right\} \quad (28)$$

где M_N — символ вычисления математического ожидания для нормального распределения (25), $n = 2, 3, \dots$, $Z_n = [Z_1^T \dots Z_n^T]^T$, $\bar{m} = [m(t_1)^T \dots m(t_n)^T]^T$,

$$\bar{K}_n = \begin{bmatrix} K(t_1, t_1) & \dots & K(t_1, t_n) \\ \vdots & & \vdots \\ K(t_n, t_1) & \dots & K(t_n, t_n) \end{bmatrix}.$$

Замечание 5.1. Необходимым условием корректности алгоритма МНА для определения стационарного стохастического режима $m = m^*$, $K = K^*$ в ЭСтС с мультипликативными шумами является асимптотическая устойчивость линеаризованных уравнений (26) и (27).

Для ЭСтС с аддитивными шумами алгоритм МНА совпадает с алгоритмом МСЛ и определяется следующими уравнениями:

$$\dot{m} = \Phi_1(m, K, t); \quad (29)$$

$$\dot{Z}^0 = \Lambda(m, K, t)Z^0 + bV; \quad (30)$$

$$\dot{K} = \Lambda(m, K, t)K + K\Lambda(m, K, t)^T + b\nu b^T; \quad (31)$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = K(t_1, t_2)\Lambda(m(t_2), K(t_2), t_2)^T (t_2 > t_1).$$

Замечание 5.2. Для нормальных ЭСтС необходима асимптотическая устойчивость матрицы

$$\Lambda(m, K) = \left[\left(\frac{\partial}{\partial m} \right) (M_N a^z(Z, t))^T \right]^T. \quad (32)$$

Асимптотическая устойчивость матрицы (32) обеспечивает устойчивость численных методов интегрирования обыкновенных дифференциальных уравнений (29) и (30).

6 Тестовые примеры

1. Рассмотрим скалярную нелинейную ЭСтС первого порядка

$$\dot{X} = a_0 + a_1 X + \int_0^t e^{-\alpha|t-\tau|} \varphi(X(\tau)) d\tau + b\dot{W}, \quad X_0 = 0,$$

где a_0, a_1, α, b — константы. Полагая

$$\int_0^t e^{-\alpha|t-\tau|} \varphi(X(\tau)) d\tau = Y,$$

приведем ЭСтС к ДСтС:

$$\dot{X} = a_0 + a_1 X + \frac{1}{\alpha} Y + b\dot{W};$$

$$\dot{Y} = -\alpha Y + \varphi(X);$$

$$X(0) = Y(0) = 0.$$

2. Рассмотрим скалярную линейную ЭСтС второго порядка

$$a\ddot{X} + b\int_0^t \dot{X}(\tau)e^{-\beta|t-\tau|} d\tau + c\int_0^t X(\tau)e^{-\gamma|t-\tau|} d\tau = Q(t), \quad X_0 = 0,$$

где a, b, c, β и γ — константы. Расширим вектор состояния системы следующим образом: $X_1 = X$, $X_2 = \dot{X}$, $\dot{X}_1 = X_2$. Полагая

$$\int_0^t e^{-\beta|t-\tau|} X_2(\tau) d\tau = Y_1; \quad \int_0^t e^{-\gamma|t-\tau|} X_1(\tau) d\tau = Y_2,$$

приведем ЭСтС к ДСтС:

$$\dot{X}_1 = X_2;$$

$$\dot{X}_2 = -\frac{b}{a}Y_1 - \frac{c}{a}Y_2 + \frac{1}{a}Q(t);$$

$$\dot{Y}_1 = -\beta Y_1 + X_2;$$

$$\dot{Y}_2 = -\gamma Y_2 + X_1;$$

$$X_1(0) = X_2(0) = Y_1(0) = Y_2(0) = 0.$$

3. Рассмотрим скалярную нелинейную ЭСтС второго порядка

$$\ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} d\tau = Q(t), \quad X_0 = 0,$$

где α — константа. Расширим вектор состояния системы следующим образом: $X_1 = X$, $X_2 = \dot{X}$, $\dot{X}_1 = X_2$.

Полагая $\int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} d\tau = Y$, придем к следующей ДСтС:

$$\dot{X}_1 = X_2,$$

$$\dot{X}_2 = -\frac{1}{\alpha} Y + Q(t),$$

$$\dot{Y} = -\alpha Y + \varphi(X_1, X_2),$$

$$X_1(0) = X_2(0) = Y(0) = 0.$$

4. Рассмотрим скалярную нелинейную ЭСтС второго порядка

$$\ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) d\tau = Q(t), \quad X_0 = 0,$$

где α и ω_0 — константы. Расширим вектор состояния системы: $X_1 = X$, $X_2 = \dot{X}$, $\dot{X}_1 = X_2$. Положим

$$\int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) d\tau = Y$$

и найдем производную \dot{Y} :

$$\begin{aligned} \dot{Y} = & \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) - \right. \\ & \left. - \omega_0 e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau + \\ & + \varphi(X_1, X_2). \end{aligned}$$

Представляя интеграл суммы суммой интегралов, получим:

$$\begin{aligned} \dot{Y} = & -\alpha Y - \\ & -\omega_0 \int_0^t e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) d\tau + \\ & + \varphi(X_1, X_2). \end{aligned}$$

Введем новую переменную:

$$I = \int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) d\tau.$$

Тогда полученное уравнение запишется в виде:

$$\dot{Y} = -\alpha Y - \omega_0 I + \varphi(X_1, X_2).$$

Выразим из последнего уравнения I :

$$I = \frac{1}{\omega_0} (-\dot{Y} - \alpha Y + \varphi(X_1, X_2)).$$

Найдем производную \dot{I} :

$$\begin{aligned} \dot{I} = & \\ = & \int_0^t \left(-\alpha e^{-\alpha|t-\tau|} \sin(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) + \right. \\ & \left. + \omega_0 e^{-\alpha|t-\tau|} \cos(\omega_0(t-\tau)) \varphi(X_1(\tau), X_2(\tau)) \right) d\tau = \\ = & -\alpha I + \omega_0 Y. \end{aligned}$$

Найдем вторую производную \ddot{Y} :

$$\begin{aligned} \ddot{Y} = & -\alpha \dot{Y} - \omega_0 \dot{I} + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = & \alpha^2 Y + 2\alpha \omega_0 I - \omega_0^2 Y - \alpha \varphi(X_1, X_2) + \\ & + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2 = \\ = & -2\alpha \dot{Y} - (\alpha^2 + \omega_0^2) Y + \alpha \varphi + \frac{\partial \varphi}{\partial X_1} \dot{X}_1 + \frac{\partial \varphi}{\partial X_2} \dot{X}_2. \end{aligned}$$

Расширим вектор состояния $Y_1 = Y, Y_2 = \dot{Y}, \dot{Y}_1 = \dot{Y}_2 = Y_2$, в итоге придем к ДСтС вида:

$$\begin{aligned} \dot{X}_1 = & X_2; \\ \dot{X}_2 = & -Y_1 + Q(t); \\ \dot{Y}_1 = & Y_2; \\ \dot{Y}_2 = & -2\alpha Y_2 - (\alpha^2 + \omega_0^2) Y_1 + \alpha \varphi + \\ & + \frac{\partial \varphi}{\partial X_1} X_2 + \frac{\partial \varphi}{\partial X_2} [Q(t) - Y]; \\ X_1(0) = & X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

5. Скалярная нелинейная стохастическая ЭСтС второго порядка

$$\begin{aligned} \ddot{X} + \int_0^t \varphi(X(\tau), \dot{X}(\tau)) e^{-\alpha|t-\tau|} (\cos(\omega_0(t-\tau)) + \\ + \gamma \sin(\omega_0(t-\tau))) d\tau = Q(t), \quad X_0 = 0 \end{aligned}$$

(α, ω_0 и γ — константы) для переменных $X_1 = X, X_2 = \dot{X}, \dot{X}_1 = X_2, Y_1 = Y, Y_2 = \dot{Y}$ и $\dot{Y}_1 = Y_2$, где

$$\begin{aligned} \int_0^t \varphi(X_1(\tau), X_2(\tau)) e^{-\alpha|t-\tau|} (\cos(\omega_0(t-\tau)) + \\ + \gamma \sin(\omega_0(t-\tau))) d\tau = Y \end{aligned}$$

имеет вид:

$$\begin{aligned} \dot{X}_1 = & X_2; \\ \dot{X}_2 = & -Y_1 + Q(t); \\ \dot{Y}_1 = & Y_2; \\ \dot{Y}_2 = & -2\alpha Y_2 - (\omega_0^2 + \alpha^2) Y_1 + (\alpha + \omega_0 \gamma) \varphi(X_1, X_2) + \\ & + \frac{\partial \varphi}{\partial X_1} X_2 + \frac{\partial \varphi}{\partial X_2} [Q(t) - Y]; \\ X_1(0) = & X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned}$$

6. Рассмотрим осциллятор Дуффинга при $X(t_0) = X_0, \dot{X} = X_0$ в эредитарной стохастической среде, описываемый уравнением

$$\begin{aligned} \ddot{X} + \omega^2 X - \mu X^3 = -\delta \dot{X} + \gamma + V + \\ + \int_{t_0}^t \left[-\omega_1^2 X(\tau) + \mu_1 X^3(\tau) - \delta_1 \dot{X}(\tau) \right] e^{-\lambda|t-\tau|} d\tau. \quad (33) \end{aligned}$$

Здесь $\omega, \omega_1, \mu, \mu_1, \delta, \delta_1, \gamma$ и λ — постоянные параметры; V — скалярный белый шум интенсивности ν . Используя метод аппроксимации экспоненциального эредитарного ядра уравнением первого порядка при нулевом начальном условии, приведем (33) для $Z = [Z_1 Z_2 Z_3]^T$ ($Z_1 = X, Z_2 = \dot{X}$) к ДСтС вида:

$$\begin{aligned} \dot{Z} = a(Z, t) + bV; \quad (34) \\ Z_1(t_0) = Z_{10}; \quad Z_2(t_0) = Z_{20}; \quad Z_3(t_0) = 0; \end{aligned}$$

$$\begin{aligned} a(Z, t) = \begin{bmatrix} Z_2 \\ -\omega^2 Z_1 + \mu Z_1^3 - \delta Z_2 + \gamma - \lambda Z_3 \\ \lambda(-\omega_1^2 Z_1 + \mu_1 Z_1^3 - \delta_1 Z_2 - Z_3) \end{bmatrix}; \\ b(Z, t) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned}$$

Применим МСЛ к (34), положив согласно [8]

$$Z_1^3 \approx (m_1^2 + 3D_1)m_1 + 3(m_1^2 + D_1)Z_1^0$$

($m_1 = M_N Z_1^0$, $D_1 = M_N Z_1^{02}$), придем к уравнениям (28)–(31). Входящие в эти уравнения количества будут определяться следующими формулами:

$$\Phi_1(m, K) = \begin{bmatrix} m_2 \\ -\omega'^2 m_1 - \delta m_2 + \gamma - \lambda m_3 \\ \lambda(-\omega''^2 m_1 - \delta_1 m_2 - m_3) \end{bmatrix};$$

$$\Lambda(m, K) = \begin{bmatrix} 0 & 1 & 0 \\ \omega'^2 & -\delta & -\lambda \\ -\lambda\omega''^2 & -\lambda\delta_1 & -\lambda \end{bmatrix}, \quad (35)$$

где

$$\omega'^2 = \omega'(m, K)^2 = \omega^2 + \omega_1^2 \mu(m_1^2 + 3D_1);$$

$$\omega''^2 = \omega''(m, K)^2 = -\omega_1^2 \mu_1(m_1^2 + 3D_1).$$

Стационарные значения m^* и K^* определяются из уравнений (29) и (31):

$$\Phi_1(m^*, K^*) = 0;$$

$$\Lambda(m^*, K^*)K^* + K^*\Lambda(m^*, K^*)^T + \nu bb^T = 0$$

при условии асимптотической устойчивости матрицы (35).

Сравнивая результаты моделирования стационарных стохастических колебаний осциллятора Дуффинга в эредитарной и неэредитарных стохастических средах, приходим к выводам:

- эредитарные свойства среды проявляются только для моментов времени λ^{-1} ;
- МСЛ для ЭСтС приводит к более точным результатам, чем МСЛ для неэредитарных сред;
- МСЛ может быть успешно применен и для осциллятора Дуффинга в негауссовской среде с автокоррелированными шумами и разрывными нелинейностями [9]. При этом точность МСЛ для эредитарных сред выше, чем для неэредитарных.

7 Заключение

Для нелинейных ЭСтС разработаны методы и алгоритмы аналитического моделирования распределений, основанные на их параметризации.

Особое внимание уделено методам и алгоритмам, основанным на приведении стохастических интегродифференциальных уравнений к стохастическим дифференциальным уравнениям с винеровскими и пуассоновскими шумами. Полученные результаты положены в основу разрабатываемого в ИПИ РАН инструментального программного обеспечения «IDStS» в среде MATLAB.

Литература

1. Колмановский В. Б., Носов В. Р. Устойчивость и периодические режимы регулируемых систем с последействием. — М.: Наука, 1981. 386 с.
2. Синецын И. Н. Stochastic hereditary control systems // Проблемы управления и теории информатики, 1986. Т. 15. № 4. С. 287–298.
3. Синецын И. Н. Конечномерные распределения процессов в стохастических интегральных и интегродифференциальных системах // 2nd Symposium (International) IFAC on Stochastic Control: Preprints. — Vilnius, 1986; Pergamon Press, 1987. Pt. 1. P. 144–153.
4. Пугачев В. С., Синецын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. [Англ. пер.: Stochastic differential systems. Analysis and filtering. — Chichester, New York: John Wiley, 1987. 549 p.]
5. Пугачев В. С., Синецын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с. [Англ. пер.: Stochastic systems. Theory and applications. — Singapore: World Scientific, 2001. 908 p.]
6. Синецын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009. 768 с.
7. Синецын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
8. Синецын И. Н., Синецын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 480 с.
9. Синецын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с разрывными характеристиками // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 3–11.

Поступила в редакцию 23.10.13

ANALYSIS AND MODELING OF DISTRIBUTIONS IN HEREDITARY STOCHASTIC SYSTEMS

I. N. Sinitsyn

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Methods and algorithms for statistical and analytical modeling of one- and multidimensional distributions in hereditary stochastic systems (HStS) with Wiener and Poisson noises are considered. Nonlinear stochastic integrodifferential equations are presented. For dying physically realizable hereditary kernels, two ways of approximation (on the basis of linear operator equations and singular kernels) are described. Basic reduction algorithms of HStS to differential StS (DStS) are given. Detailed analysis of various approaches to statistical and analytical modeling of distributions in HStS reducible to DStS is given. These approaches are based: on the direct numerical integration DStS equations and numerical integration of equations for parameters (moments, quasi-moments, etc.) of orthogonal densities expansions. The detailed consideration of the method of statistical linearization (MSL) and of the method of normal approximation (MNA) in reducible HStS to DStS is presented. Numerical stability of MSL and MNA algorithms is investigated. For MSL problems, one-step strong methods and algorithms of numerical integration (of various accuracy) for smooth and nonsmooth right hands of HStS equations are described. Test examples for the IPI RAS software tool “IDStS” in MATLAB are considered. Special attention is paid to stochastic oscillations of the Duffing oscillator and the relay oscillator in hereditary stochastic media.

Keywords: analytical and statistical modeling; differential system; hereditary kernel; hereditary system; integrodifferential system; parametrization of distribution; reducible system; singular kernel; stochastic system

DOI: 10.14357/19922264140101

Acknowledgments

The work was financially supported by the Program of the RAS Department for Nanotechnologies and Information Technologies “Intelligent information technology, systems analysis, and automation” (project 1.7).

References

1. Kolmanovskij, V. B., and V. R. Nosov. 1981. *Ustoychivost' i periodicheskie rezhimy sistem s posledeystviem* [Stability of hereditary systems]. Moscow: Nauka. 386 p.
2. Sinitsyn, I. N. 1986. Stochastic hereditary control systems. *Problems Control Inform. Theory* 15(4):287–298.
3. Sinitsyn, I. N. 1986. Finite-dimensional distributions in stochastic integral and integrodifferential systems. *2nd Symposium (International) IFAC on Stochastic Control Proceeding: Preprints*. Vilnius: Pergamon Press. Pt. 1. P. 144–153.
4. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester, New York: John Wiley. 549 p.
5. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
6. Sinitsyn, I. N. 2009. *Kanonicheskie predstavleniya sluchaynykh funktsiy i ikh primeneniye v zadachakh komp'yuternoy podderzhki nauchnykh issledovaniy* [Canonical expansions of random functions and its application to scientific computer-aided support]. Moscow: TORUS PRESS. 768 p.
7. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametrical statistical and analytical modeling of distributions in stochastic systems on manifolds]. *Informatika i ee Primeniya — Inform. Appl.* 7(2):4–16.
8. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal distributions approximations]. Moscow: TORUS PRESS. 480 p.
9. Sinitsyn, I. N. 2013. Analiticheskoe modelirovanie raspredeleniy s invariantnoy meroy v stokhasticheskikh sistemakh s razryvnymi kharakteristikami [Analytical modeling of distributions in stochastic systems with discontinuous characteristics]. *Informatika i ee Primeniya — Inform. Appl.* 7(1):3–11.

Received October 23, 2013

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, Professor, Honored scientist of Russian Federation, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

АНАЛИЗ ЗАДЕРЖЕК ПРИ ДИСПЕТЧЕРИЗАЦИИ ОДНОРОДНЫХ ЗАДАНИЙ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Ю. Е. Малашенко¹, И. А. Назарова²

Аннотация: Рассматривается проблема управления ресурсоемкими вычислительными заданиями переборного типа, допускающими распараллеливание по данным. Заявки-задачи поступают в систему в произвольные моменты времени по одной или группами, время их решения заранее не известно. Для планирования процесса обработки используется оптимизационная модель, которая базируется на текущей информации о выполнении задания: длительности пребывания в системе и объеме уже обработанных данных. С помощью модели для каждого задания определяется доля данных для обработки в плановом периоде. При расчетах используются оценки необходимых вычислительных затрат и не делается предположений о законах распределения неизвестных характеристик заявок. Предложенное правило диспетчеризации позволяет в динамике формировать порядок выполнения заданий, приоритет получают менее трудоемкие задачи.

Ключевые слова: ресурсоемкие задачи; параллельные вычисления; оптимизация управления; принцип гарантированного результата

DOI: 10.14357/19922264140102

1 Введение

Рассмотрим частный случай проблемы управления ресурсоемкими вычислениями [1]. Диспетчеру специализированной системы, состоящей из большого числа однотипных вычислительных модулей, корпоративные пользователи в случайные моменты времени предоставляют для решения однородные *situ*-задачи [2]. Поиск решения для каждой из *situ*-задач производится переборными алгоритмическими процедурами (далее — ψ -процедуры), которые сводятся к просмотру набора исходных данных, разбитых на отдельные одинаковые по размеру неделимые содержательно значимые фрагменты, и поиску среди них *уникального фрагмента*, удовлетворяющего некоторому наперед заданному критерию. *Situ*-задание завершается, если:

- (1) искомым фрагмент найден. Тогда соответствующая *situ*-задача считается решенной;
- (2) просмотрен весь массив, но ничего найти не удалось. *Situ*-задача считается нерешенной, но задание — выполненным.

Массив исходных данных для каждого из заданий становится известен после анализа на входе в систему. Заявки одинаково важны, многие из них взаимосвязаны, и по условию они должны быть завершены все вместе и каждая в отдельности как можно быстрее. В противном случае найденное

решение может потерять актуальность и не будет представлять интереса.

Все *situ*-задания являются ресурсоемкими, выполняются в режиме реального времени и допускают распараллеливание по данным. В общем случае можно указать только верхнюю оценку для объема вычислений, необходимых для выполнения конкретного задания; имеет ли решение соответствующая содержательная задача — неизвестно. Считается, что большинство *situ*-задач имеют решение.

Рассмотрим модель [1], в которой *situ*-задания выполняются высокопроизводительной специализированной вычислительной системой (СВС), состоящей из центрального управляющего устройства (ЦУП-устройства) и набора однотипных единичных вычислительных модулей (ЕВ-модулей). Под специализированной элементарной вычислительной операцией (СЭВ) операцией будем понимать просмотр ψ -процедурой отдельного неделимого содержательно значимого фрагмента данных и проверку его уникальности. Производительность каждого ЕВ-модуля определяется числом СЭВ-операций, выполняемых в единицу времени, а мощность СВС равна суммарной производительности входящих в нее ЕВ-модулей. В рамках модельного описания считается, что архитектура СВС позволяет просматривать наборы фрагментов в произвольном порядке, в том числе параллельно, т. е. любая

* Работа выполнена при финансовой поддержке РФФИ (проект 11-01-00515а).

¹ Вычислительный центр им. А. А. Дородницына Российской академии наук, malash09@ccas.ru

² Вычислительный центр им. А. А. Дородницына Российской академии наук, irina-nazar@yandex.ru

часть данных может обрабатываться независимо от других.

Для идентификации *уникального фрагмента* для каждого *situ*-задания требуется выполнить *вычислительную работу* — произвести определенное, заранее не известное число СЭВ-операций. Искомый фрагмент может быть найден в любой момент после начала просмотра, что в лучшем случае потребует небольших вычислительных затрат, в худшем — безрезультатном — полного перебора всего массива данных соответствующей заявки. Назовем *необходимой работой* [3] число СЭВ-операций, которые требуются для выделения уникального фрагмента из набора данных конкретного задания. Длительность поиска уникального фрагмента зависит как от априори не известного объема *необходимой работы*, так и от стратегии диспетчеризации в СВС.

В рассматриваемой модели предполагается, что все предоставляемые *situ*-задачи *равнозначны*, а пользователи *равноправны* в рамках правил распределения вычислительных ресурсов, не являются антагонистами и стремятся к достижению некоторой общей корпоративной цели, которая состоит в наискорейшем извлечении всей информации из имеющихся данных. Стратегии управления СВС не должны быть дискриминирующими по отношению к той или иной группе заявок.

Для анализа диспетчерских правил воспользуемся концептуальным подходом, общая схема которого предложена в [1, 3] и состоит в следующем. При организации вычислений не делается никаких гипотез о времени поступления *situ*-задач, объеме *необходимой работы* или наличии *уникальных фрагментов* в поступивших данных. В модели используются только следующие параметры и величины:

- фактическое время поступления задания в СВС;
- число обработанных фрагментов данных на текущий момент времени;
- производительность СВС.

Все остальные неизвестные параметры рассматриваются как неопределенность. Таким образом, в данной статье проблема планирования работ и распределения ограниченных вычислительных ресурсов для скорейшего выполнения требований пользователей анализируется в рамках методологии исследования операций.

В условиях неопределенности управление СВС при выполнении *работ* разбивается контрольными точками на отдельные этапы — плановые периоды, или операционные окна. Предполагается, что диспетчер СВС выбирает длительность планового периода исходя из характера данных поступающих

заявок, а планирование *работ* производится на ближайшее операционное окно.

Основными объектами предложенной модели [1, 3] являются: база данных заданий (БД заданий), буфер данных текущих работ (ТР-буфер), программный комплекс планирования и анализа (ПК-план). При моделировании процесса выполнения ресурсоемких вычислений предполагается, что любая информация о *situ*-заданиях, вновь поступивших или находящихся в обработке, помещается и далее хранится в БД заданий. Все записи из последней в реальном времени доступны для анализа ПК-планом.

В ПК-плане подготавливается список заданий, и для каждого из них определяется размер подмассива фрагментов данных для обработки на текущий плановый период. Выделенные подмассивы исходных данных формируют пакет текущих работ (ТР-пакет). В контрольный момент времени корректируется содержимое БД заданий, составляется ТР-пакет, который поступает в ТР-буфер, и в СВС начинается обработка пакета. По завершении выполнения ТР-пакета наступает следующая контрольная диспетчерская точка — момент создания очередного списка заданий, — и вся процедура, т. е. обращение к БД заданий, формирование ТР-пакета и его обработка, повторяется с учетом изменения состояния СВС и состава пакета заявок.

В ПК-плане процесс обработки *situ*-заданий в СВС описывается системой квазидинамических уравнений и неравенств. В рамках модели строятся гарантированные оценки [4] и формулируется оптимизационная задача [5], условия которой учитывают всю текущую информацию (длительность пребывания задания в системе, объем обработанных данных и др.). Решение оптимизационной задачи позволяет определить неуплощаемые относительные показатели времени пребывания каждого конкретного задания в СВС и значения параметров управления системой.

Изучение частного случая проблемы управления однородными ресурсоемкими вычислениями является продолжением исследований, начатых в [6]. В настоящей работе результаты, полученные в [7, 8] с помощью имитационного моделирования, анализируются с позиции теории исследования операций.

2 Относительные задержки при выполнении заданий

Для формального описания заданий будет использован ряд обозначений.

Пусть z_n — задание-задача с собственным идентификационным номером n .

Предполагается, что для каждого задания z_n в момент его поступления t_n^0 в СВС становится известна нормативная величина Z_n — общее число неделимых фрагментов данных задания z_n , которые необходимо будет обработать, если среди них нет уникального.

Для любого z_n неизвестной величиной является ψ_n — общее число неделимых фрагментов данных, которые необходимо будет обработать для заявки z_n , т. е. ψ_n — *необходимая работа*, которую следует произвести для поиска уникального фрагмента или полного перебора данных задания z_n объемом Z_n .

Монопольным режимом будем называть такой способ организации работы СВС, при котором на всех имеющихся ЕВ-модулях одновременно просматриваются фрагменты данных единственного задания z_n .

Далее обозначим:

τ_n — *абсолютное время* выполнения задания z_n в монопольном режиме СВС;

P_0 — мощность (суммарная производительность) СВС, которая измеряется числом СЭВ-операций, которые могут быть выполнены всей СВС в единицу времени. Тогда в соответствии с определением ψ_n для каждого задания z_n объем *необходимой работы* составит

$$\psi_n = \tau_n P_0.$$

Далее в зависимости от контекста переменная t будет обозначать либо текущий момент календарного времени, либо контрольную точку — момент принятия решения о составе нового ТР-пакета.

Для множества $Z(t)$ — заданий z_n , находящихся в СВС в момент t , — обозначим:

$\mathcal{N}(t)$ — множество номеров заданий из $Z(t)$;

$N(t) = |\mathcal{N}(t)|$ — общее число заданий в $Z(t)$;

$\Delta(t)$ — операционное окно, плановый период, промежуток (интервал) времени с началом в момент t , измеряемый в единицах календарного времени. Длительность $\Delta(t)$ определяется заранее или выбирается ПК-планом.

Обозначим через $w_n(\Delta(t))$ число неделимых фрагментов данных для задания z_n , которые войдут в состав ТР-пакета, будут помещены в ТР-буфер и в момент времени t начнут просматриваться по команде ЦУП-устройства на выделенных ЕВ-модулях. Набор подмассивов данных ТР-пакета обозначим через $\mathbf{w}(t) = \{w_1(\Delta(t)), w_2(\Delta(t)), \dots, w_{N(t)}(\Delta(t))\}$. В рамках модели предполагается, что ТР-пакет начинает обрабатываться в контрольный момент t и будет завершен до конца планового периода, т. е. не позднее $t + \Delta(t)$.

Пусть $W^\Delta(\Delta(t))$ — максимальное суммарное число СЭВ-операций, которые планируется выполнить на заданном интервале планирования $\Delta(t)$,

$$W^\Delta(\Delta(t)) = P_0 \Delta(t).$$

В рамках модельного описания $W^\Delta(\Delta(t))$ — предельно допустимое суммарное число фрагментов заданий из $Z(t)$, которые могут быть помещены в ТР-буфер в момент времени t и будут гарантированно завершены не позднее $t + \Delta(t)$.

Пусть t_n^0 — время поступления задания z_n в СВС, и предположим, что в момент t оно еще не завершено. Обозначим через $z_n^-(t)$ число фрагментов данных, которые уже были просмотрены для z_n до момента t .

При формировании ТР-пакета предполагается, что весь комплект $\mathbf{w}(t)$ будет обработан за время $\Delta(t)$. Если для задания z_n выделяется $w_n(\Delta(t))$ фрагментов данных в состав ТР-пакета в момент t , то к окончанию планового периода $\Delta(t)$ справедливо равенство:

$$z_n^-(t + \Delta(t)) = z_n^-(t) + w_n(\Delta(t)), \quad n \in \mathcal{N}(t + \Delta(t)).$$

Предположим, что в поднаборе $w_n(\Delta(t))$ либо находится уникальный фрагмент данных задачи z_n , либо z_n не имеет решения, но обработка $w_n(\Delta(t))$ завершает выполнение задания. Тогда значение $z_n^-(t + \Delta(t)) = z_n^-(t) + w_n(\Delta(t))$ является оценкой сверху для общего числа фрагментов данных, которые в действительности необходимо просмотреть для завершения задания z_n или для решения соответствующей ему задачи. В этом случае справедливо неравенство:

$$z_n^-(t) + w_n(\Delta(t)) \geq \psi_n,$$

а величина

$$\bar{\tau}_n(\Delta(t)) = \frac{z_n^-(t) + w_n(\Delta(t))}{P_0} \quad (1)$$

в момент времени $t + \Delta(t)$ является оценкой сверху для τ_n — *абсолютного времени* выполнения задания в монопольном режиме — и

$$\bar{\tau}_n(\Delta(t)) \geq \tau_n.$$

Если же $w_n(\Delta(t))$ не содержит уникального фрагмента и в момент $t + \Delta(t)$ выполнение задания не завершается, то обработка z_n будет продолжена. При этом

$$z_n^-(t) + w_n(\Delta(t)) \leq \psi_n,$$

а величина (1) в момент $t + \Delta(t)$ окажется оценкой снизу для τ_n :

$$\bar{\tau}_n(\Delta(t)) \leq \tau_n.$$

Обозначим через $T_n(t)$ длительность пребывания z_n в СВС на момент t ,

$$T_n(t) = t - t_n^0,$$

где t_n^0 — время поступления задания z_n в СВС.

Предположим, что задание z_n будет закончено к моменту $t + \Delta(t)$. Тогда время пребывания z_n в СВС составит

$$T_n(t + \Delta(t)) = t + \Delta(t) - t_n^0.$$

Введем переменную

$$\chi_n(t + \Delta(t)) = \frac{\bar{\tau}_n(\Delta(t))}{T_n(t + \Delta(t))}, \quad (2)$$

которую будем называть *показателем относительной задержки выполнения задания z_n* к моменту времени $t + \Delta(t)$. По значению $\chi_n(t + \Delta(t))$ можно оценить, какую долю составляет *абсолютное время* обработки z_n в монопольном режиме от фактического времени пребывания в СВС при условии, что оно будет завершено к моменту $t + \Delta(t)$.

Величину, обратную (2), обозначим $\varepsilon_n(t + \Delta(t)) = (\chi_n(t + \Delta(t)))^{-1}$ и назовем коэффициентом относительной задержки. Значение $\varepsilon_n(t + \Delta(t))$ показывает, во сколько раз время фактического пребывания в системе оказалось больше необходимого времени обработки, т. е. характеризует задержку выполнения z_n относительно *абсолютного времени* его просмотра в монопольном режиме. Свойства функции $\varepsilon_n(t + \Delta(t))$ подробно изучались в [7, 8], а результаты ее применения при диспетчеризации изложены ниже в разд. 4.

В ПК-плане величины выполняемых подзаданий $w_n(\Delta(t))$ выбираются так, чтобы для всех заданий z_n минимальная величина $\chi_n(t + \Delta(t))$ оказалась максимальной из возможных. Формально, ведется поиск

$$\theta^*(t) = \max_w \min_{1 \leq n \leq N(t)} \chi_n(t + \Delta(t)). \quad (3)$$

Пусть в момент времени t заданы плановый период $\Delta(t)$ и предельно допустимый объем ТР-пакета $W^\Delta(\Delta(t))$. В сделанных предположениях поиск решения (3) можно свести к следующей задаче линейного программирования:

найти

$$\theta^*(t) = \max_{\theta, w, W, \chi} \theta(t) \quad (4)$$

при ограничениях:

– на $\theta(t)$ — минимальную величину показателя относительной задержки:

$$\left. \begin{aligned} \theta(t + \Delta(t)) &\leq \chi_n(t + \Delta(t)); \\ \chi_n(t + \Delta(t)) &= \frac{z_n^-(t) + w_n(\Delta(t))}{P_0 T_n(t + \Delta(t))}, \quad n \in \mathcal{N}(t); \end{aligned} \right\} \quad (5)$$

– на размеры поднаборов фрагментов данных:

$$\left. \begin{aligned} \mathbf{Z}_n - z_n^-(t) &\geq w_n(\Delta(t)) \geq 0, \quad n \in \mathcal{N}(t); \\ \sum_{n \in \mathcal{N}(t)} w_n(\Delta(t)) &= W(\Delta(t)); \\ W^\Delta(\Delta(t)) &\geq W(\Delta(t)). \end{aligned} \right\} \quad (6)$$

Задача (4)–(6) может быть решена стандартными методами линейного программирования [9]. Пусть найдено максимальное значение функционала $\theta^*(t)$. Из множества оптимальных решений (4)–(6) для полученного $\theta^*(t)$ выделим вектор $\mathbf{w}^*(t)$ по следующему правилу:

$$w_n^*(\Delta(t)) = \begin{cases} 0, & \text{если } \theta^*(t) \leq \frac{z_n^-(t)}{P_0 \cdot T_n(t + \Delta(t))}; \\ \theta^*(t) P_0 T_n(t + \Delta(t)) - z_n^-(t), & \text{если } \theta^*(t) > \frac{z_n^-(t)}{P_0 T_n(t + \Delta(t))}. \end{cases} \quad (7)$$

Для заданного $\theta^*(t)$ и вектора $\mathbf{w}^*(\Delta(t)) = (w_1^*(\Delta(t)), w_2^*(\Delta(t)), \dots, w_{N(t)}^*(\Delta(t)))$ согласно (7) введем множества:

$$\begin{aligned} \mathcal{N}^0(t) &= \{n \mid w_n^*(\Delta(t)) = 0, n \in \mathcal{N}(t)\}, \\ \mathcal{N}^*(t) &= \{n \mid w_n^*(\Delta(t)) > 0, n \in \mathcal{N}(t)\}. \end{aligned}$$

Из определения следует

$$\mathcal{N}(t) = \mathcal{N}^*(t) \cup \mathcal{N}^0(t).$$

Вектор значений $(\theta^*(t), w_1^*(\Delta(t)), \dots, w_{N(t)}^*(\Delta(t)))$ является оптимальным решением (4)–(6). Согласно (7) размеры подзаданий, планируемых для выполнения к моменту $t + \Delta(t)$, удовлетворяют равенствам:

$$\theta^*(t) = \frac{z_n^-(t) + w_n^*(\Delta(t))}{P_0 T_n(t + \Delta(t))}, \quad n \in \mathcal{N}^*(t),$$

и согласно (6), (7)

$$\sum_{n \in \mathcal{N}(t)} w_n^*(\Delta(t)) = \sum_{n \in \mathcal{N}^*(t)} w_n^*(\Delta(t)) = W^*(\Delta(t)).$$

После преобразований получим:

$$\theta^*(t) = \frac{W^*(\Delta(t)) + \sum_{n \in \mathcal{N}^*(t)} z_n^-(t)}{P_0 \sum_{n \in \mathcal{N}^*(t)} T_n(t + \Delta(t))};$$

$$w_n^*(\Delta(t)) = \theta^*(t)P_0T_n(t + \Delta(t)) - z_n^-(t) =$$

$$= T_n(t + \Delta(t)) \frac{W^*(\Delta(t)) + \sum_{n \in \mathcal{N}^*(t)} z_n^-(t)}{\sum_{n \in \mathcal{N}^*(t)} T_n(t + \Delta(t))} - z_n^-(t),$$

$$n \in \mathcal{N}^*(t).$$

Вектор значений $\mathbf{w}^*(t) = \langle w_1^*(\Delta(t)), w_2^*(\Delta(t)), \dots, w_{N(t)}^*(\Delta(t)) \rangle$ назовем распределением выполняемых подзаданий $w_n(\Delta(t))$ по критерию (3). Для всех z_n , завершённых на интервале $\Delta(t)$, показатель относительной задержки выполнения будет не меньше $\theta^*(t)$.

Указанное распределение предполагает, что для всех заявок-заданий просматривается доля фрагментов данных с учетом оценок их *абсолютного времени* решения и фактического пребывания в СВС. Правило обработки заданий, основанное на учете показателя относительной задержки выполнения, будем обозначать аббревиатурой SWAP (*от англ. short work ahead performance*) и называть SWAP-диспетчеризацией.

3 Анализ результатов выполнения пакета текущих работ

Предположим, что осуществляется SWAP-диспетчеризация поступающих заданий. Исследуем результаты реализации такой стратегии. Будем считать, что в контрольный момент t на основании решения задачи (4)–(6) и согласно (7) был сформирован TP-пакет $\mathbf{w}^*(t) = \langle w_1^*(\Delta(t)), w_2^*(\Delta(t)), \dots, w_{N(t)}^*(\Delta(t)) \rangle$, обработка которого завершилась в течение планового периода $\Delta(t)$.

Пусть некоторая задача z_{i_+} решена на интервале $\Delta(t)$. Тогда в момент $t + \Delta(t)$ справедливы следующие неравенства:

$$z_{i_+}^-(t) + w_{i_+}^*(\Delta(t)) \geq \psi_{i_+};$$

$$\frac{z_{i_+}^-(t) + w_{i_+}^*(\Delta(t))}{P_0} \geq \tau_{i_+},$$

где τ_{i_+} — *абсолютное время* обработки задания z_{i_+} в монопольном режиме.

Значения $\theta^*(t), w_{i_+}^*(\Delta(t))$ — оптимальное решение (4)–(6) и согласно (7) удовлетворяют соотношениям:

$$\theta^*(t)(t + \Delta(t) - t_{i_+}^0) = \theta^*(t)T_{i_+}(t + \Delta(t)) =$$

$$= \frac{z_{i_+}^-(t) + w_{i_+}^*(\Delta(t))}{P_0} \geq \tau_{i_+};$$

$$P_0\theta^*(t)(t + \Delta(t) - t_{i_+}^0) \geq \psi_{i_+},$$

а также справедливы неравенства

$$\left. \begin{aligned} \theta^*(t)(t + \Delta(t)) &\geq \tau_{i_+} + t_{i_+}^0 \theta^*(t); \\ P_0\theta^*(t)(t + \Delta(t)) &\geq \psi_{i_+} + t_{i_+}^0 P_0\theta^*(t). \end{aligned} \right\} \quad (8)$$

Рассмотрим группу заданий, которые поступили в СВС раньше или одновременно с z_{i_+} , но не были завершены к моменту $t + \Delta(t)$. Обозначим множество номеров таких заданий через

$$\mathcal{N}^+(i_+, t + \Delta(t)) = \{n \mid t_n^0 \leq t_{i_+}^0, n \in \mathcal{N}(t + \Delta(t))\}.$$

Для любого задания $z_j, j \in \mathcal{N}^+(i_+, t + \Delta(t))$, к моменту $t + \Delta(t)$ уникальный фрагмент не обнаружен, поэтому справедливы оценки:

$$\psi_j \geq z_j^-(t) + w_j^*(\Delta(t));$$

$$\tau_j \geq \frac{z_j^-(t) + w_j^*(\Delta(t))}{P_0}.$$

Поскольку $w_j^*(\Delta(t))$ — оптимальное решение (4)–(6), то согласно (7)

$$\bar{\tau}_j(\Delta(t)) = \frac{z_j^-(t) + w_j^*(\Delta(t))}{P_0} = \theta^*(t)T_j(t + \Delta(t)) =$$

$$= \theta^*(t)(t + \Delta(t) - t_j^0) \leq \tau_j$$

и справедливы неравенства:

$$\left. \begin{aligned} \theta^*(t)(t + \Delta(t)) &\leq \tau_j + t_j^0 \theta^*(t); \\ P_0\theta^*(t)(t + \Delta(t)) &\leq \psi_j + t_j^0 P_0\theta^*(t). \end{aligned} \right\} \quad (9)$$

Сравнивая (8) и (9), получаем:

$$\tau_j + \theta^*(t)t_j^0 \geq \tau_{i_+} + \theta^*(t)t_{i_+}^0;$$

$$\tau_j \geq \tau_{i_+} + \theta^*(t)(t_{i_+}^0 - t_j^0).$$

Так как $j \in \mathcal{N}^+(i_+, t + \Delta(t))$, то $t_{i_+}^0 \geq t_j^0$ и

$$\tau_j \geq \tau_{i_+}; \quad \psi_j \geq \psi_{i_+}.$$

Следовательно, для задачи z_{i_+} , завершённой на интервале $\Delta(t)$, абсолютное время решения меньше, чем у незаконченных заданий, поступивших в СВС одновременно или даже раньше z_{i_+} . Другими словами, для любого задания z_{i_+} , выполненного к моменту $t + \Delta(t)$, объем необходимой работы меньше, чем у всех других заданий, поступивших в СВС одновременно или даже раньше z_{i_+} , но не завершённых к моменту $t + \Delta(t)$.

4 Одновременное поступление заданий в систему

Для иллюстрации особенностей SWAP-диспетчеризации рассмотрим процесс выполнения t^0 -комплекта — группы однородных заданий, которые поступили в СВС одновременно в момент t^0 .

Пусть в момент $t = t^0$ заданы плановый период Δ и общее число фрагментов $W^\Delta(\Delta)$, которые могут быть обработаны за этот интервал. Фактически $W^\Delta(\Delta)$ определяется вычислительными ресурсами, выделяемыми для выполнения t^0 -комплекта.

Пусть для t^0 -комплекта в момент t^0 для всех z_n , $n \in \mathcal{N}(t^0)$, $t_n^0 = t^0$, $z_n^-(t^0) = 0$, $\mathbf{Z}_n = \mathbf{Z}^0$, $N(t^0) = N^0$, $P_0 = P$, $\Delta(t) = \Delta$. Кроме того, в рамках примера, не ограничивая общности, предположим, что $N^0 \mathbf{Z}^0 \gg W^\Delta(\Delta)$. Запишем задачу (4)–(6) для этих исходных данных:

найти

$$\theta^*(t^0) = \max_{\theta, w} \theta(t^0) \quad (10)$$

при ограничениях

$$\begin{aligned} \theta(t^0) &\leq \chi_n(t^0 + \Delta), \\ \chi_n(t^0 + \Delta) &= \frac{w_n(\Delta)}{P\Delta}, \quad n \in \mathcal{N}(t^0); \\ \sum_{n \in \mathcal{N}(t^0)} w_n(\Delta) &\leq W^\Delta(\Delta); \\ \mathbf{Z}^0 \geq w_n(\Delta) \geq 0, \quad n &\in \mathcal{N}(t^0). \end{aligned}$$

Согласно разд. 3, оптимальное решение задачи (10) для всех $n \in \mathcal{N}(t^0)$:

$$w_n^*(\Delta) = w^* = \frac{W^\Delta(\Delta)}{N^0}.$$

Таким образом, суммарный вычислительный «ресурс» делится на равные части между всеми заданиями. Следовательно, для всех заданий z_n , $n \in \mathcal{N}(t^0)$, на интервале Δ предполагается обработать одинаковое число фрагментов данных — w^* , и ТР-пакет состоит из равных по размеру поднаборов данных всех задач.

В некоторый момент времени одна из задач будет решена. Пусть это произошло в момент \hat{t} и решена l -я задача. Будем считать, что \hat{t} является началом нового планового периода. При этом в момент $t = \hat{t}$ $\mathcal{N}(t) = \mathcal{N}(\hat{t}) = \{1, 2, \dots, l-1, l+1, \dots, N(t^0)\}$ и для любого $n \in \mathcal{N}(\hat{t})$ $z_n^-(t) = z^-(\hat{t})$, $T_n(t) = \hat{t} - t^0$.

Перепишем (4)–(6) для этих условий в момент $t = \hat{t}$:

найти

$$\theta^*(\hat{t}) = \max_{\theta, w} \theta(\hat{t})$$

при ограничениях

$$\begin{aligned} \theta(\hat{t}) &\leq \chi_n(\hat{t} + \Delta), \\ \chi_n(\hat{t} + \Delta) &= \frac{z_n^-(\hat{t}) + w_n(\Delta)}{P(\hat{t} + \Delta - t^0)}, \quad n \in \mathcal{N}(\hat{t}); \\ \sum_{n \in \mathcal{N}(\hat{t})} w_n(\Delta) &\leq W^\Delta(\Delta); \\ \mathbf{Z}^0 - z_n^-(\hat{t}) \geq w_n(\Delta) \geq 0, \quad n &\in \mathcal{N}(\hat{t}). \end{aligned}$$

Согласно разд. 3, оптимальное решение задачи для всех $n \in \mathcal{N}(\hat{t})$ состоит из равных по размеру поднаборов данных:

$$w_n^*(\Delta) = w^* = \frac{W^\Delta(\Delta)}{N(\hat{t})} = \frac{W^\Delta(\Delta)}{N^0 - 1}.$$

Цепочку приведенных рассуждений можно продолжить дальше, при этом оптимальный ТР-пакет после окончания очередной задачи будет состоять из равных по размеру поднаборов данных всех нерешенных на данный момент задач.

Описанная стратегия обработки t^0 -комплекта однородных situ-заданий в [7, 8] была названа *параллельной* обработкой ТР-пакета и состояла в следующем: в момент t каждому заданию z_n назначался некоторый фиксированный набор ЕВ-модулей, число которых для всех заданий комплекта одинаково. Например, программа-диспетчер формировала соответствующий список, где указывала, какие ЕВ-модули начиная с момента t будут выполнять конкретное задание z_n . При этом если некоторая задача z_l была решена в момент t , то освободившийся ресурс — работоспособные ЕВ-модули — был разделен поровну и использован ЦУП-устройством для обработки тех заданий z_j , которые еще не завершены.

В качестве примера рассмотрим выполнение t^0 -комплекта из $N^0 > 1$ заданий, в котором все задачи требуют различных объемов *необходимой работы* ψ_n . Кроме того, известно, что ровно одна из задач не имеет решения и для ее окончания потребует просмотреть все данные.

При SWAP-диспетчеризации данного t^0 -комплекта первой будет решена задача с минимальным объемом *необходимой работы*. Обозначим ее z_1 . Следующей по порядку из системы выйдет задача с минимальным значением *необходимой работы* из оставшихся, обозначим ее через z_2 и т.д. В результате все задания окажутся перенумерованы по возрастанию объемов *работы*, необходимой для выполнения. Единственное задание, которое требует полного перебора данных, завершится последним, получит номер N^0 , и для него $\psi_{N^0} = \mathbf{Z}^0$. При этом

$$\psi_1 < \psi_2 < \psi_3 < \dots < \psi_{N^0-1} < \psi_{N^0} = \mathbf{Z}^0.$$

Обозначим через ψ^* объем *необходимой работы* для завершения последней задачи, имеющей решение, т. е. положим $\psi^* = \psi_{N^0-1}$.

Пусть задача z_{N^0-1} решена в момент времени t^* . Тогда для единственного незавершенного задания z_{N^0} к этому моменту просмотрено

$$z_{N^0}^-(t^*) = \psi_{N^0-1} = \psi^*$$

фрагментов и найдены все уникальные фрагменты рассматриваемого t^0 -комплекта. Таким образом, к моменту времени t^* СВС обработала в сумме

$$\Psi_{\text{par}}^* = \sum_{n=1}^{N^0-1} \psi_n + z_{N^0}^-(t^*) = \sum_{n=1}^{N^0-1} \psi_n + \psi^* \quad (11)$$

фрагментов данных для всех задач, включая единственную незавершенную.

В [7, 8] рассматривалась также альтернативная схема выполнения, при которой задания z_n выполнялись в случайном порядке одно за другим, каждое сразу на всех работоспособных ЕВ-модулях. Данное правило предполагало, что как только очередная задача z_l завершена, сразу начинается обработка набора данных следующего задания, выбранного случайным образом из $Z(t)$, и было названо *последовательным* монопольным режимом. Исследуем, каким окажется объем *необходимой работы* для обнаружения всех уникальных фрагментов для рассмотренного выше примера t^0 -комплекта при условии, что он выполняется последовательно. Будем считать, что задания заранее пронумерованы в соответствии с объемом *необходимой работы*, т. е. в том порядке, в котором они покидают систему при параллельной дисциплине обслуживания.

Согласно приведенному в [7, 8] правилу в момент t^0 очередность выполнения заданий в монопольном режиме при последовательной обработке определяется по жребию. Тогда только в одном случае из N^0 задание z_{N^0} окажется в очереди строго на последнем месте, а все уникальные фрагменты будут идентифицированы до начала выполнения z_{N^0} . Всего при этом будет обработано

$$\Psi_{\text{seq}}^{\min} = \sum_{n=1}^{N^0-1} \psi_n \quad (12)$$

фрагментов.

Для случаев, когда задание z_{N^0} оказывается в очереди на любом месте, кроме последнего, для идентификации всех уникальных фрагментов необходимо будет обработать

$$\Psi_{\text{seq}}^{\max} = \sum_{n=1}^{N^0-1} \psi_n + Z_{N^0}$$

фрагментов данных. Сравнивая эту величину с (11) и (12), получим

$$\begin{aligned} \sum_{n=1}^{N^0-1} \psi_n = \Psi_{\text{seq}}^{\min} < \Psi_{\text{par}}^* &= \sum_{n=1}^{N^0-1} \psi_n + \psi^* < \\ < \Psi_{\text{seq}}^{\max} = \sum_{n=1}^{N^0-1} \psi_n + Z_{N^0}. \end{aligned}$$

При использовании SWAP-диспетчеризации для выполнения конкретного t^0 -комплекта объемом выполненной *необходимой работы* — значение (10) — является «постоянной» величиной. При использовании *последовательной* дисциплины обслуживания со случайной очередностью выполнения для идентификации всех уникальных фрагментов, содержащихся в t^0 -комплекте, почти всегда (за исключением одного случая из N^0) потребуется больше времени, чем при SWAP-диспетчеризации.

Для сравнения SWAP-диспетчеризации и последовательной дисциплины обслуживания при обработке различных по структуре t^0 -комплектов была проведена серия имитационных экспериментов, которые описаны в [7, 8]. Остановимся на основных результатах исследования. Последовательная стратегия при не известной заранее длительности выполнения является дискриминирующей по отношению к заявкам, имеющим время решения меньше среднего для данного t^0 -комплекта. Кроме того, случайный выбор заданий привносит дополнительную неопределенность — место в очереди на обработку оказывается произвольным. В результате задачи с небольшим объемом *необходимой работы* задерживаются в СВС на неоправданно большой промежуток времени, ожидая завершения более ресурсоемких заданий.

Схема выполнения заданий t^0 -комплекта на основе показателя относительной задержки (решения задачи (4)–(6)) обладает следующими свойствами:

- в первую очередь завершаются задания, имеющие наименьшее абсолютное время решения;
- в ходе выполнения задания фактически упорядочиваются в порядке неубывания длительности работ, а задания, требующие полного перебора всего массива данных, завершаются последними, при этом число ЕВ-модулей, выделяемое для их обработки, монотонно возрастает;
- и, наконец, самое важное: в подавляющем числе случаев суммарное время извлечения всей содержащейся в t^0 -комплекте информации, т. е. идентификации всех уникальных фрагментов, имеющихся во всех наборах данных, оказывается меньше, чем при последовательном правиле обслуживания в монопольном режиме.

5 Заключение

При реализации правила SWAP не делается никаких гипотез о времени поступления, объеме *необходимой работы* или наличии уникальных фрагментов в поступивших данных. В каждой контрольной точке для каждого задания используются только следующие параметры и величины:

- фактическое время поступления в СВС;
- объем обработанных и оставшихся непросмотренными фрагментов данных.

Анализ выполнения однородных заданий в СВС с однотипными ЕВ-модулями наглядно демонстрирует преимущества SWAP-диспетчеризации. В условиях объективной неопределенности исполнение правила SWAP позволяет в динамике упорядочивать и завершать задания с учетом объемов *необходимой работы*, при этом приоритет получают «менее трудоемкие» задачи.

Литература

1. Купалов-Ярополк И. К., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Методы оценки эффективности и директивных сроков выполнения ресурсоемких вычислительных заданий // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–12.
2. Малашенко Ю. Е., Назарова И. А. Модель управления разнородными вычислительными заданиями на основе гарантированных оценок времени выполнения // Изв. РАН. ТиСУ, 2012. № 4. С. 29–38.
3. Купалов-Ярополк И. К., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Модели и программы для системы управления ресурсоемкими вычислениями. — М.: ВЦ РАН, 2013. 72 с. http://www.ccas.ru/depart/malashen/papper/ronzhin_2012_preprint.pdf.
4. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971. 384 с.
5. Сухарев А. Г., Тимохов А. В., Федоров В. В. Курс методов оптимизации. — М.: Наука, 1986. 368 с.
6. Голосов П. С., Козлов М. В., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Модель системы управления специализированным вычислительным комплексом. — М.: ВЦ РАН, 2010. 48 с. http://www.ccas.ru/depart/malashen/papper/golosov_2010_preprint.pdf.
7. Козлов М. В., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Анализ режимов управления вычислительным комплексом в условиях неопределенности. — М.: ВЦ РАН, 2011. 60 с. http://www.ccas.ru/depart/malashen/papper/ronzhin_2011_preprint.pdf.
8. Голосов П. Е., Козлов М. В., Малашенко Ю. Е., Назарова И. А., Ронжин А. Ф. Анализ управления специализированными вычислительными заданиями в условиях неопределенности // Изв. РАН. ТиСУ, 2012. № 1. С. 50–66.
9. Данциг Дж. Б. Линейное программирование, его применения и обобщения / Пер. с англ. — М.: Прогресс, 1966. 600 с. (Dantzig, G. 1963. *Linear programming and extensions*. New Jersey: Princeton University Press. 589 p.)

Поступила в редакцию 08.11.13

ANALYSIS OF DELAYS IN SCHEDULING HOMOGENEOUS TASKS UNDER UNCERTAINTY

Yu. E. Malashenko and I. A. Nazarova

Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The problem of management of the computationally resource-intensive tasks of search type allowing parallelization by the data is considered. Tasks arrive in a system at any time one by one or in groups; their service time is not known in advance. For processing planning, the optimization model is used which is based on current information on tasks performance: the sojourn time and the amount of data already processed. Using the model for each task, the portion of data to be processed in the plan period is determined. In calculations, required computational expenses are estimated and assumptions about the distribution laws of unknown tasks characteristics are not made. The proposed scheduling rule allows to form the order of task execution in dynamics, priority being given to “less intensive” tasks.

Keywords: computationally intensive tasks; parallel computing; scheduling optimization; principle of guaranteed result

DOI: 10.14357/19922264140102

References

1. Kupalov-Yaropolk, I. K., Yu. E. Malashenko, I. A. Nazarova, and A. F. Ronzhin. 2013. Metody otsenki effektivnosti i direktivnykh srokov vypolneniya resursoemkikh vychislitel'nykh zadaniy [Methods of estimating efficiency and directive deadlines for resource-intensive computational tasks]. *Informatika i ee primeneniya — Inform. Appl.* 7(2):4–12.
2. Malashenko, Yu. E., and I. A. Nazarova. 2012. Control model for heterogeneous computational tasks based on guaranteed estimates of execution times. *J. Comput. Syst. Sci. Int.* 51:526–534.
3. Kupalov-Yaropolk, I. K., Yu. E. Malashenko, I. A. Nazarova, and A. F. Ronzhin. 2013. Modeli i programmy dlya sistemy upravleniya resursoemkimi vychisleniyami [Models and programs for intensive computing management]. Moscow: Vychisl. Tsentr Ross. Akad. Nauk Publ. 72 p. Available at: http://www.ccas.ru/depart/malashen/papper/ronzhin_2012_preprint.pdf.
4. Germeier, Yu. B. 1971. *Vvedenie v teoriyu issledovaniya operatsiy* [An introduction to operations research theory]. Moscow: Nauka Publ. 384 p.
5. Sukharev, A. G., A. V. Timokhov, and V. V. Fedorov. 1986. *Kurs metodov optimizatsii* [A course in optimization methods]. Moscow: Nauka Publ. 368 p.
6. Golosov, P. E., M. V. Kozlov, Yu. E. Malashenko, I. A. Nazarova, and A. F. Ronzhin. 2010. Model' sistemy upravleniya spetsializirovannym vychislitel'nym kompleksom [Control model for a special computer system]. Moscow: Vychisl. Tsentr Ross. Akad. Nauk Publ. 48 p. Available at: http://www.ccas.ru/depart/malashen/papper/golosov_2010_preprint.pdf.
7. Kozlov, M. V., Yu. E. Malashenko, I. A. Nazarova, and A. F. Ronzhin. 2011. Analiz rezhimov upravleniya vychislitel'nym kompleksom v usloviyakh neopredelennosti [Analysis of a computer system control modes under uncertainty]. Moscow: Vychisl. Tsentr Ross. Akad. Nauk Publ. 60 p. Available at: http://www.ccas.ru/depart/malashen/papper/ronzhin_2011_preprint.pdf.
8. Golosov, P. E., M. V. Kozlov, Yu. E. Malashenko, I. A. Nazarova, and A. F. Ronzhin. 2012. Analysis of computer job control under uncertainty. *J. Comput. Syst. Sci. Int.* 51:49–64.
9. Dantzig, G. *Linear programming and extensions*. New Jersey: Princeton University Press, 1963. 589 p.

Received November 8, 2013

Contributors

Malashenko Yuri E. (b. 1946) — Doctor of Science in physics and mathematics, Head of Department of Operations Research, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

Nazarova I. A. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

ОЦЕНКА НАДЕЖНОСТИ СЛОЖНЫХ СИСТЕМ С ВОССТАНОВЛЕНИЕМ ПО РЕЗУЛЬТАТАМ ИСПЫТАНИЙ ЭЛЕМЕНТОВ

И. В. Павлов¹

Аннотация: Рассматривается задача доверительного оценивания показателей надежности для сложных систем с сетевой структурой с восстанавливаемыми элементами. Оценка надежности системы производится по результатам испытаний ее отдельных компонентов (элементов, подсистем). Существующие методы решения данной проблемы разработаны для относительно простых последовательно-параллельных структур и в предположении, что элементы системы имеют экспоненциальные распределения времени безотказной работы. Предлагается решение этой задачи для более общей модели сложных «монотонных структур», а также для более общего случая «стареющих» (с монотонно возрастающей функцией интенсивности отказов) элементов системы. Предполагается, что в случае отказа элементы системы восстанавливаются независимо от состояния других элементов. Кроме того, решение данной проблемы получено в естественной с прикладной точки зрения асимптотике, а именно для случая высокой надежности (быстрого восстановления) элементов системы.

Ключевые слова: сложные системы; сетевые структуры; надежность; время безотказной работы; время восстановления; функция ресурса; функция интенсивности отказов

DOI: 10.14357/19922264140103

1 Введение

Обеспечение надежности является одним из ключевых требований при проектировании и эксплуатации современных информационно-вычислительных систем. Возникающая в связи с этим в инженерной практике задача оценки характеристик надежности по результатам испытаний системы или ее отдельных компонентов (элементов, подсистем) является одной из актуальных и до настоящего времени в значительной степени нерешенных проблем. При этом основной практический интерес чаще всего представляет построение гарантированных или доверительных оценок для тех или иных показателей надежности системы. Задачи подобного типа возникают при оценке характеристик надежности систем связи и управления, банковских компьютерных сетей и ряда других сложных систем с сетевой структурой (см., например, [1–8]). Аналогичные задачи возникают при развитии и наращивании сети, когда требуется дать гарантированный прогноз характеристик надежности для различных возможных вариантов системы на основе статистической информации по элементам.

Пусть в системе имеется m различных элементов. С точки зрения работоспособности состояние системы в момент времени $t \geq 0$ харак-

теризуется вектором двоичных переменных $x_t = [x_1(t), \dots, x_m(t)]$, где $x_i(t)$ — индикатор исправности i -го элемента: $x_i(t) = 1$, если в момент времени t i -й элемент исправен, $x_i(t) = 0$ в противном случае, $i = 1, \dots, m$. Работоспособность системы в состоянии $x = (x_1, \dots, x_m)$ характеризуется «структурной функцией»: $\Phi(x) = 1$, если в состоянии x система исправна, и $\Phi(x) = 0$ в противном случае. Предполагается, что эта функция удовлетворяет естественному условию монотонности $\Phi(x_1, \dots, x_m) \geq \Phi(y_1, \dots, y_m)$, если $x_i \geq y_i$ при всех $i = 1, \dots, m$, другими словами, дополнительные отказы элементов не улучшают состояние системы [2, 5]. Индикатор исправности системы в момент времени t имеет вид: $X(t) = \Phi(x_t)$.

Процесс функционирования i -го элемента $x_i(t)$ представляет собой альтернирующий процесс восстановления [1]; другими словами, последовательность сменяющих друг друга независимых между собой (и от состояния других элементов) интервалов исправной работы и восстановления.

Функции распределения интервалов исправной работы ξ_i и восстановления η_i для i -го элемента обозначим соответственно через $F_i(t)$ и $G_i(t)$, предполагая, что они имеют непрерывные плотности $f_i(t)$ и $g_i(t)$ и конечные вторые моменты. Обозначим также через $P_i(t) = 1 - F_i(t)$ функцию

¹Московский государственный технический университет им. Н. Э. Баумана, ipavlov@bmstu.ru

надежности, $u_i = \int_0^{\infty} P_i(t) dt$ — среднее время безотказной работы и $v_i = \int_0^{\infty} [1 - G_i(t)] dt$ — среднее время восстановления для i -го элемента, $i = 1, \dots, m$.

Далее предполагается, что параметры v_i — средние времена восстановления (замены) элементов — известны и малы по сравнению со средним временем безотказной работы; другими словами, выполняется следующее условие высокой надежности («быстрого восстановления») элементов системы:

$$v_i \ll u_i, \quad i = 1, \dots, m. \quad (1)$$

Задача заключается в оценке характеристик надежности системы по результатам испытаний элементов, полученным или в ходе испытаний отдельных элементов системы, или испытаний ее отдельных фрагментов, или системы в целом. При этом основной интерес для приложений чаще всего представляет построение гарантированных или доверительных оценок характеристик надежности системы. Методы решения данной проблемы в настоящее время разработаны для относительно простых последовательных и последовательно-параллельных структур для случая, когда элементы системы имеют экспоненциальные распределения времени безотказной работы [1, 4, 6]. Далее дается решение для указанной выше более общей модели «монотонных структур», а также для более общего случая, когда элементы системы имеют «стареющие» или ВФИ-распределения (с возрастающей функцией интенсивности отказов $r_i(t) = f_i(t)/P_i(t)$, $i = 1, \dots, m$) времени безотказной работы. Класс ВФИ-распределений является довольно общим и включает в себя целый ряд часто используемых параметрических семейств распределений, таких как экспоненциальное, Вейбулла, нормальное, распределение Эрланга, гамма-распределение и др. В то же время предположение о старении элементов системы является довольно естественным с физической точки зрения и часто используется в инженерной практике (см., например, [1, 2, 6]). В этом смысле выводы, основанные на предположении о старении элементов системы, являются значительно более корректными.

2 Нижняя доверительная граница для коэффициента готовности системы

Стандартный показатель надежности для систем с восстанавливаемыми элементами — коэффициент готовности (вероятность заставить систему в ис-

правном состоянии в стационарном режиме при $t \rightarrow \infty$) — для данной модели имеет вид [2, 5]:

$$K = K(u) = \sum_{i=1}^m \prod_{i=1}^m k_i^{x_i}(u_i) [1 - k_i(u_i)]^{1-x_i},$$

где $u = (u_1, \dots, u_m)$ — вектор параметров элементов, а сумма берется по всем $x = (x_1, \dots, x_m)$ таким, что $\Phi(x) = 1$, $k_i = k_i(u_i) = u_i/(u_i + v_i)$ — коэффициент готовности i -го элемента, $i = 1, \dots, m$.

Далее будем предполагать, что результаты испытаний по различным элементам системы представлены в виде стандартных статистических выборок с цензурированием типа $[N_i \text{ Б } r_i]$ (в обозначениях книги [1]), т. е. испытывались N_i элементов i -го типа до наблюдения $r_i \leq N_i$ отказов, в результате чего наблюдались последовательные моменты отказов $0 < t_1^{(i)} < \dots < t_{r_i}^{(i)}$, $i = 1, \dots, m$. В частном случае при $r_i = N_i$ данная схема испытаний содержит классическую схему полной выборки объема N_i , $i = 1, \dots, m$.

Обозначим через

$$R_i(t) = -\ln P_i(t) = \int_0^t r_i(z) dz$$

функцию ресурса для i -го элемента, V_i — множество всех функций ресурса $R_i(t)$ выпуклых вниз по $t \geq 0$ (при всех t таких, что $P_i(t) > 0$, $i = 1, \dots, m$), $R = \{R_1(t), \dots, R_m(t)\}$ — вектор функций ресурса по всем элементам, $V = V_1 \times V_2 \times \dots \times V_m$ — множество всех R таких, что функции $R_i(t)$ выпуклы вниз по $t \geq 0$. Далее будем предполагать, что $R \in V$, т. е. все элементы системы имеют ВФИ-распределения (с возрастающей функцией интенсивности отказов $r_i(t) = R_i(t)$, $i = 1, \dots, m$) времени безотказной работы.

Пусть S_i — суммарное время работы (наработка) элементов i -го типа на испытаниях:

$$S_i = t_1^{(i)} + t_2^{(i)} + \dots + t_{r_i}^{(i)} + (N_i - r_i)t_{r_i}^{(i)}, \quad i = 1, \dots, m;$$

$S = (S_1, \dots, S_m)$ и $P_R(S)$ — вероятностное распределение на множестве результатов испытаний S при данном $R \in V$. Требуется построить нижнюю γ -доверительную границу для коэффициента готовности системы $K = K(u)$, т. е. функцию результатов испытаний $\underline{K}(S)$ такую, что

$$P_R \{ \underline{K}(S) \leq K(u) \} \geq \gamma, \quad R \in V.$$

В [9, 10] ранее была получена нижняя γ -доверительная граница \underline{u}_i в классе ВФИ-распределений для параметра u_i (среднего времени безотказной работы) одного отдельно взятого i -го элемента:

$$P_R \{ \underline{u}_i \leq u_i \} \geq \gamma, \quad R \in V. \quad (2)$$

Здесь $\underline{u}_i = b_i(\gamma)S_i$, где

$$b_i(\gamma) = \frac{1 - \exp[-\Delta_{1-\gamma}(r_i - 1)/N_i]}{\Delta_{1-\gamma}(r_i - 1)};$$

$\Delta_{1-\gamma}(d)$ — решение уравнения

$$\exp(-\Delta) \sum_{j=0}^d \frac{\Delta^j}{j!} = 1 - \gamma, \quad i = 1, \dots, m.$$

Коэффициент готовности системы $K(u) = K(u_1, u_2, \dots, u_m)$ монотонно возрастает по каждому аргументу u_i , что соответствует естественному свойству — улучшению характеристик системы при улучшении параметров надежности элементов. Поэтому нижнюю границу \underline{K} , вообще говоря, можно найти, просто подставив нижние доверительные оценки для параметров отдельных элементов \underline{u}_i в функцию $K(u)$, полагая $\underline{K} = K(\underline{u}_1, \underline{u}_2, \dots, \underline{u}_m)$. Тем не менее, если число элементов системы велико, такой подход, использующий только указанное свойство монотонности, оказывается малоприменимым. Действительно, при этом справедливы неравенства (см. также [11, 12]):

$$\begin{aligned} P_R \{K(\underline{u}_1, \underline{u}_2, \dots, \underline{u}_m) \leq K(u_1, u_2, \dots, u_m)\} &\geq \\ &\geq P_R \left\{ \bigcap_{i=1}^m (\underline{u}_i \leq u_i) \right\} = \\ &= \prod_{i=1}^m P_R \{\underline{u}_i \leq u_i\} > \gamma^m, \quad R \in V, \end{aligned}$$

откуда видно, что коэффициент доверия для такой процедуры (метод Ллойда и Липова, или «метод прямоугольника» в терминологии [11–13]) быстро убывает с ростом размерности задачи (числа элементов системы) m .

Обозначим через $Q(u) = 1 - K(u)$ вероятность отказа системы и $q_i(u) = 1 - k_i(u)$ — вероятность отказа i -го элемента, $i = 1, \dots, m$. В случае высокой надежности (быстрого восстановления) элементов (1) справедлива приближенная формула [2, 5]:

$$Q(u) \cong \sum_{j=1}^M \prod_{i \in B_j} q_i(u_i) \cong \sum_{j=1}^M \prod_{i \in B_j} \left(\frac{v_i}{u_i} \right), \quad (3)$$

где $B_j \subset (1, \dots, m)$ — набор индексов j -го минимального сечения системы, $j = 1, \dots, M$ (M — число минимальных сечений). Правая часть (3) при этом дает верхнюю оценку для $Q(u)$. Доверительное оценивание коэффициента готовности системы $K(u)$ в приближении (1), таким образом, сводится к построению верхней доверительной

границы для функции вида (3) от вектора $u = (u_1, \dots, u_m)$ неизвестных параметров надежности элементов.

В соответствии с (2) для каждого типа элементов $i = 1, \dots, m$ справедливы неравенства:

$$P_R \{u_i \geq b_i(\gamma)S_i\} \geq \gamma, \quad R \in V,$$

при любом $0 < \gamma < 1$. При данных фиксированных значениях результатов испытаний $S = (S_1, \dots, S_m)$ введем (независимые) случайные величины (с.в.) $\tilde{u}_i, i = 1, \dots, m$, такие, что

$$P(\tilde{u}_i \geq b_i(\gamma)S_i) = \gamma \quad (4)$$

при всех $0 < \gamma < 1, i = 1, \dots, m$. В соответствии с (4) функция надежности с.в. \tilde{u}_i имеет вид:

$$\tilde{P}_i(t) = P\{\tilde{u}_i \geq t\} = h_i\left(\frac{t}{S_i}\right),$$

где $h_i(z)$ — функция, обратная $b_i(\gamma)$, $0 < \gamma < 1$. Случайная величина \tilde{u}_i может быть также задана при данном фиксированном S_i как $\tilde{u}_i = S_i b_i(\gamma)$, где γ — с.в., равномерно распределенная на интервале $(0, 1)$.

Пусть имеется некоторая функция $Q(u) = Q(u_1, u_2, \dots, u_m)$, монотонно убывающая по каждому аргументу $u_i > 0, i = 1, \dots, m$. При данных фиксированных значениях результатов наблюдений $S = (S_1, \dots, S_m)$ и данном фиксированном значении коэффициента доверия $0 < \gamma < 1$ определим далее верхнюю границу $\bar{Q} = \bar{Q}(S_1, \dots, S_m)$ как квантиль уровня γ для случайной величины $\bar{Q} = Q(\tilde{u}_1, \dots, \tilde{u}_m)$, т.е. из условия

$$P\{Q(\tilde{u}_1, \dots, \tilde{u}_m) \leq \bar{Q}\} = \gamma$$

(численное значение \bar{Q} далее достаточно просто может быть найдено на основе стандартного метода Монте-Карло), что является обобщением фидуциального подхода при построении доверительных границ для функций от m неизвестных параметров экспоненциальных распределений [6, 14, 15].

В одномерном случае γ -фидуциальная граница параметра одновременно является γ -доверительной, что следует непосредственно из ее определения.

В многомерном случае при $m > 1$ это, вообще говоря, не обязательно так и зависит от того или иного вида оцениваемой функции многомерного параметра.

Теорема 1 дает достаточные условия, при которых определенная таким образом величина $\bar{Q} = \bar{Q}(S_1, \dots, S_m)$ служит верхней доверительной

границей с коэффициентом доверия не меньше γ для функции $Q(u) = Q(u_1, \dots, u_m)$ в классе стареющих распределений V .

Теорема 1. Пусть выполняются условия:

- (1) функция $Q(u) = Q(u_1, u_2, \dots, u_m)$ монотонно убывает по каждому $u_i > 0, i = 1, \dots, m$;
- (2) функция $Q(e^{-z_1}, e^{-z_2}, \dots, e^{-z_m})$ выпукла вниз по $z = (z_1, \dots, z_m) \in E_m$.

Тогда справедливо неравенство $P_R \{\bar{Q} \geq Q(u)\} \geq \gamma$ при всех $R \in V$.

Доказательство. В соответствии с (4) справедливо неравенство

$$P_R \left\{ \frac{u_i}{S_i} \geq b_i(\gamma) \right\} \geq \gamma \quad (5)$$

при всех $0 < \gamma < 1, R \in V, i = 1, \dots, m$. Введем с.в. $\xi_i = u_i/S_i$, распределение которой зависит от $R \in V, i = 1, \dots, m$. Введем также независимые с.в. η_1, \dots, η_m , где $\eta_i \geq 0$ имеет функцию распределения $P\{\eta_i < t\} = 1 - h_i(t), i = 1, \dots, m$.

В соответствии с (5) с.в. ξ_i стохастически больше, чем с.в. η_i , в следующем смысле:

$$P_R \{\xi_i \geq t\} \geq P\{\eta_i \geq t\}$$

при всех $t \geq 0, R \in V, i = 1, \dots, m$. Отсюда следует, что для любой функции $\varphi(z_1, \dots, z_m) \geq 0$, монотонно убывающей по каждому аргументу $z_i \geq 0$, справедливо аналогичное неравенство

$$P_R \{\varphi(\xi_1, \dots, \xi_m) \leq t\} \geq P\{\varphi(\eta_1, \dots, \eta_m) \leq t\}$$

при всех $t \geq 0, R \in V$. Полагая

$$\varphi(z_1, \dots, z_m) = \prod_{i=1}^m \left(\frac{1}{z_i} \right)^{a_i},$$

где $a_i \geq 0, i = 1, \dots, m$, — произвольные положительные константы, получаем:

$$P_R \left\{ \prod_{i=1}^m \left(\frac{S_i}{u_i} \right)^{a_i} \leq t \right\} \geq P \left\{ \prod_{i=1}^m \left(\frac{1}{\eta_i} \right)^{a_i} \leq t \right\} \quad (6)$$

при всех $t \geq 0, R \in V$. Обозначим через $t_\gamma = t_\gamma(a)$ квантиль уровня $0 < \gamma < 1$ для с.в. $\prod_{i=1}^m (1/\eta_i)^{a_i}$, где $a = (a_1, \dots, a_m)$. Из (6) далее получаем

$$P_R \left\{ \prod_{i=1}^m \left(\frac{S_i}{u_i} \right)^{a_i} \leq t_\gamma \right\} \geq P \left\{ \prod_{i=1}^m \left(\frac{1}{\eta_i} \right)^{a_i} \leq t_\gamma \right\} = \gamma,$$

а значит, справедливо неравенство

$$P_R \left\{ \prod_{i=1}^m \left(\frac{1}{u_i} \right)^{a_i} \leq t_\gamma \prod_{i=1}^m \left(\frac{1}{S_i} \right)^{a_i} \right\} \geq \gamma \quad (7)$$

при любом $R \in V$ и любых положительных коэффициентах $a_i \geq 0, i = 1, \dots, m$.

Введем замену переменных $u_i = e^{-z_i}, i = 1, \dots, m$. Из (7) следует, что приведенная выше процедура дает верхнюю доверительную границу с коэффициентами доверия не меньше γ для любой функции вида

$$H(u) = c_0 \prod_{i=1}^m \left(\frac{1}{u_i} \right)^{a_i},$$

где $c_0 > 0, a_i \geq 0, i = 1, \dots, m$, или в переменных $z = (z_1, \dots, z_m)$ для любой линейной функции вида

$$L(z, a) = a_0 + \sum_{i=1}^m a_i z_i, \quad (8)$$

где вектор коэффициентов $a = (a_0, a_1, \dots, a_m) \in A = \{a : -\infty < a_0 < \infty, a_i \geq 0, i = 1, \dots, m\}$. Обозначим через $\bar{L} = \bar{L}(S, a)$ соответствующую верхнюю γ -доверительную границу для линейной функции $L(z, a)$: $P_R \{\bar{L}(S, a) \geq L(z, a)\} \geq \gamma$ при всех $R \in V, a \in A$. Введем функцию

$$f(z) = Q(e^{-z_1}, \dots, e^{-z_m}). \quad (9)$$

Рассмотренная выше процедура построения верхней границы $\bar{Q} = \bar{Q}(S)$ как γ -квантиля с.в. $Q(\tilde{u}_1, \dots, \tilde{u}_m)$ эквивалентна построению верхней границы $\bar{f} = \bar{f}(S)$ для функции $f(z)$ как γ -квантиля для с.в. $f(\tilde{z}_1, \dots, \tilde{z}_m)$, где $\tilde{z}_i = \ln(1/\tilde{u}_i), i = 1, \dots, m$. При этом, очевидно, $\bar{f}(S) = \bar{Q}(S)$ при любом $S = (S_1, \dots, S_m)$. В условиях теоремы функция (9) монотонно возрастает по каждому z_i и выпукла вниз по $z = (z_1, \dots, z_m) \in E_m$. Следовательно, она может быть представлена через базовую систему линейных функций (8) в виде

$$f(z) = \max_{a \in A_f} L(z, a) \quad (10)$$

в каждой точке $z \in E_m$, где максимум берется по $a \in A_f \subset A$. Отсюда следует $f(z) \geq L(z, a)$ при всех $z \in E_m, a \in A_f$ и, соответственно,

$$\bar{f}(S) \geq \bar{L}(S, a), \quad a \in A_f,$$

при любом $S = (S_1, \dots, S_m)$. Значит,

$$\bar{f}(S) \geq \bar{L}(S) = \max_{a \in A_f} \bar{L}(S, a). \quad (11)$$

Пусть $z(a) \in A_f$ — точка, в которой достигается максимум в (10):

$$f(z) = \max_{a \in A_f} L(z, a) = L[z, a(z)]. \quad (12)$$

Зафиксируем $R \in V$ и соответственно параметры $u_i = u_i(R_i) = \int_0^\infty \exp[-R_i(t)] dt$ и $z_i = z_i(R_i) = \ln [1/u_i(R_i)]$, $i = 1, \dots, m$. Из (11), (12) далее следуют неравенства

$$\begin{aligned} P_R \{ \bar{f}(S) \geq f(z) \} &\geq P_R \{ \bar{L}(S) \geq f(z) \} \geq \\ &\geq P_R \{ \bar{L}[S, a(z)] \geq f(z) \} = \\ &= P_R \{ \bar{L}[S, a(z)] \geq L[z, a(z)] \} \geq \gamma \end{aligned}$$

при любом $R \in V$. Теорема доказана.

Вероятность отказа системы (3) в переменных $z = (z_1, \dots, z_m)$ имеет вид:

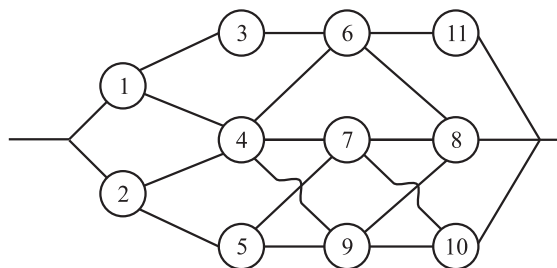
$$\begin{aligned} f(z) &= Q(e^{-z_1}, \dots, e^{-z_m}) = \\ &= \sum_{j=1}^M C_j \exp \left(\sum_{i \in B_j} z_i \right), \quad (13) \end{aligned}$$

где коэффициенты

$$C_j = \prod_{i \in B_j} v_i, \quad j = 1, \dots, M.$$

Функция (13) выпукла вниз по $z = (z_1, \dots, z_m)$ (в том числе и в случае, когда различные минимальные сечения системы B_j могут пересекаться, что имеет место для систем со сложной структурой). Тем самым функция $Q(u)$ в (3) удовлетворяет обоим условиям приведенной выше теоремы 1 (справедливость первого условия в данном случае очевидна). Таким образом, определенная выше величина \bar{Q} дает верхнюю доверительную границу с коэффициентом доверия не меньше заданной величины γ для вероятности отказа системы $Q(u)$ для общей модели сложных «монотонных структур» и при довольно общих непараметрических предположениях о том, что элементы системы — «стареющие» (с монотонно возрастающей функцией интенсивности отказов). Соответственно, величина $\underline{K} = 1 - \bar{Q}$ дает при этом нижнюю γ -доверительную границу для коэффициента готовности системы. (Заметим также, что, как видно из доказательства теоремы 1, построенная выше доверительная граница $\bar{Q} = \bar{Q}(S)$ может быть улучшена, если взять в качестве такой границы величину $\bar{L} = \bar{L}(S)$, но вычисление которой при каждом данном значении вектора результатов испытаний $S = (S_1, \dots, S_m)$ окажется значительно более сложным.)

В качестве примера рассмотрим систему из $m = 11$ элементов с сетевой структурой, изображенной на рисунке. Результаты испытаний N_i, r_i, S_i по элементам различных типов $i = 1, \dots, m$ приводятся в таблице. В этом случае нижняя γ -довери-



Система с сетевой структурой из $m = 11$ элементов

Результаты испытаний элементов

i	N_i	r_i	S_i
1	1	1	24,2
2	2	1	27,9
3	3	2	21,1
4	2	1	26,3
5	2	2	25,4
6	1	1	25,2
7	3	2	23,6
8	2	2	24,2
9	1	1	28,8
10	2	1	25,2
11	3	3	22,3

тельная граница (при $\gamma = 0,9$) для коэффициента готовности системы, вычисляемая на основе приведенной выше теоремы 1, $\underline{K} = 1 - \bar{Q} = 0,97$. При этом аналогичная нижняя доверительная граница, вычисляемая указанным выше упрощенным методом (Ллойда и Липова), основанным на непосредственном использовании частных доверительных границ для параметров отдельных элементов, равна $\underline{K} = 0,89$.

3 Заключение

Таким образом, получено решение задачи доверительного оценивания по результатам испытаний элементов системы одного из основных показателей надежности — коэффициента готовности для довольно общей модели сложных систем (с произвольной «монотонной структурой») с восстанавливаемыми «стареющими» элементами. Указанное решение получено в естественном с точки зрения приложений приближении — для случая высокой надежности («быстрого восстановления») элементов системы, а также в предположении независимого восстановления элементов. Существенный интерес с прикладной точки зрения представляет также дальнейшее обобщение полученных результатов как для более общих моделей систем (в том числе с зависимым восстановлением элементов), так и по отношению к другим часто используемым

показателям надежности систем с восстановлением — коэффициенту оперативной готовности, среднему времени безотказной работы системы и др.

Литература

1. Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д. Математические методы в теории надежности. — М.: Наука, 1965. 524 с.
2. Барлоу Р., Прошан Ф. Математическая теория надежности / Пер. с англ. — М.: Радио и связь, 1969. 488 с. (Barlow R., Proschan F. Mathematical theory of reliability. — N.Y.: John Wiley, 1965. 497 p.)
3. Васильев Н. С. Об одной модели развития сети связи // Известия Российской академии наук. Теория и системы управления, 1985. № 6. С. 227–234.
4. Павлов И. В. Приближенно оптимальные доверительные границы для показателей надежности систем с восстановлением // Известия Российской академии наук. Теория и системы управления, 1988. № 3. С. 109–116.
5. Павлов И. В., Ушаков И. А. Вычисление показателей надежности для сложных систем с восстанавливаемыми элементами // Известия Российской академии наук. Теория и системы управления, 1989. № 6. С. 170–176.
6. Gnedenko B. V., Pavlov I. V., Ushakov I. A. Statistical reliability engineering. — N.Y.: John Wiley, 1999. 517 p.
7. Коновалов М. Г. Организация работы вычислительно-го комплекса с помощью имитационной модели и адаптивных алгоритмов // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 37–48.
8. Павлов И. В. Расчет и оптимизация некоторых характеристик для модели вычислительного комплекса // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 59–62.
9. Barlow R., Proschan F. Tolerance and confidence limits for classes of distributions based on failure rate // Ann. Math. Stat., 1966. Vol. 37. No. 6. P. 1184–1195.
10. Павлов И. В. Доверительные границы в классе распределений с возрастающей функцией интенсивности отказов // Известия Российской академии наук. Теория и системы управления, 1977. № 6. С. 72–84.
11. Ллойд Д., Липов М. Надежность / Пер. с англ. — М.: Сов. радио, 1964. 668 с. (Lloyd D., Lipow M. Reliability management, methods and mathematics. — Englewood Cliffs, N.J.: Prentice-Hall, 1962. 684 p.)
12. Беляев Ю. К. Доверительные интервалы для функций от многих неизвестных параметров // Докл. АН СССР, 1967. Т. 196. № 4. С. 755–758.
13. Беляев Ю. К., Дугина Т. Н., Чепурин Е. В. Вычисление нижней доверительной оценки для вероятности безотказной работы сложных систем // Известия Российской академии наук. Теория и системы управления, 1967. № 2. С. 52–59.
14. Павлов И. В. О корректности фидуциального подхода при построении доверительных границ для показателей надежности сложных систем // Известия Российской академии наук. Теория и системы управления, 1981. № 5. С. 46–52.
15. Павлов И. В. О фидуциальном подходе при вычислении доверительных границ для функций многих неизвестных параметров // Докл. РАН, 1981. Т. 258. № 6. С. 1314–1317.

Поступила в редакцию 24.12.13

ESTIMATION OF RELIABILITY OF COMPLEX SYSTEMS WITH RENEWAL BASED ON ELEMENT TEST RESULTS

I. V. Pavlov

Bauman Moscow State Technical University, 5, 2nd Baumanskaya Str., Moscow 105005, Russian Federation

Abstract: The problem of confidence estimation of reliability of complex systems with network structure with repairable elements is considered. Estimation of reliability of a system is based on test results of its individual elements (subsystems). Existing methods for solving this problem are designed for relatively simple series-parallel structures consisting of elements with exponential distribution of time to failure. Solution of this problem is suggested for the more general model of “monotone structures” with independent renewable elements, as well as significantly more general case of “aging” system elements (with monotonically increasing function of failure rate). It is assumed that elements of the system are restored regardless of the state of other elements. In addition, the solution of this problem is obtained in the natural, from the practical point of view, asymptotic behavior for the case of high reliability (fast recovery) system elements.

Keywords: complex systems; network structures; reliability; time to failure; renewal time; resource function; failure rate function

DOI: 10.14357/19922264140103

References

1. Gnedenko, B. V., Ju. K. Beljaev, and A. D. Solov'ev. 1965. *Matematicheskie metody v teorii nadezhnosti* [Mathematical methods in reliability theory]. Moscow: Nauka. 524 p.
2. Barlow, R., and F. Proschan. 1965. *Mathematical theory of reliability*. N.Y.: John Wiley & Sons. 497 p.
3. Vasil'ev, N. S. 1985. Ob odnoy modeli razvitiya seti svyazi [On one model of network development] *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 6:227–234.
4. Pavlov, I. V. 1988. Priblizhenno optimal'nye doveritel'nye granitsy dlya pokazateley nadezhnosti sistem s vosstanovleniem [Approximately optimum confidence limits for system reliability indicators with recovery]. *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 3:109–116.
5. Pavlov, I. V., and I. A. Ushakov. 1989. Vychislenie pokazateley nadezhnosti dlya slozhnykh sistem s vosstanavliyaemyimi elementami [Calculation of reliability indices for complex systems with recoverable elements]. *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 6:170–176.
6. Gnedenko, B. V., I. V. Pavlov, and I. A. Ushakov. 1999. *Statistical reliability engineering*. N.Y.: John Wiley & Sons. 517 p.
7. Konovalov, M. G. 2012. Organizatsiya raboty vychislitel'nogo kompleksa s pomoshch'yu imitatsionnoy modeli i adaptivnykh algoritmov [Organization of work of computer complex using a simulation model and adaptive algorithms]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):37–48.
8. Pavlov, I. V. 2012. Raschet i optimizatsiya nekotorykh kharakteristik dlya modeli vychislitel'nogo kompleksa [Calculation and optimization of some characteristics of the model computer complex]. *Informatika i ee Primeneniya — Inform. Appl.* 6(2):59–62.
9. Barlow, R., and F. Proschan. 1966. Tolerance and confidence limits for classes of distributions based on failure rate. *Ann. Math. Stat.* 37(6):1184–1195.
10. Pavlov, I. V. 1977. Doveritel'nye granitsy v klasse raspredeleniy s vozrastayushchey funktsiyey intensivnosti otkazov [Confidence limits in the class of distributions with increasing failure rate function]. *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 6:72–84.
11. Lloyd, D., and M. Lipow. 1962. *Reliability management, methods and mathematics*. N.J.: Prentice-Hall, Englewood Cliffs. 684 p.
12. Beljaev, Ju. K. 1967. Doveritel'nye intervaly dlya funktsiy ot mnogikh neizvestnykh parametrov [Confidence intervals for functions of many unknown parameters]. *Dokl. AN SSSR* 196(4):755–758.
13. Beljaev, Ju. K., T. N. Dugina, and E. V. Chepurin. 1967. Vychislenie nizhney doveritel'noy otsenki dlya veroyatnosti bezotkaznoy raboty slozhnykh sistem [Calculation of the lower confidence estimates for the probability of failure-free operation of complex systems]. *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 2:52–59.
14. Pavlov, I. V. 1981. O korrektnosti fidutsial'nogo podkhoda pri postroenii doveritel'nykh granits dlya pokazateley nadezhnosti slozhnykh sistem [On the correctness of the fiducial approach when constructing confidence limits for the indicators of reliability of complex systems]. *Izvestiya Rossiyskoy Akademii Nauk. Teoriya i Sistemy Upravleniya* [Bulletin of the Russian Academy of Sciences. Theory and Control Systems] 5:46–52.
15. Pavlov, I. V. 1981. O fidutsial'nom podhode pri vychislenii doveritel'nykh granits dlya funktsiy mnogikh neizvestnykh parametrov [On fiducial approach in calculating confidence limits for functions of many unknown parameters]. *Dokl. RAN* 258(6):1314–1317.

Received December 24, 2013

Contributor

Pavlov Igor V. (b. 1945) — Doctor of Science in physics and mathematics, Professor, Bauman Moscow State Technical University, 5, 2nd Baumanskaya Str., Moscow 105005, Russian Federation; ipavlov@bmstu.ru

ИСПОЛЬЗОВАНИЕ ПРИНЦИПА РАВНОВЕСИЯ ДЛЯ УПРАВЛЕНИЯ МАРШРУТИЗАЦИЕЙ В ТРАНСПОРТНЫХ СЕТЯХ

Н. С. Васильев¹

Аннотация: Выбор алгоритмов управления передачей должен основываться на принципах функциональной эффективности (увеличить быстродействие сети) и устойчивости передачи (принцип равновесия). В сетях передачи данных имеется огромное число тяготеющих пар пользователей, каждая из которых заинтересована в скорости доставки своих сообщений. Таким образом, качество функционирования сети необходимо оценивать с помощью векторного критерия. (Существуют также и другие характеристики сетей.) Поэтому проектирование системы управления передачей осуществляется с учетом векторных целевых функций, а принимаемые (реализуемые) решения должны искажаться методами векторной оптимизации. Стремление улучшить качество передачи с целью наилучшего (по возможности) удовлетворения пользователей сети стимулирует поиск новых методов маршрутизации сообщений. В работе предложен метод маршрутизации, основанный на применении игрового принципа равновесия (по Нэшу). Игровая постановка задачи маршрутизации и указанное понятие решения (равновесие) формализуют представление об оптимальности управления передачей в распределенной системе. Использование принципа равновесия предполагает наличие ответа на следующие главные вопросы: всегда ли равновесие достижимо, устойчиво ли оно и как его найти. При общих предположениях в работе доказано существование равновесия по Нэшу. Установлено, что равновесие обладает дополнительными свойствами — вычислительной устойчивостью и эффективностью (оптимальностью) в смысле Парето. Предложен быстрый параллельный (игровой) алгоритм поиска равновесной маршрутизации и обоснована его сходимости.

Ключевые слова: пакетная сеть; потоки в сетях; метрика сети; маршрутизация; векторный критерий; многокритериальная оптимизация; игровая задача; равновесие по Нэшу; эффективность по Парето

DOI: 10.14357/19922264140104

1 Введение

Экспоненциальный рост объемов передаваемой по сетям информации стимулирует исследования, связанные с совершенствованием не только сетевого оборудования, но и алгоритмов управления передачей в пакетных сетях. В таких сетях каждая тяготеющая пара пользователей заинтересована в наискорейшей передаче своих сообщений за счет выбора оптимальных маршрутов доставки. В глобальных сетях невозможно обеспечить централизованное управление маршрутизацией. Поэтому применяются распределенные параллельные алгоритмы (сетевые протоколы).

В каждый момент времени в сети имеется огромное число тяготеющих пар. Значит, задача оптимального управления передачей (задача маршрутизации [1]) является многокритериальной. Требуется найти маршрутизацию, наилучшим образом удовлетворяющую всех пользователей сети. Сетевые задачи с векторными критериями ранее исследовались, например, в работах [2–4].

В статье поставлена и решена *игровая* многокритериальная оптимизационная задача маршрутизации.

Одним из подходов к решению задачи оптимизации векторного критерия является его сворачивание в скалярный критерий. Недостатком этого подхода является то, что распределенная модель системы заменяется на централизованную. В результате строящиеся алгоритмы поиска решения не обладают той степенью параллелизма, которая допускает распределенную реализацию.

Переход к задаче математического программирования (в случае дифференцируемости целевой функции) позволяет использовать градиентный метод поиска оптимального решения, принимаемого в качестве решения исходной задачи. Так вынуждены поступать, когда не удается найти подходящие (быстрые) алгоритмы поиска решения исходной многокритериальной задачи.

Указанный подход применен в работе [1] для решения задачи маршрутизации.

В статье предложен метод сведения игровой задачи оптимизации к *эквивалентной* задаче математического программирования с целью построения игрового (параллельного) алгоритма поиска реше-

¹Московский государственный технический университет им. Н. Э. Баумана, nik8519@yandex.ru

ния (равновесия) исходной задачи. Этот подход основан на введении новой метрики сети, модифицирующей имеющуюся. После этого на итерациях алгоритма поочередно для любой пары абонентов сети строится набор маршрутов передачи, оптимизирующих время доставки каждого сообщения. Эти вычисления основаны на использовании принципа уравнивания Ю. Б. Гермейера, применяемого при решении минимаксных задач.

В отличие от градиентного метода, вычисления маршрутов передачи проводятся поочередно для каждой тяготеющей пары в отдельности. Это позволяет реализовать алгоритм так, чтобы изменение маршрутизации сети проводить одновременно (независимо) для многих пар абонентов.

Управление потоками в пакетных телекоммуникационных системах (транспортных сетях — ТС) основано на моделях сетей с переменной метрикой [1–10]. Изменяемая метрика присуща даже однопродуктовой сети (имеется единственная пара абонентов) из-за наличия обратной связи между потоками и задержкой в передаче пакетов. Напомним, что задержки в линиях связи определяют метрику, с помощью которой оценивается быстродействие сети.

Игнорирование указанной обратной связи при построении параллельных алгоритмов маршрутизации не позволяет обеспечить устойчивое управление потоками ТС. Это наблюдается даже в сетях, имеющих кольцевую архитектуру. Так, поочередный выбор кратчайшего маршрута в текущей метрике сети для передачи сообщений между всеми тяготеющими парами может приводить к возникновению колебательного процесса [1]. В результате в сети возникают потоки, вызывающие ее перегрузку, хотя имеется маршрутизация, при которой сеть справляется с заданными входными потоками. Для поиска соответствующей допустимой маршрутизации сети достаточно применить адекватный (а не эвристический, как в [1]) игровой алгоритм маршрутизации.

Свойства ТС с переменной метрикой [9] ранее изучались в связи с поиском равновесной маршрутизации глобальной пакетной сети передачи данных [2, 4, 8]. При этом существование равновесной маршрутизации удавалось теоретически обосновать лишь для сетей с топологией, мало отличающейся от кольцевой. При численном моделировании удавалось строить равновесную маршрутизацию для весьма широкого класса сетей. Так, в результате проведения вычислительных экспериментов равновесие по Нэшу достигалось в модели сети Интернет [8]. Таким образом, численный поиск равновесного решения задачи уже прошел экспериментальную апробацию, но не по-

лучил должного теоретического обоснования.

Данная статья посвящена доказательству общей теоремы существования равновесия и установлению его свойств — вычислительной устойчивости и эффективности по Парето.

Доказана сходимости алгоритма поиска равновесной маршрутизации. Простота и параллельные свойства алгоритма позволяют надеяться на его применение при создании новых сетевых протоколов транспортного уровня [1], основанных на использовании принципа равновесия.

2 Оптимизационная модель транспортной системы

Топологию ТС будем представлять в виде связанного неориентированного графа $\Gamma = (U, V)$, вдоль ребер $l \in V, l = 1, 2, \dots, n$, которого расположены линии (каналы) передачи пакетов, а в узлах $u \in U$ размещены источники и стоки передаваемых потоков. (Требование неориентированности графа несущественно.)

Доставка сообщений для каждой *тяготеющей* k -й пары (источник–сток), $k = 1, 2, \dots, K$, осуществляется по одному или нескольким выбираемым маршрутам графа сети $\{L_j^k\}$, соединяющим эти узлы. Входные (случайные) потоки интенсивности $\lambda_k = \lambda_0^k$ поступают в узлы-источники, разделяются в них (алгоритмом маршрутизации) на маршрутные потоки величины $\{\lambda_j^k\}$ и по маршрутам $\{L_j^k\}$ передаются в соответствующие узлы-стоки, из которых покидают ТС.

Функционирование сети происходит с задержками на линиях сети $l = 1, 2, \dots, n$, равными значениям некоторой функции $f_l(z_l)$, зависящей от величин интенсивностей потоков z_l на этих линиях.

Например, в пакетных сетях передачи данных величина задержки всякого пакета на любой линии складывается из следующих величин [1, 7]:

- времени ожидания пакета в очереди;
- времени определения направления дальнейшей передачи (в транзитный узел сети) с помощью маршрутной таблицы;
- времени пересылки пакета по выбранной линии.

Определение функций задержек составляет самостоятельную задачу [1, 7, 10]. Во всяком случае, эти функции неотрицательны, монотонно и неограниченно возрастают при увеличении интенсивности потока по линии до величины ее пропускной способности. (Согласно теории массового обслуживания при совпадении интенсивностей

поступления и обслуживания заявок наблюдается неограниченный рост очередей.)

Время доставки продуктов вдоль маршрута L (или «длина» маршрута L) равно сумме задержек:

$$\rho_f(L, z) = \sum_{l \in L} f_l(z_l). \quad (1)$$

Зафиксировав векторную функцию $f = (f_1, f_2, \dots, f_n)$, опустим индекс f в обозначении метрики сети (1). Потоки в сети задаются в единицах измерения интенсивности передачи. По смыслу использованных обозначений для всех допустимых значений индексов k и l должны выполняться балансовые соотношения

$$\sum_{j=1}^{J_k} \lambda_j^k = \lambda_0^k, \quad \sum_{j, k \in L_j^k} \lambda_j^k = z_l. \quad (2)$$

Через z_l в (2) обозначен суммарный поток по l -й линии ТС, который ограничен величиной \bar{z}_l — пропускной способностью линии:

$$z_l \leq \bar{z}_l, \quad l = 1, 2, \dots, n. \quad (3)$$

Под *допустимой маршрутизацией* (ДМ) сети (для k -й тяготеющей пары) будем понимать совокупность маршрутов $M^k = \{L_j^k\}$ и маршрутных потоков $\{\lambda_j^k\}$ (интенсивностей передачи) для всех тяготеющих пар $k = 1, 2, \dots, K$ такую, что выполняются соотношения (2), (3) и $\rho(L_j^k, z) < \infty$.

Векторный входной поток $(\lambda_0^1, \lambda_0^2, \dots, \lambda_0^K)$ называется *допустимым*, если для него найдется ДМ.

Задавая маршрутизацию, будем перечислять только *применяемые* маршруты передачи, для которых маршрутные потоки положительны. В процессе управления ТС за счет (принудительного) ограничения входного потока обеспечивается его допустимость. Выбор ДМ

$$M = \{ \{L_j^k\}, \{ \lambda_j^k \}, k = 1, 2, \dots, K \}$$

однозначно задает вектор допустимых потоков по линиям ТС $z = (z_1, z_2, \dots, z_n)$, от которого зависит время передачи сообщений (вместе с вектором z изменяется метрика сети (1)).

Любая k -я тяготеющая пара «заинтересована» в уменьшении времени доставки пакетов, определяемого длинами применяемых маршрутов передачи (1). В сети с потоками z минимально возможное время доставки пакетов k -й пары равно

$$\underline{T}^k(z) = \min_{\{L^k\}} \rho(L^k, z),$$

где $\{L^k\}$ — множество всех маршрутов, соединяющих эту пару.

Маршрутизацию k -й тяготеющей пары назовем *оптимальной*, если все применяемые маршруты передачи имеют минимальную длину, равную $\underline{T}^k(z)$.

Так как вектор z зависит от выбираемой маршрутизации (см. (2)), то оптимальным решением этой задачи для k -й пары является выбор ее кратчайших маршрутов соединения, длина которых оценивается с помощью изменяющейся (вместе с сетевыми потоками) метрикой (1) [5, 6, 9].

Далее исследуется *игровая* задача об оптимальной маршрутизации ТС. Тяготеющие пары $k = 1, 2, \dots, K$ рассматриваются как игроки в бескоалиционной игре, выбирающие свою стратегию — маршрутизацию [2, 4, 8].

Равновесие по Нэшу в этой игре назовем *равновесной* (оптимальной) маршрутизацией ТС. Соответствующий вектор потоков в сети также будем называть *равновесным*.

В равновесии каждой тяготеющей паре *невыгодно* отклоняться от своей маршрутизации из-за того, что время передачи потоков только увеличится при условии, что все остальные пары придерживаются своих маршрутов и интенсивностей передачи. В этом заключается устойчивость (в игровом смысле) равновесного решения.

На практике выбор маршрутизации может проводиться не самими тяготеющими парами, а с помощью некоторого алгоритма, построенного на этом принципе. (Таковы сетевые протоколы [1].)

3 Существование равновесия

Введем вспомогательную задачу математического программирования с целью замены более трудной игровой задачи маршрутизации на однокритериальную оптимизационную задачу.

Известные стандартные методы сведения обычно приводят к сложным многоэкстремальным задачам, в которых целевые функции не являются дифференцируемыми, даже если исходная векторная целевая функция была дифференцируемой.

Особенности строения сетевых критериев качества позволяют предложить следующий способ сведения исходной игровой задачи к выпуклой оптимизационной задаче, решить которую проще, чем исходную проблему.

Определим новые функции задержек

$$t_l = t_l(z_l), \quad l = 1, 2, \dots, n,$$

решив набор не связанных между собой одномерных задач Коши для обыкновенного дифференциального уравнения (ОДУ) первого порядка:

$$z_l \frac{dt_l}{dz_l} + t_l = f_l(z_l), \quad t_l(0) = 0. \quad (4)$$

Пусть Z — многогранник допустимых потоков по линиям передачи. Ввиду равенств (2) и (3) Z — выпуклый ограниченный многогранник. Рассмотрим следующую экстремальную задачу:

$$F(z) = \sum_{l=1}^n z_l t_l(z_l) \rightarrow \min_{z \in Z}. \quad (5)$$

Теорема 1. Пусть все функции задержек монотонно возрастают, дифференцируемы и

$$(\forall l) f_l(z) \rightarrow \infty, \quad z \rightarrow \bar{z}_l.$$

Тогда если $Z \neq \emptyset$, то минимум целевой функции (5) достигается в единственной точке z^* , которой отвечает некоторая равновесная маршрутизация.

Доказательство. С помощью критерия Сильвестра проверим положительную определенность матрицы Якоби целевой функции $F(z)$. Условия теоремы и определение этой функции (см. (4) и (5)) дают:

$$\frac{\partial^2 F}{\partial z_l^2} = (z_l t_l'(z_l) + t_l(z_l))' = f_l'(z_l) > 0;$$

$$\frac{\partial^2 F}{\partial z_{l_1} \partial z_{l_2}} = 0, \quad l_1 \neq l_2.$$

(штрихом обозначено дифференцирование). Все условия этого критерия выполнены, поэтому целевая функция $F(z)$ строго выпукла. В условиях теоремы это позволяет сделать вывод о том, что решение выпуклой экстремальной задачи (5) существует и единственно.

Докажем, что z^* — точке минимума (5) — отвечает равновесная маршрутизация. Для этого применим теорему Куна–Таккера [11], которая в данном случае служит критерием оптимальности потока z^* . В записи условий оптимальности учтем, что ограничения (3) неактивны. (Согласно наложенным предположениям элемент z^* удовлетворяет строгим неравенствам в (3).)

Выразим произвольный допустимый вектор z в виде $z = \mathbf{A}\lambda$, где вектор $\lambda = (\lambda_j^k)$ входит в соотношения (2). Через \mathbf{A} обозначена $(n \times J)$ -матрица всех маршрутов, соединяющих рассматриваемую тяготеющую пару. Напомним, что \mathbf{A} — это $(0, 1)$ -матрица инцидентий ребра–маршруты. Маршруты передачи представлены столбцами матрицы \mathbf{A} . Указанное представление вектора потоков $z = \mathbf{A}\lambda$ всегда возможно: для неприменяемых маршрутов L_j^k полагаем $\lambda_j^k = 0$.

Условия оптимальности из теоремы Куна–Таккера в задаче (2)–(5) представляют собой систему соотношений, распадающуюся на подсистемы,

описывающие оптимальные решения для отдельных тяготеющих пар. Поэтому рассмотрим произвольную пару k , зафиксировав маршрутизацию остальных пар. Для упрощения записи в обозначениях опустим верхний индекс k . Итак, выпишем функцию Лагранжа [11, 12]:

$$H(\mu, \lambda) = F(\mathbf{A}\lambda) + \left\langle \mu, \lambda_0 - \sum_{j=1}^J \lambda_j \right\rangle, \quad \lambda \geq 0.$$

Тогда критерий оптимальности маршрутизации, определяемой вектором

$$\lambda = \lambda^*; \quad z^* = \mathbf{A}\lambda^*,$$

для k -й тяготеющей пары принимает следующий вид:

$$\left. \begin{aligned} \nabla_{\lambda} H = fA - \mu(1, \dots, 1) &\geq 0; \\ (\forall j) \lambda_j ((fA)_j - \mu) &= 0, \end{aligned} \right\} \quad (6)$$

причем $f = (f_1, f_2, \dots, f_n)$ — вектор задержек в (6) — вычислен в точке z^* .

Если $\lambda_j \neq 0$, то из (6) следует равенство $(fA)_j = \mu$. Если $\lambda_j = 0$, то справедливо неравенство $(fA)_j \geq \mu$. По определению матрицы маршрутов и в соответствии с определением метрики (1) имеем

$$(fA)_j = \rho(L_j, z^*).$$

Полученные соотношения означают то, что все маршруты, применяемые произвольной k -й парой, являются кратчайшими, причем их длина равна μ . Следовательно, решение задачи маршрутизации, отвечающее вектору потоков z^* , равносильно. В соответствии с определением равновесия доказана оптимальность рассматриваемой маршрутизации для всех тяготеющих пар.

Пример 1. При аппроксимации функций задержек f_i степенными функциями задержки t_i , найденные согласно уравнениям (4), с точностью до постоянного множителя совпадают с f_i . Таким образом, новая метрика сети ρ_t несущественно отличается от ρ_f .

Пример 2. В случае дробно-линейной аппроксимации

$$f(z) = \frac{az}{\bar{z} - z}, \quad 0 \leq z < \bar{z},$$

функции задержек вспомогательная метрика сети ρ_t заметно отличается от исходной, так как функция $t = t(z)$, удовлетворяющая уравнению (4), имеет вид:

$$t(z) = -a \left(1 + \frac{\bar{z}}{z} \ln(\bar{z} - z) \right), \quad 0 \leq z < \bar{z}.$$

4 Эффективность по Парето равновесной маршрутизации

На множестве допустимых маршрутизаций введем отношение эквивалентности. Маршрутизации M_1 и M_2 (всей сети или какой-нибудь отдельной тяготеющей пары) *эквивалентны*, если они приводят к одному и тому же вектору потоков z на линиях сети. Класс эквивалентности маршрутизации M будем обозначать $R_z(M)$.

Для упрощения записи будем опускать индекс z , а множество применяемых k -й парой маршрутов $\{L_j^k\}$ при маршрутизации M обозначать M^k .

Из проведенного доказательства теоремы 1 вытекает

Следствие 1. Равновесной является всякая маршрутизация, которая эквивалентна равновесной маршрутизации.

Будем считать, что в сети с потоками z может реализоваться произвольная эквивалентная маршрутизация. (Это предположение имеет место в пакетных сетях, в которых управление передачей осуществляется с помощью маршрутных таблиц [1].)

Тогда времена доставки отдельных пакетов для k -й тяготеющей пары абонентов вычисляются по следующей формуле:

$$T^k(z; M^k) = \max_{M' \in R_z(M^k)} \max_{L_j^k \in M'} \rho(L_j^k, z), \quad k = 1, 2, \dots, K. \quad (7)$$

Подстановка в (7) равновесной маршрутизации дает значения критериев (7), совпадающие с длинами применяемых (кратчайших) маршрутов (следствие 1).

Теорема 2. Пусть все пары узлов графа сети являются тяготеющими и для любых попарно смежных ребер графа сети k, l, m выполнены неравенства треугольника

$$f_k(0) \leq f_l(0) + f_m(0).$$

Тогда равновесная маршрутизация эффективна по Парето.

Доказательство. Рассуждая от противного, найдем такую допустимую маршрутизацию

$$M = \{ \{L_j^k\}, \{\lambda_j^k\}, k = 1, 2, \dots, K \},$$

для которой

$$(\forall k) T^k(z) \leq T^k(z^*). \quad (8)$$

В соотношениях (8) хотя бы одно неравенство является строгим [13], поэтому равновесные потоки z^* таковы, что $z \neq z^*$.

Из условия теоремы и неравенства треугольника следует, что в ситуации равновесия

$$z_l^* > 0, \quad l = 1, 2, \dots, n.$$

Докажем справедливость неравенства

$$z_l \leq z_l^*, \quad l = 1, 2, \dots, n. \quad (9)$$

Если в маршрутизации M линия связи l не применяется для передачи пакетов, то, очевидно, соответствующее неравенство в (9) выполнено. Пусть теперь $z_l > 0$, $l = (a, b)$. Можно считать, что у тяготеющей пары (a, b) с номером k имеется маршрут соединения

$$L \equiv \{(a, b)\} \in M^k.$$

В самом деле, если это не так, то, покажем, существует маршрутизация $M'^k \in R(M^k)$, обладающая этим свойством. (Тогда вместо маршрутизации M достаточно будет рассмотреть M' .)

Так как $z_l > 0$, то найдется такая пара k' , что $l = (a, b) \in L_{j'}^{k'} \in M_{j'}^{k'}$. Выберем величину

$$0 < \Delta < \min \{ \lambda_j^k, \lambda_j^{k'} \},$$

в которой поток λ_j^k проходит по такому маршруту L_j^k , что $l \notin L_j^k \in M^k$. Тогда часть маршрутного потока λ_j^k величиной Δ перебросим с маршрута L_j^k на L — новый для k -й пары маршрут соединения. Такую же величину Δ , являющуюся частью потока $\lambda_{j'}^{k'}$, направим по маршруту

$$L' = (L_{j'}^{k'} \setminus \{l\}) \cup L_j^k,$$

новому для пары k' . Все остальные «элементы» маршрутизации M оставим без изменения. В результате получена искомая маршрутизация M' , для которой $M' \in R(M)$.

Согласно сделанному допущению (8) относительно маршрутизаций M и M^* и определению критериев (7) имеем:

$$\begin{aligned} (\forall l = 1, 2, \dots, n) f_l(z_l) &= \\ &= \rho(L, z) \leq T^k(z) \leq T^k(z^*) = f_l(z_l^*). \end{aligned}$$

Так как функции задержек монотонно возрастают, отсюда получаем неравенства (9), $z \neq z^*$. Целевая функция задачи (5) монотонно возрастает по каждой переменной z_l . Тогда из (9) следует неравенство $F(z) < F(z^*)$, противоречащее тому, что z^* — минимум целевой функции (5). Теорема 2 доказана.

5 Вычислительная устойчивость равновесного решения задачи маршрутизации

Исходные данные модели обычно обладают некоторой неопределенностью. Возникает необходимость исследовать устойчивость искомого решения. В рассматриваемой модели будем варьировать все параметры с помощью изменения $c = (\lambda_0, \bar{z})$ — вектора, составленного из величин входных потоков и пропускных способностей линий ТС.

Задержки и целевая функция (5) есть функции переменных z и c , а многогранник потоков — значение многозначного отображения $c \rightarrow Z(c)$. Предполагается, что параметр c изменяется в пределах множества C такого, что при справедливости включения $c \in C$ выполняются все условия теоремы 1 и, кроме того, все функции задержек непрерывны по совокупности переменных. Тогда справедлива

Теорема 3. *Метрика сети непрерывно зависит от параметров модели.*

Доказательство. Из теоремы о непрерывной зависимости ОДУ (4) от параметра следует, что целевая функция $F(z, c)$, $z \in Z(c)$, $c \in C$, экстремальной задачи (5) непрерывна. Анализ линейных соотношений (2) и (3) показывает, что отображение $c \rightarrow Z(c)$ непрерывно по Хаусдорфу [11]. По теореме 1 отсюда можно заключить, что оптимальное решение задачи (5) $z^* = z^*(c)$ непрерывно зависит от параметра c . Таким образом, метрика сети также непрерывна как сумма непрерывных функций (см. (1)).

Непосредственным следствием теорем 1 и 2 является вывод о том, что решение задачи маршрутизации устойчиво к изменению параметров модели.

6 Алгоритм поиска равновесной маршрутизации

Для поиска оптимальных маршрутов передачи продуктов каждой тяготеющей пары будем применять следующую схему алгоритма.

Произвольно выберем начальную маршрутизацию (шаг $t = 0$).

Пусть на шаге $t = 0, 1, \dots$ уже построена маршрутизация, обозначаемая M^t , $M^t = (\{L_j^k\}, \{\lambda_j^k\})$.

Ей отвечает вектор потоков z^t . В сети с фиксированной метрикой $\rho(L, z^t)$ найдем кратчайший маршрут L . (Достаточно воспользоваться алгоритмом Дейкстры [14].) Пусть также в текущей маршрутизации

существует маршрут L^t , имеющий большую длину по сравнению с L . (Иначе решение задачи уже найдено.) Тогда часть потока λ^t , передаваемого по маршруту L^t , перебросим на маршрут L с целью уменьшения разности длин этих маршрутов. (Это возможно ввиду монотонности функций задержек, см. (1).) При этом либо уравниваются длины маршрутов, либо маршрут L останется по-прежнему короче, а маршрут L^t перестанет использоваться (в случае $\lambda|_L = \lambda^t$).

Определим очередную маршрутизацию ТС M^{t+1} как результат добавления к маршрутизации M^t пары $L, \lambda|_L$ и, возможно, исключения маршрута L^t в случае, когда он перестает применяться.

Указанные в алгоритме действия — применение принципа уравнивания Ю. Б. Гермейера как метода решения минимаксных задач [15].

Теорема 4. *Последовательность маршрутизаций $\{M^t, t = 0, 1, \dots\}$, построенная по принципу уравнивания, сходится к равновесному решению задачи.*

Доказательство. Вычислим производную функции $F(A\lambda)$ в текущей точке $z^t = A\lambda^t$ по направлению вектора Δ , все координаты которого равны нулю, за исключением тех, которые отвечают маршрутным потокам вдоль L и L^t , равных соответственно 1 и -1 :

$$\frac{dF(A\lambda)}{d\Delta} = \langle fA, \Delta \rangle = \rho(L, z^t) - \rho(L^t, z^t) < 0.$$

Шаг градиентного метода в направлении Δ с целью выравнивания длин этих маршрутов приводит к уменьшению значения целевой функции в задаче (5):

$$F(z^{t+1}) < F(z^t), \quad t = 0, 1, \dots$$

Монотонная числовая последовательность $\{F(z^t)\}$ сходится к $F(z^*)$, где z^* — решение задачи (5). Согласно теореме 1 $z^t \rightarrow z^*$, $t \rightarrow \infty$. По следствию 1 в пределе получаем равновесную маршрутизацию.

Литература

1. Бертсекас Д., Галлагер Р. Сети передачи данных / Пер с англ. — М.: Наука, 1989. (Bertsekas D. P., Gallager R. Data networks. — Englewood Cliffs: Prentice Hall, 1987. 544 p.)
2. Васильев Н. С., Федоров В. В. О равновесной маршрутизации в сетях передачи данных // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн., 1996. № 4. С. 39–47.
3. Korilis Y. A., Lazar A. A., Orda A. Capacity allocation under noncooperative routing // IEEE Trans. Automat. Contr., 1997. Vol. 42. No. 3. P. 309–325.

4. Vasilyev N. S. Nash equilibrium routing in ring networks // *Int. J. Math. Game Theory Algebra*, 1998. Vol. 7. No. 4. P. 221–234.
5. Васильев Н. С. О свойствах решений задачи маршрутизации сети с виртуальными каналами // *Ж. вычисл. матем. и матем. физ.*, 1997. Т. 37. № 7. С. 785–793.
6. Васильев Н. С. О свойствах решений задачи динамической маршрутизации сети // *Ж. вычисл. матем. и матем. физ.*, 1998. Т. 38. № 1. С. 42–52.
7. Соколов И. А., Шоргин С. Я. Модель и математические методы расчета характеристик сети, использующей технологии X.25 и Frame relay // *Системы и средства информатики. Спец. вып. Математические методы информатики*. — М.: Наука, 2001. С. 43–66.
8. Васильев Н. С., Федоров В. В. О построении алгоритмов маршрутизации пакетных сетей на основе векторных критериев // *Известия РАН. Теория и системы управления*, 2005. № 3. С. 36–47.
9. Васильев Н. С. Задача о кратчайших маршрутах в сетях с переменной метрикой // *Вестник МГТУ им. Н. Э. Баумана. Сер. Естеств. науки*, 2008. № 1. С. 70–75.
10. Коновалов М. Г. Оптимизация работы вычислительного комплекса с помощью имитационной модели и адаптивных алгоритмов // *Информатика и её применения*, 2012. Т. 6. Вып. 1. С. 37–48.
11. Васильев Ф. П. Численные методы решения экстремальных задач. — М.: Наука, 1980.
12. Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач. — М.: Наука, 1974. 481 с.
13. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982. 256 с.
14. Кристофидес Н. Теория графов. Алгоритмический подход / Пер с англ. — М.: Мир, 1978. (*Cristofides N. Graph theory: An algorithmic approach*. — London: Academic, 1975. 430 p.)
15. Федоров В. В. Численные методы максимина. — М.: Наука, 1979. 280 с.

Поступила в редакцию 25.05.13

EQUILIBRIUM PRINCIPLE APPLICATION TO ROUTING CONTROL IN PACKET DATA TRANSMISSION NETWORKS

N. S. Vasilyev

Bauman Moscow State Technical University, 5, 2nd Baumanskaya Str., Moscow 105005, Russian Federation

Abstract: Annual exponential growth of data flows in large scale networks impels to search not only network hardware improvements but also more perfect routing control algorithms. In networks, it is impossible to use centralized algorithms of routing control. Parallel algorithms choice must be based on the principles of functional effectiveness and stability (equilibrium). In large-scale networks, there is a huge number of users' pairs trying to achieve the maximally possible rate of data transmission by routing control. Thus, control must be based on multicriteria optimization ideas and methods. The Nash equilibrium (game formulation of the routing problem) formally presents optimality of transmission control in distributed systems. In the present paper, the equilibrium routing is proved to exist under general conditions. The solution is additionally shown to be effective in Pareto sense and computationally stable. An effective (quick and parallel) game theory algorithm is suggested and its convergence is proved.

Keywords: packet network; data flows; network metric; routing; vector criteria; multicriteria optimization; game problem; Nash equilibrium; Pareto effectiveness

DOI: 10.14357/19922264140104

References

1. Bertsecas, D. P., and R. Gallager. 1987. *Data networks*. Englewood Cliffs: Prentice Hall. 544 p.
2. Vasilyev, N. S., and V. V. Fedorov. 1996. O ravnovesnoy marshrutizatsii v setyakh peredachi dannykh [Equilibrium routing in data-transmission networks]. *Vestn. Mosk. Univ. Ser. 15: Vychisl. Mat. Kibern.* [*Bulletin of Moscow State University. Ser. 15: Comput. Math., Cybern.*] 4:47–52.
3. Korilis, Y. A., A. A. Lazar, and A. Orda. 1997. Capacity allocation under noncooperative routing. *IEEE Trans. Automat. Contr.* 42(3):309–325.
4. Vasilyev, N. S. 1998. Nash equilibrium routing in ring networks. *Int. J. Math. Game Theory Algebra* 7(4):221–234.
5. Vasilyev, N. S. 1997. O svoystvakh resheniy zadachi marshrutizatsii seti s virtual'nymi kanalami [Properties of the solutions to the problem of routing in network with virtual channels]. *Zh. Vychisl. Mat. Mat. Fiz.* [*Computational Mathematics and Mathematical Physics*] 37(7):785–793.
6. Vasilyev, N. S. 1998. O svoystvakh resheniy zadachi di-

- namicheskoy marshrutizatsii seti [Properties of the solutions to the problem of dynamic routing in networks]. *Zh. Vychisl. Mat. Mat. Fiz.* [Computational Mathematics and Mathematical Physics] 38(1):42–52.
7. Sokolov, I. A., and S. Ya. Shorgin. 2001. Model' i matematicheskie metody rascheta kharakteristik seti, ispol'zuyushchey tekhnologii X.25 i Frame relay [Models and mathematical methods of characteristics calculation for X.25 and Frame relay network]. *Sistemy i Sredstva Informatiki. Spec. vyp. "Matematicheskie metody informatiki"* [Systems and Means of Informatics. Spec. ed. "Math. Meth. of Informatics"]. Moscow: Nauka, Fizmatlit. 43–66.
 8. Vasilyev, N. S., and V. V. Fedorov. 2005. O postroenii algoritmov marshrutizatsii paketnykh setey na osnove vektornykh kriteriev [On routing algorithms in packet networks on the base of vector criterias]. *Izvestiya RAN. Teorija i sistemy upravleniya* [Bulletin of RAS. Theory and Control Systems] 3:36–47.
 9. Vasilyev, N. S. 2008. Zadacha o kratchayshikh marshrutakh v setyakh s peremennoy metrikoy [The shortest paths problem in networks with changeable metric]. *Vestnik MGTU im. N. E. Baumana, Ser. Estestv. Nauki* [Bulletin of Bauman MSTU. Ser. Natural Sciences] 1:70–75.
 10. Konovalov, M. G. 2012. Optimizatsiya raboty vychislitel'nogo kompleksa s pomoshch'yu imitatsionnoy modeli i adaptivnykh algoritmov [Optimization of computational complex work on the base of imitational models and adaptive algorithms]. *Informatika i Ee Primeneniya — Inform. Appl.* 6(1):37–48.
 11. Vasilyev, F. P. 1980. Chislennyye metody resheniya ekstremal'nykh zadach [Numerical methods for extremum problems]. Moscow: Nauka. 520 p.
 12. Ioffe, A. D., and V. M. Tihomirov. 1974. *Teoriya ekstremal'nykh zadach* [Theory of extremum problems]. Moscow: Nauka. 481 p.
 13. Podinovskij, V. V., and V. D. Nogin. 1982. *Pareto-optimal'nye resheniya mnogokriterial'nykh zadach* [Pareto optimal solutions in multicriteria problems]. Moscow: Nauka. 256 p.
 14. Cristofides, N. 1975. *Graph theory: An algorithmic approach*. London: Academic. 430 p.
 15. Fedorov, V. V. 1979. *Chislennyye metody maksimina* [Numerical methods of maximin]. Moscow: Nauka. 280 p.

Received May 25, 2013

Contributor

Vasilyev Nikolai S. (b. 1952) — Doctor of Science in physics and mathematics, professor, Bauman Moscow State Technical University, 5, 2nd Baumanskaya Str., Moscow 105005, Russian Federation; nik8519@yandex.ru

АСИМПТОТИЧЕСКИЕ СВОЙСТВА ОЦЕНКИ РИСКА ПРИ ПОРОГОВОЙ ОБРАБОТКЕ ВЕЙВЛЕТ-КОЭФФИЦИЕНТОВ В МОДЕЛИ С КОРРЕЛИРОВАННЫМ ШУМОМ*

А. А. Ерошенко¹, О. В. Шестаков²

Аннотация: Вейвлет-методы подавления шума, основанные на процедуре пороговой обработки, широко используются при анализе сигналов и изображений. Их привлекательность заключается, во-первых, в скорости алгоритмов построения оценок, а во-вторых, в возможности лучшей, чем линейные методы, адаптации к функциям, имеющим на разных участках различную степень регулярности. Анализ погрешностей этих методов представляет собой важную практическую задачу, поскольку он позволяет оценить качество как самих методов, так и используемого оборудования. В работе исследуются асимптотические свойства оценки среднеквадратичного риска при пороговой обработке коэффициентов разложения функции сигнала по вейвлет-базису в модели с коррелированным шумом. Приводятся условия, при которых имеют место состоятельность и асимптотическая нормальность несмещенной оценки риска. Полученные результаты дают возможность строить асимптотические доверительные интервалы для погрешности пороговой обработки, используя только наблюдаемые данные.

Ключевые слова: вейвлеты; пороговая обработка; несмещенная оценка риска; коррелированный шум; асимптотическая нормальность

DOI: 10.14357/19922264140105

1 Введение

Статистические методы вейвлет-анализа широко применяются при анализе и обработке зашумленных сигналов и изображений. Построение вейвлет-оценок для данных с коррелированным шумом обычно осуществляется с помощью применения мягкой пороговой обработки к эмпирическим вейвлет-коэффициентам. Порог обычно зависит от уровня разложения, и его можно выбирать различными способами, исходя из постановки задачи и целей обработки (см., например, [1–4]). Наличие шума неизбежно приводит к погрешностям в оцениваемом сигнале/изображении. Свойства оценки таких погрешностей (риска) в модели с независимым шумом исследовались в работах [1–12]. Показано, что при определенных условиях оценка риска является состоятельной и асимптотически нормальной. В данной работе исследуется асимптотическое поведение оценки риска в модели со стационарным коррелированным шумом. Для простоты изложения рассматриваются одномерные сигналы.

2 Модель данных и коэффициенты вейвлет-разложения

Вейвлет-разложение функции $f \in L^2(\mathbb{R})$, описывающей сигнал, представляет собой ряд

$$f = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{jk} \rangle \psi_{jk}, \quad (1)$$

где $\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k)$, а $\psi(t)$ — некоторая материнская вейвлет-функция (семейство $\{\psi_{jk}\}_{j,k \in \mathbb{Z}}$ образует ортонормированный базис в $L^2(\mathbb{R})$). Индекс j в (1) называется масштабом, а индекс k — сдвигом. Функция ψ должна удовлетворять определенным требованиям, однако ее можно выбрать таким образом, чтобы она обладала некоторыми полезными свойствами, например была дифференцируемой нужное число раз и имела заданное число M нулевых моментов [13], т. е.

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0, \quad k = 0, \dots, M-1.$$

* Работа выполнена при финансовой поддержке РФФИ (проект 11-01-00515а).

¹Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета вычислительной математики и кибернетики, aeroshik@gmail.com

²Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета вычислительной математики и кибернетики; Институт проблем информатики Российской академии наук, oshestakov@cs.msu.su

В дальнейшем будут рассматриваться функции сигнала $f \in L^2(\mathbb{R})$ на конечном отрезке $[a, b]$, равномерно регулярные по Липшицу с некоторым параметром $\gamma > 0$. Для таких функций известно [14], что если вейвлет-функция M раз непрерывно дифференцируема ($M \geq \gamma$), имеет M нулевых моментов и достаточно быстро убывает на бесконечности, т. е. существует такая константа $C_A > 0$, что

$$\int_{-\infty}^{\infty} (1 + |t|^\gamma) |\psi(t)| dt \leq C_A,$$

то найдется такая константа $A > 0$, что

$$\langle f, \psi_{jk} \rangle \leq \frac{A}{2^{j(\gamma+1/2)}}. \quad (2)$$

На практике функции сигнала всегда заданы в дискретных отсчетах на конечном отрезке. Не ограничивая общности, будем считать, что это отрезок $[0, 1]$ и функция f задана в точках $i/2^J$ ($i = 1, \dots, 2^J$): $f_i = f(i/2^J)$. Дискретное вейвлет-преобразование представляет собой умножение вектора значений функции f (обозначим его через \bar{f}) на ортогональную матрицу W , определяемую вейвлет-функцией ψ : $\bar{f}^W = W\bar{f}$ [14]. При этом дискретные вейвлет-коэффициенты связаны с непрерывными следующим образом: $f_{jk}^W \approx 2^{J/2} \langle f, \psi_{jk} \rangle$ (см., например, [1] или [14]). Это приближение тем точнее, чем больше J . Здесь не будут обсуждаться методы борьбы с краевыми эффектами, связанными с использованием вейвлет-разложения на конечном отрезке. Познакомиться с этими методами можно, например, в [15].

В реальных наблюдениях всегда присутствует шум. Пусть $\{e_i, i \in \mathbb{Z}\}$ — стационарный гауссовский процесс с ковариационной последовательностью $r_k = \text{cov}(e_i, e_{i+k})$. Будем полагать, что e_i имеют нулевое среднее и единичную дисперсию. Рассмотрим следующую модель данных:

$$Y_j = f_j + e_j \quad j = 1, \dots, 2^J.$$

Для $t \in [0, 1]$ определим наблюдаемый процесс

$$Y_J(t) = \frac{1}{2^J} \sum_{i=1}^{[2^J t]} Y_i = F_J(t) + \frac{1}{2^J} \sum_{i=1}^{[2^J t]} e_i,$$

где $F_J(t) = 1/2^J \sum_{i=1}^{[2^J t]} f(i/2^J)$ — «суммарный сигнал». Как и в работах [16, 17], рассмотрим два отдельных случая.

2.1 Модель краткосрочной зависимости

Пусть $\sum_{-\infty}^{+\infty} |r_k| < \infty$. Положим $\tau^2 = \sum_{-\infty}^{+\infty} r_k$. В этом случае

$$2^{J/2}(Y_J(t) - F_J(t)) \Rightarrow \tau \mathbf{B}(t), \quad t \in [0, 1], \quad (3)$$

где $\mathbf{B}(t)$ — процесс стандартного броуновского движения. Это следует, например, из [18] (лемма 5.1). Проведем масштабирование $\epsilon = \tau \cdot 2^{-J/2}$. Без ограничения общности далее будем полагать, что $\tau = 1$. Из (3) следует, что можно аппроксимировать наблюдаемый процесс $Y_J(t)$ с помощью $Y(t)$ для $t \in [0, 1]$, где

$$Y(t) = F(t) + \epsilon \mathbf{B}(t) \quad (4)$$

и $F(t) = \int_0^t f(s) ds$. Применяя к (4) вейвлет-разложение и аппроксимируя его дискретным вейвлет-преобразованием, приходим к следующей модели дискретных вейвлет-коэффициентов [16]:

$$X_{jk} = \mu_{jk} + z_{jk}, \quad (5)$$

где $\mu_{jk} = f_{jk}^W$, а z_{jk} независимы и имеют стандартное нормальное распределение.

Замечание 1. В силу краевых эффектов нормальные случайные величины z_{jk} с нулевым средним, относящиеся к фиксированному уровню j , на самом деле имеют ковариационную матрицу, удовлетворяющую условию

$$\sigma_a^2 I \leq \Gamma_j \leq \sigma_b^2 I, \quad (6)$$

где неравенства рассматриваются в смысле неотрицательно определенных матриц и константы σ_a, σ_b не зависят от j , а только от выбранного вейвлет-базиса. Таким образом, модель (5) справедлива с точностью до ограничений (6) [16].

2.2 Модель долгосрочной зависимости

Теперь предположим, что автоковариационная функция шума убывает медленно согласно модели $r_k \sim Ak^{-\alpha}$, где $0 < \alpha < 1$.

Положим $\tau^2 = 2A/((1-\alpha)(2-\alpha))$ (без ограничения общности далее будем полагать, что $\tau = 1$) и $H = 1 - \alpha/2 \in (1/2, 1)$.

Определим дробное броуновское движение $\mathbf{B}_H(t)$ как гауссовский процесс на \mathbb{R} с нулевым средним и ковариационной функцией

$$r(s, t) = \frac{V_H}{2} (|s|^{2H} + |t|^{2H} - |t-s|^{2H}), \quad s, t \in \mathbb{R},$$

где

$$V_H = D(\mathbf{B}_H(1)) = \frac{-\Gamma(2-2H) \cos(\pi H)}{\pi H(2H-1)}.$$

Теперь лемма 5.1 из [18] показывает, что

$$2^{\alpha J/2}(Y_J - F_J) \Rightarrow \tau \mathbf{B}_H(t), \quad t \in [0, 1].$$

Таким образом, полагая $\epsilon = \tau^{1/\alpha} \cdot 2^{-J/2}$, можно аппроксимировать наблюдаемый процесс $Y_J(t)$ с помощью $Y(t)$ для $t \in [0, 1]$:

$$Y(t) = F(t) + \epsilon^\alpha \mathbf{B}_H(t).$$

Так же как и в случае краткосрочной зависимости, можно перейти к следующей модели дискретных вейвлет-коэффициентов [16]:

$$X_{jk} = \mu_{jk} + 2^{(J-j)(1-\alpha)/2} z_{jk}, \quad (7)$$

где $z_{jk} = 2^{j(1-\alpha)/2} \int \psi_{jk} d\mathbf{B}_H$. Шумовые переменные z_{jk} имеют стандартное нормальное распределение, но не являются независимыми. Однако можно показать, что они обладают ограниченной зависимостью, т. е. для всех j и k

$$0 < c_0 \leq D(z_{jk}|z_{il}, i \neq j, k \neq l) \leq 1.$$

3 Пороговая обработка и оценка риска

Смысл пороговой обработки вейвлет-коэффициентов заключается в удалении достаточно маленьких коэффициентов, которые считаются шумом. Будем использовать так называемую мягкую пороговую обработку с порогом T_j , зависящим от уровня j . К каждому вейвлет-коэффициенту применяется функция $\rho_{T_j}(x) = \text{sgn}(x) (|x| - T_j)_+$, т. е. при такой пороговой обработке коэффициенты, которые по модулю меньше порога T_j , обнуляются, а абсолютные величины остальных коэффициентов уменьшаются на величину порога. Погрешность (или риск) мягкой пороговой обработки определяется следующим образом:

$$R_J(f) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} (\mu_{jk} - \rho_{T_j}(X_{jk}))^2. \quad (8)$$

В работах [1] и [3] было предложено использовать порог $T_j = \sigma_j \sqrt{2 \ln 2^J}$. Было показано, что при таком пороге риск близок к минимальному [1]. Этот порог получил название «универсальный». В дальнейшем будет использоваться именно такой вид порога. В выражении (8) присутствуют неизвестные величины μ_{jk} , поэтому вычислить значение $R_J(f)$ нельзя. Однако его можно оценить. В качестве оценки риска используется следующая величина [2]:

$$\widehat{R}_J(f) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} F [X_{jk}^2, T_j, \sigma_j], \quad (9)$$

где

$$F[x, T, \sigma] = (x - \sigma^2) \mathbf{1}(|x| \leq T^2) + (\sigma^2 + T^2) \mathbf{1}(|x| > T^2).$$

Величина $\widehat{R}_J(f)$ является несмещенной оценкой для $R_N(f)$ [14]. В работах [6–12] исследовались асимптотические свойства оценки (9) в модели с независимым шумом. Было показано, что при определенных условиях гладкости эта оценка является состоятельной и асимптотически нормальной. Поскольку для модели данных с краткосрочной зависимостью принимается модель эмпирических вейвлет-коэффициентов (5), с точностью до ограничений (6) все результаты указанных работ переносятся на этот случай. Далее будет исследовано асимптотическое поведение оценки (9) в модели данных с долгосрочной зависимостью.

4 Вспомогательные результаты

В этом разделе будут получены некоторые результаты, касающиеся характера зависимости эмпирических вейвлет-коэффициентов. Всюду далее предполагается, что используются вейвлеты Мейера [14], обладающие нужным количеством нулевых моментов и непрерывных производных. Рассмотрим ковариацию случайных величин в модели (7) [16]:

$$\begin{aligned} \text{cov}(X_{jk}, X_{il}) &= 2^{J(1-\alpha)} \mathbb{E} \int \psi_{jk} d\mathbf{B}_H \int \overline{\psi_{il}} d\mathbf{B}_H = \\ &= 2^{J(1-\alpha)} \frac{1}{2\pi} \int \widehat{\psi}_{jk}(\xi) \overline{\widehat{\psi}_{il}(\xi)} |\xi|^{-(1-\alpha)} d\xi. \end{aligned}$$

Для преобразования Фурье вейвлет-функции справедливо

$$\widehat{\psi}_{jk}(\xi) = 2^{-j/2} e^{ik \cdot 2^{-j}} \widehat{\psi}(2^{-j}\xi).$$

Рассмотрим ковариацию в пределах одного уровня j :

$$\begin{aligned} \text{cov}(X_{j0}, X_{jk}) &= \\ &= 2^{(J-j)(1-\alpha)} \frac{1}{2\pi} \int e^{ik\xi} |\widehat{\psi}(\xi)|^2 |\xi|^{-(1-\alpha)} d\xi. \end{aligned}$$

Поскольку вейвлет-функция ψ имеет M нулевых моментов и M непрерывных производных, найдется такая константа $C_M > 0$, что

$$|\text{cov}(X_{j0}, X_{jk})| \leq \frac{C_M 2^{(J-j)(1-\alpha)}}{k^M}. \quad (10)$$

Замечание 2. Для вейвлет-функции Мейера $\psi(x)$ при любом натуральном M_0 существует константа $C_{M_0} > 0$ такая, что $|\hat{\psi}(\xi)| \leq C_{M_0} |\xi|^{M_0} \mathbf{1}_{\xi \in \text{supp}(\hat{\psi})}$.

Теперь оценим ковариацию на разных уровнях $\text{cov}(X_{jk}, X_{il})$, предполагая, что $j > i$:

$$\begin{aligned} \text{cov}(X_{jk}, X_{il}) &= 2^{J(1-\alpha)} \frac{1}{2\pi} \int 2^{-j/2} e^{ik\xi \cdot 2^{-j}} \times \\ &\times \hat{\psi}(2^{-j}\xi) 2^{-i/2} \overline{e^{il\xi \cdot 2^{-i}} \hat{\psi}(2^{-i}\xi)} |\xi|^{-(1-\alpha)} d\xi = \\ &= 2^{J(1-\alpha)} \frac{1}{2\pi} 2^{-(\Delta/2)-i(1-\alpha)} \times \\ &\times \int e^{i\xi(k \cdot 2^{-\Delta}-l)} \hat{\psi}(2^{-\Delta}\xi) \overline{\hat{\psi}(\xi)} |\xi|^{-(1-\alpha)} d\xi, \end{aligned}$$

где $\Delta = j - i$. Далее, в силу замечания 2

$$\begin{aligned} |\text{cov}(X_{jk}, X_{il})| &\leq \\ &\leq 2^{J(1-\alpha)} \frac{C_{M_0}}{2\pi} 2^{-(\Delta/2)-i(1-\alpha)} \cdot 2^{-\Delta M_0} \times \\ &\times \left| \int e^{i\xi(k \cdot 2^{-\Delta}-l)} \overline{\hat{\psi}(\xi)} |\xi|^{M_0-(1-\alpha)} d\xi \right|. \end{aligned}$$

Рассмотрим два случая:

- (1) $|k \cdot 2^{-\Delta} - l| > 1$: тогда в силу свойств преобразования Фурье и гладкости выбранной вейвлет-функции найдется такая константа $C_1 > 0$, что

$$\begin{aligned} |\text{cov}(X_{jk}, X_{il})| &\leq \\ &\leq 2^{J(1-\alpha)} \frac{1}{2\pi} 2^{-(\Delta/2)-i(1-\alpha)} \cdot 2^{-\Delta M_0} \frac{C_1}{(k \cdot 2^{-\Delta} - l)^M}; \end{aligned}$$

- (2) $|k \cdot 2^{-\Delta} - l| \leq 1$: тогда найдется такая константа $C_2 > 0$, что

$$\left| \int e^{i\xi(k \cdot 2^{-\Delta}-l)} \overline{\hat{\psi}(\xi)} |\xi|^{M_0-(1-\alpha)} d\xi \right| \leq C_2$$

и

$$\begin{aligned} |\text{cov}(X_{jk}, X_{il})| &\leq \\ &\leq C_2 \cdot 2^{J(1-\alpha)} \frac{C_{M_0}}{2\pi} 2^{-(\Delta/2)-i(1-\alpha)} \cdot 2^{-\Delta M_0}. \end{aligned}$$

Следовательно, найдется такая константа $C_{Me} > 0$, что

$$|\text{cov}(X_{jk}, X_{il})| \leq \begin{cases} C_{Me} \cdot 2^{J(1-\alpha)-(\Delta/2)-i(1-\alpha)} \cdot 2^{-\Delta M_0} \times \\ \times \frac{1}{(k \cdot 2^{-\Delta} - l)^M}, & |k \cdot 2^{-\Delta} - l| > 1; \\ C_{Me} \cdot 2^{J(1-\alpha)-(\Delta/2)-i(1-\alpha)} \cdot 2^{-\Delta M_0}, & |k \cdot 2^{-\Delta} - l| \leq 1. \end{cases} \quad (11)$$

Рассмотрим теперь структуру дисперсии оценки риска. Введем обозначение для последовательностей $a_J \simeq b_J$, если $\lim(a_J/b_J) = 1$ при $J \rightarrow \infty$.

Лемма 1. Пусть $\alpha > 1/2$ и $\gamma > (4\alpha - 2)^{-1}$, тогда $D_J^2 = \text{DR}_J(f) \simeq C_\alpha \cdot 2^J$, где константа C_α зависит от α , но не зависит от функции сигнала f .

Доказательство. При выполнении условий леммы $(2\gamma + 1)^{-1} < 1 - (2\alpha)^{-1}$. Выберем p'' такое, что $(2\gamma + 1)^{-1} < p'' < 1 - (2\alpha)^{-1}$ и $p''J$ — целое число. Тогда в силу (2) $\mu_{jk} \rightarrow 0$ для всех j : $p''J \leq j < J$ при $J \rightarrow \infty$. Разобьем выражение (9) на две суммы:

$$\begin{aligned} \widehat{R}_J(f) &= \sum_{j=0}^{p''J-1} \sum_{k=0}^{2^j-1} F[X_{jk}^2, T_j, \sigma_j] + \\ &+ \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} F[X_{jk}^2, T_j, \sigma_j]. \end{aligned}$$

Так как существует такая константа $C_F > 0$, что $F[X_{jk}^2, T_j, \sigma_j] \leq C_F T_j^2 = C_F J \cdot 2^{(J-j)(1-\alpha)}$, для первой суммы имеем

$$\begin{aligned} \sum_{i=0}^{p''J-1} \sum_{l=0}^{2^i-1} F[X_{jk}^2, T_j, \sigma_j] &\leq C_F \sum_{i=0}^{p''J-1} \sum_{l=0}^{2^i-1} J \cdot 2^{(J-i)(1-\alpha)} \leq \\ &\leq C'_F J \cdot 2^{J(1-\alpha+\alpha p'')} \quad (12) \end{aligned}$$

с некоторой константой $C'_F > 0$. Далее

$$\begin{aligned} \text{DR}_J(f) &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \text{DF}[X_{jk}^2, T_j, \sigma_j] + \sum_{i=0}^{J-1} \sum_{l=0}^{2^i-1} \times \\ &\times \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \text{cov}(F[X_{il}^2, T_i, \sigma_i]^2, F[X_{jk}^2, T_j, \sigma_j]^2). \quad (13) \end{aligned}$$

Рассмотрим сумму дисперсий:

$$\begin{aligned} \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \text{DF}[X_{jk}^2, T_j, \sigma_j] &= \sum_{j=0}^{p''J-1} \sum_{k=0}^{2^j-1} \text{DF}[X_{jk}^2, T_j, \sigma_j] + \\ &+ \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{DF}[X_{jk}^2, T_j, \sigma_j]. \end{aligned}$$

В силу (12) первая сумма не превосходит $C''_F J^2 \times 2^{2J(1-\alpha+\alpha p'')}$, где C''_F — некоторая положительная константа, и, так как $p'' < 1 - (2\alpha)^{-1}$, имеем

$2(1 - \alpha + \alpha p'') < 1$. Учитывая вид порога T_j и принимая во внимание, что $\alpha > 1/2$, для второй суммы имеем:

$$\begin{aligned} & \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{DF} [X_{jk}^2, T_j, \sigma_j] \simeq \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{DX}_{jk}^2 = \\ & = \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} 2\sigma_j^2 (\sigma_j^2 + \mu_{jk}^2) \simeq \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} 2\sigma_j^4 = \\ & = \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} 2 \frac{2^{2J(1-\alpha)}}{2^{2j(1-\alpha)}} = \\ & = 2^{2J(1-\alpha)+1} \sum_{j=p''J+1}^{J-1} 2^{-j+2\alpha j} \simeq 2^J \frac{2}{2^{2\alpha-1}(2^{2\alpha-1}-1)}. \end{aligned}$$

Таким образом,

$$\begin{aligned} & \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \text{DF} [X_{jk}^2, T_j, \sigma_j] \simeq \\ & \simeq \sum_{j=p''J+1}^J \sum_{k=0}^{2^j-1} \text{DX}_{jk}^2 \simeq C'_\alpha 2^J, \quad (14) \end{aligned}$$

где C'_α — положительная константа. Рассмотрим теперь сумму ковариаций в (13). Аналогично сумме дисперсий имеем:

$$\begin{aligned} & \sum_{i=0}^{J-1} \sum_{l=0}^{2^i-1} \sum_{j=0}^{J-1} \times \\ & \times \sum_{k=0}^{2^j-1} \text{cov} (F[X_{jk}^2, T_j, \sigma_j], F[X_{lk}^2, T_l, \sigma_l]) \simeq \\ & \simeq \sum_{i=p''J+1}^{J-1} \sum_{l=0}^{2^i-1} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{cov}(X_{il}^2, X_{jk}^2). \end{aligned}$$

Известно, что если вектор (X, Y) имеет двумерное нормальное распределение, то

$$\text{cov}(X^2, Y^2) = 4\mathbb{E}X\mathbb{E}Y \text{cov}(X, Y) + 2\text{cov}^2(X, Y). \quad (15)$$

Используя (10), (11) и (15), получаем

$$\begin{aligned} & \frac{1}{2} \sum_{i=p''J+1}^{J-1} \sum_{l=0}^{2^i-1} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{cov}^2(X_{il}, X_{jk}) = \\ & = \sum_{i=p''J+1}^{J-1} \sum_{l=0}^{2^i-1} \sum_{\Delta=0}^{J-i-1} \sum_{k=\begin{cases} l+1, & \Delta=0; \\ 0, & \Delta>0. \end{cases}}^{2^{i+\Delta}-1} \text{cov}^2(X_{il}, X_{i+\Delta,k}) \leq \end{aligned}$$

$$\begin{aligned} & \leq \sum_{i=p''J+1}^{J-1} \sum_{l=0}^{2^i-1} \left(\sum_{\delta=1}^{2^i-l} C_M^2 \delta^{-2M_0} \cdot 2^{2(J-i)(1-\alpha)} + \right. \\ & \left. + \sum_{\Delta=1}^{J-i-1} \sum_{k=0}^{2^{i+\Delta}-1} 2^{2(J-i)(1-\alpha)} \cdot 2^{-\Delta(1+2M_0)} \times \right. \\ & \left. \times \frac{\mathbf{1}_{\{|k \cdot 2^{-\Delta} - l| > 1\}}}{|k \cdot 2^{-\Delta} - l|^{2M}} C_{Me}^2 + \sum_{\Delta=1}^{J-i-1} \sum_{k=0}^{2^{i+\Delta}-1} 2^{2(J-i)(1-\alpha)} \times \right. \\ & \left. \times 2^{-\Delta(1+2M_0)} \cdot \mathbf{1}_{\{|k \cdot 2^{-\Delta} - l| \leq 1\}} C_{Me}^2 \right) \leq \\ & \leq \sum_{i=p''J+1}^{J-1} \left(\sum_{l=0}^{2^i-1} \sum_{\delta=1}^{2^i-l-1} C_M^2 \delta^{-2M_0} 2^{2(J-i)(1-\alpha)} + \right. \\ & \left. + \sum_{\Delta=1}^{J-i-1} 2^{2(J-i)(1-\alpha)} \cdot 2^{-\Delta(1+2M_0)} \times \right. \\ & \left. \times \sum_{l=0}^{2^i-1} C_{Me}^2 \left(\sum_{k=0}^{2^{i+\Delta}-1} \frac{\mathbf{1}_{\{|k \cdot 2^{-\Delta} - l| > 1\}}}{|k \cdot 2^{-\Delta} - l|^{2M}} + 2^{\Delta+1} \right) \right) \leq \\ & \leq \sum_{i=p''J+1}^{J-1} 2^{2(J-i)(1-\alpha)} \left(C_M^2 \cdot 2^i H_0 + \right. \\ & \left. + \sum_{\Delta=1}^{J-i-1} (2^{i-2M_0\Delta} H_1 + 2^{i-2M_0\Delta+1} H_2) \right) \leq \\ & \leq \sum_{i=p''J+1}^{J-1} 2^{2(J-i)(1-\alpha)} \cdot 2^i H_3 \leq C''_\alpha \cdot 2^J, \quad (16) \end{aligned}$$

где H_0, H_1, H_2, H_3 и C''_α — константы, зависящие от α . Аналогично с учетом (2)

$$\begin{aligned} & \sum_{i=p''J+1}^{J-1} \sum_{l=0}^{2^i-1} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \mu_{il} \mu_{jk} \text{cov}(X_{il}, X_{jk}) \leq \\ & \leq H_\alpha \cdot 2^{J(1-\alpha(1-p''))}, \quad (17) \end{aligned}$$

где H_α — константа, зависящая от α .

Объединяя (14), (16) и (17), получаем, что $\widehat{\text{DR}}_J(f) \simeq C_\alpha \cdot 2^J$. Лемма доказана.

Докажем еще одно свойство эмпирических вейвлет-коэффициентов. Говорят, что последовательность случайных величин $\{Y_i\}_{i=1}^\infty$ обладает свойством ρ -перемешивания, если для функции

$$\rho(m) = \sup_{i,j:|i-j|>m} \text{corr}(Y_i, Y_j)$$

справедливо $\rho(m) \rightarrow 0$ при $m \rightarrow \infty$.

Лемма 2. Последовательность $\{F[X_{jk}^2, T_j, \sigma_j]\}$, $j = 0, \dots, J-1, k = 1, \dots, 2^j$, обладает свойством ρ -перемешивания. Причем для некоторой положительной константы C_ρ

$$\rho(m) \leq \begin{cases} \frac{C_\rho}{(m+1)^{2M_0}} & \text{для элементов} \\ & \text{на одном уровне } (i=j); \\ \frac{C_\rho}{2^{(m+1)\alpha}} & \text{для элементов} \\ & \text{на разных уровнях.} \end{cases}$$

Доказательство. Рассмотрим функцию перемешивания между элементами с номерами l и k на одном уровне i , $i = 0, \dots, J-1$. Для некоторой константы $C_\rho > 0$ имеем:

$$\begin{aligned} \rho(m) &= \\ &= \sup_{\substack{0 \leq i \leq J-1, \\ k, l: |k-l| > m}} \frac{\text{cov}(F[X_{il}^2, T_i, \sigma_i], F[X_{ik}^2, T_i, \sigma_i])}{\sqrt{\text{DF}[X_{il}^2, T_i, \sigma_i] \text{DF}[X_{ik}^2, T_i, \sigma_i]}} \leq \\ &\leq \sup_{\substack{0 \leq i \leq J-1, \\ k, l: |k-l| > m}} C_\rho \frac{2^{2(J-i)(1-\alpha)} |k-l|^{-2M_0}}{\sqrt{\sigma_i^8}} \leq \\ &\leq \sup_{\substack{0 \leq i \leq J-1, \\ k, l: |k-l| > m}} C_\rho \frac{2^{2(J-i)(1-\alpha)} |k-l|^{-2M_0}}{2^{2(J-i)(1-\alpha)}} \leq \\ &\leq \sup_{\substack{0 \leq i \leq J-1, \\ k, l: |k-l| > m}} C_\rho \frac{1}{|k-l|^{2M_0}} = \frac{C_\rho}{(m+1)^{2M_0}}. \end{aligned}$$

Далее обратимся к функции перемешивания для элементов $F[X_{il}^2, T_i, \sigma_i]$ и $F[X_{jk}^2, T_j, \sigma_j]$ на разных уровнях $i, j: j > i, j-i = \Delta > 0$. Рассмотрим случай $|k \cdot 2^{-\Delta} - l| > 1$. Имеем:

$$\begin{aligned} \rho(m) &= \\ &= \sup_{\substack{j-i=\Delta > m, \\ 0 \leq l \leq 2^i-1, \\ 0 \leq k \leq 2^j-1, \\ |k \cdot 2^{-\Delta} - l| > 1.}} \frac{\text{cov}(F[X_{il}^2, T_i, \sigma_i], F[X_{jk}^2, T_j, \sigma_j])}{\sqrt{\text{DF}[X_{il}^2, T_i, \sigma_i] \text{DF}[X_{jk}^2, T_j, \sigma_j]}} \leq \\ &\leq \sup_{\substack{j-i=\Delta > m, \\ 0 \leq l \leq 2^i-1, \\ 0 \leq k \leq 2^j-1, \\ |k \cdot 2^{-\Delta} - l| > 1.}} C_\rho \frac{2^{2(J-i)(1-\alpha)-\Delta} |k \cdot 2^{-\Delta} - l|^{-2M_0}}{\sqrt{\sigma_j^4 \sigma_i^4}} \leq \\ &\leq \sup_{\substack{j-i=\Delta > m, \\ 0 \leq l \leq 2^i-1, \\ 0 \leq k \leq 2^j-1, \\ |k \cdot 2^{-\Delta} - l| > 1.}} C_\rho \frac{2^{2(J-i)(1-\alpha)-\Delta} |k \cdot 2^{-\Delta} - l|^{-2M_0}}{2^{2(J-i)(1-\alpha)-\Delta(1-\alpha)}} \leq \end{aligned}$$

$$\begin{aligned} &\leq \sup_{\substack{\Delta > m, \\ 0 \leq l \leq 2^i-1, \\ 0 \leq k \leq 2^j-1, \\ |k \cdot 2^{-\Delta} - l| > 1.}} C_\rho \frac{1}{2^{\Delta\alpha}} \frac{1}{|k \cdot 2^{-\Delta} - l|^{2M_0}} \leq \\ &\leq \sup_{\Delta > m} C_\rho \frac{1}{2^{\Delta\alpha}} = \frac{C_\rho}{2^{(m+1)\alpha}}. \end{aligned}$$

Теперь рассмотрим случай $|k \cdot 2^{-\Delta} - l| \leq 1$. Имеем для некоторой константы $C'_\rho > 0$:

$$\begin{aligned} \rho(m) &= \\ &= \sup_{\substack{j-i=\Delta > m, \\ 0 \leq l \leq 2^i-1, \\ 0 \leq k \leq 2^j-1, \\ |k \cdot 2^{-\Delta} - l| \leq 1.}} \frac{\text{cov}(F[X_{il}^2, T_i, \sigma_i], F[X_{jk}^2, T_j, \sigma_j])}{\sqrt{\text{DF}[X_{il}^2, T_i, \sigma_i] \text{DF}[X_{jk}^2, T_j, \sigma_j]}} \leq \\ &\leq \sup_{j-i=\Delta > m} C'_\rho \frac{2^{2(J-i)(1-\alpha)-\Delta} C_{Me}}{\sqrt{\sigma_j^4 \sigma_i^4}} \leq \\ &\leq \sup_{j-i=\Delta > m} C'_\rho \frac{2^{2(J-i)(1-\alpha)-\Delta} C_{Me}}{2^{2(J-i)(1-\alpha)-\Delta(1-\alpha)}} \leq \\ &\leq \sup_{\Delta, k, l: \Delta > m} C_\rho \frac{1}{2^{\Delta\alpha}} = \frac{C_\rho}{2^{(m+1)\alpha}}. \end{aligned}$$

Следовательно, получаем

$$\rho(m) \leq \begin{cases} \frac{C_\rho}{(m+1)^{2M_0}} & \text{на одном уровне;} \\ \frac{C_\rho}{2^{(m+1)\alpha}} & \text{на разных уровнях.} \end{cases}$$

При $m \rightarrow \infty$ $\rho(m) \rightarrow 0$ что и доказывает утверждение леммы.

5 Основные теоремы

Докажем асимптотическую нормальность оценки риска.

Теорема 1. Пусть $\alpha > 1/2$ и функция f регулярна с параметром $\gamma > (4\alpha - 2)^{-1}$. Тогда при пороговой обработке с «универсальным» порогом T_j имеет место сходимость по распределению:

$$\frac{\hat{R}_J(f) - R_J(f)}{D_J} \Rightarrow \mathbf{N}(0, 1), \quad J \rightarrow \infty, \quad (18)$$

где $D_J^2 = C_\alpha 2^J$, а константа C_α зависит только от α и выбранного вейвлет-базиса.

Доказательство. Из леммы 1 следует, что $D\hat{R}_J(f) \simeq D_J^2 = C_\alpha 2^J$. Разобьем выражение в (18) на две суммы, как это было сделано в лемме 1:

$$\begin{aligned} \frac{\widehat{R}_J(f) - R_J(f)}{D_J} &= \\ &= \frac{\sum_{j=0}^{p''J-1} \sum_{k=0}^{2^j-1} (F[X_{jk}^2, T_j, \sigma_j] - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j])}{D_J} + \\ &+ \frac{\sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} (F[X_{jk}^2, T_j, \sigma_j] - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j])}{D_J}, \end{aligned}$$

где $p'' > (2\gamma + 1)^{-1}$. Поскольку $\gamma > (4\alpha - 2)^{-1}$, имеем $(2\gamma + 1)^{-1} < 1 - (2\alpha)^{-1}$. Следовательно, можно выбрать такое p'' , что $(2\gamma + 1)^{-1} < p'' < 1 - (2\alpha)^{-1}$. Отсюда $1 - \alpha + \alpha p'' < 1/2$, и из (12) следует, что первая сумма стремится к нулю почти всюду (п.в.).

Далее, действуя как в лемме 1, нетрудно показать, что

$$\begin{aligned} \sup_J \frac{1}{D_J^2} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}(F[X_{jk}^2, T_j, \sigma_j] - \\ - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j])^2 = \\ = \sup_J \frac{1}{D_J^2} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \text{DF}[X_{jk}^2, T_j, \sigma_j] < \infty. \end{aligned}$$

Из леммы 2 следует, что последовательность $\{F[X_{jk}^2, T_j, \sigma_j]\}$, $j = 0, \dots, J-1, k = 1, \dots, 2^j$, обладает свойством ρ -перемешивания и, следовательно, обладает свойством α -перемешивания [19].

Наконец, выполнено условие Линдберга: для любого $\epsilon > 0$

$$\begin{aligned} \frac{1}{D_J^2} \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}(F[X_{jk}^2, T_j, \sigma_j] - \\ - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j])^2 \times \\ \times \mathbf{1}(|F[X_{jk}^2, T_j, \sigma_j] - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j]| > \epsilon D_J) \rightarrow \\ \rightarrow 0, J \rightarrow \infty. \quad (19) \end{aligned}$$

Действительно, так как $F[X_{jk}^2, T_j, \sigma_j] \leq C_F J \times 2^{(J-j)(1-\alpha)}$ (см. лемму 1), а $D_J^2 = C_\alpha \cdot 2^J$, то при $\alpha > 1/2$, начиная с некоторого J , все индикаторы в (19) обращаются в ноль.

Таким образом, выполнены все условия теоремы 2.1 из работы [20] и справедлива сходимость по распределению (18). Теорема доказана.

Докажем теперь теорему о состоятельности оценки риска, справедливую при более слабых ограничениях на α и γ .

Теорема 2. Пусть $0 < \alpha < 1$, функция f регулярна с параметром $\gamma > 0$ и $b > 1 - \alpha + \alpha(2\gamma + 1)^{-1}$. Тогда при пороговой обработке с «универсальным» порогом T_j выполняется

$$\frac{\widehat{R}_J(f) - R_J(f)}{2^{bJ}} \xrightarrow{P} 0, J \rightarrow \infty. \quad (20)$$

Доказательство. Условия теоремы позволяют выбрать p'' таким, что $(2\gamma + 1)^{-1} < p'' < (b + \alpha - 1)/\alpha$. Разобьем выражение в (20) на две суммы:

$$\begin{aligned} \widehat{R}_J(f) - R_J(f) &= \\ &= \sum_{j=0}^{p''J-1} \sum_{k=0}^{2^j-1} (F[X_{jk}^2, T_j, \sigma_j] - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j]) + \\ &+ \sum_{j=p''J+1}^{J-1} \sum_{k=0}^{2^j-1} (F[X_{jk}^2, T_j, \sigma_j] - \mathbb{E}F[X_{jk}^2, T_j, \sigma_j]) \equiv \\ &\equiv R_1 + R_2. \end{aligned}$$

Повторяя рассуждения леммы 1, получаем, что $|R_1| \leq C_1 J \cdot 2^{J(1-\alpha+\alpha p'')}$ с некоторой константой $C_1 > 0$. При этом $1 - \alpha + \alpha p'' < b$ в силу выбора p'' . Следовательно, $R_1 \cdot 2^{-bJ} \xrightarrow{P} 0$. Далее, так же как в лемме 1, можно показать, что для некоторой константы C_α

$$\text{DR}_2 \simeq \begin{cases} C_\alpha \cdot 2^J, & \text{если } \alpha > \frac{1}{2}; \\ C_\alpha J \cdot 2^J, & \text{если } \alpha = \frac{1}{2}; \\ C_\alpha \cdot 2^{2J(1-\alpha)}, & \text{если } \alpha < \frac{1}{2}. \end{cases}$$

Вследствие неравенства Чебышёва имеем при любом $\delta > 0$

$$P\left(\frac{|R_2|}{2^{bJ}} \geq \delta\right) \leq \frac{\text{DR}_J(f)}{2^{2bJ}\delta^2}. \quad (21)$$

Для $1 > \alpha \geq 1/2$ правая часть (21) стремится к нулю при $b > 1/2$. Для $1/2 > \alpha > 0$ правая часть (21) стремится к нулю при $b > 1 - \alpha$.

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Donoho D., Johnstone I. M. Adapting to unknown smoothness via wavelet shrinkage // *J. Amer. Stat. Assoc.*, 1995. Vol. 90. P. 1200–1224.
3. Donoho D. L., Johnstone I. M., Kerkycharian G., Picard D. Wavelet shrinkage: Asymptopia? // *J. R. Statist. Soc. Ser. B.*, 1995. Vol. 57. No. 2. P. 301–369.

4. *Marron J. S., Adak S., Johnstone I. M., Neumann M. H., Patil P.* Exact risk analysis of wavelet regression // *J. Comput. Graph. Stat.*, 1998. Vol. 7. P. 278–309.
5. *Antoniadis A., Fan J.* Regularization of wavelet approximations // *J. Amer. Statist. Assoc.*, 2001. Vol. 96. No. 455. P. 939–967.
6. *Маркин А. В.* Предельное распределение оценки риска при пороговой обработке вейвлет-коэффициентов // *Информатика и её применения*, 2009. Т. 3. Вып. 4. С. 57–63.
7. *Маркин А. В., Шестаков О. В.* О состоятельности оценки риска при пороговой обработке вейвлет-коэффициентов // *Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн.*, 2010. № 1. С. 26–34.
8. *Шестаков О. В.* Аппроксимация распределения оценки риска пороговой обработки вейвлет-коэффициентов нормальным распределением при использовании выборочной дисперсии // *Информатика и её применения*, 2010. Т. 4. Вып. 4. С. 73–81.
9. *Шестаков О. В.* О точности приближения распределения оценки риска пороговой обработки вейвлет-коэффициентов сигнала нормальным законом при неизвестном уровне шума // *Системы и средства информатики*, 2012. Т. 22. № 1. С. 142–152.
10. *Шестаков О. В.* Асимптотическая нормальность оценки риска пороговой обработки вейвлет-коэффициентов при выборе адаптивного порога // *Доклады РАН*, 2012. Т. 445. № 5. С. 513–515.
11. *Шестаков О. В.* Зависимость предельного распределения оценки риска пороговой обработки вейвлет-коэффициентов сигнала от вида оценки дисперсии шума при выборе адаптивного порога // *Т-Comm — Телекоммуникации и транспорт*, 2012. № 1. С. 46–51.
12. *Шестаков О. В.* Центральная предельная теорема для функции обобщенной кросс-валидации при пороговой обработке вейвлет-коэффициентов // *Информатика и её применения*, 2013. Т. 7. Вып. 2. С. 40–49.
13. *Добеши И.* Десять лекций по вейвлетам. — Ижевск: НИЦ Регулярная и хаотическая динамика, 2001. 357 с.
14. *Mallat S.* A wavelet tour of signal processing. — Academic Press, 1999. 851 с.
15. *Boggess A., Narkowich F.* A first course in wavelets with Fourier analysis. — Prentice Hall, 2001. 283 с.
16. *Johnstone I. M., Silverman B. W.* Wavelet threshold estimates for data with correlated noise // *J. Roy. Statist. Soc. Ser. B*, 1997. Vol. 59. P. 319–351.
17. *Johnstone I. M.* Wavelet shrinkage for correlated data and inverse problems: Adaptivity results // *Statistica Sinica*, 1999. Vol. 9. No. 1. P. 51–83.
18. *Taqqu M. S.* Weak convergence to fractional Brownian motion and to the Rosenblatt process // *Z. Wahrscheinlichkeitsth. verw. Geb.*, 1975. Vol. 31. P. 287–302.
19. *Bradley R. C.* Basic properties of strong mixing conditions. A survey and some open questions // *Probab. Surveys*, 2005. Vol. 2. P. 107–144.
20. *Peligrad M.* On the asymptotic normality of sequences of weak dependent random variables // *J. Theoret. Prob.*, 1996. Vol. 9. No. 3. P. 703–715.

Поступила в редакцию 14.08.13

ASYMPTOTIC PROPERTIES OF WAVELET THRESHOLDING RISK ESTIMATE IN THE MODEL OF DATA WITH CORRELATED NOISE

A. A. Eroshenko¹ and O. V. Shestakov^{1,2}

¹M. V. Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Wavelet thresholding techniques of denoising are widely used in signal and image processing. These methods are easily implemented through fast algorithms; so, they are very appealing in practical situations. Besides, they adapt to function classes with different amounts of smoothness in different locations more effectively than the usual linear methods. Wavelet thresholding risk analysis is an important practical task because it allows determining the quality of techniques themselves and equipment which is being used. In the present paper, asymptotical properties of mean-square risk estimate of wavelet thresholding techniques have been studied in the model of data with correlated noise. The conditions under which the unbiased risk estimate is consistent and asymptotically normal are given. These results allow constructing asymptotical confidence intervals for wavelet thresholding risk, using only the observed data.

Keywords: wavelets; unbiased risk estimate; correlated noise; asymptotic normality

DOI: 10.14357/19922264140105

Acknowledgments

The financial support of the Russian Foundation for Basic Research is acknowledged (project 11-01-00515a).

References

1. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455.
2. Donoho, D., and I. M. Johnstone. 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.* 90:1200–1224.
3. Donoho, D. L., I. M. Johnstone, G. Kerkycharian, and D. Picard. 1995. Wavelet shrinkage: Asymptopia? *J. R. Statist. Soc. Ser. B.* 57(2):301–369.
4. Marron, J. S., S. Adak, I. M. Johnstone, M. H. Neumann, and P. Patil. 1998. Exact risk analysis of wavelet regression. *J. Comput. Graph. Stat.* 7:278–309.
5. Antoniadis, A., and J. Fan. 2001. Regularization of wavelet approximations. *J. Amer. Statist. Assoc.* 96(455):939–967.
6. Markin, A. V. 2009. Predel'noe raspredelenie otsenki riska pri porogovoy obrabotke veyvlet-koeffitsientov [Limit distribution of risk estimate of wavelet coefficient thresholding]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):57–63.
7. Markin, A. V., and O. V. Shestakov 2010. O sostoyatel'nosti otsenki riska pri porogovoy obrabotke veyvlet-koeffitsientov [Consistency of risk estimation with thresholding of wavelet coefficients]. *Vestn. Mosk. Un-ta. Ser. 15. Vychisl. Matem. i Kibern. [Herald of Moscow University, Computational Mathematics and Cybernetics]* 1:26–34.
8. Shestakov, O. V. 2010. Approksimatsiya raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov normal'nym raspredeleniem pri ispol'zovanii vyborochnoy dispersii [Normal approximation for distribution of risk estimate for wavelet coefficients thresholding when using sample variance]. *Informatika i ee Primeneniya — Inform. Appl.* 4(4):73–81.
9. Shestakov, O. V. 2012. O tochnosti priblizheniya raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov signala normal'nym zakonom pri neizvestnom urovne shuma [On the accuracy of normal approximation for risk estimate distribution when thresholding signal wavelet coefficients in case of unknown noise level]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):142–152.
10. Shestakov, O. V. 2012. Asimptoticheskaya normal'nost' otsenki riska porogovoy obrabotki veyvlet-koeffitsientov pri vybore adaptivnogo poroga [Asymptotic normality of adaptive wavelet thresholding risk estimation]. *Doklady RAN [Doklady Mathematics]* 445(5):513–515.
11. Shestakov, O. V. 2012. Zavisimost' predel'nogo raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov signala ot vida otsenki dispersii shuma pri vybore adaptivnogo poroga [The dependence of the limiting distribution of the risk assessment thresholding wavelet coefficients of the signal on the type of noise variance estimation when selecting an adaptive threshold]. *T-Comm — Telekommunikacii i Transport [T-Comm — Telecommunications and Transport]* 1:46–51.
12. Shestakov, O. V. 2013. Tsentral'naya predel'naya teorema dlya funktsii obobshchennoy kross-validatsii pri porogovoy obrabotke veyvlet-koeffitsientov [Central limit theorem for generalized cross-validation function in wavelet thresholding method]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):40–49.
13. Daubechies, I. 1992. *Ten lectures on wavelets*. SIAM. 357 p.
14. Mallat, S. 1999. *A wavelet tour of signal processing*. Academic Press. 851 p.
15. Boggess, A., and F. Narkowich. 2001. *A first course in wavelets with Fourier analysis*. Prentice Hall. 283 p.
16. Johnstone, I. M., and B. W. Silverman. 1997. Wavelet threshold estimates for data with correlated noise. *J. Roy. Statist. Soc. Ser. B.* 59:319–351.
17. Johnstone, I. M. 1999. Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica* 9(1):51–83.
18. Taqqu, M. S. 1975. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheor. verw. Geb.* 31:287–302.
19. Bradley, R. C. 2005. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surveys* 2:107–144.
20. Peligrad, M. 1996. On the asymptotic normality of sequences of weak dependent random variables. *J. Theoret. Prob.* 9(3):703–715.

Received August 14, 2013

Contributors

Eroshenko Alexander A. (b. 1989) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; aeroshik@gmail.com

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, assistant professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; oshestakov@cs.msu.su

БАЗИС РЕАЛИЗАЦИИ СУПЕР-ЭВМ ЭКСАФЛОПСНОГО КЛАССА*

И. А. Соколов¹, Ю. А. Степченков², С. Г. Бобков³, В. Н. Захаров⁴, Ю. Г. Дьяченко⁵, Ю. В. Рождественский⁶, А. В. Сурков⁷

Аннотация: Статья посвящена выбору схемотехнического базиса реализации микропроцессоров и коммуникационных сред супер-ЭВМ эксафлопсного класса. Проведен сравнительный анализ характеристик цифровых устройств различной сложности, реализованных в синхронном (С) и самосинхронном (СС, self-timed) базисе. Подтверждены основные преимущества СС-схем по сравнению с синхронными аналогами: отсутствие гонок, максимально возможный диапазон работоспособности, высокое быстродействие, относительно низкая мощность потребления. В результате перехода от синхронной к квазисамосинхронной и самосинхронной реализации зона работоспособности устройства расширяется независимо от его сложности. В наибольшей степени эти преимущества проявляются при использовании СС-схем для проектирования надежной цифровой аппаратуры. Рассмотрены различные методологии проектирования СС-схем. Проведен сравнительный анализ реализации СС-схем в обобщенном базисе схем, нечувствительных к задержкам, развиваемом авторами, и в базисе NCL (NULL Convention Logic) схем. Показано, что предлагаемый базис обеспечивает получение схем с лучшими параметрами по быстродействию, аппаратным затратам и энергопотреблению при проектировании типовых цифровых устройств, составляющих основу для построения современных вычислительных систем и комплексов.

Ключевые слова: синхронные схемы; самосинхронные схемы; нечувствительность к задержкам; NULL Convention Logic; быстродействие; энергопотребление; отказоустойчивость

DOI: 10.14357/19922264140106

1 Введение

В настоящее время потенциал инженерных технологий, опирающихся на теории микроскопического взаимодействия в рамках моделей сплошной среды, практически исчерпан. Дальнейшее развитие ядерной и термоядерной энергетики, электроники, авиастроения, биотехнологий и др. становится невозможным без проведения полномасштабных инженерных расчетов сложных технических и биологических систем с учетом атомно-молекулярного взаимодействия. А это требует уже в среднесрочной перспективе (до 2020 г.) использования супер-ЭВМ эксафлопсного класса (10^{18} оп/с). Только те страны, которые будут иметь такие ЭВМ и соответствующее программное обеспечение, способны будут создавать принципиально новые изделия. Таким образом, создание супер-ЭВМ является одним из актуальнейших направлений развития техники.

Основные трудности в достижении эксафлопсной производительности — необходимость эф-

фективного и надежного функционирования 10^8 – 10^9 процессорных ядер и преодоление физических ограничений, обусловленных энергопотреблением, надежностью и конструктивными размерами. По оценкам авторов, энергопотребление супер-ЭВМ, созданной по имеющимся технологиям и обладающей производительностью 10 PFLOPS, составит около 30 МВт. Для создания такой супер-ЭВМ эксафлопсной производительности потребуется уже строить рядом с ЭВМ небольшую электростанцию.

Решение проблем создания супер-ЭВМ эксафлопсной производительности невозможно без разработки фундаментальных основ повышения надежности и снижения потребления питания требуемых супер-ЭВМ. Используемые современные технологии не позволяют создать супер-ЭВМ эксафлопсной производительности.

Повышение надежности и снижение потребления питания требуемых супер-ЭВМ можно реализовать только за счет реализации комплекса архи-

* Исследование выполнено при финансовой поддержке РФФИ (проекты 13-07-12062 офи.м и 13-07-12068 офи.м), а также при частичной финансовой поддержке Программы фундаментальных исследований ОНИТ РАН за 2013 г. (проект 1.5).

¹ Институт проблем информатики Российской академии наук, ISokolov@ipiran.ru

² Институт проблем информатики Российской академии наук, YStepchenkov@ipiran.ru

³ Научно-исследовательский институт системных исследований Российской академии наук, bobkov@cs.niisi.ras.ru

⁴ Институт проблем информатики Российской академии наук, VZakharov@ipiran.ru

⁵ Институт проблем информатики Российской академии наук, diaura@mail.ru

⁶ Институт проблем информатики Российской академии наук, YRogdest@ipiran.ru

⁷ Научно-исследовательский институт системных исследований Российской академии наук, surkov@cs.niisi.ras.ru

тектурных, схемотехнических, технологических и конструктивных решений.

Проблемы, встающие перед разработчиками современной вычислительной техники (низкое энергопотребление, надежность работы в меняющихся условиях эксплуатации и т. д.), заставляют по-новому взглянуть на принципы проектирования цифровой аппаратуры и в первую очередь на задачу синхронизации. Синхронизация — одна из важнейших задач в цифровых системах, решающая проблему координации событий (сигналов, операций или процессов) в аппаратуре и связанная в основном с обеспечением интерфейса между физическим (естественным) и логическим (искусственным) временем [1]. Координация событий отражает причинно-следственные связи между ними и обычно определяется последовательностью множества событий, происходящих в системе.

В середине 1950-х гг. активно исследовались альтернативные методологии синхронизации элементов в аппаратуре: синхронная и самосинхронная. В С-методологии интерфейс между физическим и логическим (системным) временем определяется системными часами: все события в синхронной системе могут инициироваться только синхроимпульсами внешних часов. Действительная длительность инициированных событий никак не отслеживается. Чтобы синхронизируемая аппаратура работала корректно, период синхроимпульсов выбирается из расчета на наихудший случай — максимально возможное время переключения отдельных элементов при неблагоприятных сочетаниях условий функционирования (напряжения питания, температуры, параметров и характера распределения нагрузки и т. п.). В результате заведомо ухудшается быстродействие синхронной аппаратуры — до 130% [2] по сравнению с номинально возможным быстродействием.

Механизмы, обеспечивающие системное время в СС-подходе, включены в модель системного поведения и должны быть разработаны вместе с созданием начальной поведенческой спецификации. Корректные СС-системы базируются на механизме фиксации действительного окончания всех инициированных процессов. При этом обеспечивается их правильное функционирование независимо от задержек распространения сигналов в элементах схемы, отчего они также называются нечувствительными к задержкам — НЗ (delay-insensitive, DI).

С момента появления теории Маллера [3, 4] проектирование НЗ-схем было областью активных исследований [5–14]. Однако ввиду сравнительной сложности их проектирования популярными стали лишь С-схемы. Развитие средств автоматизации проектирования (САПР), образовательной и

технологической базы также пошло в направлении синхронной схемотехники. В конце XX в. абсолютное большинство выпускаемых серийно БИС проектировалось по синхронному маршруту.

В последние годы неуклонное развитие технологий и растущие требования выявляют все больше сложностей в проектировании С-схем, ставшая разработчиков со все большим спектром задач, ранее не изученных. Так, непрерывное увеличение производительности влечет за собой постоянную борьбу за снижение энергопотребления, а также соблюдение жестких требований к надежности и помехозащищенности схем [15]. Поиск методов и решений этих проблем возродил интерес к НЗ-схемотехнике, лишенной части недостатков С-схем, таких как, к примеру, избыточное потребление вследствие использования тактирующих цепей. Однако, несмотря на многочисленные потенциальные преимущества НЗ-схем, коммерчески выпускаемых НЗ-изделий по-прежнему мало.

В печати приводятся результаты разработки функционально-законченных нетактируемых изделий вплоть до уровня цифровых сигнальных процессоров (DSP — digital signal processor) [15, 16], сопроцессоров [17], самосинхронной машины потока данных DDM2 (MIT). Разработан широкий спектр нетактируемых микропроцессоров с архитектурами ARM (AMULET 1-3 [18], ARM996HS [19]), MIPS (MiniMIPS [20]), Intel (HT80C51 [21]), а также RISC-архитектур собственной разработки (ASPRO [22], TengYue-1 [23] и др.). Некоторые из перечисленных устройств выпускаются серийно. Относительно недавно начато производство первой в мире асинхронной ПЛИС фирмы Achronix [24]. Параллельно ведутся разработки асинхронных САПР, а в ведущих институтах Америки, Англии и Китая студентам читают образовательные курсы по асинхронным автоматам.

Однако эти реализации в действительности не являются НЗ-схемами. Они используют запростотное взаимодействие (handshake) между функциональными блоками вместо «дерева» тактового сигнала и за счет этого существенно сокращают потребление энергии и расширяют область работоспособности. Но они не содержат реального контроля окончания переходных процессов при переключениях схемы, присущего НЗ-схемам. Контроль заменен элементами задержки [19], эмулирующими реальные задержки обработки данных на соответствующих участках вычислительного тракта. Такой подход обеспечивает аппаратные затраты на уровне синхронных аналогов, но не исключает возможности появления сбоев в работе схемы при разбросе параметров транзисторов и элементов схемы, обусловленных технологическими и эксплуата-

ционными факторами, а следовательно, не может быть признан действительно нечувствительным к задержкам.

В России активным пропагандистом НЗ-подхода был коллектив под руководством д.т.н. В. И. Варшавского. В работах [1, 25–28], развивающих положения теории Маллера, убедительно доказана целесообразность внедрения принципа самосинхронизации в практику проектирования цифровых СБИС. С 1980-х гг. это направление проектирования аппаратуры активно поддерживается и развивается в ИПИ РАН [29–57].

Реализации НЗ-схем обладают рядом преимуществ по сравнению с синхронными аналогами [29]:

- устойчивая работа — отсутствие сбоев при любых возможных условиях эксплуатации;
- безопасная работа — прекращение всех переключений в момент появления отказа любого элемента (константной неисправности, при которой выход элемента «залипает» в одном состоянии);
- отсутствие периодов вынужденного простоя в ожидании очередного синхроимпульса.

Практические следствия этих преимуществ НЗ-схем:

- естественная устойчивость к параметрическим отказам, вызываемым изменением параметров элементов из-за процессов старения и неблагоприятных воздействий окружающей среды;
- естественная стопроцентная самопроверяемость и самодиагностируемость по отношению к множественным константным неисправностям;
- безопасность функционирования на основе бестестовой локализации неисправностей, т. е. прекращение работы в момент отказа элемента, исключающее выдачу недостоверной информации, с одновременной локализацией места события;
- максимально возможная область эксплуатации (диапазон работоспособности), определяемая только физическим сохранением переключаемых свойств активных элементов базиса реализации;
- максимально возможное в текущих условиях эксплуатации быстродействие;
- отсутствие накладных аппаратных и энергетических расходов, связанных с реализацией «клокового дерева» — разветвленной системы синхронизации, обеспечивающей строгую одновременность событий в разных местах проектируемой схемы.

Единственный недостаток НЗ-схем — большие аппаратные затраты. В зависимости от класса рассматриваемого цифрового устройства его НЗ статическая КМОП (комплементарная структура металл–оксид–полупроводник) реализация требует в 1,3–2,5 раза больше транзисторов, чем аналогичная синхронная реализация. Наихудшее соотношение аппаратных затрат наблюдается в комбинационных схемах из-за необходимости использования дуального представления каждой функции и добавления схемы контроля окончания всех переходных процессов при переключениях схемы.

В данной статье рассматривается возможность существенного улучшения требуемых характеристик (до 50%) высокопроизводительных вычислительных систем за счет перехода от синхронной схемотехники к самосинхронной. Возможность снижения потребления питания основывается на следующем:

- в современных высокопроизводительных микропроцессорах потребление дерева синхронизации составляет от 30% до 50% потребления всей схемы, в СС-схемах не используется дерево синхронизации;
- имеется значительный разброс параметров транзисторов на расстояниях свыше 1 мм для технологических норм 28 нм и ниже, что приводит к необходимости дополнительных технических решений, приводящих к повышению потребления питания в С-схемах, в НЗ-схемах разброс параметров учитывается автоматически в силу базовых НЗ решений;
- наибольшие производительности достигнуты в графических процессорах компаний Nvidia и AMD, где вычисления организуются как потоковые процессы; НЗ-логика наилучшим образом согласуется с потоковыми машинами, поскольку в них обоих используются состояния готовности; подобное свойство позволяет избежать дополнительных накладных расходов на организацию вычислений и снизить энергопотребление.

Ресурс повышения надежности в НЗ-схемах по сравнению с С-схемами обеспечивается их базовыми решениями, в которые закладывается дополнительная информация для их функционирования, используемая и для повышения надежности.

Схемы, нечувствительные к задержкам, органично вписываются в концепцию создания современных вычислительных систем, обеспечивая низкое энергопотребление и сохранение работоспособности в изменяющихся условиях эксплуатации оборудования. Данная работа посвящена срав-

нительному анализу вариантов реализации типовых представителей основных классов вычислительных устройств и обоснованию выбора схемотехнического базиса реализации микропроцессоров и коммуникационных сред супер-ЭВМ эксафлопсного класса.

2 Сравнительный анализ синхронных и нечувствительных к задержкам схем

С практической точки зрения наиболее интересным является подкласс схем, нечувствительных к задержкам в элементах (НЗЭ). В пределах эквифазной зоны [27] они обладают всеми свойствами и преимуществами НЗ-схем. Только при передаче информации отдаленному приемнику вне пределов эквифазной зоны необходимо предпринимать дополнительные меры по обеспечению нечувствительности к задержкам в соединительных проводах. При микронных нормах проектирования топологии микросхем эквифазная зона измерялась миллиметрами и практически покрывала всю площадь кристалла БИС, так как задержки переключения элементов превалировали над задержками распространения сигналов в соединительных проводах. Поэтому НЗЭ-схема в рамках одного кристалла БИС фактически являлась НЗ-схемой.

Однако с переходом к субмикронным нормам проектирования БИС размер эквифазной зоны существенно сократился из-за того, что задержки переключения элементов многократно уменьшились и стали сравнимы и даже меньше задержек распространения сигналов в проводах. В современных цифровых СБИС эквифазная зона покрывает лишь малую часть площади кристалла. Поэтому НЗЭ-схемы могут считаться НЗ-схемами, если связанные друг с другом функциональные блоки имеют соответствующие небольшие размеры и расположены относительно близко друг к другу.

Практически целесообразными являются также квази-НЗЭ (КНЗЭ) схемы. Основное отличие НЗЭ-схем от КНЗЭ-схем состоит в том, что НЗЭ-схемы контролируют с помощью индикаторной подсхемы окончание переключения каждого элемента в схеме, в то время как КНЗЭ-схемы имеют спекулятивную индикацию — обеспечивают индикацию только элементов, стоящих на критических путях обработки информации. За счет этого многоразрядные КНЗЭ-схемы оказываются более быстродействующими и менее сложными. Но они не

дают стопроцентной гарантии сохранения работоспособности схемы при изменении в широком диапазоне условий эксплуатации: напряжения питания, температуры — и при воздействии экстремальных факторов.

В работе [55] представлены результаты сравнительных испытаний С-, КНЗЭ- и НЗЭ-вариантов реализации цифровых устройств различной сложности:

- 4-разрядного микроядра [33], аналога ядра микроконтроллера PIC16 фирмы Microchip, США, включающего типовые арифметико-логические устройства: регистровую память, аппаратный умножитель, сдвигатель, счетчики;
- 8-разрядного отказоустойчивого последовательно-параллельного (ПП) порта, эмулирующего последовательный интерфейс между двумя цифровыми устройствами [34];
- 64-разрядного сопроцессора — устройства деления и извлечения квадратного корня [37, 38, 49, 51] в соответствии со стандартом IEEE754 [58].

Сравнение С-, КНЗЭ- и НЗЭ-вариантов реализации перечисленных цифровых устройств проводилось на основе оценки быстродействия в реальных условиях эксплуатации. Частота тактирования С-устройств устанавливалась из расчета на наилучший случай в пределах допустимой области эксплуатации. Быстродействие же НЗЭ-схем определялось реальными, а не наилучшими условиями эксплуатации. Именно поэтому НЗЭ-устройства в нормальных условиях оказываются, как правило, быстрее С-аналогов, что и было подтверждено результатами испытаний перечисленных выше вариантов цифровых устройств.

В табл. 1 приведены аппаратные затраты С- и НЗЭ-вариантов реализации микроядра и ПП-порта в вентилях базового матричного кристалла (БМК) серии 5503 (МИЭТ, Технологический центр). Микроядро НЗЭ, содержащее большую комбинационную схему — умножитель 4×4 , построенный по модифицированному алгоритму Бута, оказалось в 1,43 раза сложнее своего синхронного аналога. Благодаря заметному сокращению общего числа

Таблица 1 Аппаратные затраты

Цифровое устройство	С-вариант	НЗЭ-вариант
Микроядро	970	1390
ПП-порт	443	370

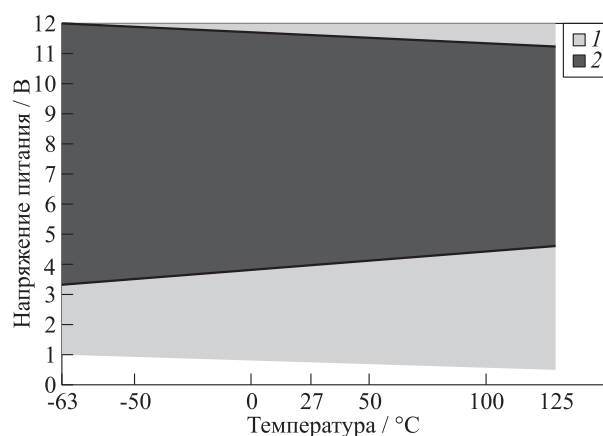


Рис. 1 Зона работоспособности НЗЭ- (1) и С-образцов (2) микроядра

устройств в составе отказоустойчивого НЗЭ-ППП-порта по сравнению с синхронным [55], его суммарные затраты оказались на 20% меньше, чем в С-варианте. Аналогичные характеристики могут быть получены и при реализации данных устройств в виде функциональных блоков заказной БИС.

На рис. 1 приведены результаты эксперимента по проверке работоспособности всех НЗЭ- и С-образцов в диапазоне изменяющихся напряжения питания и температуры при пороговых напряжениях транзисторов на уровне 0,8 В. Частота синхронизации С-образцов подбиралась для каждой пары значений «напряжение питания – температура». Из рис. 1 видно, что НЗЭ-образцы оказались работоспособными в более широком диапазоне условий эксплуатации, причем все без исключения. Синхронные же образцы продемонстрировали разброс параметров зоны работоспособности из-за флуктуации технологических параметров при изготовлении микросхем.

Производительность С-реализаций микроядра при фиксированной частоте тактирования, рассчитанной на наихудший случай, и выполнении смеси операций постоянна для всех возможных условий эксплуатации в пределах гарантированной области работоспособности и составила 4 MOPS. Быстродействие же НЗЭ-микроядра широко изменяется в зависимости от условий эксплуатации. Например, в зоне работоспособности, гарантированной изготовителем БМК, его производительность изменяется от 10,9 MOPS (5,5 В, –63 °С) до 5,2 MOPS (4,5 В, +125 °С). В среднем во всем реальном диапазоне работоспособности она оказалась выше производительности С-микроядра почти в 2 раза.

На рис. 2, а приведен график зависимости тока потребления (I_{cc}) С- и НЗЭ-вариантов реализации микроядра от величины напряжения питания при температуре $T = +27$ °С. При одном и том же напряжении питания НЗЭ-вариант потребляет несколько больше, чем синхронный, что объясняется его более высокой производительностью.

Для более корректной оценки сравнительного потребления энергии на рис. 2, б приведен график энергетической эффективности, который показывает ток потребления микроядра при выполнении операций с производительностью 1 MOPS. Чем меньше величина E , тем более эффективна реализация. Из рис. 2 видно, что НЗЭ-реализация микроядра более эффективна, чем его С-вариант. Например, при номинальном напряжении питания 5 В энергетическая эффективность составляет 1,2 мА/MOPS для НЗЭ-образца и 1,8 мА/MOPS для С-образца. При напряжении питания 12 В имеет место двукратное превосходство НЗЭ-реализаций. Сочетание возможности КМОП-НЗЭ-схем работать (и потреблять энергию) только «по требованию» с низким потреблением пассивной

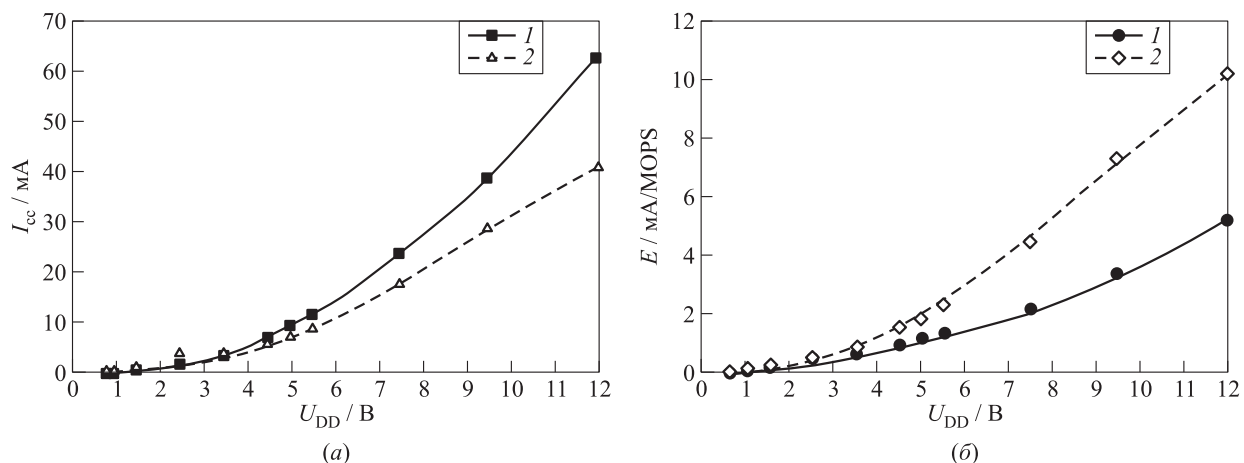


Рис. 2 Энергетические параметры НЗЭ- (1) и С-вариантов (2) микроядра при $T = 27$ °С

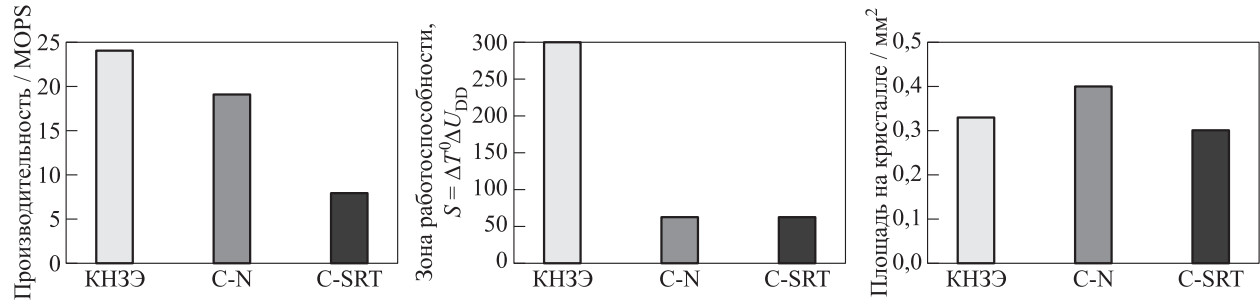


Рис. 3 Результаты испытаний вариантов сопроцессора

логики создает хорошие предпосылки для создания энергетически эффективных аппаратных решений.

Ценой повышения производительности и расширения зоны работоспособности НЗЭ-варианта микроядра является увеличение его аппаратных затрат. В качестве интегральной оценки эффективности (добротности) реализации цифрового устройства может служить отношение произведения производительности при номинальном питании на ширину зоны работоспособности к аппаратным затратам. Суммарное преимущество НЗЭ-микроядра в сравнении с синхронным аналогом по этому параметру с учетом гарантированной производителем области работоспособности С-устройств по напряжению питания (номинал $\pm 10\%$) составляет 17,6 раза.

В качестве способа построения отказоустойчивого НЗЭ-ПП-порта было выбрано дублирование его основной функциональной части — регистра сдвига, а в С-варианте — троирование регистра сдвига. Один из дубликатов изначально является рабочим, остальные — контрольными и/или резервными. Во всех случаях применяется постоянный контроль одинаковости результата, который получается независимо каждым устройством на основе общих входных данных и позволяет выявить возникшее несовпадение. «Ремонт» схемы состоит в мультиплексировании на выход заведомо исправного устройства. Такой способ обеспечивает оперативный саморемонт одного отказа в сдвиговом регистре ПП-порта и достоверность определения работоспособности всех частей схемы.

Результаты измерения показали [55], что отказоустойчивый НЗЭ-ПП-порт по сравнению с С-вариантом имеет существенно лучшие характеристики: в 2,4 раза по быстродействию; в 1,2 раза по аппаратным затратам; в 1,3 раза по энергетической эффективности; в 18 раз по добротности.

Сравнение вариантов сопроцессора проводилось на С-вариантах, реализующих алгоритмы Ньютона (C-N) и SRT Radix4 (C-SRT), и КНЗЭ-

варианте, реализующем алгоритм SRT Radix2. Все варианты сопроцессора были реализованы в составе тестовой микросхемы по стандартной 0,18-микронной КМОП-технологии с шестью слоями металлизации.

Сравнительные результаты испытаний вариантов сопроцессора показаны на рис. 3. Производительность измерялась при нормальных условиях работы (напряжение питания = 1,8 В, $T = 27^\circ\text{C}$). Зона работоспособности определялась как произведение диапазонов напряжения питания и температуры, в которых сопроцессор демонстрировал устойчивую работу.

Таким образом, реализация микроядра, ПП-порта и сопроцессора в виде НЗЭ- или КНЗЭ-устройства обеспечивает их устойчивую работу при любых допустимых условиях эксплуатации. Устройства НЗЭ и КНЗЭ экспериментально подтвердили свои неоспоримые преимущества по производительности и зоне работоспособности по сравнению с синхронными аналогами, поэтому такой схемотехнический базис целесообразно использовать и для разработки современных вычислительных устройств и комплексов.

3 Варианты методологии проектирования схем, нечувствительных к задержкам

В работе [13] рассмотрено 10 различных методологий проектирования асинхронных и, в частности, СС-схем. Их можно разделить на две группы:

- (I) методологии, базирующиеся на модели с ограниченной задержкой, например схемы Haffman'a и микроконвейеры. Некоторые из них предназначены для проектирования НЗ-схем (например, I-net), однако объединение таких схем требует использования либо линий задержки в цепях обратной

связи, либо системы локальной синхронизации. Подобно С-схемам, они вынуждены ориентироваться на наихудший случай условий работы схемы. Такие схемотехнические решения фактически являются квазисамосинхронными. Большинство наиболее известных зарубежных СС-микросхем и реализованных проектов относится именно к этому классу [15–24, 59–61];

- (II) подходы, базирующиеся на модели элементов и соединительных проводов с неограниченной задержкой до точки разветвления. При этом предполагается, что разница в задержке проводов после разветвления меньше, чем минимальная задержка элемента. Примеры таких методологий: графы сигнальных переходов (STG — signal transition graph), диаграммы изменений (CD — change diagrams) и трансляция процессов связи Martin'a [7, 9]. При необходимости эти методологии могут быть расширены для разработки схем, не удовлетворяющих требованию изохронности ветвления, путем введения индикаторов переходных процессов в непосредственной близости к приемнику сигнала.

Методология, разрабатываемая в ИПИ РАН, относится ко второй группе. Она имеет следующие особенности:

- на концептуальном уровне она базируется на теории Маллера [3, 4]. Правильная работа построенных по данной методологии схем не зависит от *задержек составляющих их элементов* (задержка любого элемента схемы, например инвертора, может быть любой, но конечной величины);
- на схемотехническом уровне использование дополнительных логических и топологических приемов позволяет обеспечить правильную работу СС-схем независимо *от задержек в соединительных проводах*;
- на уровне взаимодействия с внешней средой и другими СС-схемами используется асинхронный (запрос-ответный) принцип с фиксацией действительного окончания любого инициированного переходного процесса. Генераторы могут использоваться только для сугубо второстепенных целей, например для создания внутренних таймеров.

Перечисленные выше свойства СС-схем определяют высокую эффективность создания надежных изделий на их основе, в том числе и отказоустойчивых. Однако в полной мере данными

свойствами обладают только НЗ-схемы. Данное свойство (нечувствительность) относится к задержкам распространения сигналов через логические элементы и по соединительным проводам.

Типичными представителями НЗ-схем среди зарубежных СС-устройств являются NCL-схемы [14, 62–68]. Методология NCL была предложена в 1990-х гг. [14]. В настоящее время она развивается в основном усилиями компании Theseus Research, Inc. (TR) и университета в Арканзасе (University of Arkansas).

Ниже представлен сравнительный анализ основных принципов проектирования НЗ-схем, разрабатываемых в соответствии с методологиями ИПИ РАН и NCL. В дальнейшем будем называть первые просто НЗ-схемами, а последние — NCL-схемами.

3.1 Дисциплина сигналов

Схемотехника NCL основана на парафазном с нулевым спейсером кодировании всех информационных сигналов. Любая функция, выполняемая устройствами внутри схемы, реализуется путем ее дуального представления. Состояние каждого информационного сигнала A представляется комбинацией двух компонентов $\{A, AB\}$. В спейсерной фазе (фазе NULL) $A = AB = 0$, в рабочей фазе (фазе DATA) $\{A = 0, AB = 1\}$ или $\{A = 1, AB = 0\}$. Состояние $A = AB = 1$ запрещено и при нормальной работе схемы никогда не формируется. Переключение информационного сигнала из текущего рабочего состояния в следующее всегда происходит через спейсерное состояние, даже если текущее и следующее рабочие состояния совпадают.

Дисциплина информационных сигналов в методологии проектирования НЗ-схем более гибкая. Она включает использование как парафазных сигналов со спейсером, аналогичных сигналам NCL-схем, так и других сигналов: парафазных без спейсера, бифазных (выходов бистабильной ячейки), унарных, управляющих (аналогов тактовых сигналов в С-схемотехнике), мультифазных (например, сигналов выборки мультиплексоров). Это позволяет строить более компактные схемы.

Парафазные сигналы со спейсером своим значением отображают собственную фазу и закодированный бит данных. Остальные информационные сигналы своим статическим значением отображают только заложенный в них бит информации. Поэтому для фазового регулирования парафазные без спейсера, унарные и бифазные информационные сигналы в обязательном порядке сопровождаются сигналом управления. Сигнал управления переходит в рабочую фазу только после того, как

сопровожаемый им информационный сигнал переключился в следующее рабочее состояние, тем самым давая знать приемнику информационного сигнала, что соответствующий информационный вход можно использовать. Переключение сигнала управления в спейсер инициируется приемниками сопровождаемого им информационного сигнала и является подтверждением факта «доставки по назначению» соответствующего информационного сигнала. Информационный сигнал может изменять свое состояние только во время спейсера сопровождающего его сигнала управления.

3.2 Схмотехнический базис

Успех проектирования цифровых схем любого типа не в последнюю очередь определяется составом библиотеки элементов, на основе которой ведется проектирование. Качество СС-схем зависит от этого даже в большей степени, чем качество С-схем.

Схмотехника NCL-схем основана на использовании функциональных схем, показанных на рис. 4. Разработчики NCL-схем называют элементы схмотехнического базиса *пороговыми элементами* (threshold gates). Термин «пороговые» в данном случае относится не к потенциальному уровню сигналов на входе и выходе элемента, а к количеству входов, которые должны переключиться из спейсера (NULL) в рабочее состояние (DATA), чтобы выход элемента тоже переключился в рабочее состояние. Обратное переключение в спейсер возможно только тогда, когда все входы перейдут в спейсер.

Статическая реализация (рис. 4, а) состоит из подсхем на КМОП-транзисторах *n*- и *p*-типа прово-

димости, обеспечивающих переключение элемента в обе фазы работы и хранение текущего состояния до момента появления комбинации входов, вызывающей переключение в противоположное состояние. При этом функции, реализуемые частями «Переход в NULL» и «Переход в DATA» и обеспечивающие соединение выхода элемента с источниками активных уровней (питание и «земля»), ортогональны: при любой комбинации входов на выход элемента коммутируется только один источник логического уровня.

Полустатическая (semistatic) реализация (рис. 4, б) использует слабый инвертор в качестве обратной связи для организации хранения состояния NULL или DATA. Она обладает меньшей помехоустойчивостью и характеризуется протеканием сквозного тока при переключении NCL-схемы из текущего состояния в противоположное. В дальнейшем будем считать, что для реализации NCL-схемы используется статический вариант, показанный на рис. 4, а.

Статическая реализация NCL-схемы фактически является гистерезисным триггером [27] со сложной функциональной начинкой. В методологии проектирования СС-устройств ИПИ РАН гистерезисный триггер (Г-триггер) играет роль индикаторного элемента. Нагрузка его дополнительными функциями позволяет в ряде случаев уменьшить сложность реализации схемы, но является скорее исключением из правил, нежели типовым приемом проектирования СС-схем.

Схмотехнический базис NCL-схем состоит из фундаментальных элементов ТН_{mn}. Имя элемента обозначает «пороговый элемент с *n* входами и порогом *m*». Они воплощают в себе один из принципов проектирования NCL-схем: «реализация любого

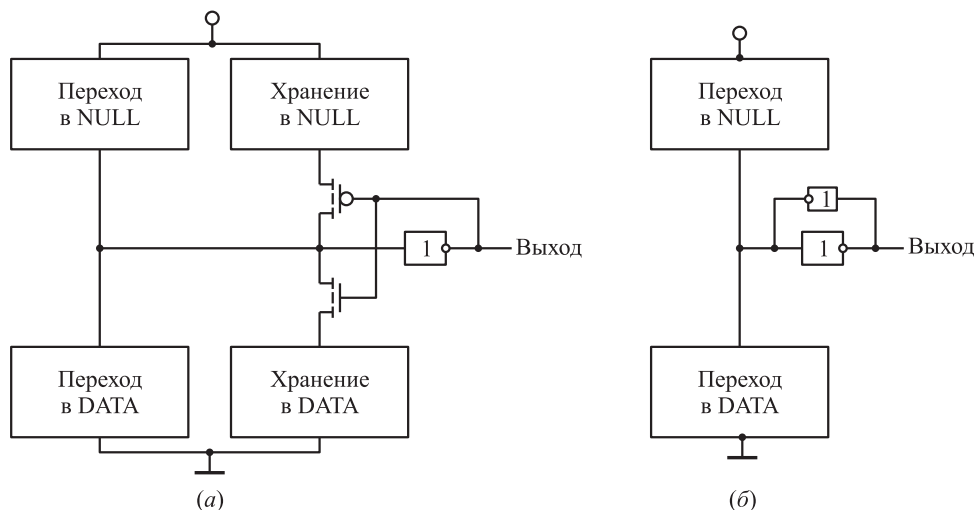


Рис. 4 Функциональная схема NCL-элемента: статическая (а) и полустатическая (б)

NCL-элемента и любой NCL-схемы должна обеспечивать переключение выходов в спейсер (NULL) только после того, как все входы перешли в спейсер, и переключение выходов в рабочее состояние (DATA) только тогда, когда не меньше порогового числа входов у элемента или схемы перешли в рабочее состояние». Некоторые элементы имеют дополнение к имени в виде окончания «wXXX». Цифры после буквы «w» показывают вес соответствующей по порядку входной переменной. Если порог элемента равен m , а вход A имеет вес k , то для переключения элемента в фазу достаточно переключения входа A и еще каких-нибудь $(m-k)$ входов, если вес остальных входов единичный, или переключения любых m входов, кроме входа A . Общее число фундаментальных элементов равно 27. Они перечислены в табл. 2 [65]. Там же указаны выполняемые ими функции и количество КМОП-транзисторов, необходимых для их реализации.

Функции, показанные в табл. 2, описывают блок «Переход в DATA» функциональной схемы на рис. 4, а. Условное графическое обозначение (УГО) фундаментальных NCL-элементов показано на рис. 5.

Таблица 2 Состав библиотеки NCL-элементов

Имя элемента	Выполняемая функция	Число транзисторов
ТН12	$A + B$	6
ТН22	AB	12
ТН13	$A + B + C$	8
ТН23	$AB + AC + BC$	18
ТН33	ABC	16
ТН23w2	$A + BC$	14
ТН33w2	$AB + AC$	14
ТН14	$A + B + C + D$	10
ТН24	$AB + AC + AD + BC + BD + CD$	26
ТН34	$ABC + ABD + ACD + BCD$	24
ТН44	$ABCD$	20
ТН24w2	$A + BC + BD + CD$	20
ТН34w2	$AB + AC + AD + BCD$	22
ТН44w2	$ABC + ABD + ACD$	23
ТН34w3	$A + BCD$	18
ТН44w3	$AB + AC + AD$	16
ТН24w22	$A + B + CD$	16
ТН34w22	$AB + AC + AD + BC + BD$	22
ТН44w22	$AB + ACD + BCD$	22
ТН54w22	$ABC + ABD$	18
ТН34w32	$A + BC + BD$	17
ТН54w32	$AB + ACD$	20
ТН44w322	$AB + AC + AD + BC$	20
ТН54w322	$AB + AC + BCD$	21
ТНxor0	$AB + CD$	20
ТНand0	$AB + BC + AD$	19
ТН24comp	$AC + BC + AD + BD$	18

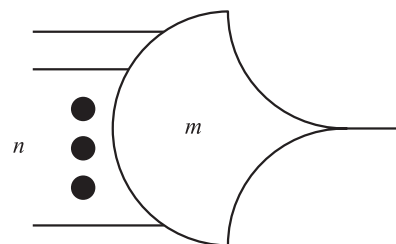


Рис. 5 Условное графическое обозначение элемента ТНmn

Фундаментальные элементы NCL-базиса служат основой для синтеза производных от них элементов, необходимых для проектирования практических цифровых устройств, например триггеров с асинхронным сбросом или установкой.

Схемотехника НЗ-схем основана на использовании классических КМОП принципиальных схем. Необходимость индцировать окончание переходных процессов в элементах НЗ-схемы делает нежелательной сильную функциональную декомпозицию. Она приводит к появлению в схеме множества «мелких» логических элементов, каждый из которых требует дополнительных аппаратных затрат для реализации его индикации. Это делает целесообразной разработку библиотеки с широкой номенклатурой логических, триггерных и индикаторных элементов. При этом функциональный состав библиотеки определяется, в первую очередь, целесообразностью использования тех или иных элементов в НЗ-схемах, а во вторую очередь, технологическим базисом реализации проектируемой БИС.

При проектировании НЗ-схем используется библиотека элементов, содержащая 260 единиц [69–71] и являющаяся самосинхронным дополнением типовых библиотек стандартных элементов. Библиотека включает логические элементы для формирования комбинационных схем, мультиплексоры, триггеры (D, RS, счетные), сумматоры, преобразователи сигналов. Условные графические обозначения элементов схемотехнического базиса НЗ-схем — традиционные для С-схемотехники, и лишь отсутствующие в последней элементы имеют характерные графические обозначения. В качестве примера на рис. 6 показаны УГО трехвходового Г-триггера (G13) и индикаторного элемента G0B3I2, индцирующего входы и состояние двух связанных друг с другом бистабильных ячеек.

Элементы, аналогичные ТН22, используются в качестве индикаторных элементов, обеспечивающих контроль окончания переключений в НЗ-схеме.

Ряд триггеров разработанной библиотеки имеет уникальные свойства, обеспечивающие эффектив-

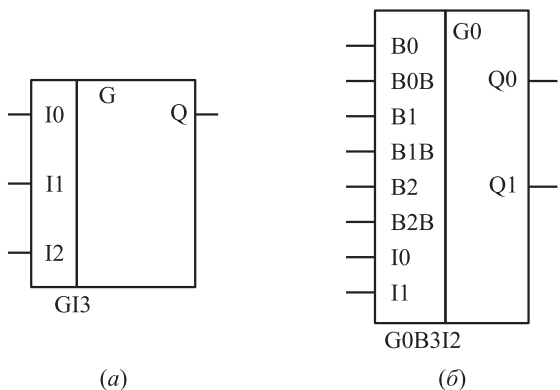


Рис. 6 Условные графические обозначения Г-триггера GI3 (а) и индикаторного элемента G0B3I2 (б)

ное решение корректными средствами двух проблем:

- (1) большой нагрузочной способности выходов триггеров;
- (2) упрощенной реализации входного интерфейса с С-формирователями входных данных.

Использование таких триггеров в составе НЗ-схем гарантирует сохранение их свойств в полном объеме при приемлемых потребительских характеристиках: быстродействию и сложности реализации. К числу таких триггеров относятся, в первую очередь, триггер с мощными инверторами на инфор-

мационных выходах [72] и триггер с единичным непарафазным информационным входом [48].

Библиотека элементов для проектирования НЗ-схем включена в состав САПР «Ковчег» (Технологический центр, МИЭТ) и позволяет разрабатывать НЗ-схемы различной степени сложности с использованием БМК серий 5503, 5507, 5508, 5509.

Разработанная библиотека внедрена также в САПР фирмы Cadence для КМОП-технологии 0,18 мкм [73] и 65 нм (для проектирования заказных БИС). Она включает схематехнические, топологические представления, а также Verilog и VHDL-модели. Топология элементов для проектирования заказных БИС может быть отмасштабирована до уровня 45 нм. Для их характеристики использовались стандартные программные средства фирмы Cadence и разработанный в ИПИ РАН программный комплекс СТЕРХ [74]. Разработанная библиотека успешно прошла апробацию на ряде полужаказных и заказных БИС.

3.3 Принципы построения схемы

В NCL-методологии любая схема строится в виде конвейера (рис. 7). Элементы TH22 на входах каждого комбинационного блока формируют парафазный код на основе парафазного выхода предыдущей ступени конвейера и сигнала разрешения, выдаваемого индикатором следующей ступ-

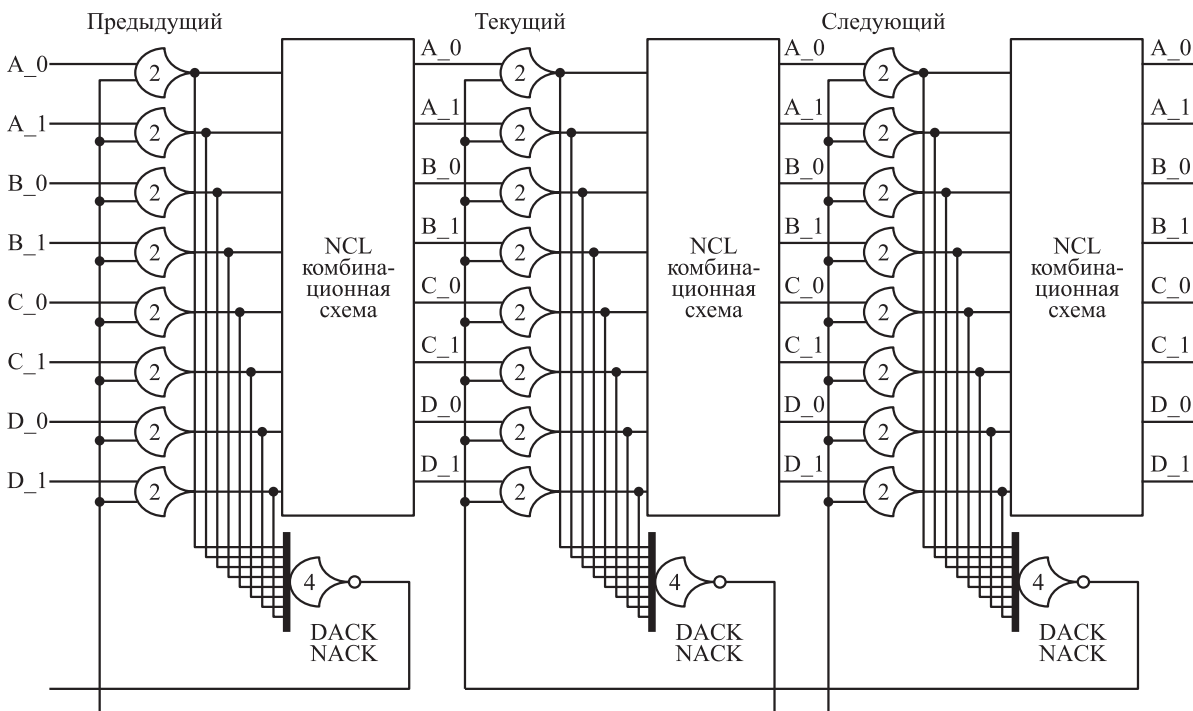


Рис. 7 Конвейер NCL

ни. Схема индикации, функционально эквивалентная элементу ТН48, формирует фазовый сигнал, обеспечивающий взаимодействие соседних ступеней конвейера.

Рисунок 8 демонстрирует реализацию последовательной NCL-схемы. Здесь регистры присутствуют и на входе схемы, и на ее выходе. Кроме того, используется регистр обратной связи, обеспечивающий корректное хранение состояния переменных памяти в схеме.

Спейсер NULL одинаков для всех элементов и устройств NCL-схемы: нулевое значение обеих составляющих каждого парафазного сигнала. Все элементы схемы, кроме элементов индикаторной подсхемы, имеют неинверсные выходы. С одной стороны, это упрощает согласование соседних устройств в тракте обработки данных; с другой стороны, создает дополнительную задержку, которая не всегда оправдана.

В N3-методологии элементы библиотеки объединяются в более сложные N3-схемы в соответствии с дисциплиной формирования и согла-

сования информационных, управляющих и индикаторных сигналов N3-схем:

- информационные сигналы формируются с использованием одного из видов СС-кодирования (парафазного со спейсером, бифазного с управляющим сигналом и т. д.). Число вариантов кодирования в пределах одной N3-схемы не ограничено;
- каждый рабочий набор кодированного сигнала в обязательном порядке чередуется со специальным самосинхронным промежуточным кодом — спейсером. Тип спейсера (нулевой или единичный) может быть произвольным, управляющие сигналы схемы и ее окружения подчиняются запросу-ответной дисциплине;
- все информационные и управляющие сигналы схемы должны индексироваться на ее выходах, т. е. любое переключение каждого сигнала должно в конечном итоге приводить к переключению одного или нескольких выходов схемы.

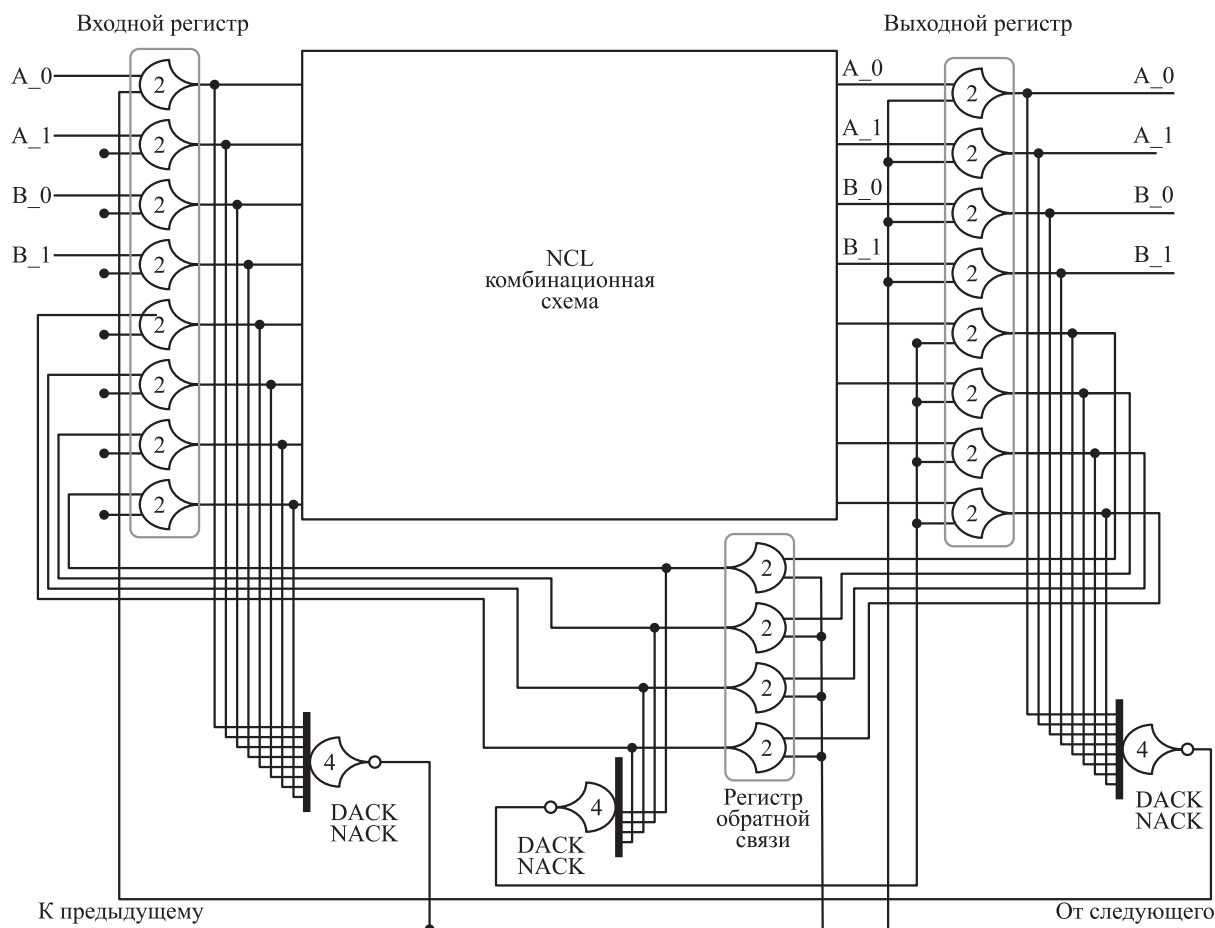


Рис. 8 Последовательная NCL-схема

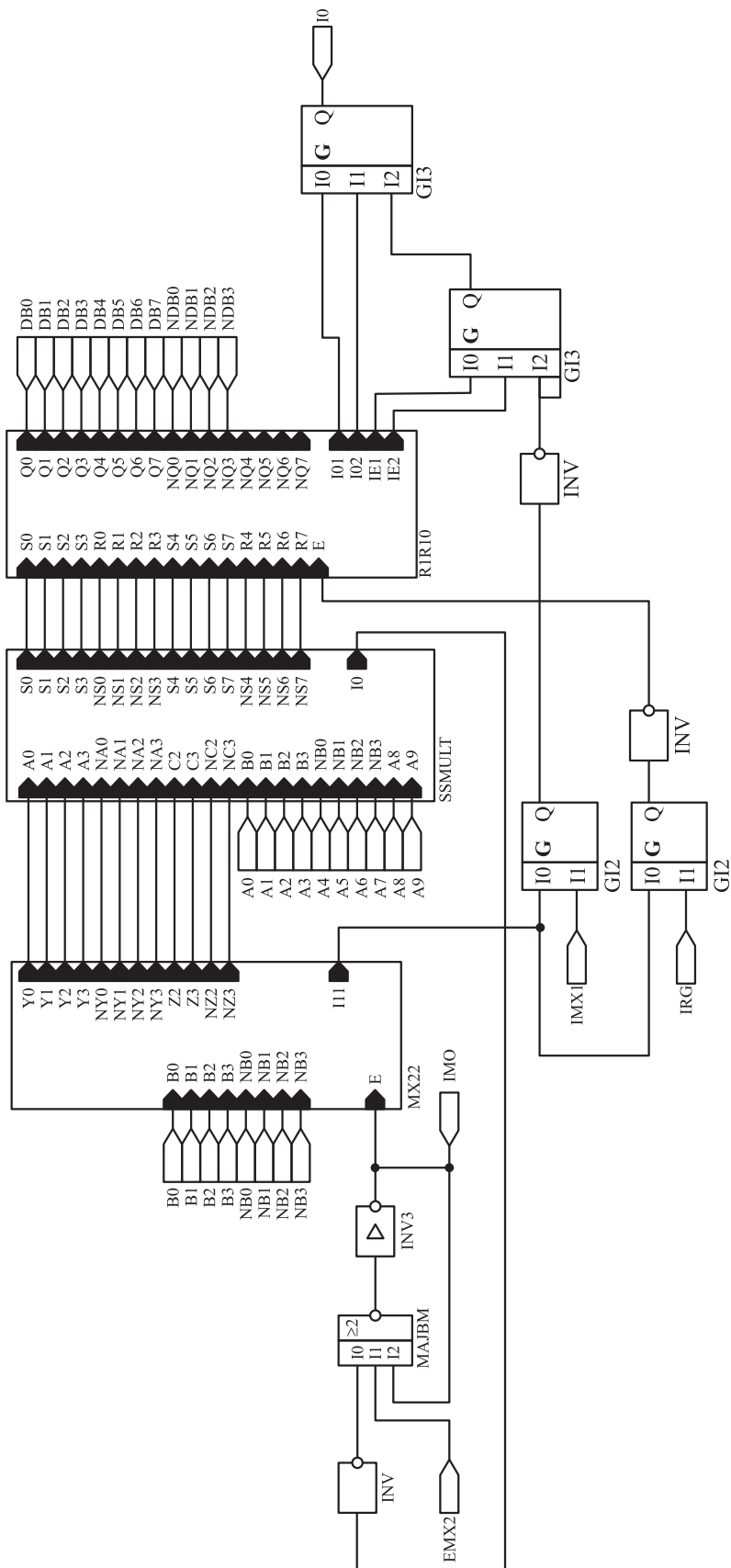


Рис. 9 Умножитель $НЗ 4 \times 4$

На рис. 9 показана реализация умножителя 4×4 [55] в виде НЗ-схемы. Здесь регистр R1R10 используется только на выходе схемы. Основной блок SSMULT является комбинационной схемой, так же как и схема MX22 преобразования бифазного сигнала $\{B[3..0], NB[3..0]\}$ в парафазный. Остальные элементы схемы обеспечивают индцирование схемы умножителя и запрос-ответное взаимодействие блоков умножителя с его окружением.

Использование разных типов кодирования сигналов и произвольного типа спейсера позволяют в большинстве случаев получить менее сложную реализацию НЗ-схемы с более высоким быстродействием.

3.4 Сложность реализации

Сложность реализации NCL-схемы определяется характером индцируемости (наблюдаемости —

observability) и ограниченной номенклатурой схемотехнического базиса. На рис. 10 показаны две реализации схемы, выполняющей функцию $Z = X \oplus Y$ («неравнозначность»): (а) без индцируемости входов на выходе и (б) с индцируемостью [65]. С точки зрения булевой алгебры обе эти реализации оказываются избыточными даже с учетом парафазного кодирования входов и выходов. Сложность их составляет 68 (см. рис. 10, а) и 76 (см. рис. 10, б) КМОП-транзисторов в соответствии с табл. 2.

Для сравнения на рис. 11 приведена функциональная схема НЗ-элемента, идентичного по выполняемым функциям и степени индцируемости входов и выходов схеме на рис. 10, а, и его реализация на КМОП-транзисторах.

Как видно из рис. 11, б, сложность адекватного НЗ-элемента составляет всего лишь 16 транзисторов. Функциональная схема НЗ-элемента, аналогичного схеме на рис. 10, б, показана на рис. 12.

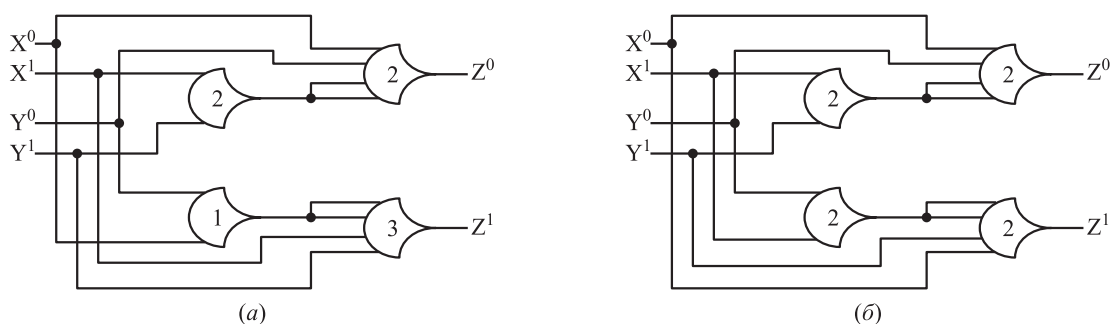


Рис. 10 Схема NCL-элемента «неравнозначность»: без индцируемости входов на выходах (а) и с индцируемостью входов (б)

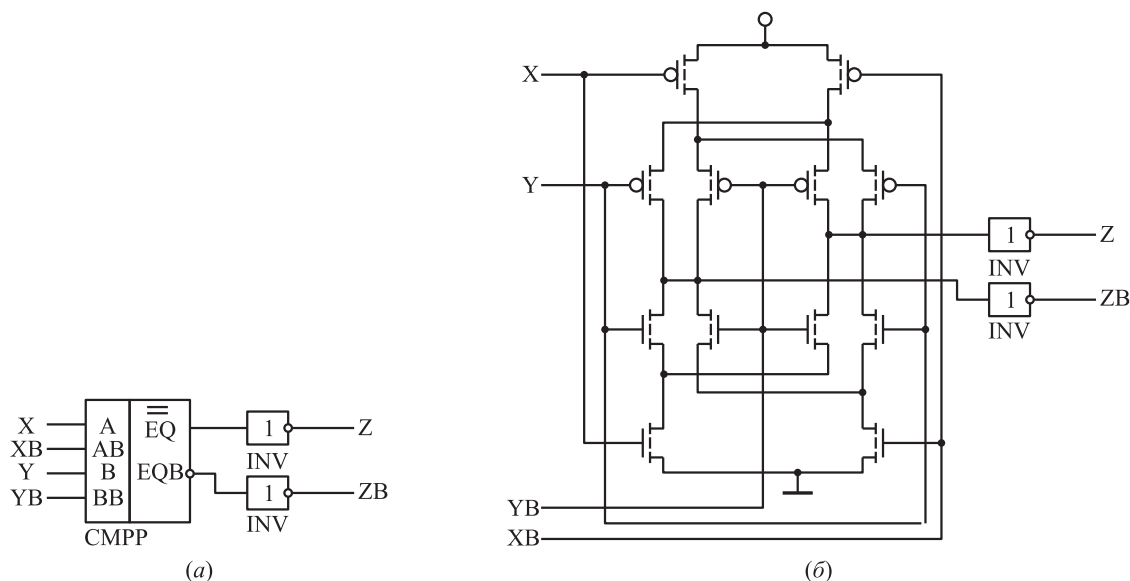


Рис. 11 Функциональная (а) и принципиальная (б) НЗ-схемы «неравнозначность» без индикации входов и выходов

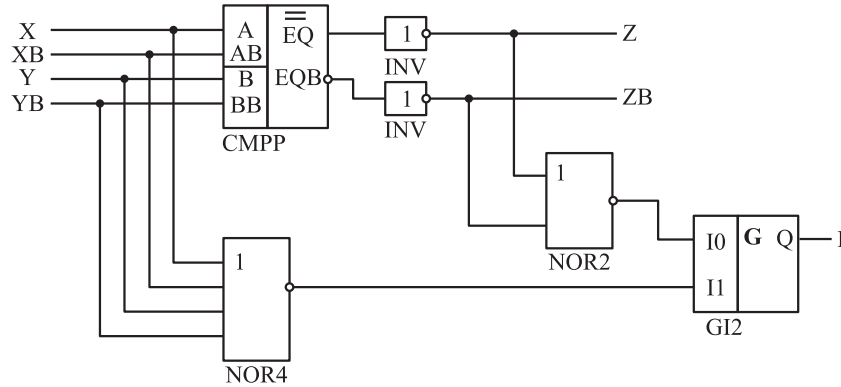


Рис. 12 Схема НЗ «неравнозначность» с индикацией входов и выходов

Сложность ее реализации равна 40 КМОП-транзисторам. Но по сравнению со схемой рис. 10, б она формирует дополнительный выход I, индицирующий все ее входы и выходы. Если схему на рис. 10, б дополнить аналогичной индикацией (элемент TH12), то ее сложность возрастет до 82 транзисторов.

Схемы на рис. 11 и 12 имеют парафазные входы и выходы. Спейсеры входов и выходов совпадают. Спейсер входов и выходов схемы на рис. 11 может быть любым: единичным или нулевым. Спейсер сигналов в схеме на рис. 12 может быть только нулевым. Это определяется элементами NOR4 и NOR2, индицирующими входы и выходы схемы. Если эти элементы заменить элементами NAND4 и NAND2 соответственно, то тип спейсера станет единичным.

Таким образом, НЗ-реализация элемента «неравнозначность» оказывается проще NCL-реализации в 2,05–4,25 раза в зависимости от степени индицируемости входов и выходов.

На рис. 13 представлена оптимизированная функциональная NCL-схема одноразрядного сумматора [65]. Она реализуется схемой из 80 транзисторов, в то время как НЗ-схема однораз-

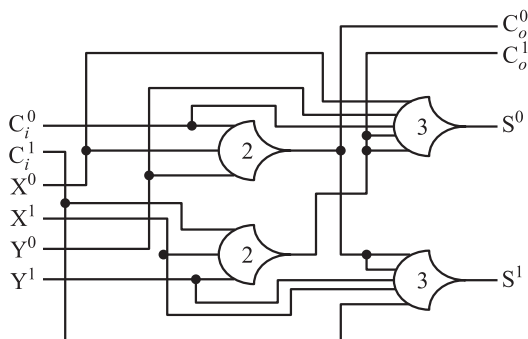


Рис. 13 Функциональная NCL-схема одноразрядного сумматора

рядного полного сумматора (рис. 14) реализуется на 40 транзисторах, что в 2 раза меньше, чем в NCL-схеме сумматора.

Следует отметить, что НЗ-реализация сумматора на рис. 14 индицирует на своих выходах только рабочие состояния входных сигналов. Практика проектирования показала, что этого в ряде случаев достаточно для обеспечения нечувствительности к задержкам схемы с сумматором, так как зачастую входы сумматора в полном объеме или в фазе спейсера индицируются в другом месте.

На рис. 15 показаны соответственно NCL- и НЗ-реализации полного сумматора, имеющие выход I, индицирующий все входы и выходы схемы в полном объеме. Их сравнение показывает, что NCL-реализация содержит 100 транзисторов против 84 у НЗ-реализации.

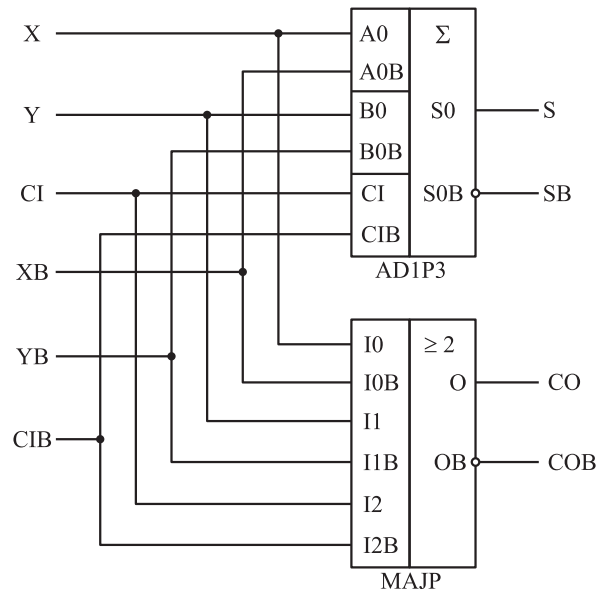


Рис. 14 Функциональная НЗ-схема одноразрядного сумматора

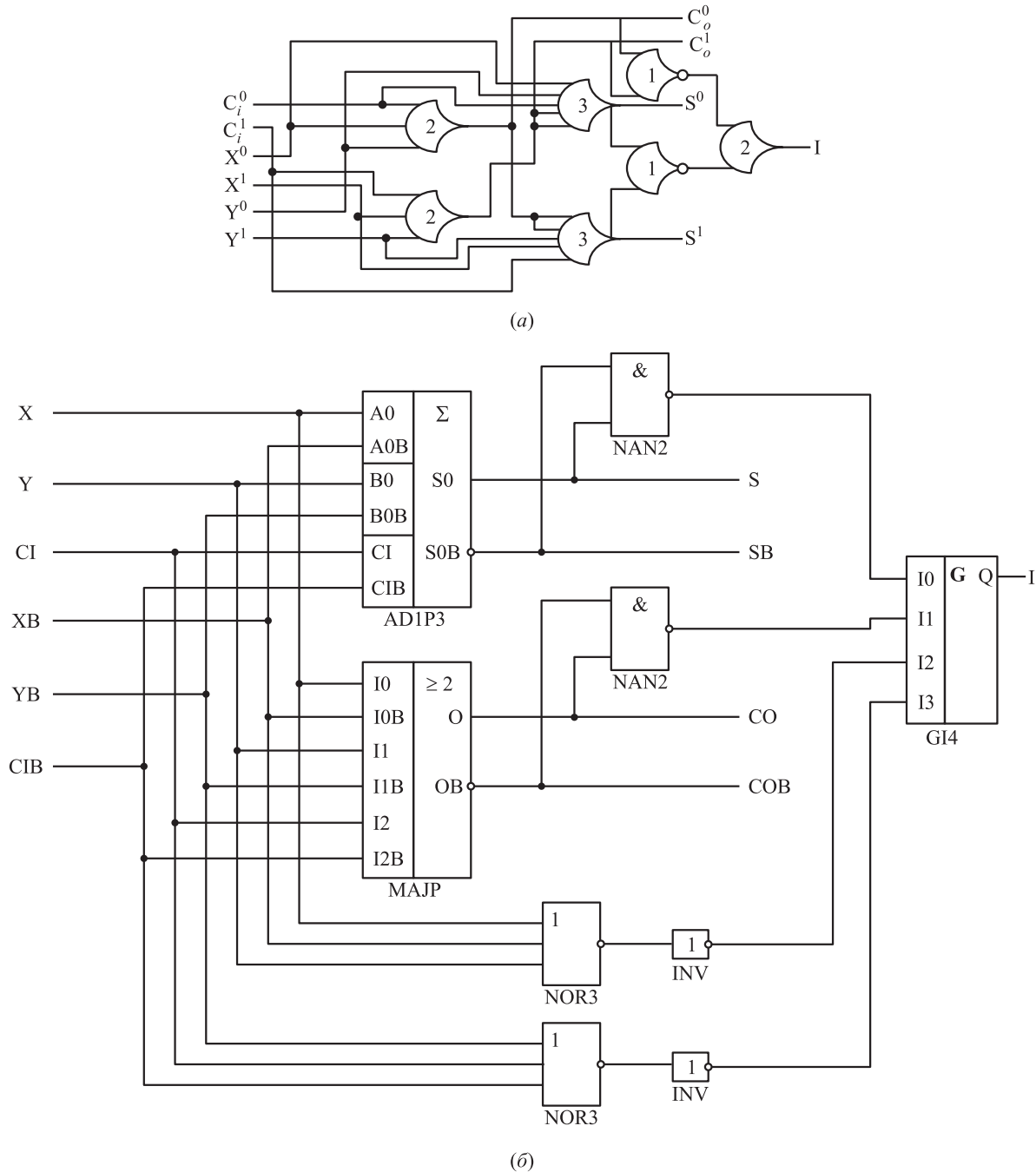


Рис. 15 Схемы сумматора с полной индикацией: (а) NCL; (б) N3

Таким образом, реализация полного однорядного сумматора в виде N3-схемы оказывается проще NCL-аналога в 1,19–2 раза в зависимости от степени индицируемости входов и выходов в самой схеме.

Комбинационные схемы в NCL-базисе наиболее избыточны в случае, если не требуется индицировать входы на выходах элемента. На практике в комбинационных схемах все или часть входных сигналов служат входами сразу нескольких элемен-

тов (разветвляются). Поэтому гораздо эффективнее индицировать их именно как парафазные входы схемы, а не как входы отдельных элементов. Это дает неоспоримое преимущество N3-схемам, поскольку их элементный базис не избыточен в этом отношении.

Рассмотрим NCL- и N3-реализации последовательных схем. Однотактный триггер в NCL-схемах реализуется на двух 2-входовых С-элементах со сбросом и элементе 2ИЛИ-НЕ в качестве индикато-

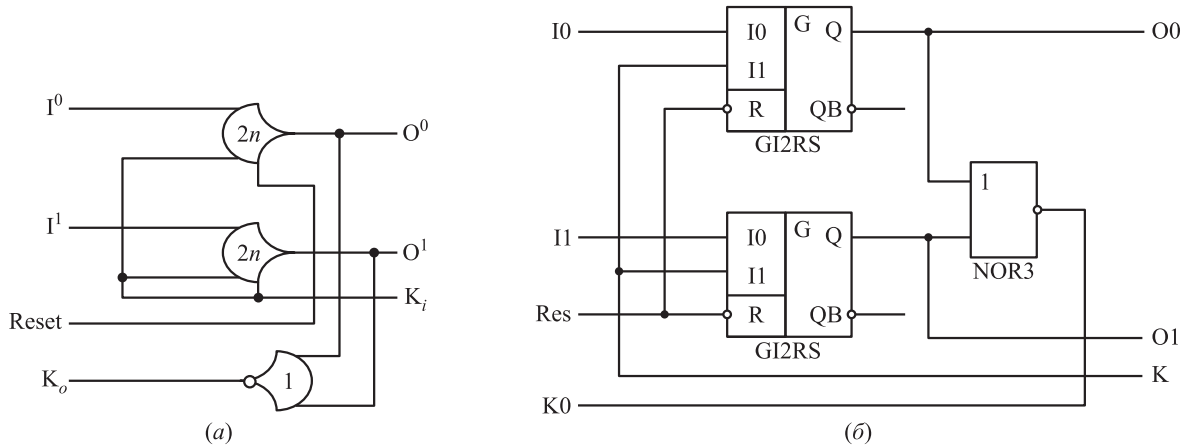


Рис. 16 Функциональная схема однотактного NCL-триггера (а) и H3-триггера с парафазными входами и выходами (б)

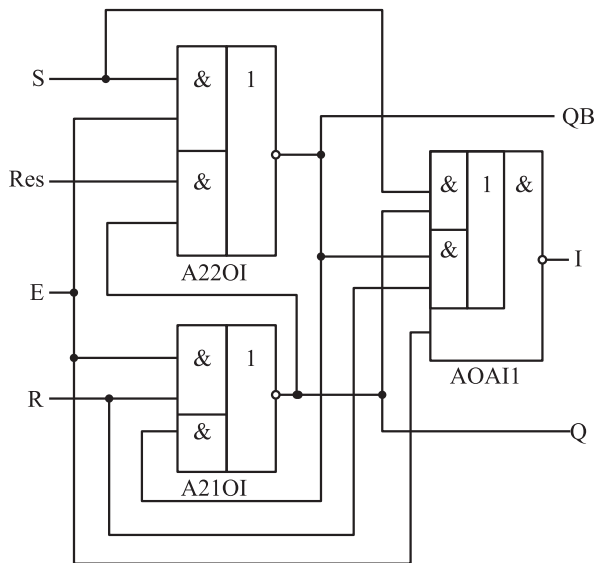


Рис. 17 Функциональная схема однотактного H3-триггера с бифазными входами и выходами

ра, всего 32 транзистора в статическом исполнении (рис. 16, а). Однотактный триггер H3 с аналогичными свойствами показан на рис. 16, б. Его реализация содержит 32 транзистора, как и NCL-аналог.

Однако в H3-схемах обычно используется другой принцип хранения и передачи информации между соседними устройствами: бифазные или унарные информационные сигналы с сопровождающим их сигналом управления. Поэтому однотактный триггер имеет другую схемотехническую реализацию, представленную на рис. 17. У него бифазные информационные входы (R, S), вход асинхронного сброса (Res) и вход управления (E), бифазный информационный выход (Q, QB) и индикаторный выход (I). Такая реализация однотакт-

ного триггера требует 24 транзистора — в 1,33 раза меньше, чем NCL-реализация.

Двухтактный NCL-триггер реализуется на двух однотактных триггерах. И в этом случае H3-реализация оказывается в 1,33 раза проще по числу транзисторов. Следовательно, регистры хранения и сдвига, реализуемые на однотактных и двухтактных триггерах, в H3-исполнении будут примерно на треть проще, чем в NCL-исполнении.

Рисунок 18 демонстрирует NCL-схему двоичного счетчика [75, рис. 36]. Она вынужденно включает в себя комбинационную схему увеличения текущего состояния счетчика на «1» и три регистра на однотактных триггерах с асинхронным сбросом.

На рис. 19 показана оптимизированная NCL-реализация комбинационной части счетчика. Отсутствие счетных триггеров в составе библиотеки NCL-элементов приводит к существенным аппаратным затратам при реализации счетчика. Аналогичный четырехразрядный двоичный H3-счетчик показан на рис. 20. В нем используется счетный триггер C0R [69], функциональная схема которого показана на рис. 21. Сравнительный анализ схем NCL- и H3-счетчиков показывает, что по числу транзисторов, требующихся для реализации счетчика, H3-вариант проще NCL-варианта в 4,49 раза (134 транзистора против 602), так как использование триггера C0R в составе счетчика исключило необходимость применения регистров для накопления и хранения результата. Следовательно, и по энергопотреблению он будет намного эффективнее.

Сравнение реализаций аппаратного однотактного умножителя 4×4 без знака в NCL [75, рис. 59] и H3 [55] базисах также подтверждает преимущество H3-варианта, сложность которого составляет

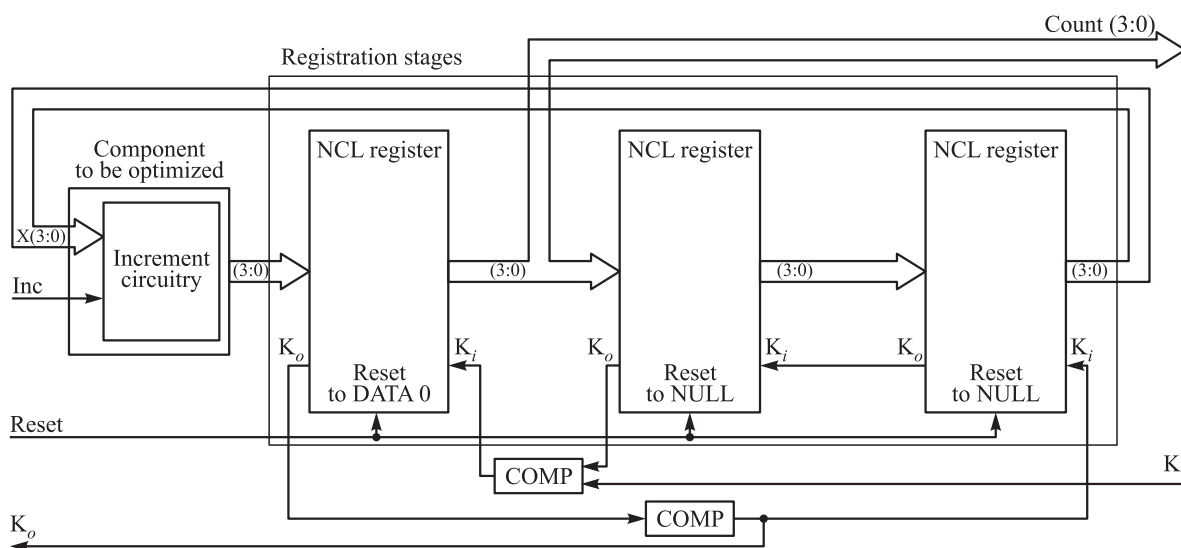


Рис. 18 Функциональная NCL-схема четырехразрядного счетчика

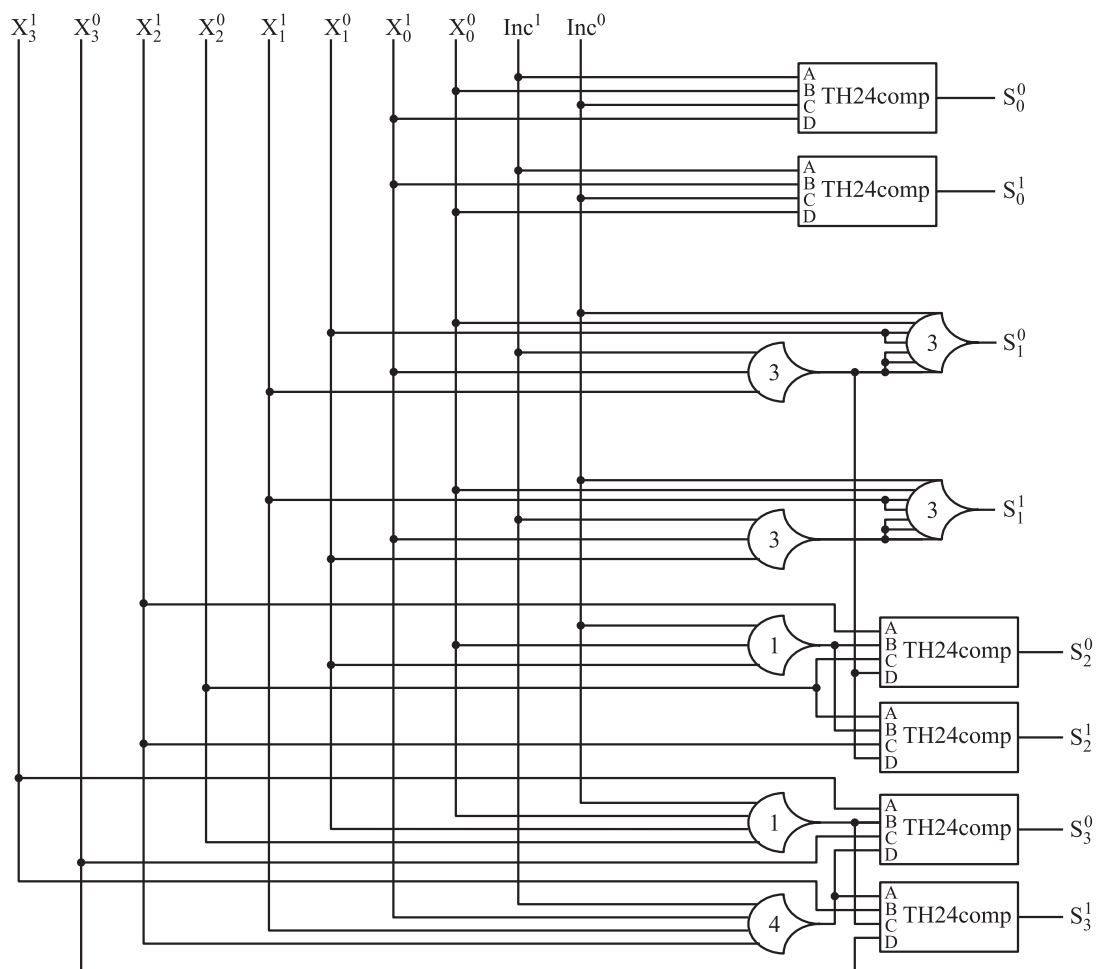


Рис. 19 Комбинационная часть NCL-счетчика (блок Increment circuitry)

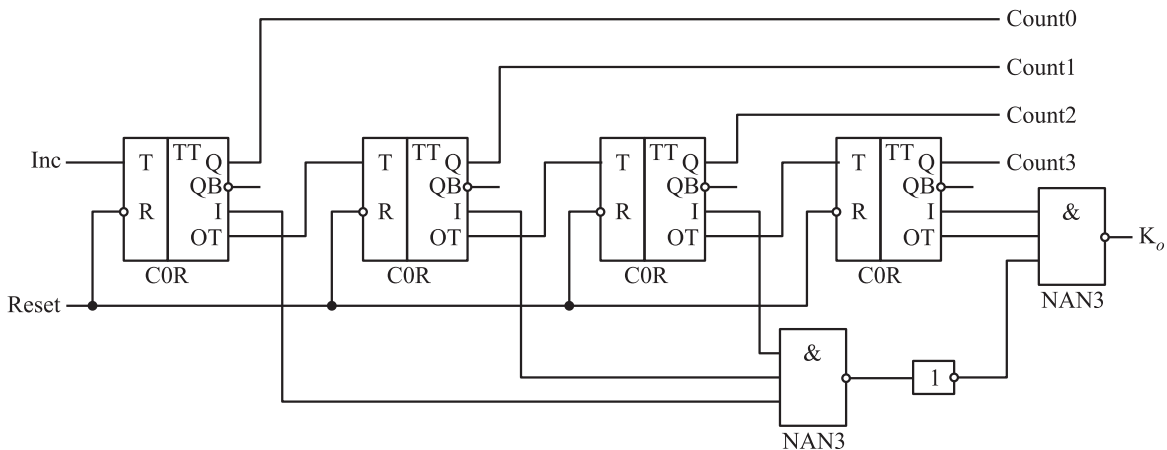


Рис. 20 Функциональная схема НЗ-счетчика

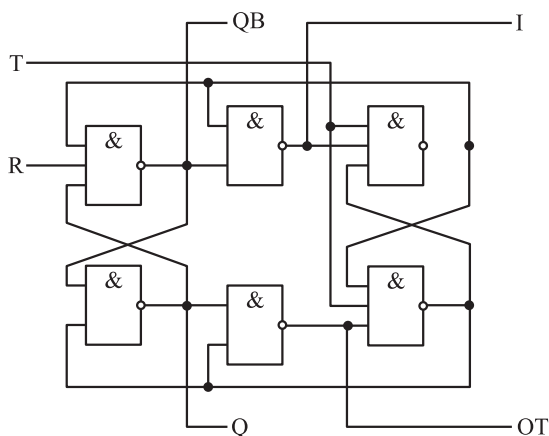


Рис. 21 Функциональная схема НЗ-элемента C0R

1558 транзисторов, в то время как сложность NCL-варианта равна 1766 транзисторам.

Таким образом, проектирование арифметических устройств в НЗ-базисе оказывается намного эффективнее, чем в NCL-базисе. Из-за ограниченности функционального элементного базиса и типов кодирования информационных сигналов NCL-схемы получаются более сложными (четырёхразрядный счетчик — в 4,49 раза, умножитель 4×4 без знака — в 1,13 раза), а следовательно, и более энергопотребляющими.

Схемы NCL имеют неоспоримые преимущества в сравнении с НЗ-схемами:

1. При реализации комбинационных схем они не требуют индикации каждого элемента схемы. Достаточно проиндексировать только ее последние ярусы, если каждый элемент полностью индицирует все свои входы на своих выходах.
2. Строгое соблюдение парафазной дисциплины с нулевым спейсером (NULL) существенно упрощает построение сложных NCL-схем.

3. Благодаря использованию единственного способа кодирования информационных сигналов и единственного спейсера процесс проектирования NCL-схем легче поддается формализации и автоматизации. В настоящее время уже существует как минимум два программных средства синтеза NCL-схем по формальному описанию на специальном языке — BALSА [76] и UNCLE [77].

Однако NCL-схемы обладают и существенными недостатками по сравнению с НЗ-схемами:

1. Индикация входов на выходах в каждом элементе приводит к большой избыточности аппаратных затрат.
2. Ограниченность элементного базиса, использование единственного способа кодирования информационных сигналов и единственного спейсера не позволяют получать более компактные реализации последовательных схем.
3. Вследствие аппаратной избыточности и наличия инвертора на выходе каждого элемента ухудшается быстродействие и увеличивается энергопотребление.

4 Заключение

Несмотря на изначально более сложную аппаратную реализацию НЗ-схем по сравнению с синхронными аналогами (до 2,1 раза для регистровых структур и до 2,5 раза для комбинационных структур), НЗ-схемы обеспечивают более высокое быстродействие аппаратуры в реальных условиях. В ряде случаев они обладают и существенно меньшим энергопотреблением. Поэтому применение НЗ-схемотехники может быть оправдано даже в

областях, где высокая надежность функционирования не является определяющей, но требуется высокое реальное быстродействие или низкое энергопотребление.

Типовые вычислительные устройства, реализованные в базисе НЗ-схем, оказываются в 1,5–2 раза лучше своих синхронных аналогов по энергетической эффективности (отношению энергии потребления к производительности) и в 1,7–2,6 раза лучше по производительности в реальных условиях. По добротности, учитывающей энергию потребления, производительность и допустимые диапазоны напряжения питания и температуры окружающей среды, НЗ-схемы оказываются лучше синхронных аналогов в 15–18 раз.

Наиболее предпочтительно применение НЗ-схемотехники в высоконадежных отказоустойчивых системах реального времени. Результаты испытаний отказоустойчивых вариантов исполнения ПП-порта показали, что НЗ-исполнение по сравнению с синхронной реализацией характеризуется лучшими показателями по всем параметрам: в 1,2 раза по быстродействию и по аппаратным затратам, в 1,3 раза по энергетической эффективности и в 18 раз по добротности.

Независимо от сложности реализации НЗ-схемы зона ее работоспособности определяется физическими характеристиками транзисторов. Она гораздо шире зоны работоспособности традиционных С-схем с фиксированной частотой синхронизации и превышает аналогичную зону С-схем с адаптивной частотой синхронизации.

Маршрут проектирования НЗ-схем поддерживается разработанными в ИПИ РАН программными средствами:

- синтеза относительно простых НЗ-схем (СИНТАБИБ, СИНКОМБ);
- анализа разрабатываемой схемы на возможное нарушение принципов построения НЗ-схем (АСИАН [78], АСПЕКТ [50], САМАН, ФАЗАН).

Эти программные средства обеспечивают безошибочное проектирование НЗ-устройств и гарантируют принадлежность разрабатываемой схемы к классу НЗ-схем. Программы анализа способны обработать достаточно сложные цифровые устройства, например 64-разрядное АЛУ.

Результаты практических исследований представителей различных подклассов СС-схем подтвердили декларированные теоретически преимущества НЗ-схем по зоне работоспособности, быстродействию и энергетической эффективности по сравнению с синхронными аналогами.

Схемы НЗ, разрабатываемые в соответствии с методологией, продвигаемой ИПИ РАН, обладают меньшими аппаратными затратами (в 4,49 раза при реализации двоичного счетчика, в 1,13 раза при реализации умножителя 4×4 , до 2 раз при реализации более простых логических схем), большей производительностью и меньшим энергопотреблением по сравнению с NCL-схемами. Поэтому именно их целесообразно использовать в качестве схемотехнического базиса для проектирования и изготовления супер-ЭВМ эксафлопсного класса: они обеспечат пониженное энергопотребление и высокую надежность проектируемых цифровых устройств любой сложности.

Литература

1. *Varshavsky V.* Time, timing and clock in massively parallel computing systems // Conference (International) on Massively Parallel Computing Systems Proceedings. — Colorado Springs, 1998. P. 100–106.
2. *Beerel P., Cortadella J., Kondratyev A.* Bridging the gap between asynchronous design and designers (Tutorial) // VLSI Design Conference Proceedings. — Mumbai, 2004. P. 18–20.
3. *Muller D., Bartky W.* A theory of asynchronous circuits // Annals of Computation Laboratory of Harvard University, 1959. Vol. 29. P. 204–243.
4. *Muller D. E.* Asynchronous logics and application to information processing // Switching theory in space technology. — Stanford, CA: Stanford University Press, 1963. P. 289–297.
5. *Seitz C. L.* System timing // Introduction to VLSI Systems. — Reading, MA: Addison-Wesley, 1980. P. 218–262.
6. *Singh N. P.* A design methodology for self-timed systems. Master's Thesis. MIT/LCS/TR-258. — MIT, Laboratory for Computer Science, 1981. 98 p.
7. *Martin A. J.* Compiling communicating processes into delay-insensitive VLSI circuits // Distrib. Comput., 1986. Vol. 1. No. 4. P. 226–234.
8. *Anantharaman T. S.* A delay insensitive regular expression recognizer // IEEE VLSI Techn. Bull., 1986. Vol. 1. No. 2. P. 4.
9. *Martin A. J.* Programming in VLSI // Development in concurrency and communication. — Reading, MA: Addison-Wesley, 1990. P. 1–64.
10. *Van Berkel K.* Beware the isochronic fork // Integration, VLSI J., 1992. Vol. 13. No. 2. P. 103–128.
11. *David I., Ginosar R., Yoeli M.* An efficient implementation of Boolean functions as self-timed circuits // IEEE Trans. Comput., 1992. Vol. 41. No. 1. P. 2–10.
12. *Sparso J., Staunstrup J., Dantzer-Sorensen M.* Design of delay insensitive circuits using multi-ring structures // European Design Automation Conference Proceedings, 1992. P. 15–20.

13. *Hauck S.* Asynchronous design methodologies: An overview // Proc. IEEE, 1995. Vol. 83. No. 1. P. 69–93.
14. *Fant K. M., Brandt S. A.* NULL convention logic: A complete and consistent logic for asynchronous digital circuit synthesis // Conference (International) on Application Specific Systems, Architectures, and Processors Proceedings, 1996. P. 261–273.
15. *Paver N. C., Day P., Farnsworth C., Jackson D. L., Lien W. A., Liu J.* A low-power, low-noise, configurable self-timed DSP // ASYNC'98: 4th Symposium (International) on Advanced Research in Asynchronous Circuits and Systems Proceedings, 1998. P. 32–42.
16. *Laiho M., Vianio O.* A full-custom self-timed DSP processor implementation // European Solid-State Circuits Conference Proceedings, 1997. <http://www.imec.be/esscirc/papers-97/172.pdf>.
17. *Matsubara G., Ide N., Tago H., Suzuki S., Goto N.* 30-m 55-b shared Radix 2 Division and square root using a self-timed circuit // ARITH'95: 12th Symposium on Computer Arithmetic Proceedings, 1995. P. 98–105.
18. *Garside J. D., Bainbridge W. J., Bardsley A., et al.* AMULET3i — an asynchronous system-on-chip // ASYNC-2000 Proceedings. — Eilat, Israel, 2000. P. 162–175.
19. *Bink A., York R.* ARM996HS: The first licensable, clockless 32-bit processor core // IEEE Micro, 2007. Vol. 27. No. 2. P. 58–68.
20. *Martin A. J., Nyström M., Wong C. G.* Three generations of asynchronous microprocessors // IEEE Des. Test Comput., 2003. Vol. 20. No. 6. P. 9–17.
21. Handshake Solutions HT80C51 User Manual. http://www.keil.com/dd/docs/datashts/handshake/ht80c51_um.pdf.
22. TIMA Laboratory Annual Report 2006. 2007. http://tima.imag.fr/publications/files_reports/ann-rep-06.pdf.
23. *Gang J., Lei W., Zhiying W.* The design of asynchronous microprocessor based on optimized NCLX design-flow // IEEE Conference (International) on Networking, Architecture and Storage Proceedings, 2009. P. 357–364.
24. *Ramaswamy S., Rockett L., Patel D., Danziger S., Manohar R., Kelly C. W., Holt J. L., Ekanayake V., Elftmann D.* A radiation hardened reconfigurable FPGA // IEEE Aerospace Conference Proceedings, 2009. P. 1–10.
25. Аperiodические автоматы / Под ред. В. И. Варшавского. — М.: Наука, 1976. 424 с.
26. Автоматное управление асинхронными процессами в ЭВМ и дискретных системах / Под ред. В. И. Варшавского. — М.: Наука, 1986. 400 с.
27. *Varshavsky V., Kishinevsky M., Marakhovskiy V., et al.* Self-timed control of concurrent processes. — Dordrecht, The Netherlands: Kluwer Acad. Publs., 1990. 245 p.
28. *Kishinevsky M., Kondratyev A., Taubin A., Varshavsky V.* Concurrent hardware: The theory and practice of self-timed design. — N.Y.: John Wiley & Sons, 1994. 368 p.
29. *Филин А. В., Степченков Ю. А.* Компьютеры без синхронизации // Системы и средства информатики, 1999. Вып. 9. С. 247–261.
30. *Степченков Ю. А., Дьяченко Ю. Г., Петрухин В. С., Филин А. В.* Цена реализации уникальных свойств самосинхронных схем // Системы и средства информатики, 1999. Вып. 9. С. 261–292.
31. *Степченков Ю. А., Дьяченко Ю. Г., Петрухин В. С., Филин А. В.* Самосинхронная схемотехника — альтернатива синхронной // Электронный сборник научных трудов сотрудников ОИВТА РАН. Разд. Элементная база, 1999. 10 с. http://samosinhron.ru/files/articles/native/sss_alternative_1999.DOC.
32. *Плеханов Л. П., Степченков Ю. А.* Экспериментальная проверка некоторых свойств строго самосинхронных электронных схем // Системы и средства информатики, 2006. Вып. 16. С. 476–485.
33. *Степченков Ю. А., Петрухин В. С., Дьяченко Ю. Г.* Опыт разработки самосинхронного ядра микроконтроллера на базовом матричном кристалле // Нано- и микросистемная техника, 2006. № 5. С. 29–36.
34. *Степченков Ю. А., Дьяченко Ю. Г., Петрухин В. С.* Самосинхронные последовательностные схемы: опыт разработки и рекомендации по проектированию // Системы и средства информатики, 2007. Вып. 17. С. 503–529.
35. *Соколов И. А., Степченков Ю. А., Петрухин В. С., Дьяченко Ю. Г., Захаров В. Н.* Самосинхронная схемотехника — перспективный путь реализации аппаратуры // Системы высокой доступности, 2007. Т. 3. № 1-2. С. 61–72.
36. *Степченков Ю. А., Дьяченко Ю. Г., Петрухин В. С., Плеханов Л. П.* Самосинхронные схемы — ключ к построению эффективной и надежной аппаратуры долговременного действия // Системы высокой доступности, 2007. Т. 3. № 1-2. С. 73–88.
37. *Дьяченко Ю. Г., Степченков Ю. А., Бобков С. Г.* Квазисамосинхронный вычислитель: методологические и алгоритмические аспекты // Проблемы разработки перспективных микро- и наноэлектронных систем: Мат-лы конф. — М.: ИППМ РАН, 2008. С. 441–446.
38. *Stepchenkov Y., Diachenko Y., Zakharov V., Rogdestvenski Y., Morozov N., Stepchenkov D.* Quasi-delay-insensitive computing device: Methodological aspects and practical implementation // PATMOS'2009: Workshop (International) on Power and Timing Modeling, Optimization and Simulation Proceedings. — Delft, The Netherlands, 2009. P. 276–285.
39. *Степченков Ю. А., Дьяченко Ю. Г., Плеханов Л. П., Гринфельд Ф. И., Степченков Д. Ю.* Самосинхронный двухтактный D-триггер с высоким активным уровнем сигнала управления: Патент РФ № 2365031 // Офиц. бюлл. «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2009. № 23. 9 с.
40. *Степченков Ю. А., Дьяченко Ю. Г., Рождественские А. В., Морозов Н. В., Петрухин В. С.* Самосинхронный двухтактный D-триггер с низким активным уровнем сигнала управления: Патент на изобретение № 2366080 // Офиц. бюлл. «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2009. № 24. 9 с.

41. Дьяченко Ю. Г., Степченков Ю. А., Гринфельд Ф. И. Г-триггер с парафазными входами с нулевым спейсером: Патент на изобретение № 2366081 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2009. № 24. 7 с.
42. Степченков Ю. А., Дьяченко Ю. Г., Плеханов Л. П., Денисов А. Н., Филимонок О. П. Самосинхронный триггер для связи с удаленным приемником: Патент РФ № 2382487 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 5. 7 с.
43. Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. Г., Петрухин В. С. Однотактный самосинхронный RS-триггер с предустановкой: Патент РФ № 2390092 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 14. 18 с.
44. Степченков Ю. А., Дьяченко Ю. Г., Захаров В. Н., Гринфельд Ф. И. Двухтактный самосинхронный RS-триггер с предустановкой и входом управления: Патент РФ № 2390093 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 14. 20 с.
45. Степченков Ю. А., Дьяченко Ю. Г., Степченков Д. Ю., Плеханов Л. П. Двухтактный самосинхронный RS-триггер с предустановкой: Патент РФ № 2390923 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 15. 20 с.
46. Степченков Ю. А., Дьяченко Ю. Г., Морозов Н. В., Филин А. В. Однотактный самосинхронный RS-триггер с предустановкой и входом управления: Патент РФ № 2391772 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 16. 18 с.
47. Степченков Ю. А., Дьяченко Ю. Г., Плеханов Л. П. Двоичный самосинхронный счетчик с предустановкой: Патент РФ № 2392735 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 17. 11 с.
48. Соколов И. А., Степченков Ю. А., Дьяченко Ю. Г. Самосинхронный триггер с однофазным информационным входом: Патент № 2405246 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2010. № 33. 32 с.
49. Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченков Д. Ю. Разработка вычислителя, не зависящего от задержек элементов // Системы и средства информатики, 2010. Вып. 20. № 1. С. 5–23.
50. Рождественский Ю. В., Морозов Н. В., Рождественская А. В. АСПЕКТ: Подсистема событийного анализа самосинхронных схем // Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всероссийский науч.-технич. конф. (МЭС-2010): Сб. науч. тр. — М.: ИППМ РАН, 2010. С. 26–31.
51. Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченков Д. Ю. Самосинхронный вычислитель для высоконадежных применений // Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всероссийский науч.-технич. конф. (МЭС-2010): Сб. науч. тр. — М.: ИППМ РАН, 2010. С. 418–423.
52. Плеханов Л. П. Разработка самосинхронных схем: функциональный подход // Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всероссийский науч.-технич. конф. (МЭС-2010): Сб. науч. тр. — М.: ИППМ РАН, 2010. С. 424–429.
53. Соколов И. А., Степченков Ю. А., Дьяченко Ю. Г. Самосинхронный RS-триггер с повышенной помехоустойчивостью (варианты): Патент РФ № 2427955 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2011. № 24. 42 с.
54. Степченков Ю. А., Дьяченко Ю. Г., Плеханов Л. П., Петрухин В. С., Степченков Д. Ю. Комбинированный Г-триггер с единичным спейсером: Патент РФ № 2434318 // Официальный бюллетень «Изобретения (заявки и патенты)». — М.: ВНИИПИ, 2011. № 32. 10 с.
55. Степченков Ю. А., Дьяченко Ю. Г., Горелкин Г. А. Самосинхронные схемы — будущее микроэлектроники // Вопросы радиоэлектроники, 2011. Вып. 2. С. 153–184.
56. Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В. Анализ на самосинхронность некоторых типов цифровых устройств // Системы и средства информатики, 2011. Вып. 21. № 1. С. 74–83.
57. Плеханов Л. П. Основы самосинхронных электронных схем. — М.: Бином, 2013. 208 с.
58. IEEE Computer Society. IEEE Standard for Floating-Point Arithmetic IEEE Std 754-2008. doi:10.1109/IEEESTD.2008.4610935.
59. Karthik S., de Souza I., Rahmeh J., Abraham J. Interlock schemes for micropipelines: Application to a self-timed rebound sorter // Conference (International) on Computer Design Proceedings. — Cambridge, 1991. P. 393–396.
60. Liebchen A., Gopalakrishnan G. Dynamic reordering of high latency transactions using a modified micropipeline // Conference (International) on Computer Design Proceedings. — Cambridge, 1992. P. 336–340.
61. Payne R. Self-timed FPGA systems // 5th Workshop (International) on Field Programmable Logic and Applications Proceedings. — Berlin/Heidelberg: Springer, 1995. P. 21–35.
62. Sobelman G. E., Fant K. CMOS circuit design of threshold gates with hysteresis // Symposium (International) on Circuits and Systems Proceedings, 1998. P. 61–64.
63. Weng N., Yuan J. S., DeMara R. F., Ferguson D., Hagedorn M. Glitch power reduction for low power IC design // 9th Annual NASA Symposium on VLSI Design Proceedings. — Albuquerque, 2000. P. 7.5.1–7.5.7.
64. Smith S. C. Completion-completeness for NULL convention digital circuits utilizing the bit-wise completion strategy // Conference (International) on VLSI Proceedings. — Las Vegas, 2003. P. 143–149.
65. Smith S. C., DeMara R. F., Yuan J. S., Ferguson D., Lamb D. Optimization of NULL convention self-timed circuits // Integration, VLSI J., 2004. Vol. 37. No. 3. P. 135–165.

66. *Fant K. M.* Logically determined design: Clockless system design with NULL convention logic. — N.Y.: John Wiley & Sons, 2005. 292 p.
67. *Smith S. C.* Development of a large word-width high-speed asynchronous multiply and accumulate unit // *Integration, VLSI J.*, 2005. Vol. 39. No. 1. P. 12–28.
68. *Smith S. C., Jia Di.* Designing asynchronous circuits using NULL Convention Logic (NCL) // *Synthesis Lectures Digital Circuits Syst.*, 2009. Vol. 4. No. 1. P. 61–73.
69. *Степченко Ю. А., Денисов А. Н., Дьяченко Ю. Г., Гриinfeld Ф. И., Филимонок О. П., Фомин Ю. П.* Библиотека элементов БМК для критических областей применения // *Системы и средства информатики*, 2004. Вып. 14. С. 318–361.
70. *Степченко Ю. А., Денисов А. Н., Дьяченко Ю. Г. и др.* Библиотека самосинхронных элементов для технологии БМК // *Проблемы разработки перспективных микроэлектронных систем* — 2006. — М.: ИППМ РАН, 2006. С. 259–264.
71. *Морозов Н. В., Степченко Ю. А., Дьяченко Ю. Г., Степченко Д. Ю.* Функциональная полузаказная библиотека самосинхронных элементов ML03: Свид. № 2010611908 от 12.03.10.
72. *Sokolov I. A., Stepchenkov Y. A., Dyachenko Y. G.* Self-timed RS-trigger with the enhanced noise immunity: U.S. Patent No. 8232825. 31 p.
73. Artisan Components. Chartered Semiconductor 0.18 μm IB Process 1.8-Volt SAGE-X™ Standard Cell Library Databook. Release 1.0. 2003. 313 p.
74. *Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю.* Характеризация псевдодинамических элементов // *Проблемы разработки перспективных микро- и нанoeлектронных систем: IV Всеросс. науч.-технич. конф. (МЭС-2010): Сб. науч. тр.* — М.: ИППМ РАН, 2010. С. 32–35.
75. Gate and throughput optimizations for null convention self timed digital circuits. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.7825&rep=rep1&type=pdf>.
76. *Edwards D., Bardsley A., Jani L., Plana L., Toms W.* Balsa: A tutorial guide. Version V3.5 — Manchester, 19/5/06. 157 p. <ftp://ftp.cs.man.ac.uk/pub/apt/balsa/3.5/BalsaManual3.5.pdf>.
77. *Reese R. B.* UNCLE (Unified NCL Environment): Technical Report MSU-ECE-10-001. <http://www.ece.msstate.edu/~reese/uncle/UNCLE.pdf>.
78. *Рождественский Ю. В., Морозов Н. В., Степченко Ю. А., Рождественскене А. В.* Универсальная подсистема анализа самосинхронных схем // *Системы и средства информатики*. — М.: Наука, 2006. Вып. 16. С. 463–475.

Поступила в редакцию 29.08.13

IMPLEMENTATION BASIS OF EXAFLOPS CLASS SUPERCOMPUTER

I. Sokolov¹, Y. Stepchenkov¹, S. Bobkov², V. Zakharov¹, Y. Diachenko¹, Y. Rogdestvenski¹, and A. Surkov²

¹Institute of Informatics Problems, Russian Academy of Sciences, Moscow 119333, 44-2 Vavilov Str., Russian Federation

²Scientific Research Institute for System Studies, Russian Academy of Sciences, 36 bld. 1, Nakhimovsky Prosp., Moscow 117218, Russian Federation

Abstract: The paper deals with choice of a circuitry basis for implementation of microprocessors and communication environment of exaflops supercomputers. A comparative analysis of the characteristics of the digital circuits with different complexity which are implemented in the synchronous basis as well as in the self-timed (ST) one was performed. It has proved the fundamental advantages of ST circuits comparing to synchronous analogues: absence of hazards, a maximum reachable operability range, high performance, and relatively low power consumption. Transforming any synchronous circuit into its quasi-ST or ST implementation leads to extension of its operability range independently of its complexity. The advantages of ST circuits show up to the maximum extent when they are used for designing reliable equipment. Various methodologies of ST circuits development are discussed. A comparative analysis of ST circuit implementation in the generic basis of the delay-insensitive circuits that is suggested by the authors and in the NULL Convention Logic circuit basis is performed. It is demonstrated that the suggested basis makes it possible to synthesize the circuits with the best parameters of performance, complexity, and power consumption while developing standard digital circuits serving as the basis for designing high end computing systems and hardware.

Keywords: synchronous circuits; self-timed circuits; delay-insensitivity; NULL Convention Logic; performance; power consumption; fault tolerance

DOI: 10.14357/19922264140106

Acknowledgments

This project was financially supported by the Russian Foundation for Basic Research (projects 13-07-12062 ofi_m and 13-07-12068 ofi_m) and partially supported by the Program of Basic Research of the RAS Department for Nanotechnologies and Information Technologies in 2013 (project 1.5).

References

- Varshavsky, V. 1998. Time, timing and clock in massively parallel computing systems. *Conference (International) on Massively Parallel Computing Systems Proceedings*. Colorado Springs. 100–106.
- Beerel, P., J. Cortadella, and A. Kondratyev. 2004. Bridging the gap between asynchronous design and designers (Tutorial). *VLSI Design Conference Proceedings*. Mumbai. 18–20.
- Muller, D., and W. Bartky. 1959. A theory of asynchronous circuits. *Annals of Computation Laboratory of Harvard University*. 29:204–243.
- Muller, D. E. 1963. Asynchronous logics and application to information processing. *Switching theory in space technology*. Stanford, CA: Stanford University Press. 289–297.
- Seitz, C. L. 1980. System timing. *Introduction to VLSI Systems*. Addison-Wesley. 218–262.
- Singh, N. P. 1981. A design methodology for self-timed systems. Cambridge: MIT Laboratory for Computer Science, MIT. M.Sc. Thesis. 98 p.
- Martin, A. J. 1986. Compiling communicating processes into delay-insensitive VLSI circuits. *Distrib. Comput.* 1(4):226–234.
- Anantharaman, T. S. 1986. A delay insensitive regular expression recognizer. *IEEE VLSI Technical Bulletin* 1(2):4.
- Martin, A. J. 1990. Programming in VLSI. *Development in concurrency and communication*. Reading, MA: Addison-Wesley. 1–64.
- Van Berkel, K. 1992. Beware the isochronic fork. *Integration, VLSI J.* 13(2):103–128.
- David, I., R. Ginosar, and M. Yoeli. 1992. An efficient implementation of Boolean functions as self-timed circuits. *IEEE Trans. Comput.* 41(1):2–11.
- Sparso, J., J. Staunstrup, and M. Dantzer-Sorensen. 1992. Design of delay insensitive circuits using multi-ring structures. *European Design Automation Conference Proceedings*. Hamburg. 15–20.
- Hauck, S. 1995. Asynchronous design methodologies: An overview. *Proc. IEEE* 83(1):69–93.
- Fant, K. M., and S. A. Brandt. 1996. NULL convention logic: A complete and consistent logic for asynchronous digital circuit synthesis. *Conference (International) on Application Specific Systems, Architectures, and Processors Proceedings*. Chicago. 261–273.
- Paver, N. C., P. Day, C. Farnsworth, D. L. Jackson, W. A. Lien, and J. Liu. 1998. A low-power, low-noise, configurable self-timed DSP. *4th Symposium (International) on Advanced Research in Asynchronous Circuits and Systems Proceedings*. San-Diego. 32–42.
- Laiho, M., and O. Vianio. 1997. A full-custom self-timed DSP processor implementation. *European Solid-State Circuits Conference Proceedings*. Available at: <http://www.imec.be/esscirc/papers-97/172.pdf> (accessed August 18, 2013).
- Matsubara, G., N. Ide, H. Tago, S. Suzuki, and N. Goto. 1995. 30-ns 55-b shared Radix 2 Division and square root using a self-timed circuit. *12th Symposium on Computer Arithmetic Proceedings*. 98–105.
- Garside, J. D., W. J. Bainbridge, A. Bardsley, et al. 2000. AMULET3i — an asynchronous system-on-chip. *6th IEEE Symposium (International) on Asynchronous Circuits and Systems Proceedings*. Eilat. 162–175.
- Bink, A., and R. York. 2007. ARM996HS: The first licensable, clockless 32-bit processor core. *IEEE Micro* 27(2):58–68.
- Martin, A. J., M. Nystrom, and C. G. Wong. 2003. Three generations of asynchronous microprocessors. *IEEE Des. Test Comput.* 20(6):9–17.
- Handshake solutions. HT80C51 User Manual. Available at: http://www.keil.com/dd/docs/datashts/handshake/ht80c51_um.pdf (accessed August 27, 2013).
- TIMA Laboratory Annual Report 2006. 2007. Available at: http://tima.imag.fr/publications/files_reports/ann-rep-06.pdf (accessed August 18, 2013).
- Jin, G., L. Wang, and Z. Wang. 2009. The design of asynchronous microprocessor based on optimized NCL_X design-flow. *IEEE Conference (International) on Networking, Architecture and Storage Proceedings*. 357–364.
- Ramaswamy, S., L. Rockett, D. Patel, S. Danziger, R. Manohar, C. W. Kelly, J. L. Holt, V. Ekanayake, and D. Elftmann. 2009. A radiation hardened reconfigurable FPGA. *IEEE Aerospace Conference Proceedings*. 1–10.
- Varshavsky, V. I., ed. 1976. *Aperiodicheskie avtomaty [Aperiodic machines]*. Moscow: Nauka Publ. 424 p.
- Varshavsky, V. I., ed. 1986. *Avtomatnoe upravlenie asinkhronnymi processami v EVM i diskretnykh sistemakh [Automata control of concurrent processes in computers and discrete systems]*. Moscow: Nauka Publ. 400 p.
- Varshavsky, V., M. Kishinevsky, V. Marakhovsky, et al. 1990. *Self-timed control of concurrent processes*. Kluwer Acad. Publ. 245 p.
- Kishinevsky, M., A. Kondratyev, A. Taubin, and V. Varshavsky. 1994. Concurrent hardware: The theory and practice of self-timed design. New York: John Wiley & Sons. 368 p.
- Filin, A. V., and Y. A. Stepchenkov. 1999. Komp'yutery bez sinkhronizatsii [Clockless computers]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 9:247–261.
- Stepchenkov, Y. A., Y. G. Diachenko, V. S. Petruhin, and A. V. Filin. 1999. Tsena realizatsii unikal'nykh svoystv

- samosinkhronnykh skhem [The penalty of self-timed circuit's unique features implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 9:261–292.
31. Stepchenkov, Y. A., Y. G. Diachenko, V. S. Petruhin, and A. V. Filin. 1999. Samosinkhronnaya skhemotekhnika — al'ternativa sinkhronnoy [Self-timed circuitry as an alternative of synchronous one]. Available at: http://samosinhron.ru/files/articles/native/sss_alternative_1999.DOC (accessed August 18, 2013).
 32. Plehanov, L. P., and Y. A. Stepchenkov. 2006. Eksperimental'naya proverka nekotorykh svoystv strogo samosinkhronnykh elektronnykh skhem [Experimental test of some features of strictly self-timed electronic circuits]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:476–485.
 33. Stepchenkov, Y. A., V. S. Petruhin, and Y. G. Diachenko. 2006. Opyt razrabotki samosinkhronnogo yadra mikrokontrollera na bazovom matrichnom kristalle [The experience in microcontroller's self-timed core design on FPGA]. *Nano- i Mikrosistemnaya Tekhnika [Nano- and Microsystem Technology]* 5:29–36.
 34. Stepchenkov, Y. A., Y. G. Diachenko, and V. S. Petruhin. 2007. Samosinkhronnye posledovatel'nostnye skhemy: Opyt razrabotki i rekomendatsii po proektirovaniyu [Self-timed sequential logic: An experience and design guidelines]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 17:503–529.
 35. Sokolov, I. A., Y. A. Stepchenkov, V. S. Petruhin, Y. G. Diachenko, and V. N. Zakharov. 2007. Samosinkhronnaya skhemotekhnika — perspektivnyy put' realizatsii apparatury [Timed circuitry — perspective method of hardware development]. *Sistemy Vysokoy Dostupnosti [High Availability Systems]* 3(1-2):61–72.
 36. Stepchenkov, Y. A., Y. G. Diachenko, V. S. Petruhin, and L. P. Plehanov. 2007. Samosinkhronnye skhemy — klyuch k postroeniyu effektivnoy i nadezhnoy apparatury dolgovremennogo deystviya [Self-timed circuits are a key for designing the efficient and reliable hardware with permanent operation]. *Sistemy Vysokoy Dostupnosti [High Availability Systems]* 3(1-2):73–88.
 37. Diachenko, Y. G., Y. A. Stepchenkov, and S. G. Bobkov. 2008. Kvazisamosinkhronnyy vychislitel': Metodologicheskie i algoritmicheskie aspekty [Quasi-self-timed coprocessor: The methodological aspects]. *Trudy Mezhdunarodnoy Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" [Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2008] Proceedings*. Moscow. 441–446.
 38. Stepchenkov, Y., Y. Diachenko, V. Zakharov, Y. Rogdestvenski, N. Morozov, and D. Stepchenkov. 2009. Quasi-delay-insensitive computing device: Methodological aspects and practical implementation. *The Workshop (International) on Power and Timing Modeling, Optimization and Simulation Proceedings*. Delft. 276–285.
 39. Stepchenkov, Y. A., Y. G. Diachenko, L. P. Plehanov, F. I. Grinfel'd, and D. Y. Stepchenkov. 2009. Samosinkhronnyy dvukhtaktnyy D-trigger s vysokim aktivnym urovnem signala upravleniya [Self-timed D flip-flop with high level control signal]. Patent RF No. 2365031. *Byulleten' Izobreteniy [Bulletin of Inventions]* 23. 9 p.
 40. Stepchenkov, Y. A., Y. G. Diachenko, A. V. Rozhdestvenskene, N. V. Morozov, and V. S. Petruhin. 2009. Samosinkhronnyy dvukhtaktnyy D-trigger s nizkim aktivnym urovnem signala upravleniya [Self-timed D flip-flop with low level control signal]. Patent RF No. 2366080. *Byulleten' Izobreteniy [Bulletin of Inventions]* 24. 9 p.
 41. Diachenko, Y. G., Y. A. Stepchenkov, and F. I. Grinfel'd. 2009. G-trigger s parafaznymi vkhodami s nulevym speyserom [G-trigger with null spacer dual-rail inputs]. Patent RF No. 2366081. *Byulleten' Izobreteniy [Bulletin of Inventions]* 24. 7 p.
 42. Stepchenkov, Y. A., Y. G. Diachenko, L. P. Plehanov, A. N. Denisov, and O. P. Filimonenko. 2010. Samosinkhronnyy trigger dlya svyazi s udalennym priemnikom [Self-timed trigger for connection to remote receiver]. Patent RF No. 2382487. *Byulleten' Izobreteniy [Bulletin of Inventions]* 5. 7 p.
 43. Stepchenkov, Y. A., Y. G. Diachenko, Y. G. Rogdestvenski, and V. S. Petruhin. 2010. Odnotaktnyy samosinkhronnyy RS-trigger s predustanovkoy [Self-timed RS-latch with preset]. Patent RF No. 2390092. *Byulleten' Izobreteniy [Bulletin of Inventions]* 14. 18 p.
 44. Stepchenkov, Y. A., Y. G. Diachenko, V. N. Zakharov, and F. I. Grinfel'd. 2010. Dvukhtaktnyy samosinkhronnyy RS-trigger s predustanovkoy i vkhodom upravleniya [Self-timed RS flip-flop with preset and control input]. Patent RF No. 2390093. *Byulleten' Izobreteniy [Bulletin of Inventions]* 14. 20 p.
 45. Stepchenkov, Y. A., Y. G. Diachenko, D. Y. Stepchenkov, and L. P. Plehanov. 2010. Dvukhtaktnyy samosinkhronnyy RS-trigger s predustanovkoy [Self-timed RS flip-flop with preset]. Patent RF No. 2390923. *Byulleten' Izobreteniy [Bulletin of Inventions]* 15. 20 p.
 46. Stepchenkov, Y. A., Y. G. Diachenko, N. V. Morozov, and A. V. Filin. 2010. Odnotaktnyy samosinkhronnyy RS-trigger s predustanovkoy i vkhodom upravleniya [Self-timed RS-latch with preset and control input]. Patent RF No. 2391772. *Byulleten' Izobreteniy [Bulletin of inventions]* 16. 18 p.
 47. Stepchenkov, Y. A., Y. G. Diachenko, and L. P. Plehanov. 2010. Dvoichnyy samosinkhronnyy schetchik s predustanovkoy [Self-timed binary counter with preset]. Patent RF No. 2392735. *Byulleten' Izobreteniy [Bulletin of Inventions]* 17. 11 p.
 48. Sokolov, I. A., Y. A. Stepchenkov, and Y. G. Diachenko. 2010. Samosinkhronnyy trigger s odnofaznym informatsionnym vkhodom [Self-timed trigger with single-phase data input]. Patent RF No. 2405246. *Byulleten' Izobreteniy [Bulletin of Inventions]* 33. 32 p.
 49. Stepchenkov, Y. A., Y. G. Diachenko, Y. V. Rogdestvenski, N. V. Morozov, and D. Y. Stepchenkov. 2010. Razrabotka vychislitelya, nezavisyashchego ot zaderzhek elementov [The design of a cell delay-insensitive coprocessor]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 20:5–23.

50. Rogdestvenski, Y. V., N. V. Morozov, and A. V. Rozhdestvenskene. 2010. ASPEKT: Podsystema sobytijnogo analiza samosinkhronnykh skhem [ASPECT: A suite of self-timed event-driven analysis]. *Trudy Mezhdunarodnoj Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" ["Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2010"] Proceedings*. Moscow. 26–31.
51. Stepchenkov, Y. A., Y. G. Diachenko, Y. V. Rogdestvenski, N. V. Morozov, and D. Y. Stepchenkov. 2010. Samosinkhronnyy vychislitel' dlya vysokonadezhnykh primeneniy [Self-timed coprocessor for high-reliable applications]. *Trudy Mezhdunarodnoj Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" ["Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2010"] Proceedings*. Moscow. 418–423.
52. Plehanov, L. P. 2010. Razrabotka samosinkhronnykh skhem: Funktsional'nyy podkhod [Self-timed circuits design: A functional approach]. *Trudy Mezhdunarodnoj Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" ["Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2010"] Proceedings*. Moscow. 424–429.
53. Sokolov, I. A., Y. A. Stepchenkov, and Y. G. Diachenko. 2011. Samosinkhronnyy RS-trigger s povyshennoy pomekhoustoychivost'yu (varianty) [Self-timed RS-trigger with the enhanced noise immunity]. Patent RF No. 2427955. *Byulleten' Izobreteniy [Bulletin of Inventions]* 24. 42 p.
54. Stepchenkov, Y. A., Y. G. Diachenko, L. P. Plehanov, V. S. Petruhin, and D. Y. Stepchenkov. 2011. Kombinirovanny G-trigger s edinichnym speyserom [Composite G-trigger with the unit spacer]. Patent RF No. 2434318. *Byulleten' Izobreteniy [Bulletin of Inventions]* 32. 10 p.
55. Stepchenkov, Y. A., Y. G. Diachenko, and G. A. Gorelkin. 2011. Samosinkhronnye skhemy — budushchee mikroelektroniki [Self-timed circuits are the future of microelectronics]. *Voprosy Radioelektroniki [The Problems of Radio Electronics]* 2:153–184.
56. Stepchenkov, Y. A., Y. G. Diachenko, Y. V. Rogdestvenski, and N. V. Morozov. 2011. Analiz na samosinkhronnost' nekotorykh tipov tsifrovyykh ustroystv [Self-timed analysis of the few types of the digital units]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 21(1):74–83.
57. Plehanov, L. P. 2013. *Osnovy samosinkhronnykh elektronnykh skhem [The base of the self-timed electronic circuits]*. Moscow: Binom Publ. 208 p.
58. IEEE Computer Society. 2008. IEEE Standard for Floating-Point Arithmetic IEEE Std 754-2008. doi:10.1109/IEEESTD.2008.4610935.
59. Karthik, S., I. de Souza, J. Rahmeh, and J. Abraham. 1991. Interlock schemes for micropipelines: Application to a self-timed rebound sorter. *Conference (International) on Computer Design Proceedings*. Cambridge. 393–396.
60. Liebchen, A., and G. Gopalakrishnan. 1992. Dynamic reordering of high latency transactions using a modified micropipeline. *Conference (International) on Computer Design Proceedings*. Cambridge. 336–340.
61. Payne, R. 1995. Self-timed FPGA systems. *5th Workshop (International) on Field Programmable Logic and Applications Proceedings*. Berlin/Heidelberg. 21–35.
62. Sobelman, G. E., and K. Fant. 1998. CMOS circuit design of threshold gates with hysteresis. *Symposium (International) on Circuits and Systems Proceedings*. 61–64.
63. Weng, N., J. S. Yuan, R. F. DeMara, D. Ferguson, and M. Hagedorn. 2000. Glitch power reduction for low power IC design. *9th Annual NASA Symposium on VLSI Design Proceedings*. Albuquerque. 7.5.1–7.5.7.
64. Smith S. C. 2003. Completion-completeness for NULL convention digital circuits utilizing the bit-wise completion strategy. *Conference (International) on VLSI Proceedings*. Las Vegas. 143–149.
65. Smith, S. C., R. F. DeMara, J. S. Yuan, D. Ferguson, and D. Lamb. 2004. Optimization of NULL convention self-timed circuits. *Integration, VLSI J.* 37(3):135–165.
66. Fant, K. M. 2005. Logically determined design: Clockless system design with NULL convention logic. New York: John Wiley & Sons. 292 p.
67. Smith, S. C. 2005. Development of a large word-width high-speed asynchronous multiply and accumulate unit. *Integration, VLSI J.* 39(1):12–28.
68. Smith, S. C., and J. Di. 2009. Designing asynchronous circuits using NULL Convention Logic (NCL). *Synthesis Lectures Digital Circuits Syst.* 4(1):61–73.
69. Stepchenkov, Y. A., A. N. Denisov, Y. G. Diachenko, F. I. Grinfel'd, O. P. Filimonenko, and Y. P. Fomin. 2004. Biblioteka elementov BMK dlya kriticheskikh oblastey primeneniya [The gate array cell library for critical applications]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 14:318–361.
70. Stepchenkov, Y. A., A. N. Denisov, Y. G. Diachenko, et al. 2006. Biblioteka samosinkhronnykh elementov dlya tekhnologii BMK [Self-timed cell library for gate array technology]. *Trudy Mezhdunarodnoj Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" ["Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2006"] Proceedings*. Moscow. 259–264.
71. Morozov, N. V., Y. A. Stepchenkov, Y. G. Diachenko, and D. Y. Stepchenkov. 2010. Funktsional'naya poluzakaznaya biblioteka samosinkhronnykh elementov ML03 [The functional semicustom library of the self-timed cells]. Certificate on official registration of the computer program No. 2010611908. (In Russian, unpublished.)
72. Sokolov, I. A., Y. A. Stepchenkov, and Y. G. Dyachenko. 2010. Self-timed RS-trigger with the enhanced noise immunity. U.S. Patent No. 8232825. 31 p.
73. Artisan Components. Chartered Semiconductor 0.18 μm IB Process 1.8-Volt SAGE-XTM Standard Cell Library Databook. 2003. Release 1.0. 313 p.
74. Diachenko, Y. G., N. V. Morozov, and D. Y. Stepchenkov. 2010. Kharakterizatsiya psevdodinamicheskikh elementov

- [The characterization of the pseudodynamic cells]. *Trudy Mezhdunarodnoy Konferentsii "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh sistem"* ["Problems of the Perspective Micro- and Nanoelectronic Systems Development — 2010" Proceedings]. Moscow. 32–35.
75. Gate and throughput optimizations for null convention self timed digital circuits. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.7825&rep=rep1&type=pdf> (accessed August 18, 2013).
76. Edwards, D., A. Bardsley, L. Jani, L. Plana, and W. Toms. 2006. Balsa: A tutorial guide. Manchester. 157 p. Available at: <ftp://ftp.cs.man.ac.uk/pub/apt/balsa/3.5/BalsaManual3.5.pdf> (accessed August 18, 2013).
77. Reese, R. B. 2011. UNCLE (Unified NCL Environment). Technical Report MSU-ECE-10-001. Available at: <http://www.ece.msstate.edu/~reese/uncle/UNCLE.pdf> (accessed August 18, 2013).
78. Rogdestvenski, Y. V., N. V. Morozov, Y. A. Stepchenkov, and A. V. Rozhdestvenskene. 2006. Universal'naya pod-sistema analiza samosinkhronnykh skhem [The universal suite for self-timed circuit analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:463–475.

Received August 29, 2013

Contributors

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ISokolov@ipiran.ru

Stepchenkov Yuri A. (b. 1951) — Candidate of Sciences (PhD) in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; YStepchenkov@ipiran.ru

Bobkov Sergey G. (b. 1951) — Doctor of Science in technology, Head of Department, Scientific Research Institute for System Studies, Russian Academy of Sciences, 36 bld. 1, Nakhimovsky Prosp., Moscow 117218, Russian Federation; bobkov@cs.niisi.ras.ru

Zakharov Victor N. (b. 1948) — Doctor of Science (PhD) in technology, associate professor; Scientific Secretary, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VZakharov@ipiran.ru

Diachenko Yuri G. (b. 1958) — Candidate of Sciences (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; diaura@mail.ru

Rogdestvenski Yuri V. (b. 1952) — Candidate of Sciences (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; YRogdest@ipiran.ru

Surkov Alexei V. (b. 1978) — senior scientist, Scientific Research Institute for System Studies, Russian Academy of Sciences, 36 bld. 1, Nakhimovsky Prosp., Moscow 117218, Russian Federation; surkov@cs.niisi.ras.ru

ИНФОРМАЦИОННАЯ МОДЕЛЬ ТЕХНОЛОГИИ ПРЕДСТАВЛЕНИЯ НАТУРНОГО ОБЪЕКТА И ИЗМЕНЕНИЯ ЕГО ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ

О. П. Архипов¹, Ю. А. Маньяков², Д. О. Сиротинин³

Аннотация: Приводится описание информационной модели технологии представления натурального объекта (НО) и изменения его пространственного положения с последующей интеграцией полученных данных в трехмерные (3D) изображения. Рассматриваются этапы технологии, а также основные структуры и потоки данных. Предлагаемая технология подразумевает простой способ создания сложной анимации 3D объектов, содержащий как изменение положения различных объектов, так и их формы, не предъявляющий высоких требований к квалификации пользователя и не подразумевающий наличия сложного дорогостоящего оборудования. Технология может быть использована для создания обучающих программ и энциклопедий, рекламных видеороликов, клипов, фильмов, а также для реализации систем представления и передачи визуальной информации, для решения задач дистанционного управления, в системах дополненной реальности, пользовательских интерфейсах, системах поддержки принятия решений, мониторинга, контроля качества.

Ключевые слова: стереоскопия; 3D-модель; обработка изображений; цветовые характеристики; опорные точки; трехмерная реконструкция; анимация

DOI: 10.14357/19922264140107

1 Введение

В настоящее время при создании видеороликов, фильмов, рекламы и другого визуального контента широко используется 3D анимация. Среди методов создания такой анимации выделяют технологии захвата движения (Motion capture), которая изменяет виртуальную модель на основе данных, получаемых с реального объекта, и «ручное» создание анимации в пакетах 3D графики. Среди недостатков технологии Motion capture можно выделить необходимость размещения различных датчиков, затрудняющих перемещения, а также отсутствие возможности их размещения на определенных объектах, например человеческом лице. Кроме того, обе рассмотренные технологии требуют наличия довольно большого количества высококвалифицированного технического персонала и дорогостоящего оборудования.

Вследствие этого актуальна разработка бесконтактных способов 3D реконструкции и анимации, использующих распространенные и недорогие технические средства (веб-камеры, фотоаппараты).

Целью предлагаемой работы является создание информационной модели технологии представле-

ния натурального объекта и изменения его пространственного положения с последующей интеграцией полученных данных в 3D изображения.

Основной идеей предлагаемой технологии является поиск особых опорных точек (ОТ) на изображениях объекта, составляющих стереопары. Поиск производится на основе сопоставления цветовых характеристик различных областей изображений. В отличие от известных, данная технология подразумевает простой способ создания сложной анимации 3D объектов, содержащий изменение как положения различных объектов, так и их формы, не предъявляющий высоких требований к квалификации пользователя.

Информационная модель, представляющая данную технологию, подразумевает ряд этапов:

1. Построение сцены и съемка НО.
2. Поиск границ объекта на изображениях.
3. Фрагментация изображений объекта.
4. Построение локальных систем координат (ЛСК).
5. Генерация ОТ:
 - (а) аппроксимация цветового пространства изображения;

¹Орловский филиал Института проблем информатики Российской академии наук, arhipov12@yandex.ru

²Орловский филиал Института проблем информатики Российской академии наук, maniaikov_yuri@mail.ru

³Орловский филиал Института проблем информатики Российской академии наук, vespert@mail.ru

- (б) сегментация изображения на основе цветовых характеристик;
- (в) структурирование сегментированных областей.
- 6. Индексирование.
- 7. Поиск сопряженных и соответственных ОТ.
- 8. Фильтрация ОТ.

- 9. Вычисление координат ОТ в 3D пространстве.
- 10. Вычисление векторов смещения ОТ.
- 11. Построение 3D воксельной модели.
- 12. Текстурирование воксельной модели.

Обобщенная схема информационной модели представлена на рисунке.

Рассмотрим все этапы более подробно.

2 Построение сцены и съемка натурального объекта

Съемка натурального объекта должна обеспечивать получение полной информации о поверхности объекта. Для этого необходимо получить информацию со всех сторон объекта. Эта задача решается путем круговой съемки одной камерой или несколькими камерами.

Съемка объекта характеризуется следующими параметрами:

$$SP = \langle d, \alpha, f, fov \rangle,$$

где d — расстояние до объекта; α — угол между оптическими осями камер; f — фокусное расстояние; fov — угловое поле объектива.

В результате получаются серии изображений НО, полученные через определенные временные интервалы, — I^{LR} :

$$I^{LR} = \begin{bmatrix} \langle I^L, I^R \rangle_{1,1} & \dots & \langle I^L, I^R \rangle_{1,r} \\ \vdots & \ddots & \vdots \\ \langle I^L, I^R \rangle_{t,1} & \dots & \langle I^L, I^R \rangle_{t,r} \end{bmatrix}.$$

Здесь r — индекс ракурса; t — индекс временного этапа съемки; $\langle I^L, I^R \rangle$ — стереопара, где I^L и I^R являются подмножествами I :

$$I = \begin{bmatrix} px_{1,1} & \dots & px_{1,w} \\ \vdots & \ddots & \vdots \\ px_{h,1} & \dots & px_{h,w} \end{bmatrix}, \quad px = (r, g, b),$$

w и h — соответственно ширина и высота изображения в пикселах.

3 Поиск границ объекта на изображениях

Первым этапом обработки полученных изображений является выделение границ объекта. Главной

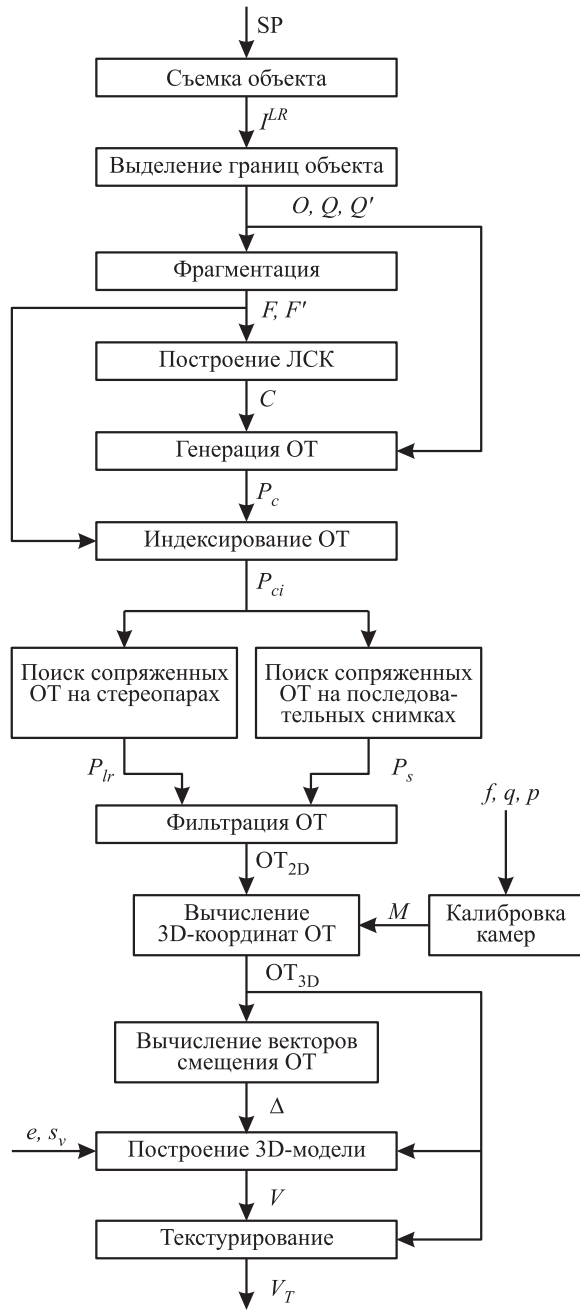


Схема информационной модели технологии представления НО и изменения его пространственного положения

информационной единицей для этого этапа служит маска, представляющая собой битовую матрицу и определяющая принадлежность каждой точки изображения объекту:

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,w} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,w} \\ \cdots & \cdots & \cdots & \cdots \\ q_{h,1} & q_{h,2} & \cdots & q_{h,w} \end{bmatrix},$$

где $q_{ij} = [1, 0]$ — переменная, равная единице в том случае, когда данный пиксел изображения принадлежит объекту.

Результатом выделения границ объекта является множество точек, принадлежащих объекту на изображении:

$$O = I \cap Q, \quad O \in I,$$

где Q — маска (растровое бинарное изображение, обозначающее, является ли каждый его пиксел частью объекта).

У совокупности пикселов изображения, принадлежащих объекту, выделяются пикселы, являющиеся граничными. Для этого применяется еще одна битовая маска:

$$Q' = \begin{bmatrix} q_{11} & \cdots & q_{1w} \\ \vdots & \ddots & \vdots \\ q_{h1} & \cdots & q_{hw} \end{bmatrix},$$

где $q_{ij} = [1, 0]$ — переменная, равная единице в том случае, если данный пиксел изображения является граничным пикселом объекта.

4 Фрагментация изображений объекта

Этап фрагментации осуществляется пользователем в интерактивном режиме по принципу определения движущихся частей объекта. Цель фрагментации — создание групп ОТ, к которым относятся ОТ, принадлежащие неделимой движущейся части (например, руке, ноге и т. п.). Множество фрагментов на изображении можно представить в виде

$$F = \{f_1, \dots, f_{10}\}.$$

Здесь

$$f_i = \langle (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) \rangle,$$

где (x_1, y_1) , (x_2, y_2) , (x_3, y_3) и (x_4, y_4) — координаты четырех вершин прямоугольника, описывающего фрагмент.

Множество точек объекта, принадлежащих фрагментам:

$$F' = f(O, F), \quad F' \subset O.$$

5 Построение локальных систем координат

Для оптимизации процесса обработки изменений положения точек в каждом из фрагментов их координаты рассчитываются относительно ЛСК фрагмента. Для этого используется матрица преобразования координат точек изображения к ЛСК:

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \Delta x & \Delta y & 1 \end{bmatrix},$$

где Δx и Δy — смещение начала ЛСК относительно точки начала мировой системы координат.

6 Генерация опорных точек

Основными этапами генерации ОТ являются:

- (а) аппроксимация цветового пространства изображения:

$$I \rightarrow A, \quad A = \begin{bmatrix} R_1 & G_1 & B_1 \\ \vdots & \vdots & \vdots \\ R_l & G_l & B_l \end{bmatrix},$$

где l — количество цветов палитры изображения после аппроксимации; R , G и B — цветовые характеристики красного, зеленого и синего каналов цвета;

- (б) сегментация изображения на основе цветовых характеристик:

$$SP = \begin{bmatrix} tlx_1 & tly_1 & brx_1 & bry_1 \\ tlx_2 & tly_2 & brx_2 & bry_2 \\ \vdots & \vdots & \vdots & \vdots \\ tlx_n & tly_n & brx_n & bry_n \end{bmatrix},$$

где tlx и tly — координаты верхней левой граничной точки сегмента; brx и bry — координаты нижней правой граничной точки сегмента;

- (в) структурирование сегментированных областей:

$$S = \begin{bmatrix} tlx_1 & tly_1 & brx_1 & bry_1 & s \\ tlx_2 & tly_2 & brx_2 & bry_2 & s \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ tlx_n & tly_n & brx_n & bry_n & s \end{bmatrix},$$

где s — состояние структурирующего прямоугольника, $s \in \{0, 1\}$.

В результате получаются две совокупности ОТ для левых (P_l) и правых (P_r) изображений стереопар:

$$P = P_l \cup P_r .$$

Здесь

$$P_l = \begin{bmatrix} p_{1,1} & \cdots & p_{1,s} \\ \vdots & \ddots & \vdots \\ p_{t,1} & \cdots & p_{t,s} \end{bmatrix} ; P_r = \begin{bmatrix} p_{1,1} & \cdots & p_{1,q} \\ \vdots & \ddots & \vdots \\ p_{t,1} & \cdots & p_{t,q} \end{bmatrix} ,$$

где

$$p_{i,j} = \langle x, y, r, g, b \rangle ,$$

(x, y) — координаты ОТ на изображении; (r, g, b) — цветовые координаты ОТ в пространстве RGB; s и q — количество ОТ на левых и правых изображениях соответственно; t — индекс временного этапа съемки.

Сразу после генерации ОТ производится преобразование их координат к локальным:

$$P_c = PC .$$

7 Индексирование

Индексирование ОТ осуществляется по принципу принадлежности к тому или иному фрагменту. В результате каждый элемент (p_{ij}) матрицы P преобразуется к виду:

$$p_{ci} = \langle x, y, r, g, b, F_i \rangle ,$$

где F_i — индекс фрагмента, которому принадлежит ОТ.

В результате получается преобразованная матрица P_{ci} , аналогичная P_c , с элементами, содержащими индексированные ОТ (p_c).

8 Поиск сопряженных и соответственных опорных точек

Исходными данными являются сегменты изображения S , полученные на этапе сегментации. Результатом поиска будут множества точек, полученные в результате применения функций сбора окружения F_n , проверки соответствия F_{sp} и сопряженности F_p :

$$P_{lr} = F_n(P_c) * F_{sp}(P_c) ; \\ P_s = F_n(P_c) * F_p(P_c) ,$$

где P_{lr} — совокупность сопряженных ОТ; P_s — совокупность соответственных ОТ.

Учитывая, что в результате поиска сопряженных точек каждой ОТ однозначно соответствует опре-

деленная точка на изображении, можно каждой ОТ поставить в соответствие определенное значение цветовых координат точки на изображении. Таким образом, j -ю сопряженную ОТ можно представить кортежем

$$P_{lr} = \{p_1^{lr}, \dots, p_k^{lk}\} .$$

Здесь

$$p_j^{lr} = \langle x_1, y_1, x_2, y_2, r, g, b, F_i \rangle , j = 1, \dots, k ,$$

где (x_1, y_1) — координаты ОТ на левом изображении стереопары; (x_2, y_2) — координаты ОТ на правом изображении стереопары; (r, g, b) — усредненные цветовые координаты ОТ в пространстве RGB; F_i — индекс фрагмента.

Тогда соответственные точки можно представить как

$$P_s = \{P_{lr_1}, \dots, P_{lr_t}\} .$$

9 Фильтрация опорных точек

Фильтрация ОТ производится с целью уменьшения количества рассматриваемых ОТ. Данная возможность позволяет достаточно гибко варьировать вычислительную нагрузку, изменяя количество ОТ в зависимости от требуемой степени детализации. Результатом фильтрации является множество опорных точек

$$OT_{2D} \in P_s .$$

10 Вычисление координат опорных точек в трехмерном пространстве

Вычисление координат ОТ в 3D пространстве основано на модели стереоскопической системы общего назначения [1]. Главной особенностью такой системы является наличие модели камеры, содержащей матрицы параметров внешних и внутренних камер.

Матрица внутренних параметров камеры содержит только параметры оптической системы и фотоприемника камеры и имеет следующий вид:

$$A = \begin{bmatrix} \frac{f}{q} & 0 & u_0 \\ 0 & \frac{f}{p} & v_0 \\ 0 & 0 & 1 \end{bmatrix} ,$$

где (u_0, v_0) — координаты главной точки относительно начала координат фотоприемника (в естественных координатах фотоприемника); q и p —

расстояния между сенсорами матрицы камеры вдоль строк и столбцов.

Внешние параметры описывают взаимное расположение камер в стереоскопической системе. Данные параметры определяются матрицами аффинных преобразований, посредством которых осуществляется переход от системы координат одной камеры к системе координат другой, и выражаются соотношением

$$M_l = RM_r + T,$$

где R — матрица поворота, описывающая ориентацию системы координат левой камеры относительно правой (поворот оптической оси); T — вектор, определяющий положение оптического центра левой камеры в системе координат правой.

Основываясь на полученной на предыдущем этапе модели камеры, можно вычислить координаты ОТ в 3D пространстве:

$$OT_{3D} = \langle x, y, z, F_i \rangle.$$

11 Построение и преобразование трехмерной воксельной модели

Воксельное представление подразумевает определенную упорядоченность объемных примитивов — вокселей. Построение воксельной модели основано на методе воксельной аппроксимации [2], исходной информацией для которой служит совокупность ОТ в 3D пространстве, полученная на предыдущем этапе, а результатом — воксельная модель вида

$$V = (v_1, v_2, \dots, v_n),$$

$$V = \langle xs_v, ys_v, zs_v \rangle,$$

где s_v — размер вокселя, определяющий 3D разрешение сцены.

Преобразование модели осуществляется на основе изменений положения ОТ, которые в свою очередь описываются векторами смещения (Δ):

$$\Delta = OT_{3D_t} - OT_{3D_{t-1}}.$$

Для отображения преобразования на модели при изменении положения ОТ необходимо вычислить смещение окружающих данную ОТ вокселей в определенной области. После этого для каждой ОТ, изменившей свое положение, осуществить воксельную аппроксимацию.

12 Текстурирование воксельной модели

Процесс текстурирования в рамках рассматриваемой информационной технологии заключается в присвоении вокселям модели цветов путем интерполяции на основе цветовых характеристик ОТ.

Исходной информацией на данном этапе является совокупность ОТ в 3D пространстве.

На основе исходной информации о цветах ближайших ОТ осуществляется линейная интерполяция цветов вокселей в области, ограниченной данными ОТ. В результате получается совокупность вокселей, образующих 3D модель и содержащих информацию о цвете:

$$V_T = \langle x, y, z, r, g, b, F_i \rangle.$$

13 Заключение

Таким образом, выше описана информационная модель технологии представления НО и изменения его пространственного положения с последующей интеграцией полученных данных в 3D изображения. Приведено описание основных этапов информационной технологии, структур и потоков данных.

Предлагаемая технология может быть использована для создания обучающих программ и энциклопедий, рекламных видеороликов, клипов, фильмов. Кроме того, может быть использована для реализации систем представления и передачи визуальной информации для решения задач дистанционного управления, систем дополненной реальности, пользовательских интерфейсов, систем поддержки принятия решений, мониторинга, контроля качества.

Литература

1. Грузман И.С., Киричук В.С., Косых В.П., Перетягин Г.И., Спектор А.А. Цифровая обработка изображений в информационных системах. — Новосибирск: НГТУ, 2000.
2. Маньяков Ю.А. Технология регистрации поведения объектов в трехмерном пространстве // Информационные технологии в науке, образовании и производстве (ИТНОП): Мат-лы Междунар. науч.-технич. конф. — Орел: ОрелГТУ, 2010. Т. 3. С. 182–186.

Поступила в редакцию 02.10.13

INFORMATION MODEL OF FULL-SCALE OBJECT AND ITS ATTITUDE CHANGES REPRESENTATION TECHNOLOGY

O. P. Arkhipov, Y. A. Maniakov, and D. O. Sirotinin

Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation

Abstract: The article describes the information model of full-scale object and its attitude changes representation with the following integration of the obtained data into three-dimensional (3D) images technology. The proposed technology provides a simple way to create complex animations of 3D objects including changes of positions of various objects and of their form. The technology does not require highly qualified users or expensive equipment. It can be used to create educational software and encyclopedias, commercials, videos, movies, as well as for implementation presentation and transmission of visual information solutions, remote control systems, augmented reality, user interfaces, decision support systems, monitoring systems, and quality control.

Keywords: stereoscopy; 3D model; image processing; color characteristics; markers; 3D reconstruction; animation

DOI: 10.14357/19922264140107

References

1. Gruzman, I. S., V. S. Kirichuk, V. P. Kosyh, G. I. Peretjagin, and A. A. Spektor. 2000. *Tsifrovaya obrabotka izobrazheniy v informatsionnykh sistemakh* [Digital image processing in information systems]. Novosibirsk: Novosibirsk State Technical University. 168 p.
2. Maniakov, Y. A. 2010. Tekhnologiya registratsii povedeniya ob"ektov v trekhmernom prostranstve [Three-dimensional space object behavior tracking technology]. *Informatsionnye Tekhnologii v Nauke, Obrazovanii i Proizvodstve (ITNOP): Materialy Konferentsii* [Information Technologies in Science, Education and Production: Science and Technology Conference (International) Proceedings]. Orel. 3:182–186.

Received October 2, 2013

Contributors

Arkhipov Oleg P. (b. 1948) — Candidate of Science (PhD) in technology, Director, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; arkipov12@yandex.ru

Maniakov Yury A. (b. 1984) — Candidate of Science (PhD) in technology, scientist, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; maniakov_yuri@mail.ru

Sirotinin Denis O. (b. 1984) — Candidate of Science (PhD) in technology, scientist, Orel Branch of Institute of Informatics Problems, Russian Academy of Sciences, 137 Moskovskoe Highway, Orel 302025, Russian Federation; vespert@mail.ru

ДИНАМИЧЕСКИЕ КОНТЕКСТЫ БАЗЫ ДАННЫХ РЕЛЯЦИОННОГО ТИПА*

С. В. Зыкин¹

Аннотация: Предложена технология динамического формирования представления данных. Эта технология является развитием методов аналитической обработки данных (OLAP — online analytical processing). Источником данных служит реляционная база данных (БД) с произвольной схемой (не обязательно иерархической). Целевое представление данных — композиционная таблица, которая позволяет представлять многомерные данные на плоскости. Эта таблица предполагает раздельное формирование размерностей с последующим сопоставлением мер размерностям в таблице. Основой промежуточных представлений данных является таблица связанных соединений, удовлетворяющая контекстным и логическим ограничениям. Предложены алгоритмы формирования таких таблиц и исследованы их свойства. Особое внимание уделено рассмотрению контекстов, используемых при формировании таблиц связанных соединений. Для создания контекстов предложен алгоритм направленного перебора и на примере выполнен сравнительный анализ работы алгоритмов формирования контекстов. Исследованные свойства контекстов и предложенные алгоритмы предназначены для автоматизации работы пользователя при формировании новых представлений данных.

Ключевые слова: реляционная база данных; контекст; соединение без потерь информации; композиционная таблица

DOI: 10.14357/19922264140108

1 Введение

В работах, посвященных OLAP, значительное внимание уделяется исследованию свойств моделей многомерных представлений данных (гиперкубов) [1–3] и операциям их преобразования [2, 4] с целью получения представления, необходимого для анализа данных. Особое внимание уделяется построению иерархий в размерностях [2, 3, 5–7], что позволяет гарантировать корректность операций агрегации данных. В работах [3, 5, 7] рассматриваются нормальные формы для многомерных моделей данных, которые позволяют контролировать значения NULL в иерархиях размерностей. Во всех перечисленных работах предполагается ручное формирование и анализ корректности размерностей гиперкубов.

В данной работе предлагается исследовать проблему автоматизации формирования размерностей с использованием свойств исходной реляционной БД. Это особенно актуально для систем, в которых не предполагается хранение многомерного представления данных целиком (технология MOLAP — multidimensional OLAP). Кроме того, считается недопустимым преобразование операционной БД с целью получения иерархической схемы «звезда»

или «снежинка» (технология ROLAP — relational OLAP). Отправной точкой служит предположение о том, что основой аналитической работы пользователя является необходимость формирования новых гиперкубов из исходного реляционного представления данных, а не многократное манипулирование одним и тем же гиперкубом. Новые гиперкубы нужны при выявлении скрытых закономерностей в данных и проведения анализа данных, не предусмотренного при проектировании складов данных по технологии MOLAP либо ROLAP.

Рассмотрим содержательную сторону постановки задачи и обсудим подходы к ее решению. Основной целью данной работы является повышение уровня автоматизации работы пользователя при формировании представления нового гиперкуба с возможностью его визуализации. Для демонстрации предлагаемых подходов к решению поставленной задачи рассмотрим пример.

Пример 1. Рассмотрим упрощенный фрагмент схемы БД учебного заведения [8], где подчеркнуты ключевые атрибуты отношений:

$R_1 = \text{Студенты}$ (№ студента, № группы, ФИО студента);

$R_2 = \text{Список групп}$ (№ группы, Код группы, № специальности, № курса);

* Работа выполнена при поддержке РФФИ (грант № 12-07-00066-а).

¹ Институт математики им. С. Л. Соболева Сибирского отделения Российской академии наук, szykin@mail.ru

Таблица 1 Фрагмент расписания занятий

№ курса		2					
Код группы		М-210			М-220		
День недели	Время начала занятия	Предмет	ФИО преподавателя	№ аудитории	Предмет	ФИО преподавателя	№ аудитории
Понедельник	8-00	Физика	Чанышев О. Г.	1-330			
	9-40	Химия	Чигишев О. М.	1-330	История	Дергачев А. С.	2-110
	11-30	История	Дергачев А. С.	1-330	Химия	Чигишев О. М.	2-110
	13-15	Литература	Арутян В. А.	1-330			
Вторник	8-00						
	9-40	Философия	Дергачев А. С.	1-330	Философия	Дергачев А. С.	1-330

$R_3 = \text{Предметы}$ (№ предмета, Предмет);
 $R_4 = \text{Преподаватели}$ (№ преподавателя, ФИО преподавателя);
 $R_5 = \text{Неделя}$ (№ дня недели, День недели);
 $R_6 = \text{Начало занятий}$ (№ занятия, Время начала занятия);
 $R_7 = \text{Оценки}$ (№ студента, № группы, № предмета, Оценка);
 $R_8 = \text{Расписание}$ (№ группы, № дня недели, № занятия, № предмета, № преподавателя, № аудитории);
 $R_9 = \text{Специальности}$ (№ специальности, Наименование специальности);
 $R_{10} = \text{Нагрузка}$ (№ специальности, № предмета, Количество часов, Контроль успеваемости);
 $R_{11} = \text{Аттестация}$ (№ студента, № группы, № предмета, Вид аттестации, Балл).

Один из возможных гиперкубов представлен в табл. 1.

Атрибуты размерностей в табл. 1 представлены жирным шрифтом, атрибуты мер — курсивом, значения атрибутов — обычным шрифтом.

Другой гиперкуб представлен в табл. 2.

В табл. 1 заданы две размерности: по горизонтали {День недели, Время начала занятий} и по вертикали {№ курса, Код группы}. К раз-

мерности по вертикали приписаны меры Предмет, ФИО преподавателя, № аудитории. В табл. 2 имеются две вертикальных размерности: {Предмет}, к которой приписана мера {Оценка}, и размерность {Предмет, Вид аттестации}, к которой приписана мера {Балл}. В этих размерностях объединены значения общего атрибута {Предмет}. Горизонтальная размерность в табл. 2 {№ студента, ФИО студента} является общей для обеих вертикальных размерностей. Кроме того, на значения в табл. 2 навешено логическое ограничение: «Код группы = М-220».

Представление данных в табл. 1 и 2 будем называть *композиционной таблицей* (композиция вертикальных размерностей). Далее общую размерность будем обозначать символом X . Вертикальные размерности обозначим Y_1, Y_2 и т. д.

Предложенная компоновка данных в табл. 1 и 2 является удобной для визуального анализа многомерных данных, поскольку все данные располагаются на плоскости, значения мер сопоставлены размерностям и представление данных не является разреженным.

Для автоматизации построения композиционной таблицы предлагается следующая последовательность формирования ее представления:

Таблица 2 Фрагмент сводной ведомости

Ограничение: Код группы = М-220.

Предмет		Физика		Философия		Химия	
Вид аттестации		Реферат	Оценка	Доклад	Оценка	Лаб. раб.	Оценка
№ студента	ФИО студента	Балл		Балл		Балл	
1	Алексенко С. В.	44	4	46	4		
2	Белоусов П. О.	52	5	43	4		
3	Бессараб О. П.	40	4	53	5		
4	Вяткин М. С.	42	4				
5	Драница А. А.						
6	Ефимов Е. С.	52	5	52	5		

1. Пользователь из списка атрибутов БД формирует множества атрибутов: размерности X, Y_1, Y_2, \dots, Y_N и соответствующие им меры Z_1, Z_2, \dots, Z_N . Мера Z_1 соответствует размерностям (X, Y_1) , мера Z_2 соответствует размерностям (X, Y_2) и т. д. Естественными являются ограничения: $X \cap Y_i = \emptyset$, $(X \cup Y_i) \cap Z_i = \emptyset$, $i = 1, 2, \dots, N$. Дополнительным технологическим ограничением является запрет на использование атрибута в качестве меры, если на него установлено ограничение в логическом выражении. Основанием для такого ограничения является возможность наличия значений размерностей в представлении и отсутствия соответствующих значений мер, хотя они есть в БД. Это может служить причиной для неверной интерпретации результатов.
2. Формирование иерархий размерностей для множеств атрибутов X, Y_1, Y_2, \dots, Y_N . Иерархии формируются автоматически по правилам, рассмотренным в работе [9], и пользователю предлагается только их модифицировать.
3. По специальному шаблону задаются логические ограничения на размерности $F_0(X), F_1(Y_1), F_2(Y_2), \dots, F_N(Y_N)$. По умолчанию каждая формула есть конъюнкция условий определенности (IS NOT NULL) для атрибутов размерности.
4. Формирование контекстов размерностей C_0, C_1, \dots, C_N (некоторые контексты могут быть пустыми, а некоторые — псевдоконтекстами). Далее будут представлены соответствующие определения и алгоритмы формирования контекстов.
5. Формирование контекста приложения C и соответствующей реализации таблицы связанных соединений s . С учетом структуры композиционной таблицы очевидно, что логическое ограничение на кортежи из s имеет следующий вид:

$$F(C) = F_0(X) \wedge (F_1(Y_1) \vee F_2(Y_2) \vee \dots \vee F_N(Y_N)).$$
6. Формирование реализаций размерностей X, Y_1, Y_2, \dots, Y_N с сортировкой значений в соответствии с иерархией. Если контекст размерности не пуст, то он используется для формирования, в противном случае реализация размерности является проекцией s .
7. Формирование реализации (представления) композиционной таблицы (заполнение значений мер на соответствующих местах таблицы).

Пользователь вручную выполняет шаги 1 и 3 и осуществляет выбор предложенных вариантов в

шагах 2, 4 и 5. Все остальные операции выполняются автоматически.

Заметим, что в предложенной последовательности шагов формирования композиционной таблицы исключается необходимость каким-либо образом модифицировать исходную операционную БД, что делает возможным реализовать все принципы проектирования БД [10, 11], в том числе самый важный — принцип независимости данных.

В работах, посвященных построению гиперкубического представления данных, в качестве промежуточных моделей используются SQL-таблицы. Интерфейс между БД и хранилищем данных программируется. Предлагаемый в данной работе подход исключает затраты на программирование.

В предлагаемой статье рассмотрена общая постановка задачи, включающая шаги от построения схемы композиционной таблицы до ее реализации. Особое внимание уделено правилам и алгоритмам формирования контекстов.

2 Контекст

2.1 Определения

Пусть $\mathfrak{R} = \{R_1, R_2, \dots, R_k\}$ — реляционная БД, определенная на множестве атрибутов $U = \{A_1, A_2, \dots, A_n\}$, где R_i — отношение, $[R_i]$ — схема отношения (подмножество атрибутов, на которых определено отношение R_i). Рассмотрим базовые зависимости, используемые при проектировании схемы БД [10–12]. Пусть DEP — множество зависимостей (функциональных, многозначных, включения, соединения), определенных на множестве атрибутов U и множестве отношений \mathfrak{R} . Пусть R — отношение, определенное на множестве атрибутов U (универсальное реляционное отношение).

Следующие четыре определения являются компонентами традиционной теории БД.

Определение 2.1 (ФЗ). Пусть X и Y — некоторые подмножества из множества атрибутов U . Будем говорить, что X функционально определяет Y : $X \rightarrow Y$, если в любой реализации R не могут присутствовать два кортежа $t, u \in R$, такие что $t[X] = u[X]$ и $t[Y] \neq u[Y]$.

Пусть заданы множества атрибутов $X \subseteq U$, $Y \subseteq U$ и $X \cap Y = \emptyset$, $Z = R \setminus (X \cup Y)$.

Определение 2.2 (МЗ). Множество X мультиопределяет множество Y в контексте Z : $X \rightarrow Y(Z)$ (многозначная зависимость), когда выполнено условие, что если для произвольной реализации R существу-

ет два кортежа $t_1, t_2 \in R$ таких, что $t_1[X] = t_2[X]$, то существует кортеж t_3 , для которого

$$t_3[X] = t_1[X]; \quad t_3[Y] = t_1[Y]; \quad t_3[Z] = t_2[Z],$$

и в силу симметрии существует кортеж t_4 :

$$t_4[X] = t_1[X]; \quad t_4[Y] = t_2[Y]; \quad t_4[Z] = t_1[Z].$$

Определение 2.3 (ЗС). Отношение $R(V_1, V_2, \dots, V_p)$ удовлетворяет зависимости соединения на множествах атрибутов V_1, V_2, \dots, V_p тогда и только тогда, когда R удовлетворяет свойству соединения без потерь информации (СБПИ):

$$R = R[V_1] \bowtie R[V_2] \bowtie \dots \bowtie R[V_p],$$

где \bowtie — операция естественного соединения; $R[V_i]$ — операция проекции отношения R по атрибутам V_i [10].

Заметим, что многозначная зависимость является частным случаем зависимости соединения, а функциональная зависимость является частным случаем многозначной зависимости [10, 11].

Формальным основанием для установления связей на схеме БД являются зависимости включения [12].

Определение 2.4 (ЗВ). Пусть $R_i[A_1, \dots, A_m]$ и $R_j[B_1, \dots, B_p]$ — схемы отношений (не обязательно различные), $V \subseteq \{A_1, \dots, A_m\}$ и $W \subseteq \{B_1, \dots, B_p\}$, $|V| = |W|$. Тогда между отношениями R_i и R_j существует зависимость включения по атрибутам V и W соответственно, если $R_i[V] \subseteq R_j[W]$, где $|V|$ — мощность множества V .

Если выполнено условие $V = W$, то такой вид ЗВ называется типизированным (typed) [13, 14]. Это дополнительное ограничение вполне согласуется с общепринятым свойством связей на схеме БД: связи отражают количественное соотношение кортежей в отношениях и не обладают какой-либо семантикой. Необходимость связывания различных по смыслу атрибутов, скорее всего, является признаком потери какой-либо ФЗ для связываемых атрибутов.

В традиционной теории БД рассматривается свойство сохраненных зависимостей при декомпозиции отношений [1]. Однако добиться реализации этого свойства на практике не всегда удается. Введем усиленный вариант этого свойства. Пусть $C = \{R_1^*, R_2^*, \dots, R_m^*\}$ — произвольное подмножество отношений реляционной БД.

Определение 2.5. Зависимость $\text{dep}_j \in \text{DEP}$ будем считать реализованной на C , когда операция дополнения, удаления или модификации кортежа в произвольном отношении $R_i^* \in C$ будет заблокирована организационно-техническими средствами, если при этом нарушается зависимость dep_j .

Под организационными средствами подразумевается способ проектирования схемы БД с указанием ограничений целостности на данные, под техническими средствами — возможности системы управления базами данных (СУБД) по поддержке этих ограничений целостности.

Совокупность отношений, по которым строится гиперкуб, должна удовлетворять свойству СБПИ [15], поскольку лишние кортежи в промежуточном представлении данных дают лишние значения в рабочей области гиперкуба. Для реализации этого свойства будем применять понятие контекста, которое впервые было использовано в работе [16].

Определение 2.6. Множество C будем называть контекстом, если оно удовлетворяет свойству СБПИ на зависимостях DEP, реализованных в C .

Замечание. В основе контекста лежит операция естественного соединения, которая собирает из различных отношений БД связанные друг с другом по значению данные. Затем эти данные (кортежи) участвуют в формировании новых структур, естественным образом дополняя и ограничивая друг друга, что делает уместным использование термина «контекст» для совокупности таких значений. Проверка свойства СБПИ осуществляется по алгоритму, представленному в [10].

2.2 Формирование контекста

Первоначальный выбор размерностей и мер гиперкуба предлагается сделать в расширенном виде: R_i, A_j , где R_i — наименование отношения из исходной реляционной БД и A_j — наименование атрибута в этом отношении. Таким образом, будет задано начальное множество отношений $R^0 = \{R_1^0, R_2^0, \dots, R_q^0\}$, участвующее в обязательном порядке в формируемом контексте, так как контекст далее используется в качестве основы для формирования размерностей и всего многомерного представления данных. Дальнейшая задача состоит в дополнении R^0 отношениями из \mathcal{R} , чтобы результирующее множество удовлетворяло свойству СБПИ на реализованных зависимостях, т. е. являлось контекстом. В общем случае таких вариантов дополнения существует несколько. Каждый из вариантов (контекстов) имеет свою смысловую нагрузку, поэтому окончательный выбор контекста может выполнить только пользователь. Задача алгоритма заключается в последовательной генерации контекстов без заикливания. Для сокращения числа перебираемых вариантов при формировании контекстов, ближайших к множеству R^0 , предлагается сделать этот перебор направленным.

Сформулируем критерии, которые позволят сделать перебор отношений направленным.

1. Замыкание первичного ключа нового отношения R_i совпадает со всем множеством атрибутов в выбранных отношениях. По теореме 5.8 [10] полученное множество отношений обладает свойством СБПИ, что исключает необходимость проверки этого свойства по алгоритму. В этом случае достаточно построить замыкание первичного ключа R_i . Такое отношение получает приоритет 3.
2. Для отношения R_i выполнено условие существования связи, соответствующей ЗВ $R_i[X] \subseteq R_j[X]$ с уже выбранными отношениями R_j , где множество атрибутов X является первичным ключом отношения R_j . Отношение R_i является «связующим» и позволяет объединить несколько несвязанных отношений в иерархию. Однако это условие не гарантирует выполнения свойства СБПИ, и его надо проверять отдельно. Отношение R_i получает приоритет 2.
3. Дополняемое отношение R_i должно иметь непустое пересечение с уже выбранными отношениями. В работе [16] показано, что это условие является необходимым для выполнения свойства СБПИ, если в ДЕР отсутствует реализованная декомпозицией (теорема Фейджина) многозначная зависимость с пустой левой частью. Такое отношение получает приоритет 1. Остальные отношения получают приоритет 0.
4. Формируемый контекст не должен содержать лишних отношений, наличие которых обусловлено только порядком присоединения отношений к контексту в алгоритме.

Перечисленные критерии увеличивают вероятность более быстрого достижения результата.

Введем обозначения. Пусть $R^1 = \{R_1^1, R_2^1, \dots, R_p^1\}$ — множество отношений, не входящих в исходное множество R^0 : $R^1 = \mathfrak{R} \setminus R^0$.

Рассмотрим схему алгоритма, удовлетворяющего сформулированным критериям (*алгоритм направленного перебора*).

1. Подсчет весов для отношений R^1 и их упорядочение по убыванию весов.
2. Формируются сочетания без повторений из отношений R^1 , сначала по одному, затем по два и т.д. Сочетания начинаются с наименьших значений и далее последовательно увеличиваются, например сочетания по два элемента: (1, 2), (1, 3), ..., (2, 3), ... Текущее сочетание отношений совместно с R^0 проверяется на выполнение свойства СБПИ. Если свойство вы-

полнено, то полученное сочетание дополняет множество контекстов.

3. В процессе выполнения алгоритма пользователю предлагается выбрать нужный контекст.

Замечание. Такая схема алгоритма на ближайших итерациях находит дополнительные отношения, с наибольшей вероятностью образующие наименьший контекст с исходным множеством отношений R^0 .

В табл. 1 размерности Y_1 соответствует контекст $C_1 = \{R_2\}$ ($C_0 = \{R_5, R_6\}$ для размерности X является псевдоконтекстом), контекст приложения $C = \{R_2, R_3, R_4, R_5, R_6, R_8\}$. В табл. 2 размерности X соответствует контекст $C_0 = \{R_1, R_2\}$ (отношение R_2 дополнено в контекст за счет логического ограничения). Размерности Y_1 соответствует контекст $C_1 = \{R_2, R_3, R_9, R_{10}\}$. В этом контексте пользователь определил, что ему требуется только список предметов из учебной нагрузки по специальности, а не все предметы из отношения R_3 . По аналогичным рассуждениям размерности Y_2 соответствует контекст $C_2 = \{R_2, R_3, R_9, R_{10}, R_{11}\}$. Оба контекста удовлетворяют свойству СБПИ с учетом многозначной зависимости: «№ специальности» → «№ группы» («№ предмета»). Поскольку в обеих размерностях присутствует отношение R_2 , то для них используется логическое ограничение, указанное в заголовке таблицы. Контекст приложения в табл. 2: $C = \{R_1, R_2, R_3, R_7, R_9, R_{10}, R_{11}\}$.

3 Реализация контекста

В качестве реализации контекста будем использовать представление данных в виде таблицы связанных соединений, являющейся частным случаем модели «таблица соединений» [17]. Совокупность свойств этой таблицы, которые будут рассмотрены ниже, позволяет получить необходимое представление для формирования многомерных представлений данных.

Рассмотрим преобразование представления реляционной БД: $\mathfrak{R} = \{R_1, R_2, \dots, R_k\}$ в таблицу связанных соединений (S, l) , где S — схема, определенная на множестве атрибутов $U = \{A_1, A_2, \dots, A_n\}$, l — вектор вхождения кортежей отношений длины k . Определим принцип формирования кортежей $t \in s$, где s — реализация (множество кортежей) схемы отношения S . Рассмотрим все возможные сочетания без повторений отношений R_1, R_2, \dots, R_k , удовлетворяющие свойству СБПИ. Пусть множество отношений $C' = \{R_{m(1)}, R_{m(2)}, \dots, R_{m(p)}\}$ — контекст, где $m(p)$ — целочисленный массив из p номеров отношений текущего сочетания, и c' — его

реализация, ограниченная операцией селекции σ_F с логической формулой F :

$$c' = \sigma_F(R_{m(1)} \bowtie R_{m(2)} \bowtie \dots \bowtie R_{m(p)}).$$

Для каждого кортежа $u \in c'$ формируем кортеж t по следующим правилам: $t[A_j] = u[A_j]$, если атрибут A_j принадлежит хотя бы одному отношению соединения, и $t[A_j] = \text{emp}$ в противном случае, где emp — пустое значение. Каждому кортежу поставим в соответствие битовый вектор $l(t) = (l_1(t), l_2(t), \dots, l_k(t))$, где $l_j(t) = 1$, если отношение R_j участвует в текущем соединении, и $l_j(t) = 0$ в противном случае.

Рассмотрим отношение частичного порядка над кортежами $t \in s$.

Определение 3.1. Кортеж $t \in s$ является менее определенным или равным кортежу $t' \in s$, когда для любого атрибута A_i выполнено условие: если $t[A_i] \neq t'[A_i]$, то $t[A_i] = \text{emp}$ и $l_j(t') \geq l_j(t)$, $j = 1, \dots, k$, причем $t[A_i] = t'[A_i]$, если A_i принимает значение NULL в обоих кортежах. В этом случае назовем кортеж t подчиненным кортежу t' и будем писать $t < t'$.

В представлении s достаточно хранить только кортеж t' , который содержит в себе все менее определенные либо равные кортежи. Следовательно, завершающим этапом построения представления s является удаление в нем всех подчиненных кортежей. Равенство неопределенных значений в определении 3.1 позволяет избавиться от подчиненных кортежей t , которые получены из тех же кортежей БД, что и t' . Отличие значений NULL и emp в том, что первое указывает на неопределенное значение атрибута, а второе — на отсутствие соответствующего кортежа в текущем соединении. Очевидно, что отношение $<$ является транзитивным.

Реализации контекстов для рассмотренного примера 1 являются громоздкими, поэтому рассмотрим самую простую из них: $C_0 = \{R_1, R_2\}$ для табл. 2 (см. табл. 3).

Пусть $X(J) = ([R_{j(1)}] \cup [R_{j(2)}] \cup \dots \cup [R_{j(m)}])$, где $J = (j(1), j(2), \dots, j(m))$, и $[R_{j(i)}]$ — множество

атрибутов отношения $R_{j(i)}$. Определим операцию проекции на множестве s .

Определение 3.2. Проекция $\pi_{X(J)}(s)$ есть совокупность кортежей $u[X(J)]$, определенных на множестве атрибутов $X(J)$, где для каждого $u[X(J)]$ существует кортеж $t \in s$ такой, что $u[X(J)] = t[X(J)]$ и $l_{j(i)}(t) = 1, i = 1, 2, \dots, m$.

Логическое ограничение $F(t)$ будем представлять в виде дизъюнктивной нормальной формы, что удобно реализовать в пользовательском интерфейсе в виде шаблонов

$$F = F_1 \vee F_2 \vee \dots \vee F_m,$$

где каждая элементарная формула является конъюнкцией предикатов сравнения языка SQL:

$$F_i = F_{i,1} \wedge F_{i,2} \wedge \dots \wedge F_{i,p},$$

где $F_{i,j} = [\text{NOT}]A_q \Theta (\text{выражение})$, $\langle \text{выражение} \rangle$ — константа либо атрибут A_s , Θ — операция сравнения, $[\text{NOT}]$ — необязательный параметр. Если какой-либо предикат $F_{i,j}$ не определен на кортеже t (атрибуты A_q и/или A_s имеют значение emp или NULL), то $F_{i,j}$ заменяется значением TRUE независимо от операции Θ . Такая подстановка позволяет оставить в s кортежи, для которых пока не определены некоторые атрибуты или отсутствуют связанные по значениям кортежи в других отношениях, что также является предметом анализа информации с использованием композиционной таблицы. Формула F после подстановки будет принимать только два значения: TRUE и FALSE.

Реализация этого свойства требует определенной последовательности формирования s . Если его формировать сверху вниз (начиная с соединения всех отношений контекста, затем подмножеств отношений и т. д. до одиночных отношений) и сразу отфильтровывать кортежи, используя формулу F , то в s останутся кортежи, подчиненные удаленным кортежам на предыдущих итерациях. Корректное формирование таблицы s достигается в следующем алгоритме.

A1: Вход: $C = \{R_1, R_2, \dots, R_k\}$ — контекст, DEP — реализованные зависимости, \mathfrak{R} — БД.

Таблица 3 Пример реализации контекста

№ студента	№ группы	ФИО студента	Код группы	№ специальности	№ курса	l_1	l_2
1	2	Алексенко С. В.	М-220	5	2	1	1
2	2	Белоусов П. О.	М-220	5	2	1	1
3	2	Бессараб О. П.	М-220	5	2	1	1
4	2	Вяткин М. С.	М-220	5	2	1	1
5	2	Драница А. А.	М-220	5	2	1	1
6	2	Ефимов Е. С.	М-220	5	2	1	1

Выход: s — таблица связанных соединений.

```

 $s = \emptyset$ 
for  $i = k$  to 1 step  $-1$ 
     $m = (0, 0, \dots, 0)$ 
    do while Comb( $i, k, m$ )
        if Llj( $C, m, DEP$ ) then
             $r = R_{m(1)} \bowtie R_{m(2)} \bowtie \dots \bowtie R_{m(i)}$ 
             $s'' = \text{Transfor}(r)$ 
             $s = s \cup s''$ 
        endif
    enddo
endfor
for each  $t_i$  in  $s$ 
    for each  $t_j$  in  $s$ 
        if  $t_i \neq t_j$  then
            if  $t_i \prec t_j$  then  $s = s - t_i$ 
            if  $t_j \prec t_i$  then  $s = s - t_j$ 
        endif
    endfor
endfor
 $s = \sigma_F(s)$ 
    
```

где Comb(i, k, m) — сочетания без повторений из k элементов по i , результат помещается в массиве m , функция принимает значение FALSE, если текущий набор сочетаний исчерпан, в противном случае — TRUE; Llj(C, m, DEP) — проверка свойства СБПИ для отношений с ненулевыми номерами в массиве m , Transfor(r) — преобразование соединения в таблицу связанных соединений.

Размер s перед фильтрацией может оказаться огромным, поэтому предлагается последовательность формирования s снизу вверх. Сначала просматриваются одиночные отношения контекста, из их кортежей формируются кортежи s с фильтрацией по формуле F . Затем просматриваются соединения пар отношений контекста, удовлетворяющие свойству СБПИ. Сформированные новые кортежи в s сначала используются для удаления подчиненных кортежей, а затем подвергаются фильтрации. Далее по аналогичным правилам просматриваются соединения трех отношений контекста и т. д. до соединения всех отношений контекста.

Формализуем эти рассуждения в виде алгоритма.

A2: Вход: $C = \{R_1, R_2, \dots, R_k\}$ — контекст, DEP — реализованные зависимости, \mathfrak{R} — БД.

Выход: s' — таблица связанных соединений.

```

 $s' = \emptyset$ 
for  $i = 1$  to  $k$ 
     $m = (0, 0, \dots, 0)$ 
    do while Comb( $i, k, m$ )
        if Llj( $C, m, DEP$ ) then
             $r = R_{m(1)} \bowtie R_{m(2)} \bowtie \dots \bowtie R_{m(i)}$ 
             $s'' = \text{Transfor}(r)$ 
        endif
    enddo
endfor
    
```

```

for each  $t_i$  in  $s'$ 
    for each  $t_j$  in  $s''$ 
        if  $t_i \prec t_j$  then  $s' = s' - t_i$ 
    endfor
endfor
 $s' = s' \cup \sigma_F(s'')$ 
endif
enddo
endfor
    
```

Алгоритмы вычисления s и s' совпадают по количеству вновь генерируемых кортежей, т. е. до места, где выполнена операция Transfor(r). Однако размер s' в процессе формирования меньше s и, как следствие, количество сравнений кортежей также меньше. Это достигается за счет фильтрации ненужных кортежей в процессе формирования s' и за счет выбранной последовательности формирования промежуточных соединений. Можно получить оценки для количества кортежей и операций в обоих алгоритмах в предположении о распределении атрибутов, как это сделано в работе [18] для таблицы соединений. Однако формулы для оценок будут громоздкими, хотя очевидно, что алгоритм формирования для s' является более экономичным по памяти и по числу операций. Осталось показать следующее.

Теорема 3.1. $s = s'$.

Доказательство. Пусть $t \in s$ и $t \neq s'$. Следовательно, существует кортеж $t' \in s'$: $t \prec t'$, а по предположению $t \neq t'$. Поскольку s и s' формируются из совпадающих наборов соединений, то существует кортеж $t'' \in s$: $t' \prec t''$. Учитывая транзитивность операции \prec , имеем $t \prec t''$. Следовательно, кортеж t должен быть удален из s : $s' \subseteq s$. В алгоритме A2 не проверяется подчиненность кортежей внутри одного соединения s'' . В этом случае подчиненными могут быть только совпадающие кортежи. С одной стороны, это не нарушает соотношения $s' \subseteq s$, с другой — совпадающие кортежи могут появиться в s'' , только если в каком-либо отношении $R_{m(j)}$ есть совпадающие кортежи, что недопустимо для реляционной БД.

Пусть $t \in s'$ и $t \notin s$. Следовательно, существует кортеж $t' \in s$: $t \prec t'$, а $t \neq t'$. По построению кортеж t' на некоторой итерации должен появиться в s'' алгоритма A2. Кортежи t и t' принадлежат следующим соединениям: $t \in R_{m(1)} \bowtie R_{m(2)} \bowtie \dots \bowtie R_{m(j)}$, $t' \in R_{q(1)} \bowtie R_{q(2)} \bowtie \dots \bowtie R_{q(i)}$. По определению операции \prec для множеств индексов выполнено: $\{m(1), m(2), \dots, m(j)\} \subseteq \{m(1), m(2), \dots, m(j)\}$. Это означает, что кортеж t' появится в s'' (алгоритм A2) позднее, чем t в s' , и t будет удален из s' . Следовательно, $s \subseteq s'$. Теорема доказана.

Основываясь на способе формирования таблицы s , сформулируем ее важные свойства.

Теорема 3.2. Для любого множества отношений $R^* = \{R_1^*, R_2^*, \dots, R_q^*\} \subseteq C'$, удовлетворяющего свойству СБПИ, где $C' = \{R_{m(1)}, R_{m(2)}, \dots, R_{m(p)}\}$ — контекст, выполнено:

$$\pi_{R^*}(s) = \sigma_F(R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*).$$

Доказательство. Пусть кортеж $t \in \sigma_F(R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*)$. Покажем, что этот кортеж принадлежит $\pi_{R^*}(s)$. По определению операции селекции $F(t) = \text{TRUE}$. Поскольку $R^* \subseteq C'$, то сочетание, состоящее из всех отношений R^* , также будет участвовать в формировании s с теми же ограничениями, задаваемыми формулой F . Кортеж t может быть удален из s , если существует кортеж $t' \in s: t \prec t'$. В этом случае возможны два варианта:

- (1) $F(t') = \text{FALSE}$;
- (2) $F(t') = \text{TRUE}$.

В первом варианте кортеж t' будет удален из s до реализации $\sigma_F(R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*)$ при формировании s и не сможет воспрепятствовать появлению кортежа t в s . Во втором варианте кортеж t' останется в s , а кортеж t будет удален. Однако по правилу получения проекции $\pi_{R^*}(t') = t$, что доказывает соотношение $\pi_{R^*}(s) \supseteq \sigma_F(R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*)$.

Докажем включение в обратную сторону. Пусть $t \in \pi_{R^*}(s)$. По правилу построения проекции существует кортеж $u \in s$ такой, что u определен для всех отношений множества R^* и по построению представления s выполнено $F(u) = \text{TRUE}$. Поскольку u определен для всех отношений R^* , то каждое отношение $R_i^* \in R^*$ содержит кортеж $u_i = u[R_i^*] = t[R_i^*]$. По правилу выполнения операции естественного соединения из совокупности кортежей u_i будет сформирован кортеж t , который принадлежит $R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*$. Поскольку $F(u) = \text{TRUE}$, то по правилу вычисления $F(t)$ получим $F(t) = \text{TRUE}$. Следовательно, кортеж t принадлежит $\sigma_F(R_1^* \bowtie R_2^* \bowtie \dots \bowtie R_q^*)$. Теорема доказана.

Теорема 3.3. Представление s всегда существует и единственно для любой схемы реляционной БД.

Доказательство. Существование s следует из его построения. Для доказательства единственности предположим, что существует $s' \neq s$. Заметим, что таблицы s и s' сформированы для одного состояния БД, по одинаковым сочетаниям отношений и с одинаковым ограничением F . В этом случае возможны два варианта:

- (1) существует кортеж $t \in s$ и $t \notin s'$;
- (2) существует кортеж $t' \in s'$ и $t' \notin s$.

Рассмотрим первый вариант. Для того чтобы кортеж t отсутствовал в s' , необходимо, чтобы существовал кортеж $t_1 \in s$ такой, что $t \prec t_1$. Поскольку $t \neq t_1$, то t_1 сформирован по большему количеству отношений, чем t . Если кортеж t_1 есть в s , то в s не должно быть кортежа t . Пусть t_1 отсутствует в s , тогда должен существовать кортеж $t_2 \in s$, для которого выполнено $t_1 \prec t_2$. С учетом свойства транзитивности отношения \prec получим, что $t \prec t_2$ и кортеж t должен быть удален из s .

Аналогичные рассуждения могут быть приведены и для второго варианта. Полученное противоречие доказывает единственность представления s . Теорема доказана.

Рассмотрим дополнительное обоснование выбора контекстов для реализации размерностей в виде таблицы связанных соединений. Кроме отображения частично заполненных данных в размерности таблицы можно ограничить присутствие только тех данных, которые связаны с другими объектами БД. Например, при формировании сводной ведомости по специальности в учебном заведении к размерности «Предметы» можно дополнить отношение «Учебный план». Тогда в размерности будут отображены только те предметы, которые есть в учебном плане, а не весь список предметов учебного заведения, причем, если по какому-либо предмету из учебного плана оценки пока отсутствуют, он все равно будет отображен в размерности таблицы. Такие свойства размерностей не предусмотрены во всех известных моделях многомерных данных.

Определение 3.3. Псевдоконтекстом C_i называется множество отношений, для которых не гарантируется выполнение свойства СБПИ. Реализацией размерности, соответствующей псевдоконтексту, является соединение отношений C_i . Ограничения на такую размерность задаются только на атрибутах уже выбранных отношений.

Псевдоконтексты соответствуют размерностям, в которых требуются все допустимые комбинации значений атрибутов. Например, значения атрибутов «День недели» и «Время начала занятия» в табл. 1 должны быть сопоставлены каждый каждому.

Определение 3.4. Контекст C_i называется пустым, если его реализация формируется в виде проекции от существующего контекста.

Такие контексты соответствуют размерностям, в которых присутствуют только те значения атрибутов, которым сопоставлено непустое множество значений меры. Заметим, что аналогичный результат дают все известные продукты технологии ROLAP.

После того как сформированы контексты и псевдоконтексты размерностей и общий контекст приложения, задача построения композиционной таблицы может быть решена с использованием алгоритма, рассмотренного в работе [19]. Основная идея алгоритма заключается в раздельном формировании размерностей с установлением иерархий и общей таблицы приложения. Затем значениям размерностей сопоставляются значения меры, и те и другие первоначально должны находиться в одном кортеже общей таблицы приложения.

4 Сравнительный анализ алгоритмов формирования контекстов

В работе [20] рассмотрен метод формирования контекстов по множествам атрибутов $X, Y_1, Y_2, \dots, Y_N, Z_1, Z_2, \dots, Z_N$, задающих размерности и меры композиционной таблицы, причем множества атрибутов задаются без указания отношений, которым они принадлежат (без расширения). Автором этого метода П. Г. Редревым разработаны два алгоритма. Кроме формирования собственно контекстов в алгоритмах для каждого контекста требуется еще определить начальное множество отношений исходной БД, в которых присутствуют атрибуты размерностей и мер.

Первый алгоритм автором назван *алгоритмом полного перебора*, второй — *эвристическим алгоритмом*. Схема функционирования первого алгоритма понятна из его названия, во втором алгоритме к исходному множеству отношений дополняются только те отношения, которые имеют связи с исходным множеством. Сравнительный анализ этих двух алгоритмов выполнен Редревым для схемы БД примера 1, ограниченной отношениями R_1-R_8 .

Для каждого алгоритма подсчитано число итераций, выполненных для нахождения минимального контекста при заданном множестве атрибутов.

Пример 2. Пусть для пользовательского приложения «Предметы» заданы следующие множества атрибутов композиционной таблицы:

$$\begin{aligned} X &= \{\text{№ дня недели, День недели}\}; \\ Y &= \{\text{Код группы}\}; \\ Z &= \{\text{Предмет}\}. \end{aligned}$$

Минимальным контекстом для данного набора атрибутов является следующее множество отношений: $\{R_2, R_3, R_5, R_8\}$. В алгоритме полного перебора по заданному набору атрибутов будет сформировано 2 начальных множества. В результате работы алгоритма полного перебора минимальный контекст найден за **2 итерации**.

Для тех же начальных данных в работе эвристического алгоритма минимальный контекст будет найден за **1 итерацию**.

Пример 3. Пусть для пользовательского приложения «Предметы» заданы следующие множества атрибутов композиционной таблицы:

$$\begin{aligned} X &= \{\text{№ дня недели}\}; \\ Y &= \{\text{№ группы}\}; \\ Z &= \{\text{№ предмета}\}. \end{aligned}$$

Минимальным контекстом для данного набора атрибутов является аналогичное множество отношений $\{R_2, R_3, R_5, R_8\}$.

В результате работы алгоритма полного перебора будет сформировано 24 начальных множества. Минимальный контекст при этом будет найден за **71 итерацию**.

Для тех же начальных данных в работе эвристического алгоритма минимальный контекст будет найден за **2 итерации**.

Рассмотрим работу алгоритма направленного перебора для рассмотренных примеров. Начальное множество отношений будет сформировано пользователем: $\{R_2, R_3, R_5\}$. Далее для обоих примеров будут присвоены веса отношениям, которые не вошли в начальное множество: $R_8 - 3, R_1 - 2, R_7 - 2, R_4 - 0, R_6 - 0$. Необходимый минимальный контекст будет найден на *первой итерации* для обоих примеров без проверки свойства СБПИ, так как замыкание первичного ключа отношения R_8 содержит все атрибуты отношений контекста.

5 Заключение

Рассмотренный в данной работе подход к формированию представлений данных является развитием технологии OLAP. Он ориентирован на работу аналитика, где не требуется быстрая (за доли секунд) реакция системы на запросы, поскольку в большинстве случаев аналитик должен вдумчиво выполнять различные виды анализа над различными представлениями данных с участием ИТ-специалиста. Основным методологическим принципом в данной работе является то, что операционная база данных должна удовлетворять принципам независимости, избыточности, непротиворечивости и т.д. Эта БД является ядром приложений для множества пользователей, а не только для отдельно взятого аналитика.

Сравнительный анализ алгоритмов формирования контекстов показал, что направленный перебор имеет преимущество перед алгоритмом полного перебора, хотя и содержит незначительные дополнительные расходы, связанные с определением весов отношений и упорядочением этих отношений. Это

преимущество обусловлено тем, что выбор отношений вместе с атрибутами на начальном этапе существенно уточняет семантику приложения и сокращает число допустимых вариантов. Эвристический алгоритм может не найти существующий контекст, так как отсутствие непосредственных связей между отношениями не гарантирует отсутствие свойства СБПИ. Направленный перебор рассмотрит все варианты сочетания отношений и последовательно вычислит все существующие контексты аналогично алгоритму полного перебора. Следовательно, направленный перебор предпочтителен при динамическом формировании контекстов.

Литература

1. *Vassiliadis P., Sellis T.* A survey of logical models for OLAP databases // SIGMOD Record, 1999. Vol. 28. No. 4. P. 64–69.
2. *Pedersen T. B., Jensen C. S., Dyreson C. E.* A foundation for capturing and querying complex multidimensional data // Inform. Syst., 2001. Vol. 26. No. 5. P. 383–423.
3. *Lechtenborger J., Vossen G.* Multidimensional normal forms for data warehouse design // Inform. Syst., 2003. Vol. 28. No. 5. P. 415–434.
4. *Li H.-G., Yu H., Agrawal D., Abadi A. E.* Progressive ranking of range aggregates // Data Knowl. Eng., 2007. Vol. 63. No. 1. P. 4–25.
5. *Lehner W., Albrecht J., Wedekind H.* Normal forms for multidimensional databases // 10th Conference (International) on Scientific and Statistical Database Management Proceedings. — Capri: IEEE Computer Society, 1998. P. 63–72.
6. *Giorgini P., Rizzi S., Garzetti M.* Goal-oriented requirement analysis for data warehouse design // DOLAP'05: 8th ACM Workshop (International) on Data Warehousing and OLAP Proceedings. — Bremen: ACM, 2005. P. 47–56.
7. *Mazon J.-N., Trujillo J., Lechtenborger J.* Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms // Data Knowl. Eng., 2007. Vol. 63. No. 3. P. 725–751.
8. *Кукин А. В., Зыкин С. В.* Построение математической модели учебного процесса для долгосрочного планирования // Математические структуры и моделирование, 2002. Вып. 10. С. 77–86.
9. *Редреев П. Г.* Построение иерархий в многомерных моделях данных // Изв. Саратовского ун-та. Сер. Математика. Механика. Информатика, 2009. Т. 9. Вып. 4. Ч. 1. С. 84–87.
10. *Ульман Дж.* Основы систем баз данных / Пер с англ. — М.: Финансы и статистика, 1983. 334 с. (Ullman J. Principles of database systems. — Computer Science Press, 1980. 379 p.)
11. *Мейер Д.* Теория реляционных баз данных / Пер с англ. — М.: Мир, 1987. 608 с. (Maier D. The theory of relational databases. — Computer Science Press, 1983. 637 p.)
12. *Casanova M., Fagin R., Papadimitriou C.* Inclusion dependencies and their interaction with functional dependencies // J. Comput. Syst. Sci., 1984. Vol. 28. No. 1. P. 29–59.
13. *Missaoui R., Godin R.* The implication problem for inclusion dependencies: A graph approach // SIGMOD Record, 1990. Vol. 19. No. 1. P. 36–40.
14. *Levene M., Vincent M. W.* Justification for inclusion dependency normal form // IEEE Trans. Knowl. Data Eng., 2000. Vol. 12. No. 2. P. 281–291.
15. *Miller L., Nilakanta S.* Data warehouse modeler: A CASE tool for warehouse design // 31st Annual Hawaii Conference (International) on System Sciences. — Kohala Coast: IEEE Computer Society, 1998. Vol. 6. P. 42–48.
16. *Зыкин С. В., Полуянов А. Н.* Формирование представлений данных с контекстными ограничениями // Омский научный вестник. Сер. Приборы, машины и технологии, 2008. № 1(64). С. 141–144.
17. *Зыкин С. В.* Построение отображения реляционной базы данных в списковую модель данных // Управляющие системы и машины, 2001. № 3. С. 42–63.
18. *Зыкин С. В.* Оценка мощности списка для отображения типа «частичное соединение» // Кибернетика и системный анализ, 1993. № 6. С. 142–152.
19. *Зыкин С. В.* Формирование гиперкубического представления данных со списочными компонентами // Информационные технологии и вычислительные системы, 2010. № 4. С. 38–46.
20. *Редреев П. Г.* Построение табличных приложений со списочными компонентами // Информационные технологии, 2009. № 5. С. 7–12.

Поступила в редакцию 09.01.13

DYNAMIC CONTEXTS OF RELATIONAL-TYPE DATABASE

S. V. Zykin

Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 4 Acad. Koptyug Av., Novosibirsk 630090, Russian Federation

Abstract: The technology of dynamic formation of data presentation is suggested. This technology is the development of online analytical processing. Data source is a relational database with any scheme (not necessarily hierarchical). Target data presentation is a composite table which allows to present multivariate data on a plane. This table assumes separate formation of dimensions with the following juxtaposition of measures to dimensions in the table. The foundation of data presentation is a table of connected joins, which satisfies contextual and logic restrictions. The algorithms used to form such tables are suggested and their properties are investigated. Special attention is given to contexts which are used to form tables of connected joins. The algorithm of directed search for creation of contexts is proposed and comparative analysis of algorithms of contexts formation is performed on an example. The investigated properties of contexts and the offered algorithms are intended to automate user work to form new data presentations.

Keywords: relational database; context; lossless join; composite table

DOI: 10.14357/19922264140108

Acknowledgments

The financial support of the Russian Foundation for Basic Research (grant No. 12-07-00066-a) is acknowledged.

References

- Vassiliadis, P., and T. Sellis. 1999. A survey of logical models for OLAP databases. *SIGMOD Rec.* 28(4):64–69. doi: 10.1145/344816.344869.
- Pedersen, T. B., C. S. Jensen, and C. E. Dyreson. 2001. A foundation for capturing and querying complex multidimensional data. *Inform. Syst.* 26(5):383–423. doi: 10.1016/S0306-4379(01)00023-0.
- Lechtenborger, J., and G. Vossen. 2003. Multidimensional normal forms for data warehouse design. *Inform. Syst.* 28(5):415–434. doi: 10.1016/S0306-4379(02)00024-8.
- Li, H.-G., H. Yu, D. Agrawal, and A. E. Abbadi. 2007. Progressive ranking of range aggregates. *Data Knowl. Eng.* 63(1):4–25. doi: 10.1016/j.datak.2006.10.008.
- Lehner, W., J. Albrecht, and H. Wedekind. 1998. Normal forms for multidimensional databases. *10th Conference (International) on Scientific and Statistical Database Management Proceedings*. Capri: IEEE Computer Society. 63–72.
- Giorgini, P., S. Rizzi, and M. Garzetti. 2005. Goal-oriented requirement analysis for data warehouse design. *8th ACM Workshop (International) on Data Warehousing and OLAP Proceedings*. Bremen: ACM. 47–56.
- Mazon, J.-N., J. Trujillo, and J. Lechtenborger. 2007. Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms. *Data Knowl. Eng.* 63(3):725–751. doi: 10.1016/j.datak.2007.04.004.
- Kukin, A. V., and S. V. Zykin. 2002. Postroenie matematicheskoy modeli uchebnogo protsessa dlya dolgosrochnogo planirovaniya [Creation of the mathematical model of organization of the study process for long-term planning]. *Matematicheskie struktury i modelirovanie [Mathematical Structures and Modeling]* 10:77–86.
- Redreev, P. G. 2009. Postroenie ierarkhiy v mnogomernykh modelyakh dannykh [Construction of hierarchies in multidimensional data models]. *Izvestiya Saratovskogo Universiteta. Seriya Matematika. Mekhanika. Informatika [News of Saratov University. Mathematics. Mechanics. Computer Science ser.]* 9(4-1):84–87.
- Ullman, J. 1980. *Principles of database systems*. Computer Science Press. 379 p.
- Maier, D. 1983. *The theory of relational databases*. Computer Science Press. 637 p.
- Casanova, M., R. Fagin, and C. Papadimitriou. 1984. Inclusion dependencies and their interaction with functional dependencies. *J. Comput. Syst. Sci.* 28(1):29–59. doi: 10.1145/588111.588141.
- Missaoui, R., and R. Godin. 1990. The implication problem for inclusion dependencies: A graph approach. *SIGMOD Record* 19(1):36–40. doi: 10.1145/382274.382402.
- Levene, M., and M. W. Vincent. 2000. Justification for inclusion dependency normal form. *IEEE Trans. Knowl. Data Eng.* 12(2):281–291. doi: 10.1109/69.842267.
- Miller, L., and S. Nilakanta. 1998. Data warehouse modeler: A CASE tool for warehouse design. *31st Annual Hawaii Conference (International) on System Sciences*. Kohala Coast: IEEE Computer Society. 42–48.

16. Zykin, S. V., and A. N. Poluyanov. 2008. Formirovanie predstavleniy dannykh s kontekstnymi ogranicheniyami [Formation of the data representations with contextual restrictions]. *Omskiy Nauchnyy Vestnik [Omsk Scientific Bulletin]* 1(64):141–144.
17. Zykin, S. V. 2001. Postroenie otobrazheniya relyatsionnoy bazy dannykh v spiskovuyu model' dannykh [Construction of mapping of a relational database at the list model of data]. *Upravlyayushchie Sistemy i Mashiny [Control Systems and Machines]* 3:42–63.
18. Zykin, S. V. 1993. Otsenka moshchnosti spiska dlya otobrazheniya tipa “chastichnoe soedinenie” [The list size estimation for mapping of “partial join” type]. *Kibernetika i Sistemnyy Analiz [Cybernetics and Systems Analysis]* 6:142–152.
19. Zykin, S. V. 2010. Formirovanie giperkubicheskogo predstavleniya dannykh so spisochnymi komponentami [Formation of the hypercubic data representation with list components]. *Informatsionnye tekhnologii i vychislitel'nye sistemy [Information Technologies and Computing Systems]* 4:38–46.
20. Redreev, P. G. 2009. Postroenie tablitsnykh prilozheniy so spisochnymi komponentami [Construction of applications with the list components]. *Informatsionnye Tekhnologii [Information Technologies]* 5:7–12.

Received January 9, 2013

Contributor

Zykin Sergey V. (b. 1959) — Doctor of sciences in technology, professor, Head of laboratory, Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 4 Acad. Koptyug Av., Novosibirsk 630090, Russian Federation, szykin@mail.ru

ИНТЕГРАЛЬНОЕ МОДЕЛИРОВАНИЕ ЯЗЫКОВЫХ СТРУКТУР В ЛИНГВИСТИЧЕСКИХ ПРОЦЕССОРАХ СИСТЕМ ОБРАБОТКИ ЗНАНИЙ И МАШИННОГО ПЕРЕВОДА*

Е. Б. Козеренко¹

Аннотация: Данная статья посвящена проблемам исследования и интегрального моделирования когнитивно-лингвистических представлений языковых структур и механизмов разрешения синтаксической неоднозначности в процессе создания лингвистических процессоров интеллектуальных систем обработки знаний и машинного перевода. Методика представления языковых структур и разрешения их неоднозначности построена на основе логико-лингвистических правил и векторных пространств. Проведено сравнительное исследование методов классификации применительно к лингвистическим задачам; выработан эффективный метод отображения вектора естественно-языковых структур в расширенное пространство признаков для классификации новых языковых объектов и структур; сформирована фокусная выборка параллельных текстов деловых и научных документов на русском, английском и французском языках по различным отраслям науки и техники; сформирована расширенная система новых категорий, повышающая изобразительные возможности исходного варианта унификационно-порождающей грамматики; выработаны пути развития базовых представлений на основе аппарата расширенных семантических сетей и метод применения векторных пространств, обеспечивающих разрешение неоднозначности ключевых языковых структур в процессе синтаксического анализа текста для извлечения знаний и машинного перевода. Разработан грамматический формализм и алгоритмические представления парсера, в котором учитываются реальные трудности перевода, такие как языковые трансформации.

Ключевые слова: параллельные тексты; векторные пространства; синтаксис; семантика; фразовые структуры; гибридные модели; машинный перевод

DOI: 10.14357/19922264140109

1 Введение

В данной работе рассматривается проблема создания целостной математико-лингвистической модели языковых структур для лингвистических процессоров информационных систем на основе синергетического подхода, в котором применяются лингвистические знания, логико-семантические представления, статистические методы и механизмы машинного обучения для извлечения новых грамматических правил из текстовых корпусов и разрешения неоднозначности. Для формализации лингвистических знаний используется когнитивная трансферная грамматика (КГТ) [1, 2], учитывающая многовариантность и неоднозначность языковых структур, языковую синонимию, включая семантические соответствия на различных языковых уровнях и их вероятностные характеристики. Проведены эксперименты с грамматикой категориального типа [3].

Методика построения интегрированной когнитивно-лингвистической модели представления языковых структур и разрешения их неоднознач-

ности основана на сочетании логико-лингвистических правил и векторных пространств семантических характеристик фразовых структур, отражающих категориально-функциональные значения. Для этого было сделано следующее:

- проведено сравнительное исследование методов классификации применительно к лингвистическим задачам;
- выработан эффективный метод отображения вектора естественно-языковых структур в расширенное пространство признаков для классификации новых языковых объектов и структур;
- сформирована фокусная выборка параллельных текстов деловых и научных документов на русском, английском и французском языках по различным отраслям науки и техники;
- сформирована расширенная система новых категорий для повышения изобразительных возможностей исходного варианта унификационно-порождающей грамматики;
- выработаны пути расширения базовых представлений на основе аппарата расширенных семантических сетей [4] и результатов приме-

* Работа частично поддержана Российским фондом фундаментальных исследований, грант № 11-06-00476-а.

¹ Институт проблем информатики Российской академии наук, kozerenko@mail.ru

нения метода векторных пространств, направленного на разрешение неоднозначности языковых структур для синтаксического разбора при распознавании текста в процессе извлечения знаний и машинного перевода.

Особое внимание в представленных исследованиях уделяется явлениям синтаксической неоднозначности, вызывающим основные затруднения при создании систем обработки естественного языка, созданию оптимальных методик разрешения неоднозначности на основе статистического подхода и семантического выравнивания параллельных текстовых корпусов. Целью исследований является построение эффективных методов структурного анализа и компьютерного моделирования полнотекстовых научных и патентных документов, а также документов деловой сферы коммуникации, таких как финансы и управление. Разработана теоретическая концепция и алгоритмы автоматического выравнивания параллельных текстов для пополнения многоязычных фразеологических словарей и развития грамматических компонент систем машинного перевода и обработки знаний. Основным результатом исследований — модель лингвистической составляющей интеллектуальных информационно-систем, работающих в многоязычном пространстве для поиска информации, обеспечения оптимальных аналитических и управленческих решений.

Сформирована инструментальная среда исследования параллельных текстов и подготовки синтаксико-семантических представлений для проектирования лингвистических процессоров интеллектуальных систем. Результаты исследований применяются в логико-семантических и статистических процедурах обработки слабоструктурированной текстовой информации, при разработке технологии и инструментальных средств построения лингвистических компонент интеллектуальных систем и систем машинного перевода.

2 Моделирование динамического аспекта языковых преобразований

Процедуры анализа и синтеза естественно-языковых высказываний отражают динамический характер языка как деятельности; соответственно, в модели, которая кладется в основу проекта системы обработки естественно-языковых высказываний, должен быть заложен механизм, позволяющий строить представления движения.

Семантическое моделирование на основе правил служит основой для автоматического извлече-

ния лингвистических данных с помощью машинного обучения. Исследования, представленные в данной статье, основаны на традициях сильного семантического подхода отечественной лингвистики: построение модели от значения — к форме. Представленный в данной работе подход на основе КТГ дает возможность компактного представления структуры составляющих предложения (грамматика фразовых структур), с одной стороны, а с другой стороны, учитывает механизмы зависимости между узлами дерева предложения. Ядро КТГ составляют прототипические структуры исследуемых языков (русского, английского и французского), их наиболее вероятные позиции в предложении, а также статистические данные о дистрибутивных характеристиках структур (т. е. информация о контекстных условиях употребления исследуемых объектов — о структурных контекстах), схемы полного разбора предложений. Методы машинного обучения на основе векторных моделей развиваются и используются в различных областях знаний, применительно к лингвистическим задачам эти методы вполне эффективны для разрешения лексической многозначности [5–10].

Более сложной задачей и новым направлением исследований возможности применения векторных моделей для представления и обработки лингвистических данных является моделирование грамматических преобразований на основе векторных пространств и тензоров. Тензор (от *лат.* *tensus* — напряженный) — объект линейной алгебры, преобразующий элементы одного линейного пространства в элементы другого. Часто тензор представляют как многомерную таблицу, заполненную числами — компонентами тензора $d \times d \times \dots \times d$, где d — размерность, над которым задан тензор, а число сомножителей совпадает с так называемой валентностью, или рангом тензора. Важно, что такое представление (кроме скаляров, т. е. тензоров валентности ноль) возможно только после выбора базиса (или системы координат): при смене базиса компоненты тензора меняются определенным образом. Сам тензор как «геометрическая сущность» от выбора базиса не зависит, компоненты вектора меняются при смене координатных осей, но сам вектор, образом которого может быть просто нарисованная стрелка, от этого не изменяется. Тензор обычно обозначают некоторой буквой с совокупностью верхних (контрвариантных) и нижних (ковариантных) индексов: $X_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r}$. При смене базиса ковариантные компоненты меняются так же, как и базис (с помощью того же преобразования), а контрвариантные — обратным изменению базиса (обратным преобразованием). Тензор является сущностью любой системы реального мира и со-

храняется, несмотря на происходящие изменения в этой системе [11]. Эта особенность тензора чрезвычайно актуальна для моделирования языковых преобразований в лингвистических процессорах, когда необходимо выявлять сходные значения, выраженные многочисленными способами, системой разнородных языковых средств

В работе, представленной в данной статье, рассмотрены два основных подхода к представлению смысла в вычислительной лингвистике: символичный подход [12, 13] и подход на основе дистрибутивной семантики [5–7, 10], поставлен основополагающий вопрос о том, как эти подходы могут быть объединены для достижения оптимального результата. Решение заключается в сочетании методов компьютерной лингвистики и когнитивной науки, в котором символическое и «коннекционистское» (т. е. основанное на модели машинного обучения, например *нейронных сетей*) представления объединяются с помощью тензорных произведений. Исследованы возможные применения данного метода для обработки синтаксических структур и контекстов в русском, английском и французском языках и проведены межъязыковые сопоставления.

Изучены потенциально плодотворные связи между вычислительной лингвистикой и другими смежными областями, например поиском информации и машинным переводом. Для систем обработки естественного языка наиболее сложной проблемой является установление семантической эквивалентности языковых структур [14–17], имеющих сходное значение, но по-разному оформленных. В частности, для машинного перевода наиболее сложной проблемой является реализация языковых трансформаций, которые необходимо производить при переводе с одного языка на другой. Текущий этап развития систем машинного перевода характеризуется междисциплинарными исследованиями в области когнитивной семантики и вероятностных языковых моделей и разработкой семантико-синтаксических представлений, учитывающих многозначность и неоднозначность синтаксических структур [1, 2, 4, 6, 8–10, 15–18].

3 Математические модели структуры предложения на основе категориальных грамматик

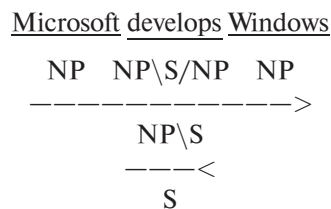
Модели векторных пространств находят все более широкое применение в исследованиях, связанных с семантическими представлениями естественного языка, и имеют разнообразный спектр потенциальных и действующих приложений [6, 10,

19]. Основными сферами применения дистрибутивных моделей являются: разрешение лексической неоднозначности, информационный поиск, кластеризация документов, автоматическое формирование словарей (словарей семантических отношений, двуязычных словарей), создание семантических карт, моделирование перифраз, определение тематики документа, определение тональности высказывания, выявление именованных сущностей. В качестве вычислительного инструмента и способа представления моделей используется линейная алгебра.

Информация о дистрибуции лингвистических единиц представляется в виде многомерных векторов, а семантическая близость между лингвистическими единицами вычисляется как расстояние между векторами. Многомерные векторы образуют матрицу, где каждый вектор соответствует лингвистической единице (слову или словосочетанию), а каждое измерение вектора соответствует контексту (документ, параграф, предложение, словосочетание, слово). Для вычисления меры близости между векторами могут использоваться различные формулы: расстояние Минковского, расстояние Манхэттена, евклидово расстояние, расстояние Чебышёва, скалярное произведение, косинусная мера. Наиболее популярной является косинусная мера [6, 10].

Формально-грамматической основой [20–25] исчисления семантики предложения с учетом его структурных составляющих в векторных моделях [26–30] служат различные варианты категориальных грамматик, которые позволяют отразить комбинаторные свойства фразовых структур в их линейной реализации во фразе. В категориальных грамматиках в качестве базовых понятий используются синтаксические типы, или *категории* (приблизительно соответствуют тому, что в традиционных грамматиках называется *частями речи*), а механизм *наложения* (*juxtaposition, superposition*) позволяет выразить функциональные значения структурных компонентов, которые реализуются в конкретном высказывании.

Методы исчисления структуры предложения на основе категориальных грамматик восходят к работам Айдукевича [20] и Бар-Хиллела [21]. Дери́вация предложения в категориальных грамматиках [22, 23] выглядит следующим образом:



В работах [22, 24] был представлен аппарат грамматики на основе предгрупп. Предгруппа — это такой частично упорядоченный моноид, в котором каждый элемент a^l имеет левый адьюнкт a^l и правый адьюнкт a^r такие, что

$$a^l \cdot a \rightarrow 1; a \cdot a^r \rightarrow 1.$$

Такой моноид — это множество грамматических типов $(NP, NP^l, NP^r, NP^{ll}, NP^{rr}, S, PP, \dots)$ с оператором *наложения* (\cdot) , которые используются для деривации сложных типов и пустой цепочки как единицы (1): $NP \cdot (NP^r \cdot S \cdot NP^l) \cdot NP$.

Частичное упорядочение кодирует отношение деривации для более ранних дериваций или редукций:

$$NP \cdot (NP^r \cdot S \cdot NP^l) \cdot NP \rightarrow 1; \\ (S \cdot NP^l) \cdot NP \rightarrow 1; S \cdot 1 \rightarrow S.$$

На абстрактном математическом уровне *теории категорий* алгебра предгрупп и векторные пространства могут рассматриваться как эквивалентные.

В целом категориальные грамматики ориентированы на алгебраический подход к моделированию естественного языка, что делает их вполне органичной основой для интегрального представления семантики предложений естественного языка в виде семантических векторных пространств (СВП). Важным понятием в категориальных грамматиках является *композициональность*: правила деривации предложения в целом по его составляющим с учетом их линейного расположения во фразе. В современных векторных моделях семантики предложения с учетом его синтаксиса наиболее популярны комбинаторная категориальная грамматика Стивмана (CCG) [23], грамматика Ламбека [22, 24], применяются также формализмы Муртгата [25], аппликативная грамматика Шаумяна [3].

Тензоры используются для выражения преобразований синтаксической структуры.

4 Семантический подход к выявлению статистических характеристик языковых структур в параллельных корпусах

Категориальные грамматики активно применяются для сопоставления текстовых корпусов и извлечения перифраз. Параметры моделей статистического выравнивания оптимизируются с учетом

критерия максимального сходства, который далеко не всегда отражает качество выравнивания. В этой связи чрезвычайную важность имеет проблема перифразирования, и ей уделяется все больше внимания в работах ведущих исследовательских групп в области компьютерной лингвистики.

Извлечение перифраз из двуязычных параллельных корпусов было описано в работах [14–17], которые выводили перифразы с использованием методов статистического машинного перевода на основе фраз. Затем путем введения комплексных синтаксических меток вместо использования только нетерминальных символов из деревьев разбора авторы смогли добиться существенного улучшения по сравнению с основным методом. Синтаксические ограничения значительно улучшают качество этого метода перифразы.

В работе [16] представлен новый подход к перифразе на основе двуязычного корпуса. Автор демонстрирует, что более высокое качество может быть достигнуто, если ввести ограничение, что перифраза должна иметь такой же синтаксический тип, что и исходная фраза. В работе предложены ограничения на перифразы на двух этапах: когда они выводятся на основе разобранных параллельных корпусов и когда они подставляются в разобранные тестовые предложения. Автор ввел синтаксические ограничения, пометив все фразы и перифразы (даже не входящие в число составляющих) с помощью слэш-категорий, используемых в CCG. Однако автор не дает ни формального определения некоторой синхронной грамматики, ни предлагает систему декодирования, поскольку его система представлений не содержит иерархических правил (деревьев разбора). Правила синхронной грамматики для перевода извлекаются из пар предложений, автоматически разобранных и выравненных по словам. Методы извлечения варьируют в зависимости от того, извлекают ли они только минимальные правила для фраз, у которых есть доминирующие узлы в дереве разбора, или более сложные правила, включающие фразы, не входящие в число составляющих [17].

Главной мотивацией для использования синтаксических перифраз наряду с их исходными фразовыми соответствиями является их потенциальная возможность в более общем виде отразить лингвистические трансформации, сохраняющие значение.

Необходимость моделирования языковых трансформаций для систем машинного перевода и извлечения знаний из текстов обусловлена тем, что до сих пор эти явления мало исследованы с точки зрения возможностей их компьютерной реализации и, соответственно, недостаточно учтены в действующих системах машинного перевода и обработки знаний. В КТГ [1] функциональ-

ные значения языковых структур определяются категориальными значениями головных вершин, а именно представляются как наложение категорий (*juxtaposition*), что имеет сходство с механизмами категориальных грамматик. Вероятностные характеристики вводятся в правила унификационной грамматики в виде весов, присваиваемых деревьям разбора. Когнитивная трансферная грамматика, разработанная автором данной статьи, содержит механизмы категориальных грамматик, такие как категоризация и наложение категориальных значений для выражения функциональных значений фразовых структур, а также механизмы унификационно-порождающих грамматик для анализа и порождения не только отдельных типов фразовых структур, но и предложения в целом.

По определению, КТГ исходно является двуязычной грамматикой. В КТГ элементарными структурами являются трансфемы. Трансфема — это единица когнитивного переноса, устанавливающая функционально-семантическое соответствие между структурами исходного языка и структурами целевого языка. Для процедур анализа и перевода в лингвистических процессорах трансфемы задаются как правила переписывания, в которых в левой части стоит нетерминальный символ, а в правой — выравненные пары цепочек терминальных и нетерминальных символов, принадлежащих исходному и целевому языкам.

Процессы синтаксических трансформаций очень сходны в ряде европейских языков. Включение статистических данных в системы, основанные на правилах, позволяет отразить динамику и разнообразие языковых форм и значений, порождаемых в процессе речевой деятельности. Функциональная и когнитивная мотивация правил исходной грамматики позволяет увеличить точность соответствий в среднем на 34% в зависимости от типа сопоставляемых текстов.

Применяется интегральный подход на основе правил формальной грамматики (грамматика когнитивного трансфера — унификационно-порождающая грамматика на основе фразовых структур) и дистрибутивной семантики. Дистрибутивная семантика — область научных исследований, занимающаяся вычислением степени семантической близости между лингвистическими единицами на основании их дистрибутивных (контекстных) признаков в больших массивах лингвистических данных. Информация о дистрибуции лингвистических единиц представляется в виде многоуровневых векторов, а семантическая близость между лингвистическими единицами вычисляется как расстояние между векторами. Многоуровневые векторы образуют СВП, где каждый вектор соответствует

лингвистической единице (слово, словосочетание, языковая структура), а каждое измерение вектора соответствует контексту (документ, параграф, предложение, словосочетание, слово, языковая структура). Одной из главных целей исследования является применение модели СВП для извлечения переводных соответствий языковых структур из параллельных текстов.

Для разрешения неоднозначности языковых объектов и структур используются векторные пространства [18, 19].

Концепция СВП впервые была реализована в информационно-поисковой системе SMART. Идея СВП состоит в представлении каждого документа из коллекции в виде точки в пространстве, т. е. вектора в векторном пространстве. Точки, расположенные ближе друг к другу в этом пространстве, считаются более близкими по смыслу. Пользовательский запрос рассматривается как псевдодокумент и тоже представляется как точка в этом же пространстве. Документы сортируются в порядке возрастания расстояния, т. е. в порядке уменьшения семантической близости от запроса, и в таком виде предоставляются пользователю. Впоследствии концепция СВП была успешно применена для других семантических задач.

Для выделения значимых словосочетаний в компьютерной лингвистике используются различные статистические меры (меры ассоциации, меры ассоциативной связанности — *англ.* association measures), вычисляющие силу связи между элементами в составе коллокации. Мера MI (mutual information), введенная в работе [31], сравнивает зависимые контекстно-связанные частоты с независимыми частотами слов в тексте. Если значение MI превосходит определенное пороговое значение, то словосочетание считают статистически значимым.

5 Создание интегральной кросслингвистической модели предложения на основе расширяемой грамматики

Основная цель исследований, проводимых автором данной работы, — извлечение из параллельных текстов на разных языках таких фразовых структур, которые выражают одинаковые значения, и включение их в систему правил расширяемой грамматики для решения задач машинного перевода и извлечения знаний из многоязычных предметно-ориентированных текстов. Расширяемая грамматика, которая при этом используется,

содержит когнитивные и функциональные характеристики фразовых структур и базируется на формализме КТГ [1].

В статистической исследовательской парадигме термин «фраза» означает произвольный сегмент предложения, выделенный статистическим инструментом и вовсе не являющийся нетерминальным символом (синтаксической единицей) какой-либо формальной грамматики. В отличие от такого подхода в представленных исследованиях фраза — это синтаксически значимая единица в составе предложения, которая рассматривается в парадигматическом и синтагматическом аспектах. Основные проблемы выравнивания параллельных текстов и, соответственно, обучения статистических процессоров естественного языка обусловлены наличием значительных трансформаций предложений исходного и целевого текстов, поскольку каждый язык использует свои специфические механизмы описания референтной ситуации. Создание и исследование параллельных корпусов в рамках представленной здесь работы направлено на выявление наиболее типичных (частотных) и значимых трансформаций в изучаемых параллельных текстах и развитие типологии трансформаций, представленных в исходном формализме КТГ. Таким образом, необходимо разработать и применить такие стратегии и инструменты, которые обеспечивают наиболее адекватные средства для описания и выявления сопоставимых языковых структур. При этом сочетаются подходы КТГ и категориальной грамматики *SUG* (*semiotic universal grammar*) [2, 3]. Динамика выравнивания реализуется в соответствиях M , которые фиксируются посредством механизма категориальной грамматики *SUG*.

Семантически-ориентированный этап выравнивания параллельных текстов проводится в двух режимах:

- (1) сопоставления на уровне трансфем, при этом выявляются языковые структуры, выражающие сходные функциональные значения в параллельных текстах;
- (2) сопоставления на уровне концептов и отношений.

Режим выравнивания первого типа базируется на понятиях трансфемы T (*transfeme*) и соответствия M (*match*).

Определение. Трансфема T — это единица парадигматического плана, относящаяся к языку как системе, соответствие M — единица синтагматического плана, относящаяся к речи (дискурсу); таким образом, трансфемы T реализуются в соответствиях M .

Соответствие M может быть шире, чем трансфема T , и часто включает контекст.

Выравнивание на основе концептов (сущностей) и отношений (связей) параллельных и концептуально-сопоставимых текстов на различных языках направлено на выявление языковых реализаций структур знаний и формирования многоязычных баз знаний, которые затем будут применяться в интеллектуальных аналитических системах. Такой режим выравнивания будем называть концептуально-ориентированным, инструментом для него служит лингвистический процессор *Semantix*. Механизмы концептуально-ориентированного выравнивания основаны на аппарате расширенных семантических сетей (РСС) [4], обладающих достаточной выразительной мощностью для представления естественно-языковых структур с высокой степенью вложенности и выполняющих роль языка-посредника. Все множество языковых объектов задается в виде предикатно-аргументных структур. Анализ предложения производится на основе унификационной грамматики. Модели управления и трансформационные свойства задаются в словаре в рамках словарных статей глаголов. В результате трансформаций происходит сдвиг моделей управления.

Особое внимание в исследованиях уделяется номинализации и изменению предложного управления на беспредложное, например: *стрелять по уткам* — *стрелять уток*.

Таким образом, концептуально-ориентированное выравнивание — это процесс извлечения знаний в многоязычном режиме и наполнения базы лингвистических знаний для последующего использования в системах машинного перевода и аналитической обработки текстовых знаний.

6 Заключение

Успешность вероятностной модели в значительной степени зависит от адекватного определения события. Обычный тип события в вероятностной обработке естественного языка — совместное появление одного или нескольких слов в определенном контексте. Для построения интегральной модели предложения соответствия, или события, составляют множество хорошо оформленных *нетерминалов* (т. е. грамматических категорий и типов синтаксических структур), а динамически выявляемые контекстные структуры представляются в нотации категориальной грамматики (*SUG*) [2, 3]. Для разрешения неоднозначности языковых объектов и структур используются векторные пространства. Основным прием выравнивания параллельных тек-

стов по трансфемам — установление соответствий между головными вершинами фразовых структур, в настоящее время осуществляется в полуавтоматическом режиме, автоматическое сопоставление находится в стадии разработки: выравнивание по головным элементам фраз и сравнение фразовых структур. Методы машинного обучения обеспечивают системе возможность извлекать значения, которые ожидаются в рамках определенных контекстов. Для вычисления вероятности реализации определенного значения в соответствующем контексте используются статистические данные, полученные из больших массивов текстовой информации (параллельных текстовых корпусов).

В результате использования логико-лингвистических правил анализа и перевода фразовых структур, представленных на основе аппарата расширенных семантических сетей, и метода векторных пространств построено системное межязыковое представление лингвистических объектов, отражающее взаимосвязи типа структура—свойство; сформирована представительная обучающая выборка семантически выровненных параллельных текстов; собраны и систематизированы статистические данные о типах смысловых соответствий языковых структур в русском, английском и французском деловом и научном тексте; построены базовые векторные пространства признаков неоднозначных языковых структур; выявлены и формализованы различные типы линейаризации структур при разборе и переводе; проведено сравнительное исследование статистических методов анализа языковых структур с целью формирования оптимального аппарата вероятностных расширений логико-лингвистических правил и выработки методов создания логико-статистических алгоритмов разбора и трансфера языковых структур.

Литература

1. *Kozerenko E. B.* Cognitive approach to language structure segmentation for machine translation algorithms // Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA: CSREA Press, 2003. P. 49–55.
2. *Козеренко Е. Б.* Стратегии выравнивания параллельных текстов: семантические аспекты // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 82–89.
3. *Shaumyan S.* Categorical grammar and semiotic universal grammar // IC-AI'03: Conference (International) on Artificial Intelligence Proceedings. — Las Vegas, USA: CSREA Press, 2003. P. 623–629.
4. *Kuznetsov I. P., Kozerenko E. B., Matskevich A. G.* Intelligent extraction of knowledge structures from natural language texts // 2011 IEEE/WIC/ACM Joint Conferences (International) on Web Intelligence and Intelligent Agent Technology — Workshops WI-IAT 2011 Proceedings. — Washington, DC, USA: IEEE Computer Society, 2011. Vol. 03. P. 269–272.
5. *Dempster A. P., Laird N. M., Rubin D. B.* Maximum likelihood from incomplete data via the EM algorithm // J. Roy. Stat. Soc. B, 1977. Vol. 39. No. 1. P. 1–22.
6. *Lund K., Burgess C.* Producing high-dimensional semantic spaces from lexical co-occurrence // Behav. Res. Meth. Instr. Comp., 1996. Vol. 28. No. 2. P. 203–208.
7. *Curran J. R.* From distributional to semantic similarity. PhD Thesis. — Edinburgh: University of Edinburgh, 2004. 177 p.
8. *McCarthy D., Koeling R., Weeds J., Carroll J.* Finding predominant senses in untagged text // 42nd Annual Meeting of the Association for Computational Linguistics Proceedings. — Barcelona, Spain: ACL, 2004. P. 280–287.
9. *Clark S., Pulman S.* Combining symbolic and distributional models of meaning // AAAI Spring Symposium on Quantum Interaction Proceedings. — Palo Alto, CA, USA: AAAI Press, 2007. P. 52–55. <http://www.cl.cam.ac.uk/~sc609/pubs/aaai07.pdf>.
10. *Morozova Yu.* Method for extracting translation correspondences from a parallel corpus // ICAI'13, WORLD-COMP'13 Proceedings. — Las Vegas, USA: CSREA Press, 2013. Vol. II. P. 65–69.
11. *Danielson D. A.* Vectors and tensors in engineering and physics. — 2nd ed. — Boulder, CO: Westview (Perseus), 2003. 282 p.
12. *Montague R.* Universal grammar // Theoria, 1970. Vol. 36. P. 373–398. (Reprinted: Formal Philosophy: Selected papers of Richard Montague / Ed. R. H. Thomason. — New Haven — London: Yale University Press, 1974. P. 222–246.)
13. *Partee B.* Compositionality // Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee. — Oxford: Blackwell, 2004. P. 153–181.
14. *Pang B., Knight K., Marcu D.* Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences // NAACL'03: 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings. — Stroudsburg, PA, USA: ACL, 2003. Vol. 1. P. 102–109.
15. *Bannard C., Callison-Burch C.* Paraphrasing with bilingual parallel corpora // 43rd Annual Meeting of the ACL Proceedings. — Stroudsburg, PA, USA: ACL, 2005. P. 597–604.
16. *Callison-Burch C.* Syntactic constraints on paraphrases extracted from parallel corpora // EMNLP-2008 Proceedings. — Stroudsburg, PA, USA: ACL, 2008. P. 196–205.
17. *Ganitkevitch Ju., Callison-Burch C., Napoles C., Van Durme B.* Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation // 2011 Conference on Empirical Methods in Natural Language Processing Proceedings. — Stroudsburg, PA, USA: ACL, 2011. P. 1168–1179.

18. *Bogatyrev K.* In defense of symbolic NLP // MLMTA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA: CSREA Press, 2006. P. 63–68.
19. *Malkov K. V., Tunitsky D. V.* On extreme principles of machine learning in anomaly and vulnerability assessment // MLMTA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA: CSREA Press, 2006. P. 24–29.
20. *Ajdukiewicz K.* Die Syntaktische Konnexitat // Stud. Philos., 1935. Vol. 1. No. 1. P. 1–27.
21. *Bar-Hillel Y.* A quasiarithmetical notation for syntactic description // Language, 1953. Vol. 29. No. 1. P. 47–58.
22. *Lambek J.* The mathematics of sentence structure // Am. Math. Mon., 1958. Vol. 65. No. 3. P. 154–170.
23. *Steedman M.* Surface structure and interpretation. — Massachusetts: MIT Press, 1996. 140 p.
24. *Lambek J.* From word to sentence: A computational algebraic approach to grammar. — Monza, Italy: Polimetrica Publ., 2008. 154 p.
25. *Moortgat M.* Symmetric categorial grammar // J. Philos. Logic, 2009. Vol. 38. No. 6. P. 681–710.
26. *Gazdar G.* Paradigm merger in natural language processing // Computing tomorrow: Future research directions in computer science / Eds. R. Milner, I. Wand. — Cambridge, U.K.: Cambridge University Press, 1996. P. 88–109.
27. *Clark S., Curran J. R.* Wide-coverage efficient statistical parsing with CCG and log-linear models // Comput. Linguist., 2007. Vol. 33. No. 4. P. 493–552.
28. *Baroni M., Zamparelli R.* Nouns are vectors, adjectives are matrices: Representing adjective–noun constructions in semantic space // 2010 Conference on Empirical Methods in Natural Language Processing Proceedings. — Stroudsburg, PA, USA: ACL, 2010. P. 1183–1193.
29. *Grefenstette E., Sadrzadeh M.* Experimental support for a categorical compositional distributional model of meaning // Conference on Empirical Methods in Natural Language Processing Proceedings. — Stroudsburg, PA, USA: ACL, 2011. P. 1394–1404.
30. *Hermann K. M., Blunsom P.* The role of syntax in vector space models of compositional semantics // 51st Annual Meeting of the Association for Computational Linguistics Proceedings. — Stroudsburg, PA, USA: ACL, 2013. P. 894–904.
31. *Church K., Hanks P.* Word association norms, mutual information, and lexicography // Comput. Linguist., 1996. Vol. 16. No. 1. P. 22–29.

Поступила в редакцию 26.12.13

INTEGRATED MODELING OF LANGUAGE STRUCTURES FOR LINGUISTIC PROCESSORS OF KNOWLEDGE MANAGEMENT AND MACHINE TRANSLATION SYSTEMS

E. B. Kozerenko

Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper is dedicated to research and integrated modeling of cognitive linguistic representations of language structures and the mechanisms for resolution of syntactic ambiguity in the process of creating linguistic processors for intelligent knowledge processing and machine translation systems. The technique of constructing the hybrid cognitive linguistic model for representation of language structures and resolution of their ambiguity on the basis of logical linguistic rules and vector spaces has been developed and specified. The method presented is new and rests on the modern level of development in science and technology. The following tasks have been carried out: comparative research of classification methods has been made with respect to linguistic problems; the effective method of mapping the vector of natural language structures into the expanded space of attributes for classification of new language objects and structures has been worked out; the focal sample of parallel texts of business and scientific documents in Russian, English, and French has been developed; the expanded system of new categories enhancing the representational power of the initial grammar variant has been formed; the extended semantic networks were employed for unified representation of the matched language structures and the experiments of vector spaces method application for resolution of syntactic ambiguity of the key language structures were performed; the grammatical formalism and the algorithmic representation have been designed of the parser in which the difficulties of translation including the language transformations are taken into account.

Keywords: parallel texts; vector spaces; syntax; semantics; phrase structures; integrated models; machine translation; knowledge management

DOI: 10.14357/19922264140109

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research, Grant No. 11-06-00476-a.

References

1. Kozerenko, E. B. 2003. Cognitive approach to language structure segmentation for machine translation algorithms. *Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings*. Las Vegas, USA: CSREA Press. 49–55.
2. Kozerenko, E. B. 2013. Strategii vyravnivaniya parallel'nykh tekstov: Semanticheskie aspekty [Parallel texts alignment strategies: The semantic aspects]. *Informatika i ee primeneniya — Inform. Appl.* 7(1):82–89.
3. Shaumyan, S. 2003. Categorial grammar and semiotic universal grammar. *IC-AI'03: Conference (International) on Artificial Intelligence Proceedings*. Las Vegas, USA: CSREA Press. 623–629.
4. Kuznetsov, I. P., E. B. Kozerenko, and A. G. Matskevich. 2011. Intelligent extraction of knowledge structures from natural language texts. *2011 IEEE/WIC/ACM Joint Conferences (International) on Web Intelligence and Intelligent Agent Technology Proceedings*. Washington, DC, USA: IEEE Computer Society. 03:269–272. doi:10.1109/WI-IAT.2011.235
5. Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B* 39(1):1–22.
6. Lund, K., and C. Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Meth. Instr. Comp.* 28(2):203–208.
7. Curran, J. R. 2004. From distributional to semantic similarity. PhD Thesis. Edinburgh: University of Edinburgh. 177 p. Available at: <http://sydney.edu.au/engineering/it/~james/pubs/pdf/phdthesis.pdf>.
8. McCarthy, D., R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant senses in untagged text. *42nd Annual Meeting of the Association for Computational Linguistics Proceedings*. Barcelona, Spain: ACL. 280–287.
9. Clark, S., and S. Pulman. 2007. Combining symbolic and distributional models of meaning. *AAAI Spring Symposium on Quantum Interaction Proceedings*. Palo Alto, CA, USA: AAAI Press. 52–55. Available at: <http://www.cl.cam.ac.uk/~sc609/pubs/aaai07.pdf>.
10. Morozova, Yu. 2013. Method for extracting translation correspondences from a parallel corpus. *ICAI'13, WORLDCOMP'13 Proceedings*. Las Vegas, USA: CSREA Press. II:65–69.
11. Danielson, D. A. 2003. *Vectors and tensors in engineering and physics*. 2nd ed. Boulder, CO: Westview (Perseus). 282 p.
12. Montague, R. 1970. Universal grammar. *Theoria*. 36:373–398. (Reprinted: Thomason, R. H., ed. 1974. *Formal philosophy: Selected papers of Richard Montague*. New Haven – London: Yale University Press. 222–246.)
13. Partee, B. 2004. *Compositionality. Compositionality in formal semantics: Selected papers by Barbara H. Partee*. Malden, MA: Blackwell. 153–181. doi: 10.1002/9780470751305.ch7.
14. Pang, B., K. Knight, and D. Marcu. 2003. Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences. *NAACL'03: 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings*. Stroudsburg, PA, USA: ACL. 1:102–109.
15. Bannard, C., and C. Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. *43rd Annual Meeting of the ACL Proceedings*. Stroudsburg, PA, USA: ACL. 597–604.
16. Callison-Burch, C. 2008. Syntactic constraints on paraphrases extracted from parallel corpora. *EMNLP-2008 Proceedings*. Stroudsburg, PA, USA: ACL. 196–205.
17. Ganitkevitch, Ju., C. Callison-Burch, C. Napoles, and B. Van Durme. 2011. Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation. *2011 Conference on Empirical Methods in Natural Language Processing Proceedings*. Stroudsburg, PA, USA: ACL. 1168–1179.
18. Bogatyrev, K. 2006. In defense of symbolic NLP. *MLM-TA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings*. Las Vegas, USA: CSREA Press. 63–68.
19. Malkov, K. V., and D. V. Tunitsky. 2006. On extreme principles of machine learning in anomaly and vulnerability assessment. *MLMTA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings*. Las Vegas, USA: CSREA Press. 24–29.
20. Ajdukiewicz, K. 1935. Die Syntaktische Konnexitat. *Stud. Philos.* 1(1):1–27.
21. Bar-Hillel, Y. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29(1):47–58. Available at: <http://ling.umd.edu/~alxndrw/CGReadings/bar-hillel-53.pdf>.
22. Lambek, J. 1958. The mathematics of sentence structure. *Am. Math. Mon.* 65(3):154–170.
23. Steedman, M. 1996. Surface structure and interpretation. *Linguistic inquiry monographs*. Massachusetts: MIT Press. 140 p.
24. Lambek, J. 2008. *From word to sentence: A computational algebraic approach to grammar*. Monza, Italy: Polimetrica Publ. 154 p.
25. Moortgat, M. 2009. Symmetric categorial grammar. *J. Philos. Logic* 38(6):681–710.
26. Gazdar, G. 1996. Paradigm merger in natural language processing. *Computing tomorrow: Future research directions in computer science*. Eds. R. Milner and I. Wand. Cambridge, U.K.: Cambridge University Press. 88–109.

27. Clark, S., and J. R. Curran. 2007. Wide-coverage efficient statistical parsing with CCG and log-linear models. *Comput. Linguist.* 33(4):493–552.
28. Baroni, M., and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective–noun constructions in semantic space. *2010 Conference on Empirical Methods in Natural Language Processing Proceedings*. Stroudsburg, PA, USA: ACL. 1183–1193.
29. Grefenstette, E., and M. Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. *Conference on Empirical Methods in Natural Language Processing Proceedings*. Stroudsburg, PA, USA: ACL. 1394–1404.
30. Hermann, K. M., and P. Blunsom. 2013. The role of syntax in vector space models of compositional semantics. *51st Annual Meeting of the Association for Computational Linguistics Proceedings*. Stroudsburg, PA, USA: ACL. 894–904.
31. Church, K., and P. Hanks. 1996. Word association norms, mutual information, and lexicography. *Comput. Linguist.* 16(1):22–29.

Received December 26, 2013

Contributor

Kozerenko Elena B. (b. 1959) — Candidate of Science (PhD) in linguistics, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; kozerenko@mail.ru

РАЗРАБОТКА МОДЕЛИ УПРАВЛЕНИЯ ПРОЦЕССОМ ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ КОГНИТИВНЫХ ТЕХНОЛОГИЙ

В. А. Маренко¹, О. Н. Лучко², О. С. Лупенцов³

Аннотация: Приведена когнитивная карта «Процесс обучения» в виде ориентированного графа. На дуги ориентированного графа нанесены согласованные экспертные оценки. Объекты на когнитивной карте делятся на целевой фактор и управляющие факторы. Целевым фактором является качество обучения. Управляющие факторы используются для корректировки образовательного процесса. Приведена информация, которая используется при построении модели управляющего фактора «когнитивная готовность студента». Формализация экспериментальных данных осуществлена с применением метода семантического дифференциала и аппарата нечетких множеств. Приведена когнитивная модель образовательного процесса в виде функционального ориентированного графа. Показаны результаты имитационного эксперимента «управление образовательным процессом». Показано, что при увеличении управляющего фактора «уровень стабильности внешней среды» увеличивается и целевой фактор «качество обучения».

Ключевые слова: когнитивная карта; когнитивная модель; процесс обучения; имитационный эксперимент; нечеткое множество; семантический дифференциал

DOI: 10.14357/19922264140110

1 Введение

Современное общество нуждается в компетентных и конкурентоспособных специалистах, которые должны сочетать в себе как профессиональные знания, так и личностные качества, позволяющие в короткие сроки решать профессиональные задачи. Реализация таких потребностей требует применения современных подходов в управлении образованием, например таких, как когнитивные технологии. Их используют в управленческой деятельности для проведения оперативного анализа устойчивости социально-экономических процессов. Методологию когнитивного подхода успешно применяют сотрудники ИПУ РАН А. А. Кулинич, Е. К. Корноушенко, С. В. Качаев и др. [1].

Управлению в сфере образования посвящены работы многих ученых. Среди них Д. А. Новиков, М. М. Поташник, В. М. Филиппов, которые рассматривали развитие образовательных систем различных уровней в целом с точки зрения общекибернетического подхода [2].

В данной работе решается частная задача: повышение эффективности процесса обучения в Омском государственном институте сервиса с использованием когнитивной методологии. Применение когнитивной методологии дает возможность с помощью активизации интеллекта объективизировать знания экспертов о процессах, формализо-

вать изучаемую социально-экономическую проблему и провести ее исследование с помощью имитационного эксперимента.

2 Разработка когнитивной модели управления

Процесс разработки когнитивной модели состоит из последовательности взаимосвязанных шагов: проведения PEST (Political, Economic, Social, and Technological) анализа, построения когнитивной карты, формирования когнитивной модели в виде функционального графа для проведения имитационного эксперимента.

PEST-анализ помогает исследователю увидеть картину внешней среды, выделить наиболее важные влияющие факторы. Внешняя среда включает экономическую, политическую, правовую, социальную и технологическую составляющие, которые оказывают различное по степени, характеру и периодичности влияние на исследуемый процесс.

В работах ученых Б. П. Мартиросяна, О. А. Уткина, С. В. Ивановой и др. учитывается воздействие внешней среды как источника ресурсов для образовательного учреждения [3]. В предлагаемой статье воздействие внешней среды в рамках PEST-анализа оценивается экспертами степенью стабильности.

¹Институт математики им. С. Л. Соболева Сибирского отделения Российской академии наук, marenko@ofim.oscsbras.ru

²Омский государственный институт сервиса, o_luchko@rambler.ru

³Омский государственный институт сервиса, lupentsov@mail.ru

Базисные факторы

№	Фактор	Значения	Интервал измерений
Целевой фактор			
1	Качество образовательного процесса	Высокое, низкое, не очень высокое. . .	[0, 1]
Управляющие факторы			
2	Содержание обучения	Сложное, несложное, очень сложное, не очень сложное. . .	[0, 1]
3	Квалификация преподавателей	Высокая, не очень высокая. . .	[0, 1]
4	Когнитивная готовность студента	Высокая, низкая. . .	[0, 1]
5	Мотивация	Сильная, умеренная, слабая. . .	[0, 1]
6	Потребность общества в специалистах	Высокая, низкая. . .	[0, 1]
7	Уровень стабильности внешней среды	Высокий, невысокий. . .	[0, 1]

2.1 Построение когнитивной карты

Авторы [4] считают, что когнитивная карта — это ориентированный взвешенный граф, в котором вершины взаимно однозначно соответствуют базисным факторам. Взаимосвязи между базисными факторами определяются путем рассмотрения причинно-следственных цепочек, описывающих распространение влияний одного из них на другие. Влияние вершин может быть положительным, отрицательным или нулевым [4].

По мере накопления знаний в виде текстов, таблиц, математических структур становится возможным более детально раскрыть как характер связей между базисными факторами, так и содержание самих базисных факторов.

В данном исследовании когнитивная карта в виде ориентированного графа $G = \langle V, E \rangle$, где V — множество вершин, $V_i \in V, i = 1, 2, \dots, k$; E — множество дуг, $e_{ij} \in E, i, j = 1, 2, \dots, n$, состоит из

базисных факторов, наиболее значимых для управления образовательным процессом (см. таблицу).

Между базисными факторами устанавливаются причинно-следственные связи. На рис. 1 приведен фрагмент когнитивной карты с оценками специалистов в исследуемой предметной области.

Совокупность базисных факторов, влияющих на состояние исследуемой системы, можно разделить на группу управляющих факторов и группу целевых факторов. Для их количественного описания использовались лингвистические переменные и данные, полученные в результате экспериментов.

Для установления причинно-следственных связей между управляющими факторами и целевым фактором привлечено несколько экспертов. Их оценки согласовывались с использованием средств математической статистики.

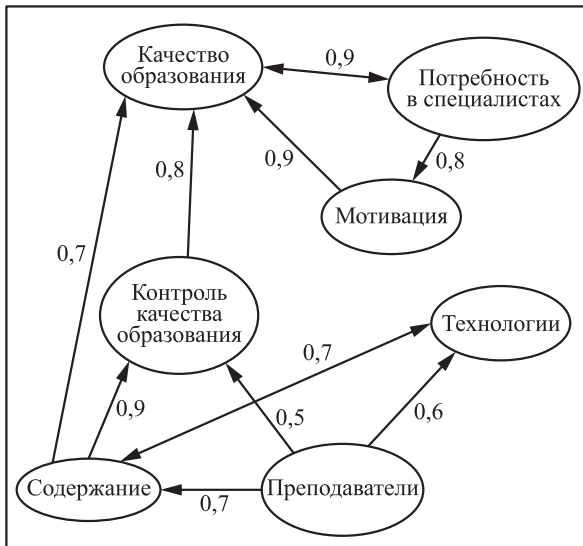


Рис. 1 Фрагмент когнитивной карты

2.2 Имитационный эксперимент

Следующий шаг исследования — построение когнитивной модели для проведения имитационного эксперимента. Когнитивная модель $\Phi = (G, X, F)$, где $G = \langle V, E \rangle$ — ориентированный граф; X — множество параметров вершин $V, X = \{x^{(v_i)}\}, i = 1, 2, \dots, k; x^{(v_i)} = \{x_g^{(i)}\}, g = 1, 2, \dots, n; x_g^{(i)}$ — параметр вершины V_i , если $g = 1$, то $x_g^{(i)} = x_i$; $X : V \rightarrow R, R$ — множество вещественных чисел; $F = F(X, E) = F(x_i, x_j, e_{ij})$ — функционал преобразования дуг, ставящий в соответствие каждой дуге знак, весовой коэффициент ω_{ij} или функцию $f(x_i, x_j, e_{ij}) = f_{ij}$ [5]. На рис. 2 представлена упрощенная когнитивная модель, которая задается матрицей смежности. Элементы матрицы — экспертные оценки.

На следующем этапе исследовался процесс распространения возмущений на графе.

OpenOffice.org Calc

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Аrial Cyr 10 Ж К Ч

C24 $=((B24+100)*(1+C16/100)-100)*(-1)$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3	Матрица влияния факторов																			
4	Качество обучения	(A)	0	0,6	0	0,7	0,7	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Преподавательский состав	(Б)	0,7	0	0,6	0,7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Содержание обучения	(В)	0	0	0	0,5	0,1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Методы обучения (технологии)	(Г)	0,2	0	0,5	0	0	0,8	0	0	0	0	0	0	0	0	0	0	0	0
8	Потребность в специалистах	(Д)	0	0	0	0	0	0	-0,8	0	0	0	0	0	0	0	0	0	0	0
9	Мотивация	(Е)	-0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Изменения в % пошагово																			
12	Качество обучения	0,0	0,0	0,0	3,4	2,3	2,6	2,1	1,5	1,0	0,5	0,2	-0,1	-0,2	-0,3	-0,3	-0,2	-0,2	-0,1	-0,1
13	Преподавательский состав	0,0	0,0	5,6	1,9	3,5	2,5	1,6	1,3	0,6	0,2	-0,1	-0,3	-0,3	-0,3	-0,3	-0,2	-0,1	-0,1	0,0
14	Содержание обучения	0,0	0,0	3,2	0,0	0,8	0,3	-0,1	0,0	-0,2	-0,2	-0,2	-0,1	-0,1	0,0	0,0	0,0	0,0	0,0	0,0
15	Методы обучения (технологии)	0,0	6,0	0,0	1,6	0,7	-0,5	-0,2	-0,7	-0,6	-0,5	-0,4	-0,3	-0,2	-0,1	0,0	0,0	0,1	0,1	0,1
16	Потребность в специалистах	0,0	-8,0	0,0	0,0	0,0	1,3	0,9	1,0	0,8	0,6	0,4	0,2	0,1	0,0	-0,1	-0,1	-0,1	-0,1	-0,1
17	Мотивация	10,0	0,0	0,0	0,0	-1,7	-1,1	-1,3	-1,1	-0,7	-0,5	-0,3	-0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1
19	Изменения в % нар. итогом																			
20	Качество обучения	0	0	0	3	6	8	11	12	13	14	14	14	14	14	13	13	13	13	13
21	Преподавательский состав	0	0	6	8	11	14	16	18	18	18	18	18	18	17	17	17	16	16	16
22	Содержание обучения	0	0	3	3	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3
23	Методы обучения (технологии)	0	8	8	10	10	10	10	9	8	8	7	7	7	7	7	7	7	7	7
24	Потребность в специалистах	0	8	8	8	8	9	10	12	13	13	14	14	14	14	14	14	14	13	13
25	Мотивация	10	10	10	10	8	7	6	4	4	3	3	3	3	3	3	3	3	3	3

Рис. 2 Упрощенная когнитивная модель

Параметры когнитивной модели $x_i(t)$, $t = 1, \dots, n$, зависят от времени. Если в момент времени $t - 1$ в вершину поступал импульс $p_j \in P$, то переход системы из состояния $t - 1$ в t осуществлялся по правилу

$$x_i(t) = x_i(t - 1) + \sum_{j=1}^{k-1} f(x_i, x_j, e_{ij})p_j(t - 1)$$

при известных начальных значениях [5].

Возмущение поступало в одну из вершин графа и актуализировало всю систему показателей в большей или меньшей степени. Если между двумя

управляющими факторами связь имела величину, например, 0,5 и значение одного управляющего фактора увеличивалось на 10%, то величина другого управляющего фактора возрастала на 5%.

Кривые на рис. 3 визуализируют изменения, происходящие на графе. Уменьшение значений управляющего фактора «степень стабильности внешней среды» (2) приводит к уменьшению значений целевого фактора «качество образовательного процесса» (1). На рис. 3 показано изменение значений целевого фактора до (а) и после (б) уменьшения значений управляющего фактора «степень стабильности внешней среды».

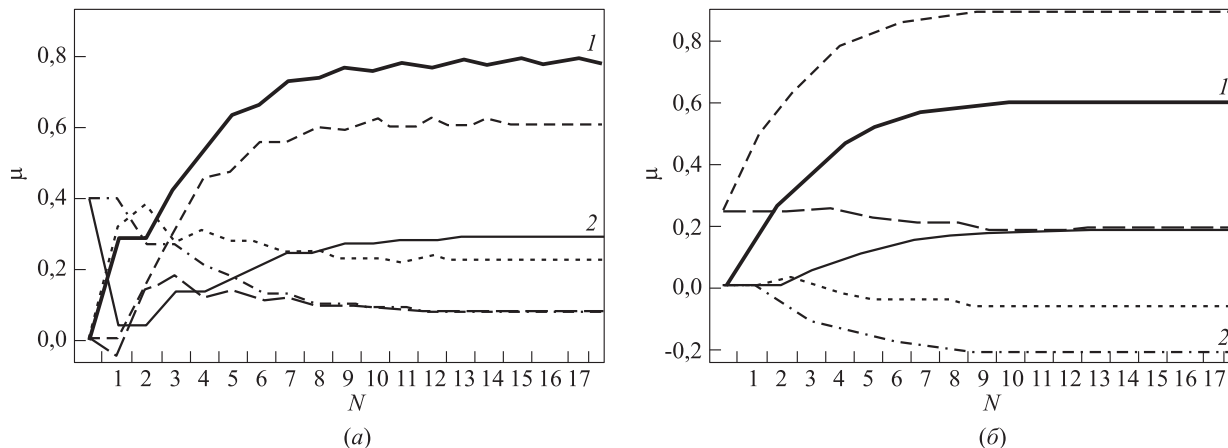


Рис. 3 Визуализация расчетов

В настоящее время реализован простейший вариант имитационного эксперимента, позволяющий наблюдать различные состояния образовательного процесса по значениям целевого фактора.

2.3 Модель «Когнитивная готовность студента» как управляющий фактор

Авторами разработана модель «Когнитивная готовность студента», которая служит управляющим фактором. Модель имеет три составляющие: интеллектуальную, рефлексивную и информационную. Эти составляющие являются личностными характеристиками студентов.

Интеллектуальная характеристика (A_z) оценивается баллами по результату интернет-теста. Данные интернет-теста для группы студентов исполь-

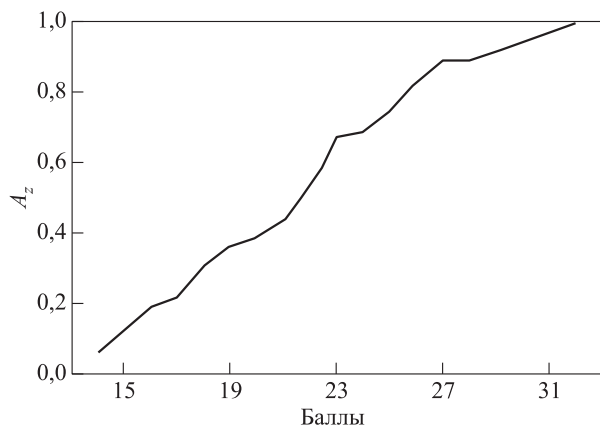


Рис. 4 Нечеткое множество «интеллектуальная характеристика»

зовались для формирования нечеткого множества «интеллектуальная характеристика», которое характеризует степень выраженности исследуемой характеристики у субъекта образовательного процесса (рис. 4).

Рефлексивная характеристика (A_x) помогает студенту осознать содержание обучения за счет размышления над собственным познавательным процессом. Для выявления численных значений этой характеристики в рамках компетентностного подхода применялось анкетирование. Вопросы анкеты составлены в соответствии с названиями компетенций по направлению подготовки 230700.68 «Прикладная информатика» (всего 22 компетенции). Задача опрашиваемого состоит в том, чтобы зафиксировать свою оценку степени владения компетенцией на шкале семантического дифференциала какой-либо меткой [6]. В результате возникает субъективный семантический профиль — ломаная линия, соединяющая все метки, поставленные испытуемым. Ось абсцисс — номера компетенций, отражающих общепрофессиональные умения и навыки. Ось ординат — степень владения компетенцией (рис. 5).

Для уточнения рефлексивного аспекта, состоящего в том, что субъект размышляет над своей образовательной деятельностью, анкетирование для каждой личности проводилось в трех вариантах в соответствии с моделью В. А. Лефевра [7]. Сначала каждый испытуемый осуществлял самооценку знаний по соответствующей компетенции в баллах. Затем выставлялась балльная оценка знаний испытуемому другим субъектом. А потом также в баллах оценивал себя сам испытуемый, став на позицию стороннего наблюдателя. Из полученных

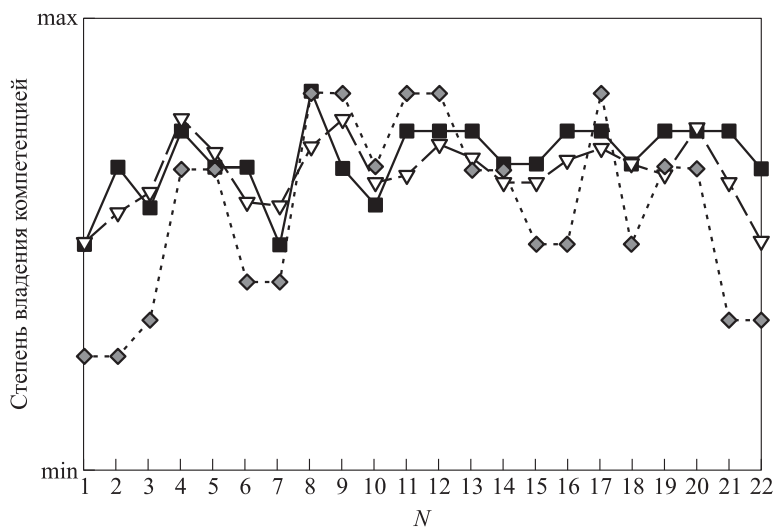


Рис. 5 Индивидуальные семантические профили

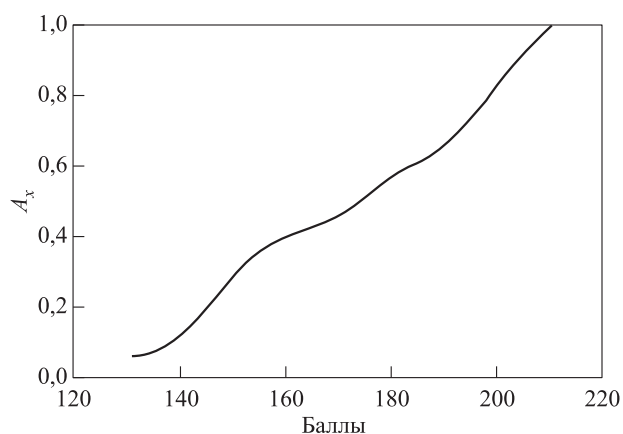


Рис. 6 Нечеткое множество «рефлексивная характеристика»

трех чисел выбиралась медиана. Интегральный показатель — сумма баллов по всем компетенциям. По результатам анкетирования группы студентов построено нечеткое множество «рефлексивная характеристика» (рис. 6).

Информационная характеристика (A_y) выявляет совокупность способностей студента рационально добывать информацию, превращая ее в знания и компетенции, умение овладевать новыми технологиями переработки информации. Численное значение информационной характеристики устанавливается по каждому субъекту образовательного процесса в баллах экспертами — членами методического совета кафедры. Интегральный показатель — сумма баллов по всем компетенциям. По результатам анкетирования группы студентов строится также нечеткое множество «информационная характеристика», по которому можно определять степень выраженности исследуемой характеристики у каждого студента.

Три характеристики субъекта процесса обучения составляют сформированную авторами статьи нечеткую модель «Когнитивная готовность студента» (рис. 7). На координатных осях модели откладываются степени выраженности исследуемых характеристик (интеллектуальной, рефлексивной и информационной) в интервале от 0 до 1.

Таким образом, каждый студент представлен точкой A в трехмерном пространстве. Длина вектора OA — численное значение управляющего фактора «когнитивная готовность студента». В перспективе планируется расширение разработанной нечеткой модели «Когнитивная готовность студента» до n -мерного уровня для более полного учета индивидуальных характеристик личности.

Модель «Когнитивная готовность студента» можно применять не только в качестве управля-

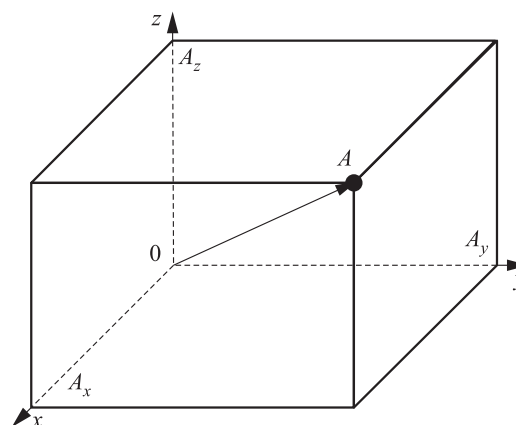


Рис. 7 Нечеткая модель «Когнитивная готовность студента»

ющего фактора при когнитивном моделировании, но и для классификации студентов в соответствии с их характеристиками. Например, классификация по интеллектуальной характеристике может использоваться для рекомендации технологий обучения с различной степенью интерактивности, а также для диагностики пригодности личности к определенной профессиональной деятельности.

3 Заключение

Проведенное моделирование позволяет прогнозировать качество процесса обучения с учетом индивидуальных характеристик студентов, квалификации преподавательского состава, содержания учебного материала, мотивации и других управляющих факторов.

Заключительный этап исследований состоит в том, чтобы сформировать систему научно обоснованных рекомендаций руководящему звену образовательного учреждения для улучшения процесса обучения.

Литература

1. Максимов В. И., Тер-Егизарова Н. В. IV Междунар. конф. «Когнитивный анализ и управление развитием ситуаций» CASC'2004 // Проблемы управления, 2005. № 1. С. 83–87.
2. Новиков Д. А. Теория управления образовательными системами. — М.: Народное образование, 2009. 416 с.
3. Иванова С. В. Образование в организационно-гуманистическом измерении. — М.: РУДН, 2007. 236 с.
4. Максимов В. И., Корноушенко Е. К., Качаев С. В. Когнитивные технологии для поддержки принятия управ-

- ленческих решений. <http://emag.iis.ru/arc/infosoc/emag.nsf/bpa/092aa276c601a997c32568c0003ab839>.
5. Горелова Г. В., Радченко С. А. Когнитивные технологии поддержки управленческих решений в социально-экономических системах // Известия Южного федерального университета. Технические науки, 2003. Т. 34. № 5. С. 95–104.
 6. Лупенцов О. С., Лучко О. Н., Маренко В. А. Применение семантического дифференциала для реализации компетентностного подхода в вузе // Информатизация образования и науки, 2012. № 3(15). С. 128–134.
 7. Лефевр В. А. Формула человека: Контуры фундаментальной психологии / Пер. с англ. — М.: Прогресс, 1991. 108 с.

Поступила в редакцию 21.02.13

DEVELOPMENT OF LEARNING PROCESS CONTROL MODEL WITH COGNITIVE TECHNOLOGIES

V. A. Marenko¹, O. N. Luchko², and O. S. Lupentsov²

¹Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 4 Acad. Koptyug Av., Novosibirsk 630090, Russian Federation

²Omsk State Institute of Service, 13 Petrov Str., Omsk 644099, Russian Federation

Abstract: The paper describes the “Learning process” cognitive map in the form of a directed graph. Arcs of the directed graph are labeled with the agreed expert estimates. Objects on the cognitive map are divided into the target factor and controlling factors. The quality of education is the target factor. Controlling factors are used to adjust the educational process. The paper provides information used to devise the model of the control factor of cognitive readiness of a student. Formalization of experimental data is carried out using the semantic differential method and fuzzy sets. The cognitive model of the educational process, which is a functional directed graph, is presented. The results of experiments simulating educational process management are shown. It is also shown that while the level of stability of external environment (viewed as a control factor) increases, the level of the quality of education (viewed as the target factor) also increases.

Keywords: control; cognitive model; cognitive map; educational process; simulation experiment; semantic differential; fuzzy set

DOI: 10.14357/19922264140110

References

1. Maksimov, V. I., and N. V. Ter-Egiazarova. 2005. IV Mezhdunarodnaya konferentsiya “Kognitivnyy analiz i upravlenie razvitiem situatsiy” CASC’2004 [IV International Conference “Cognitive Analysis and Management of Development of Situations” CASC’2004]. *Problemy Upravleniya [Management Problems]* 1:83–87.
2. Novikov, D. A. 2009. *Teoriya upravleniya obrazovatel’nymi protsessami [Theory of management of educational systems]*. Moscow: National Education. 416 p.
3. Ivanova, S. V. 2007. *Obrazovanie v organizatsionno-gumanisticheskoy izmerenii [Education in organizational and humanistic measurement]*. Moscow: RUDN. 236 p.
4. Maksimov, V. I., E. K. Kornoushenko, and S. V. Kachayev. Kognitivnye tekhnologii dlya podderzhki prinyatiya upravlencheskikh resheniy [Cognitive technologies for support of adoption of administrative decisions]. Available at: <http://emag.iis.ru/arc/infosoc/emag.nsf/bpa/092aa276c601a997c32568c0003ab839> (accessed December 26, 2013).
5. Gorelova, G. V., and S. A. Radchenko. 2003. Kognitivnye tekhnologii podderzhki upravlencheskikh resheniy v sotsial’no-ekonomicheskikh sistemakh [Cognitive technologies of support of administrative decisions in social and economic systems]. *Izvestiya Yuzhnogo Federal’nogo Universiteta. Tekhnicheskie Nauki [News of the Southern Federal University. Technical Sciences]*. 34(5):95–104.
6. Lupentsov, O. S., O. N. Luchko, and V. A. Marenko. 2012. Primenenie semanticheskogo differentsiala dlya realizatsii kompetentnostnogo podkhoda v vuze [Application of semantic differential for realization of competence-based approach in higher education institution]. *Informatizatsiya Obrazovaniya i Nauki [Science and Education Informatization]* 3(15):128–134.
7. Lefevre, V. A. 1982. *The formula of man: An outline of fundamental psychology*. Irvine: School of Social Sciences, University of California. 290 p.

Received February 21, 2013

Contributors

Marenko Valentina A. (b. 1951) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 4 Acad. Koptug Av., Novosibirsk 630090, Russian Federation; marenko@ofim.oscsbras.ru

Luchko Oleg N. (b. 1961) — Candidate of Science (PhD) in education, professor; Head of Department, Omsk State Institute of Service, 13 Petrov Str., Omsk 644099, Russian Federation; o_luchko@rambler.ru

Lupentsov Oleg S. (b. 1986) — PhD student, Omsk State Institute of Service, 13 Petrov Str., Omsk 644099, Russian Federation; lupentsov@mail.ru

ОБЩИЕ ОЦЕНКИ УСТОЙЧИВОСТИ ДЛЯ НЕСТАЦИОНАРНЫХ МАРКОВСКИХ ЦЕПЕЙ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ*

А. И. Зейфман¹, В. Ю. Королев², А. В. Коротышева³, С. Я. Шоргин⁴

Аннотация: Рассмотрен общий метод получения оценок устойчивости нестационарных марковских цепей с непрерывным временем с использованием специальных весовых норм, связанных с полной вариацией. Метод основан на оценках, получаемых при помощи логарифмической нормы оператора линейного дифференциального уравнения, и специальных преобразованиях редуцированной матрицы интенсивностей процесса. Доказаны утверждения, дающие точные оценки возмущений вероятностных характеристик моделей в случае отсутствия эргодичности в равномерной операторной топологии. В качестве возможных приложений рассмотрены системы массового обслуживания, описываемые процессами рождения и гибели с катастрофами, а также системы массового обслуживания с групповым поступлением и обслуживанием требований. Исследованы также конкретные классы таких систем, для которых применим предлагаемый метод получения оценок их устойчивости, в частности система обслуживания $M_t/M_t/S$ с катастрофами, а также простейшая система обслуживания с групповым поступлением и обслуживанием заявок. Рассмотрен численный пример построения предельных характеристик возмущенной системы обслуживания.

Ключевые слова: марковские цепи и модели с непрерывным временем; нестационарные марковские цепи; оценки устойчивости; специальные нормы; модели массового обслуживания

DOI: 10.14357/19922264140111

1 Введение

Исследования устойчивости различных характеристик стохастических моделей активно проводятся начиная с 1970-х гг. [1–3]. В это же время В. М. Золотарев предложил подход, согласно которому предельные теоремы теории вероятностей рассматривались как теоремы устойчивости. Этот подход был затем развит в работах В. М. Золотарева, В. В. Калашникова, В. Ю. Королева и их коллег в рамках традиционных международных семинаров по проблемам устойчивости стохастических моделей, основанного В. М. Золотаревым (см., например, [4–8]). Этот подход оказался очень продуктивным для изучения случайных сумм в теории массового обслуживания, теории восстановления, ветвящихся процессов [9].

Проблемы, связанные с оценками устойчивости марковских цепей по отношению к возмущениям их характеристик, в разных ситуациях и для разных классов процессов детально исследовались начиная с работ Н. В. Карташова и (параллельно) А. И. Зейфмана в 1980-х гг. [10–12].

В настоящей работе рассматриваются только цепи с непрерывным временем, поэтому дальнейшие замечания относятся в основном к таким цепям. Следует отметить прежде всего два следующих важных момента.

Во-первых, равномерная эргодичность цепи (т. е. ее эргодичность в равномерной операторной топологии) позволяет сравнительно легко получать точные оценки возмущений предельных характеристик процесса.

Во-вторых, для получения явных и точных оценок устойчивости цепи необходимо наличие оценок скорости сходимости цепи к предельным характеристикам в виде явных неравенств. Более того, чем точнее удастся получить оценки скорости сходимости, тем точнее будут и оценки устойчивости.

Такого рода оценки проще всего удастся получить для конечных стационарных (однородных по времени) марковских цепей, так что наибольшее число результатов относится именно к этой ситуации (см., например, [13–18]).

* Исследование поддержано РФФИ (гранты 12-07-00109, 12-07-00115, 13-07-00223, 14-07-00041).

¹ Вологодский государственный университет; Институт проблем информатики Российской академии наук, Институт социально-экономического развития территорий Российской академии наук, a_zeifman@mail.ru

² Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vkorolev@cs.msu.ru

³ Вологодский государственный университет, a_korotysheva@mail.ru

⁴ Институт проблем информатики Российской академии наук, SShorgin@ipiran.ru

В случае же счетного пространства состояний (и тем более для нестационарных цепей) возникают очень серьезные трудности, связанные в первую очередь с отсутствием равномерной эргодичности для наиболее интересных, с точки зрения приложений, процессов. Так, процессы рождения и гибели, используемые при описании систем массового обслуживания, биологии, химии, статистической физики, как правило, *не* являются равномерно эргодичными.

Следуя идеям Н. В. Карташова (см. детальное описание в [19]), большинство исследователей применяет вероятностные методы для изучения стационарных цепей (с конечным, счетным или общим фазовым пространством), их эргодичности и устойчивости в различных нормах [20–22]. Близкий подход для широкого класса стационарных цепей с дискретным временем рассмотрен в [23]. Обзор основных результатов и подходов, а также новейшие достижения были представлены в [24].

В работах авторов настоящей статьи проблемы устойчивости исследовались для нестационарных конечных и счетных цепей с непрерывным временем другими методами.

Первые исследования, в которых рассматривались нестационарные модели теории массового обслуживания, появились в 1970-х гг. ([25, 26], а также более поздняя работа [27]). Более того, тогда же в [28] была отмечена принципиальная возможность использования логарифмической нормы матрицы для изучения скорости сходимости марковских цепей с непрерывным временем. Соответствующий общий подход, использующий теорию линейных дифференциальных уравнений в банаховом пространстве, был развит в серии работ [12, 29–32] (см. также подробное описание в [33–35]). А метод исследования устойчивости вектора вероятностей состояний марковской цепи с непрерывным временем по норме полной вариации (l_1 -норме) относительно возмущений инфинитезимальных характеристик цепи был впервые предложен в [12] (см. также [29, 36]). В [37] оценки устойчивости по отношению к условно малым возмущениям были подробно исследованы для существенно нестационарных процессов рождения и гибели. Проблемы устойчивости и соответствующие оценки для одного нового класса процессов были недавно рассмотрены в [38, 39].

Как правило, изучаемые цепи не являются равномерно эргодичными, и поэтому для получения содержательных оценок устойчивости приходится накладывать дополнительные условия на структуру инфинитезимальной матрицы возмущенного процесса.

В настоящей же работе предложен подход, позволяющий получать общие оценки устойчивости в терминах специальных «взвешенных» норм, связанных с естественной (полной вариацией, l_1 -нормой) без наложения отмеченных дополнительных условий.

2 Предварительные сведения

Пусть $X = X(t)$, $t \geq 0$, — нестационарная, вообще говоря, марковская цепь с непрерывным временем и счетным пространством состояний $0, 1, \dots$. Обозначим через $p_{ij}(s, t) = \Pr \{X(t) = j | X(s) = i\}$, $i, j \geq 0$, $0 \leq s \leq t$, переходные вероятности для $X = X(t)$. Пусть $p_i(t) = \Pr \{X(t) = i\}$ — вероятности состояний цепи, а $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$ — соответствующий вектор вероятностей состояний. Далее предполагается, что

$$\Pr \{X(t+h) = j | X(t) = i\} = r = \begin{cases} q_{ij}(t)h + \alpha_{ij}(t, h) & \text{при } j \neq i; \\ 1 - \sum_{k \neq i} q_{ik}(t)h + \alpha_i(t, h) & \text{при } j = i, \end{cases}$$

где все $\alpha_i(t, h)$ есть $o(h)$ равномерно по i , т. е. $\sup_i |\alpha_i(t, h)| = o(h)$.

Кроме того, как обычно, считаем, что в нестационарном случае все интенсивности (т. е. функции $q_{ij}(t)$) являются линейными комбинациями конечного числа локально интегрируемых на $[0, \infty)$ неотрицательных функций.

Положим $a_{ij}(t) = q_{ji}(t)$ при $j \neq i$ и $a_{ii}(t) = -\sum_{j \neq i} a_{ji}(t) = -\sum_{j \neq i} q_{ij}(t)$.

Далее, для обеспечения возможности получения более обзримых оценок (см. детали в [32–34, 40]) предположим, что матрица интенсивностей существенно ограничена, т. е.

$$|a_{ii}(t)| \leq L < \infty \quad (1)$$

почти при всех $t \geq 0$.

Тогда для вероятностей состояний справедлива прямая система Колмогорова

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}(t), \quad (2)$$

где $A(t)$ — транспонированная матрица интенсивностей процесса.

Через $\|\cdot\|$ будем обозначать l_1 -норму, т. е. $\|\mathbf{x}\| = \sum |x_i|$, и при $B = (b_{ij})_{i,j=0}^\infty$ соответственно $\|B\| = \sup_j \sum_i |b_{ij}|$. Пусть Ω — множество всех стохастических векторов, т. е. l_1 -векторов с неотрицательными координатами и единичной нормой.

Тогда имеем

$$\|A(t)\| = 2 \sup_k |a_{kk}(t)| \leq 2L$$

почти при всех $t \geq 0$. Следовательно, оператор-функция $A(t)$ из l_1 в себя ограничена почти при всех $t \geq 0$ и локально интегрируема на $[0, \infty)$. А значит, систему (2) можно отождествить с дифференциальным уравнением в пространстве l_1 с ограниченным оператором. Как известно [41], тогда задача Коши для уравнения (2) имеет единственное решение при любом начальном условии и, кроме того, если $\mathbf{p}(s) \in \Omega$, то и $\mathbf{p}(t) \in \Omega$ при $t \geq s \geq 0$.

Значит, можно положить $p_0(t) = 1 - \sum_{i \geq 1} p_i(t)$.

Тогда из (2) получаем следующее уравнение (см. подробное обсуждение в [32–34, 40]):

$$\frac{d\mathbf{z}}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t), \quad (3)$$

где $\mathbf{f}(t) = (a_{10}, a_{20}, \dots)^T$;

$$B = \begin{pmatrix} a_{11} - a_{10} & a_{12} - a_{10} & \dots & a_{1r} - a_{10} & \dots \\ a_{21} - a_{20} & a_{22} - a_{20} & \dots & a_{2r} - a_{20} & \dots \\ a_{31} - a_{30} & a_{32} - a_{30} & \dots & a_{3r} - a_{30} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{r1} - a_{r0} & a_{r2} - a_{r0} & \dots & a_{rr} - a_{r0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (4)$$

Через $\bar{X} = \bar{X}(t)$ обозначим «возмущенную» марковскую цепь с вероятностями состояний $\bar{p}_i(t)$, транспонированной матрицей интенсивностей $\bar{A}(t) = (\bar{a}_{ij}(t))_{i,j=0}^\infty$ и т. д.

Через $E(t, k) = E\{X(t) | X(0) = k\}$ будем обозначать далее среднее (математическое ожидание) процесса в момент t при начальном условии $X(0) = k$.

3 Основные результаты

Пусть $\{d_i\}, i = 1, 2, \dots$, — возрастающая последовательность положительных чисел, причем $d_1 = 1$.

Рассмотрим треугольную матрицу D следующего вида:

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \dots \\ 0 & d_2 & d_2 & \dots \\ 0 & 0 & d_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix},$$

а также соответствующее пространство последовательностей $l_{1D} = \{\mathbf{z} = (p_1, p_2, \dots)^T | \|\mathbf{z}\|_{1D} \equiv \|\mathbf{Dz}\| < \infty\}$. Отметим, что $\|B\|_{1D} = \|DBD^{-1}\|$. Положим вдобавок $\|\mathbf{p}\|_{1D} = \|\mathbf{z}\|_{1D}$.

Отметим, что введенная матрица не имеет прямой вероятностной интерпретации. С другой стороны, она оказывается очень полезной в проводимых исследованиях.

Будем далее предполагать выполненными следующие условия:

$$\|B(t)\|_{1D} \leq \mathfrak{B} < \infty; \quad \|\mathbf{f}(t)\|_{1D} \leq f < \infty$$

почти при всех $t \geq 0$.

Определение 1. Марковская цепь $X(t)$ называется $1D$ -экспоненциально слабо эргодичной, если существуют положительные M и a такие, что при любых s, t таких, что $t \geq s \geq 0$, и любых начальных условиях $\mathbf{p}^*(s) \in l_{1D}, \mathbf{p}^{**}(s) \in l_{1D}$ выполнено неравенство:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1D} \leq M e^{-a(t-s)} \|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\|_{1D}.$$

Замечание 1. Если у цепи, кроме того, существует стационарное распределение вероятностей $\pi \in \Omega$, то $X(t)$ является $1D$ -экспоненциально сильно эргодичной, т. е.

$$\|\mathbf{p}^*(t) - \pi\|_{1D} \leq M e^{-at} \|\mathbf{p}^*(0) - \pi\|_{1D} \quad (5)$$

при любом $t \geq 0$ и любом начальном условии $\mathbf{p}^*(0)$.

Определение 2. Марковская цепь $X(t)$ имеет предельное математическое ожидание (предельное среднее) $\phi(t)$, если $|E(t, k) - \phi(t)| \rightarrow 0$ при $t \rightarrow \infty$ для любого k .

Пусть почти при всех $t \geq 0$ выполнены условия:

$$\left. \begin{aligned} \|B(t) - \bar{B}(t)\|_{1D} &\leq \mathfrak{B} - \bar{\mathfrak{B}}; \\ \|\mathbf{f}(t) - \bar{\mathbf{f}}(t)\|_{1D} &\leq f - \bar{f}. \end{aligned} \right\} \quad (6)$$

Теорема 1. Пусть марковская цепь $X(t)$ $1D$ -экспоненциально слабо эргодична. Тогда при достаточно малых возмущениях (6) возмущенная цепь $\bar{X}(t)$ также $1D$ -экспоненциально слабо эргодична и справедлива следующая оценка устойчивости в $1D$ -норме:

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| \leq \frac{M(M(\mathfrak{B} - \bar{\mathfrak{B}})f + a(f - \bar{f}))}{a(a - M(\mathfrak{B} - \bar{\mathfrak{B}}))}.$$

Если, кроме того, $W = \inf_{i \geq 1} d_i/i > 0$, то обе цепи

$X(t)$ и $\bar{X}(t)$ имеют предельные средние и справедлива оценка:

$$\begin{aligned} \limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| &\leq \\ &\leq \frac{M(M(\mathfrak{B} - \bar{\mathfrak{B}})f + a(f - \bar{f}))}{W a(a - M(\mathfrak{B} - \bar{\mathfrak{B}}))}. \end{aligned} \quad (7)$$

Доказательство. Обозначим через $V(t, s)$ оператор Коши для уравнения (3), а через $\bar{V}(t, s)$ — оператор Коши соответствующего уравнения для

возмущенного процесса. Тогда из $1D$ -экспоненциальной эргодичности, в соответствии с (5), вытекает при всех $t \geq s \geq 0$ оценка

$$\|V(t, s)\|_{1D} \leq M e^{-a(t-s)}.$$

Теперь, применяя лемму 3.2.3 из [41], получаем при всех $t \geq s \geq 0$ оценку нормы «возмущенного» оператора Коши

$$\|\bar{V}(t, s)\|_{1D} \leq M e^{-(a-M(\mathfrak{B}-\bar{\mathfrak{B}}))(t-s)},$$

откуда вытекает $1D$ -экспоненциальная эргодичность $\bar{X}(t)$ при достаточно малых $\mathfrak{B} - \bar{\mathfrak{B}}$. Перепишем уравнение (3) в виде:

$$\frac{d\mathbf{z}}{dt} = \bar{B}(t)\mathbf{z}(t) + \mathbf{f}(t) + (B(t) - \bar{B}(t))\mathbf{z}(t).$$

Отсюда

$$\begin{aligned} \mathbf{z}(t) &= \bar{V}(t, 0)\mathbf{z}(0) + \int_0^t \bar{V}(t, \tau)\mathbf{f}(\tau) d\tau + \\ &+ \int_0^t \bar{V}(t, \tau) (B(\tau) - \bar{B}(\tau))\mathbf{z}(\tau) d\tau; \\ \bar{\mathbf{z}}(t) &= \bar{V}(t, 0)\bar{\mathbf{z}}(0) + \int_0^t \bar{V}(t, \tau)\bar{\mathbf{f}}(\tau) d\tau. \end{aligned}$$

Значит, при одинаковых начальных условиях исходного и возмущенного процессов получаем следующее неравенство в $1D$ -норме:

$$\begin{aligned} \|\mathbf{z}(t) - \bar{\mathbf{z}}(t)\| &\leq \\ &\leq \int_0^t \|\bar{V}(t, \tau)\| (\|B(\tau) - \bar{B}(\tau)\|\|\mathbf{z}(\tau)\| + \\ &+ \|\mathbf{f}(\tau) - \bar{\mathbf{f}}(\tau)\|) d\tau \leq \\ &\leq \int_0^t M e^{-(a-M(\mathfrak{B}-\bar{\mathfrak{B}}))(t-\tau)} ((\mathfrak{B} - \bar{\mathfrak{B}})\|\mathbf{z}(\tau)\| + \\ &+ (\mathbf{f} - \bar{\mathbf{f}})) d\tau. \end{aligned}$$

Используя условия теоремы, можно оценить при всех $t \geq 0$

$$\begin{aligned} \|\mathbf{z}(t)\|_{1D} &\leq \|V(t)\|_{1D}\|\mathbf{z}(0)\|_{1D} + \\ &+ \int_0^t \|V(t, \tau)\|_{1D}\|\mathbf{f}(\tau)\|_{1D} d\tau \leq M e^{-at}\|\mathbf{z}(0)\|_{1D} + \frac{M}{a}\mathbf{f}. \end{aligned}$$

А тогда получаем в $1D$ -норме при любом начальном условии $\mathbf{z}(0) \in l_{1D}$ оценку

$$\begin{aligned} \|\mathbf{z}(t) - \bar{\mathbf{z}}(t)\| &\leq M \left((\mathfrak{B} - \bar{\mathfrak{B}}) \frac{M}{a}\mathbf{f} + (\mathbf{f} - \bar{\mathbf{f}}) \right) \times \\ &\times \int_0^t e^{-(a-M(\mathfrak{B}-\bar{\mathfrak{B}}))(t-\tau)} d\tau + \\ &+ M \int_0^t e^{-(a-M(\mathfrak{B}-\bar{\mathfrak{B}}))(t-\tau)} (\mathfrak{B} - \bar{\mathfrak{B}}) \times \\ &\times M e^{-a\tau}\|\mathbf{z}(0)\| d\tau \leq \\ &\leq \frac{M(M(\mathfrak{B} - \bar{\mathfrak{B}})\mathbf{f} + a(\mathbf{f} - \bar{\mathbf{f}}))}{a(a - M(\mathfrak{B} - \bar{\mathfrak{B}}))} + o(1) \quad (8) \end{aligned}$$

и первое утверждение теоремы доказано.

Введем теперь пространство последовательностей l_{1E} таких, что

$$l_{1E} = \left\{ z = (p_1, p_2, \dots)^T / \|z\|_{1E} \equiv \sum n|p_n| < \infty \right\}.$$

Тогда второе утверждение и оценка (7) вытекают из известного неравенства $\|z\|_{1E} \leq W^{-1}\|z\|_{1D}$ (см., например, [40]) и оценки (8).

Следствие 1. Пусть $X(t)$ — стационарная $1D$ -экспоненциально сильно эргодичная марковская цепь. Тогда при достаточно малых возмущениях цепь $\bar{X}(t)$ также $1D$ -экспоненциально сильно эргодична и справедлива следующая оценка в норме $1D$:

$$\|\pi - \bar{\pi}\| \leq \frac{M(M\|B - \bar{B}\|\|\mathbf{f}\| + a\|\mathbf{f} - \bar{\mathbf{f}}\|)}{a(a - M\|B - \bar{B}\|)}.$$

Если вдобавок $W = \inf_{i \geq 1} (d_i/i) > 0$, то обе марковских цепи $X(t)$ и $\bar{X}(t)$ имеют предельные средние и в норме $1D$ выполнено неравенство:

$$|\phi - \bar{\phi}| \leq \frac{M(M\|B - \bar{B}\|\|\mathbf{f}\| + a\|\mathbf{f} - \bar{\mathbf{f}}\|)}{W a(a - M\|B - \bar{B}\|)}.$$

Можно получить и оценки устойчивости в естественной l_1 -норме. Для этого достаточно воспользоваться известным соотношением (см., например, [33] или [40]):

$$\|\mathbf{p}^* - \mathbf{p}^{**}\| \leq 2\|\mathbf{z}^* - \mathbf{z}^{**}\| \leq 4\|\mathbf{z}^* - \mathbf{z}^{**}\|_{1D}.$$

Следствие 2. Пусть $X(t)$ — стационарная $1D$ -экспоненциально сильно эргодичная марковская цепь. Тогда справедлива следующая оценка устойчивости:

$$\|\pi - \bar{\pi}\| \leq \frac{4M(M\|B - \bar{B}\|_{1D}\|\mathbf{f}\|_{1D} + a\|\mathbf{f} - \bar{\mathbf{f}}\|_{1D})}{a(a - M\|B - \bar{B}\|_{1D})}.$$

Следствие 3. Пусть $X(t)$ — нестационарная 1D-экспоненциально слабо эргодичная марковская цепь. Тогда справедлива следующая оценка устойчивости:

$$\limsup_{t \rightarrow \infty} \|p(t) - \bar{p}(t)\| \leq \frac{4M(M(\mathfrak{B} - \bar{\mathfrak{B}})f + a(f - \bar{f}))}{a(a - M(\mathfrak{B} - \bar{\mathfrak{B}}))}.$$

4 Процессы рождения и гибели с катастрофами

Пусть $X(t)$ — процесс рождения и гибели с катастрофами (ПРГК). Тогда ненулевыми являются интенсивности переходов: $q_{i,i+1}(t) = \lambda_i(t)$ — интенсивности рождения, $q_{i,i-1}(t) = \mu_i(t)$ — интенсивности гибели, $q_{i,0}(t) = \xi_i(t)$ — интенсивности катастроф, а все остальные $q_{ij}(t) \equiv 0$ при $i \neq j$.

Условия ограниченности (1) будут выглядеть следующим образом:

$$\sup_{k \geq 0} (\lambda_k(t) + \mu_k(t) + \xi_k(t)) \leq L < \infty$$

почти при всех $t \geq 0$.

Для элементов матрицы (4) получаем теперь следующие выражения:

$$b_{ij} = \begin{cases} -(\lambda_0 + \lambda_1 + \mu_1 + \xi_1) & \text{при } i = j = 1; \\ \mu_2 - \lambda_0 & \text{при } i = 1, j = 2; \\ -\lambda_0 & \text{при } i = 1, j > 2; \\ -(\lambda_j + \mu_j + \xi_j) & \text{при } i = j > 1; \\ \mu_j & \text{при } i = j - 1 > 1; \\ \lambda_j & \text{при } i = j + 1 > 1; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Следовательно,

$$DBD^{-1} = \begin{pmatrix} -(\lambda_0 + \mu_1 + \xi_1) & \frac{d_1}{d_2}\mu_2 & 0 & \dots & \dots \\ \frac{d_2}{d_1}\lambda_1 & -(\lambda_1 + \mu_2 + \xi_2) & \frac{d_2}{d_3}\mu_3 & 0 & \dots \\ 0 & \frac{d_3}{d_2}\lambda_2 & -(\lambda_2 + \mu_3 + \xi_3) & \frac{d_3}{d_4}\mu_4 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix};$$

$$\|B(t)\|_{1D} = \sup_{k \geq 0} (\lambda_k(t) + \mu_{k+1}(t) + \xi_{k+1}(t) + \frac{d_{k+1}}{d_k}\lambda_{k+1}(t) + \frac{d_{k-1}}{d_k}\mu_k(t)).$$

Отметим, что простое дополнительное условие

$$\sup_{k \geq 0} \frac{d_{k+1}}{d_k} = m < \infty$$

гарантирует справедливость оценки

$$\|B(t)\|_{1D} \leq (1 + m)L < \infty$$

почти при всех $t \geq 0$. Так как в этом разделе $X(t)$ — ПРГК, то $f(t) = (\lambda_0, 0, 0, \dots)^T$ и

$$\|f(t)\|_{1D} = d_1\lambda_0(t) \leq L < \infty$$

почти при всех $t \geq 0$.

В предыдущих работах изучалась 1D-эргодичность для ПРГ и ПРГК, соответствующие точные оценки получены в [31, 32, 38, 42]. Здесь только сформулированы основные результаты, которые позволяют получить соответствующие оценки устойчивости.

Рассмотрим выражения:

$$\alpha_k(t) = \lambda_k(t) + \mu_{k+1}(t) + \xi_{k+1}(t) - \frac{d_{k+1}}{d_k}\lambda_{k+1}(t) - \frac{d_{k-1}}{d_k}\mu_k(t), \quad k \geq 0;$$

$$\alpha(t) = \inf_{k \geq 0} \alpha_k(t).$$

Теорема 2. Пусть

$$\int_0^\infty \alpha(t) dt = +\infty. \tag{9}$$

Тогда $X(t)$ слабо эргодичен. Если же, кроме того, существуют положительные M и a такие, что при всех $0 \leq s \leq t$

$$e^{-\int_s^t \alpha(\tau) d\tau} \leq Me^{-a(t-s)},$$

то $X(t)$ 1D-экспоненциально слабо эргодичен.

Следствие 4. Пусть $X(t)$ — стационарный ПРГК и пусть теперь вместо (9) соответствующее $\alpha > 0$. Тогда $X(t)$ 1D-экспоненциально (сильно) эргодичен. При этом можно выбрать $M = 1$ и $a = \alpha$.

Следствие 5. Пусть теперь все интенсивности рождения, гибели и катастроф 1-периодичны по t . Пусть

$$\int_0^1 \alpha(t) dt = \alpha^* > 0$$

вместо (9). Тогда $X(t)$ 1D-экспоненциально слабо эргодичен,

$$a = \alpha^*, \quad M = \sup_{|t-s| \leq 1} e^{\int_s^t \alpha(\tau) d\tau}.$$

4.1 Система обслуживания $M_t/M_t/S$ с катастрофами

Это очень часто используемая модель системы массового обслуживания. Число требований в системе $X(t)$ — ПРГК с интенсивностями: $\lambda_n(t) = \lambda(t)$ — интенсивность поступления требования (заявки) в систему; $\mu_n(t) = \min(n, S)\mu(t)$ — интенсивность обслуживания требования, если в системе находится n требований; наконец, $\xi_n(t) = \xi(t)$ — интенсивность катастрофы (т.е. обнуления числа заявок).

Здесь рассмотрим только случай существенной интенсивности обслуживания. Соответствующие оценки получены в [38]. Используя доказанные теоремы, получаем для этой системы следующий результат.

Утверждение 1. Пусть существует $\delta \in (1, S/(S-1)]$, функция $\theta(t)$ и положительные числа R и ν такие, что при любых $s, t, 0 \leq s \leq t$,

$$S\mu(t) - \delta\lambda(t) \geq \theta(t); \quad e^{-\int_s^t (1-\delta^{-1})\theta(u) du} \leq Re^{-\nu(t-s)}.$$

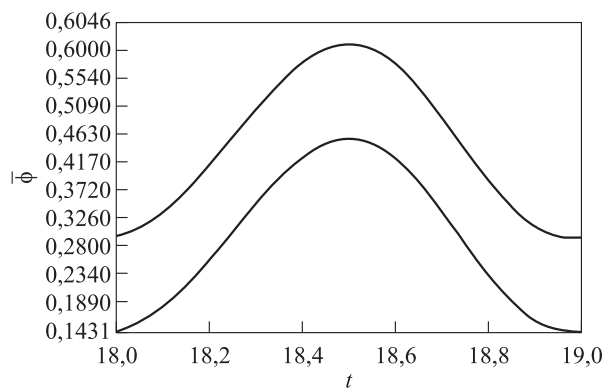


Рис. 1 Предельное среднее $\bar{\phi}(t)$

Тогда при $d_k = \delta^{k-1}$, $M = R$, $a = \nu$ процесс $X(t)$ 1D-экспоненциально слабо эргодичен и, следовательно, справедливы следующие оценки устойчивости:

$$\limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\|_{1D} \leq \frac{R(R(\mathfrak{B} - \bar{\mathfrak{B}})\mathbf{f} + \nu(\mathbf{f} - \bar{\mathbf{f}}))}{\nu(\nu - R(\mathfrak{B} - \bar{\mathfrak{B}}))}; \quad (10)$$

$$\limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| \leq \frac{R(R(\mathfrak{B} - \bar{\mathfrak{B}})\mathbf{f} + \nu(\mathbf{f} - \bar{\mathbf{f}}))}{W\nu(\nu - R(\mathfrak{B} - \bar{\mathfrak{B}}))}, \quad (11)$$

где $W = \inf_{i \geq 1} \delta^{i-1}/i > 0$. Если, кроме того, процесс $X(t)$ стационарен (т.е. интенсивности λ, μ и ξ не зависят от t), то $R = 1, \nu = (1 - \delta^{-1})(S\mu - \delta\lambda)$, $X(t)$ 1D-экспоненциально сильно эргодичен и выполняются соответствующие оценки устойчивости.

Рассмотрим теперь более конкретный пример. Пусть $\bar{X}(t)$ — число требований в системе обслуживания $M_t/M_t/2$ без катастроф, а интенсивности поступления требований $\bar{\lambda}_n(t) = \bar{\lambda}(t) = 2 + \sin 2\pi t + \varepsilon_1 \cos t^2$, обслуживания — $\bar{\mu}_n(t) = \min(n, 2)\bar{\mu}(t) = \min(n, 2)(3 + \cos 2\pi t + \varepsilon_2 \sin \sqrt{t})$, где $|\varepsilon_i| \leq \epsilon^*$. Выберем невозмущенный процесс $X(t)$ с интенсивностями $\lambda_n(t) = \lambda(t) = 2 + \sin 2\pi t$ и $\mu_n(t) = \min(n, 2)\mu(t) = \min(n, 2)(3 + \cos 2\pi t)$. Положим $d_k = 2^{k-1}$, тогда

$$\alpha(t) = 1 + \cos 2\pi t - \sin 2\pi t.$$

При этом $\mathbf{f} = 3, W = 1, a = \alpha^* = \int_0^1 \alpha(t) dt = 1$,

$M = \sup_{|t-s| \leq 1} e^{\int_s^t \alpha(\tau) d\tau} \leq 2$ и справедливы следующие оценки устойчивости:

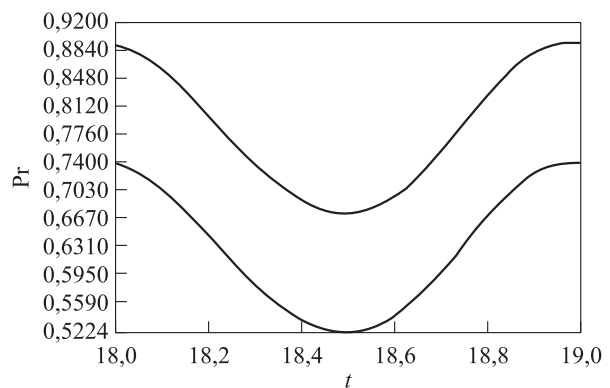


Рис. 2 Вероятность $\text{Pr}\{\bar{X}(t) = 0\}$

$$\begin{aligned} \limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\|_{1D} &\leq \frac{74\epsilon^*}{1 - 12\epsilon^*}; \\ \limsup_{t \rightarrow \infty} |\phi(t) - \bar{\phi}(t)| &\leq \frac{74\epsilon^*}{1 - 12\epsilon^*}; \\ \limsup_{t \rightarrow \infty} \|\mathbf{p}(t) - \bar{\mathbf{p}}(t)\| &\leq \frac{296\epsilon^*}{1 - 12\epsilon^*}. \end{aligned}$$

С учетом оценок скорости сходимости получаем, что для построения с точностью 10^{-6} при $\epsilon^* = 10^{-3}$ предельных характеристик исходного процесса достаточно брать $t \geq 18$. А используя прием, подробно изученный в [40], находим, что для достижения нужной точности достаточно рассмотреть «усеченный» процесс с пространством состояний от 0 до 40.

Рисунки 1 и 2 дают «полосы», показывающие границы для приближенных предельных характеристики $\bar{p}_0(t)$ и $\bar{\phi}(t)$.

4.2 Простой процесс рождения, гибели и иммиграции с конечным числом состояний

Эта модель рассмотрена в недавней статье [42], где получены и оценки скорости сходимости в разных ситуациях (см. также описание в [43] и [44]). Здесь $\lambda_n(t) = (n + 1)\lambda(t)$ — интенсивности рождения, $\mu_n(t) = n\mu(t)$ — интенсивности гибели, $\xi_n(t) = \xi(t)$ — интенсивности катастрофы (мгновенного исчезновения популяции), а пространство возможных состояний — $\{0, 1, \dots, S\}$.

Утверждение 2. Пусть существуют положительные R и ν такие, что при любых $s, t, 0 \leq s \leq t$,

$$e^{-\int_s^t (\mu(u) - \lambda(u) + \xi(u)) du} \leq Re^{-\nu(t-s)}.$$

Тогда $X(t)$ $1D$ -экспоненциально слабо эргодичен, причем $d_k = 1, M = R, a = \nu, W = 1/S$. Следовательно, выполняются оценки устойчивости (10) и (11). Если же, кроме того, $X(t)$ стационарен (т. е. интенсивности λ, μ и ξ не зависят от t), то $R = 1, \nu = \mu - \lambda + \xi$ и $X(t)$ будет $1D$ -экспоненциально сильно эргодичным.

Оценки скорости сходимости при конкретных интенсивностях для этой модели рассмотрены в [42].

5 Система массового обслуживания с групповым поступлением и обслуживанием заявок

Пусть теперь $X(t)$ — число требований в нестационарной марковской модели массового обслу-

живания с групповым поступлением и обслуживанием требований. Будем предполагать, что интенсивности поступления $\chi_k(t)$ и обслуживания $\nu_k(t)$ требований не зависят от числа заявок в системе. Дополнительно считаем, что при всех k и почти всех $t \geq 0$ выполняются неравенства $\chi_{k+1}(t) \leq \chi_k(t)$ и $\nu_{k+1}(t) \leq \nu_k(t)$. Тогда процессу $X(t)$ соответствует следующая транспонированная матрица интенсивностей:

$$A(t) = \begin{pmatrix} a_{00}(t) & \nu_1(t) & \nu_2(t) & \cdots & \nu_r(t) & \cdots \\ \chi_1(t) & a_{11}(t) & \nu_1(t) & \cdots & \nu_{r-1}(t) & \ddots \\ \chi_2(t) & \chi_1(t) & a_{22}(t) & \cdots & \nu_{r-2}(t) & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ \chi_r(t) & \chi_{r-1}(t) & \cdots & \chi_1(t) & a_{rr}(t) & \ddots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

где $a_{ii}(t) = -\sum_{k=1}^i \nu_k(t) - \sum_{k=1}^{\infty} \chi_k(t)$.

Условие ограниченности (1) приводит к неравенству

$$\sum_{k=1}^{\infty} \nu_k(t) + \sum_{k=1}^{\infty} \chi_k(t) \leq L$$

почти при всех $t \geq 0$.

Основные свойства таких систем изучены в работах [39, 45–48].

Элементы матрицы (4) здесь выглядят следующим образом:

$$b_{ij} = \begin{cases} a_{jj} - \chi_j & \text{при } i = j \geq 1; \\ \chi_{i-j} - \chi_i & \text{при } 1 \leq j < i; \\ \nu_{j-i} - \chi_i & \text{при } 1 \leq i < j. \end{cases}$$

Следовательно,

$$DBD^{-1} = \begin{pmatrix} a_{11} & \frac{d_1}{d_2}(\nu_1 - \nu_2) & \frac{d_1}{d_3}(\nu_2 - \nu_3) & \cdots \\ \frac{d_2}{d_1}\chi_1 & a_{22} & \frac{d_2}{d_3}(\nu_1 - \nu_3) & \ddots \\ \frac{d_3}{d_1}\chi_2 & \frac{d_3}{d_2}\chi_1 & a_{33} & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix};$$

$$\begin{aligned} \|B(t)\|_{1D} &\leq \\ &\leq \sup_{k \geq 0} \left(|a_{kk}(t)| + \sum_{i=1}^{\infty} \frac{d_{k+i}}{d_k} \chi_i(t) \right) + \sum_{i=1}^{\infty} \nu_i(t). \end{aligned}$$

А значит, если потребовать, чтобы

$$\sum_{i=1}^{\infty} \frac{d_{k+i}}{d_k} \chi_i(t) \leq K_1 < \infty,$$

то получим

$$\|B(t)\|_{1D} \leq 2L + K_1$$

почти при всех $t \geq 0$.

Предположим еще, что почти при всех $t \geq 0$

$$\|\mathbf{f}(t)\|_{1D} = d_1 \chi_1(t) + (d_1 + d_2) \chi_2(t) + \dots \leq K_2 < \infty.$$

Рассмотрим выражения

$$\beta_i(t) = -a_{ii}(t) - \sum_{k=1}^{i-1} (\nu_{i-k}(t) - \nu_i(t)) \frac{d_k}{d_i} - \sum_{k \geq 1} \chi_k(t) \frac{d_{k+i}}{d_i}, \quad k \geq 0;$$

$$\beta(t) = \inf_{k \geq 0} \beta_k(t).$$

Тогда справедливо следующее утверждение.

Теорема 3. Пусть

$$\int_0^\infty \beta(t) dt = +\infty. \quad (12)$$

Тогда $X(t)$ слабо эргодичен. Если же, кроме того, при некоторых положительных M и a и почти всех s, t ($0 \leq s \leq t$)

$$e^{-\int_s^t \beta(\tau) d\tau} \leq M e^{-a(t-s)},$$

то $X(t)$ $1D$ -экспоненциально слабо эргодичен и справедливы соответствующие оценки устойчивости.

Следствие 6. Если вдобавок процесс $X(t)$ стационарен и вместо (12) выполняется соответствующее неравенство $\beta > 0$, то $X(t)$ $1D$ -экспоненциально сильно эргодичен. При этом $M = 1$ и $a = \beta$.

Следствие 7. Пусть все интенсивности 1-периодичны по t , а вместо (12) выполнено условие:

$$\int_0^1 \beta(t) dt = \beta^* > 0.$$

Тогда $X(t)$ $1D$ -экспоненциально слабо эргодичен и

$$a = \beta^*, \quad M = \sup_{|t-s| \leq 1} e^{\int_s^t \beta(\tau) d\tau}.$$

6 Простая система обслуживания с групповым поступлением и обслуживанием требований

Рассмотрим здесь, следуя [49], модель, в которой возможно поступление групп требований размером не более m и обслуживание групп требований размером не более n . Интенсивность поступления группы k заявок ($k \leq m$) равна $\lambda(t)$, а интенсивность обслуживания группы k заявок ($k \leq n$) равна $\mu(t)$.

Утверждения об эргодичности и скорости сходимости для этой модели были получены в [49]. Сформулируем здесь основной результат.

Утверждение 3. Пусть найдется $\delta > 1$ такое, что

$$\int_0^\infty \beta^*(t) dt = +\infty, \quad (13)$$

где

$$\beta^*(t) = (\delta - 1) \left(\left\{ \frac{1}{\delta} + \frac{1 + \delta}{\delta^2} + \dots \dots + \frac{1 + \delta + \dots + \delta^{n-1}}{\delta^n} \right\} \mu(t) - \left\{ 1 + (1 + \delta) + \dots + (1 + \delta + \dots + \delta^{m-1}) \right\} \lambda(t) \right).$$

Тогда процесс $X(t)$ слабо эргодичен.

Если найдутся положительные R^* и ν^* такие, что при всех s, t , $0 \leq s \leq t$,

$$e^{-\int_s^t \beta^*(u) du} \leq R^* e^{-\nu^*(t-s)},$$

то при $d_k = \delta^{k-1}$ процесс $X(t)$ $1D$ -экспоненциально слабо эргодичен, $M = R^*$, $a = \nu^*$ и справедливы соответствующие оценки устойчивости.

Замечание 2. Если интенсивности поступления и обслуживания требований постоянны, то условие (13) эквивалентно неравенству

$$m(m+1)\lambda - n(n+1)\mu < 0.$$

В этом случае $X(t)$ $1D$ -экспоненциально сильно эргодичен, $M = R^* = 1$, $a = \nu^* = \beta^*$.

Если интенсивности 1-периодичны по t , то условие (13) эквивалентно неравенству

$$m(m+1) \int_0^1 \lambda(t) dt - n(n+1) \int_0^1 \mu(t) dt < 0,$$

которое гарантирует $1D$ -экспоненциально слабую эргодичность $X(t)$.

Литература

1. Штойян Д. Качественные свойства и оценки стохастических моделей / Пер с нем. — М.: Мир, 1979. 272 с. (Stoyan D. Qualitative Eigenschaften und Abschätzungen stochastischer Modelle. — Berlin: Akademie-Verlag, 1977. 198 p.)
2. Zolotarev V. M. Quantitative estimates for the continuity property of queueing systems of type $G/G/\infty$ // Theory Probab. Appl., 1977. Vol. 22. P. 679–691.
3. Калашиников В. В. Качественный анализ сложных систем методом пробных функций. — М.: Наука, 1978. 248 с.
4. Stability problems for stochastic models // Lecture notes in mathematics / Eds. V. V. Kalashnikov, V. M. Zolotarev. — N.Y.: Springer Verlag, 1983. Vol. 982. 295 p.
5. Stability problems for stochastic models // Lecture notes in mathematics / Eds. V. V. Kalashnikov, B. Penkov, V. M. Zolotarev. — N.Y.: Springer Verlag, 1987. Vol. 1233. 224 p.
6. Stability problems for stochastic models // Lecture notes in mathematics / Eds. V. V. Kalashnikov, V. M. Zolotarev. — N.Y.: Springer Verlag, 1989. Vol. 1412. 380 p.
7. Stability problems for stochastic models // Stability problems for stochastic models: 15th Perm Seminar Proceedings / Eds. V. M. Zolotarev, V. M. Kruglov, V. Yu. Korolev. — Perm, Russia, 1992. 312 p.
8. Stability problems for stochastic models // Lecture notes in mathematics / Eds. V. V. Kalashnikov, V. M. Zolotarev. — N.Y.: Springer Verlag, 1993. Vol. 1546. 229 p.
9. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996. 267 p.
10. Карташов Н. В. Сильно устойчивые цепи Маркова // Проблемы устойчивости стохастических моделей: Труды семинара. — М.: ВНИИСИ, 1981. С. 54–59.
11. Карташов Н. В. Неравенства в теоремах эргодичности и устойчивости цепей Маркова с общим фазовым пространством. I // Теория вероятностей и ее применения, 1985. Т. 30. Вып. 2. С. 230–240.
12. Zeifman A. I. Stability for continuous-time nonhomogeneous Markov chains // Lecture notes in mathematics / Eds. V. V. Kalashnikov, B. Penkov, V. M. Zolotarev. — N.Y.: Springer Verlag, 1985. Vol. 1233. P. 401–414.
13. Aldous D. J., Fill J. Reversible Markov chains and random walks on graphs. Ch. 8. <http://www.stat.berkeley.edu/users/aldous/RWG/book.html>.
14. Diaconis P., Stroock D. Geometric bounds for eigenvalues of Markov chains // Ann. Appl. Probab., 1991. Vol. 1. P. 36–61.
15. Mitrophanov A. Yu. Stability and exponential convergence of continuous-time Markov chains // J. Appl. Probab., 2003. Vol. 40. P. 970–979.
16. Mitrophanov A. Yu. The spectral gap and perturbation bounds for reversible continuous-time Markov chains // J. Appl. Probab., 2004. Vol. 41. P. 1219–1222.
17. Mitrophanov A. Yu. Ergodicity coefficient and perturbation bounds for continuous-time Markov chains // Math. Inequal. Appl., 2005. Vol. 8. P. 159–168.
18. Митрофанов А. Ю. Оценки устойчивости для конечных однородных цепей Маркова с непрерывным временем // Теория вероятностей и ее применения, 2005. Т. 50. Вып. 2. С. 371–379.
19. Kartashov N. V. Strong stable Markov chains. — Utrecht, Kiev: VSP, TBiMC, 1996. 138 p.
20. Altman E., Avrachenkov K., Núñez-Queija R. Perturbation analysis for denumerable Markov chains with application to queueing models // Adv. Appl. Probab., 2004. Vol. 36. P. 839–853.
21. Mouhoubi Z., Aïssani D. New perturbation bounds for denumerable Markov chains // Linear Algebra and Its Applications, 2010. Vol. 432. P. 1627–1649.
22. Ferre D., Herve L., Ledoux J. Regular perturbation of V-geometrically ergodic Markov chains // J. Appl. Probab., 2013. Vol. 50. P. 184–194.
23. Meyn S. P., Tweedie R. L. Computable bounds for geometric convergence rates of Markov chains // Ann. Appl. Probab., 1994. Vol. 4. P. 981–1012.
24. Yuanyuan L. Perturbation bounds for the stationary distributions of Markov chains // SIAM J. Matrix Anal. Appl., 2012. Vol. 33. P. 1057–1074.
25. Gnedenko D. B. On a generalization of Erlang formulae // Zastosow. Mat., 1972. Vol. 12. P. 239–242.
26. Gnedenko B., Soloviev A. On the conditions of the existence of final probabilities for a Markov process // Math. Oper. Stat., 1973. Vol. 4. P. 379–390.
27. Massey W. A., Whitt W. Uniform acceleration expansions for Markov chains with time-varying rates // Ann. Appl. Probab., 1998. Vol. 8. P. 1130–1155.
28. Гнеденко Б. В., Макаров И. П. Свойства решений задачи с потерями в случае периодических интенсивностей // Дифф. уравнения, 1971. Т. 7. Вып. 9. С. 1696–1698.
29. Зейфман А. И. Качественные свойства неоднородных процессов рождения и гибели // Проблемы устойчивости стохастических моделей. — М.: ВНИИСИ, 1988. С. 32–40.
30. Зейфман А. И. Некоторые свойства системы с потерями в случае переменных интенсивностей // Автоматика и телемеханика, 1989. Вып. 1. С. 107–113.
31. Zeifman A. I. Some estimates of the rate of convergence for birth and death processes // J. Appl. Probab., 1991. Vol. 28. P. 268–277.
32. Zeifman A. I. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // Stoch. Proc. Appl., 1995. Vol. 59. P. 157–173.
33. Granovsky B. L., Zeifman A. I. Nonstationary queues: Estimation of the rate of convergence // Queueing Syst., 2004. Vol. 46. P. 363–388.

34. Зейфман А. И., Бенинг В. Е., Соколов И. А. Марковские цепи и модели с непрерывным временем. — М.: Элекс-КМ, 2008. 168 с.
35. Ван Доорн Э. А., Зейфман А. И., Панфилова Т. Л. Оценки и асимптотика скорости сходимости для процессов рождения и гибели // Теория вероятностей и ее применения, 2009. Т. 54. С. 18–38.
36. Zeifman A. I., Isaacson D. On strong ergodicity for nonhomogeneous continuous-time Markov chains // Stoch. Proc. Appl., 1994. Vol. 50. P. 263–273.
37. Zeifman A. I. Stability of birth and death processes // J. Math. Sci., 1998. Vol. 91. P. 3023–3031.
38. Zeifman A. I., Korotysheva A. Perturbation bounds for $M_t/M_t/N$ queue with catastrophes // Stochastic Models, 2012. Vol. 28. P. 49–62.
39. Zeifman A. I., Korolev V., Korotysheva A., Satin Y., Benning V. Perturbation bounds and truncations for a class of Markovian queues // Queueing Syst., 2014 (in press).
40. Zeifman A. I., Leorato S., Orsingher E., Satin Ya., Shilova G. Some universal limits for nonhomogeneous birth and death processes // Queueing Syst., 2006. Vol. 52. P. 139–151.
41. Далецкий Ю. Л., Крейн М. Г. Устойчивость решений дифференциальных уравнений в банаховом пространстве. — М.: Наука, 1970. 386 с.
42. Zeifman A. I., Satin Ya., Panfilova T. Limiting characteristics for finite birth–death–catastrophe processes // Math. Biosci., 2013. Vol. 245. P. 96–102.
43. Thorne J. L., Kishino H., Felsenstein J. An evolutionary model for maximum-likelihood alignment of DNA sequences // J. Mol. Evol., 1991. Vol. 33. P. 114–124.
44. Mitrophanov, A. Yu., Borodovsky M. Convergence rate estimation for the TKF91 model of biological sequence length evolution // Math. Biosci., 2007. Vol. 209. P. 470–485.
45. Сатин Я. А., Зейфман А. И., Коротышева А. В., Шоргин С. Я. Об одном классе марковских систем обслуживания // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 18–24.
46. Сатин Я. А., Зейфман А. И., Коротышева А. В. О скорости сходимости и усечениях для одного класса марковских систем обслуживания // Теория вероятностей и ее применения, 2012. Т. 57. Вып. 3. С. 611–621.
47. Zeifman A. I., Korotysheva A., Satin Ya., Shilova G., Panfilova T. On a queueing model with group services // Lecture notes in communications in computer and information science. 2013. Vol. 356. P. 198–205.
48. Zeifman A. I., Satin Y., Shilova G., Korolev V., Benning V., Shorgin S. On $M_t/M_t/S$ type queue with group services // ECMS 2013: 27th Conference (European) on Modeling and Simulation Proceedings. — Alesund, Norway, 2013. P. 604–609.
49. Зейфман А. И., Коротышева А. В., Сатин Я. А., Шоргин С. Я. Оценки в нуль-эргодическом случае для некоторых систем обслуживания // Информатика и её применения, 2012. Т. 6. Вып. 4. С. 27–33.

Поступила в редакцию 27.08.13

GENERAL BOUNDS FOR NONSTATIONARY CONTINUOUS-TIME MARKOV CHAINS

A. I. Zeifman^{1,2}, V. Yu. Korolev^{2,3}, A. V. Korotysheva¹, and S. Ya. Shorgin²

¹Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: A general approach for obtaining perturbation bounds of nonstationary continuous-time Markov chains is considered. The suggested approach deals with a special weighted norms related to total variation. The method is based on the notion of a logarithmic norm of a linear operator function and respective bounds for the Cauchy operator of a differential equation. Special transformations of the reduced intensity matrix of the process are applied. The statements are proved which provide estimates of perturbation of probability characteristics for the case of absence of ergodicity in uniform operator topology. Birth–death–catastrophe queueing models and queueing systems with batch arrivals and group services are also considered in the paper. Some classes of such systems are studied, and bounds of perturbations are obtained. Particularly, such bounds are given for the $M_t/M_t/S$ queueing system with possible catastrophes and a simple model of a queueing system with batch arrivals and group services is analyzed. Moreover, approximations of limiting characteristics are considered for the queueing model.

Keywords: nonstationary continuous-time chains and models; nonstationary Markov chains; perturbation bounds; special norms; queueing models

DOI: 10.14357/19922264140111

Acknowledgments

The research is supported by the Russian Foundation for Basic Research (grants Nos. 12-07-00109, 12-07-00115, 13-07-00223, and 14-07-00041).

References

1. Stoyan, D. 1977. Qualitative Eigenschaften und Abschätzungen stochastischer Modelle. Berlin: Akademie-Verlag. 198 p.
2. Zolotarev, V. M. 1978. Quantitative estimates for the continuity property of queueing systems of type $G/G/\infty$. *Theory Probab. Appl.* 22(4):679–691. doi: 10.1137/1122083.
3. Kalashnikov, V. V. 1978. Kachestvennyy analiz slozhnykh sistem metodom probnykh funktsiy [Qualitative analysis of complex systems behavior by test functions method]. Moscow: Nauka. 248 p.
4. Kalashnikov, V. V., and V. M. Zolotarev, eds. 1983. Stability problems for stochastic models. *Lecture notes in mathematics*. N.Y.: Springer Verlag. Vol. 983. 295 p.
5. Kalashnikov, V. V., V. Penkov, and V. M. Zolotarev, eds. 1987. Stability problems for stochastic models. *Lecture notes in mathematics*. N.Y.: Springer Verlag. Vol. 1233. 224 p.
6. Kalashnikov, V. V., and V. M. Zolotarev, eds. 1989. Stability problems for stochastic models. *Lecture notes in mathematics*. N.Y.: Springer Verlag. Vol. 1412. 380 p.
7. Zolotarev, V. M., V. M. Kruglov, and V. Yu. Korolev, eds. 1994. Stability problems for stochastic models. *15th Perm Seminar Proceedings*. Perm, Russia. 312 p.
8. Kalashnikov, V. V., and V. M. Zolotarev, eds. 1993. Stability problems for stochastic models. *Lecture notes in mathematics*. N.Y.: Springer Verlag. Vol. 1546. 229 p.
9. Gnedenko, B. V., and V. Yu. Korolev. 1996. Random summation: Limit theorems and applications. Boca Raton: CRC Press. 267 p.
10. Kartashov, N. V. 1981. Sil'no ustoychivye tsepi Markova [Strong stable Markov chains]. *Problemy ustoychivosti stokhasticheskikh modeley: Trudy seminarov* ["Stability Problems for Stochastic Models" Seminar Proceedings]. Moscow: VNIISI. 54–59.
11. Kartashov, N. V. 1986. Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Theory Probab. Appl.* 30(2):247–259. doi: 10.1137/1122083.
12. Zeifman, A. I. 1985. Stability for continuous-time non-homogeneous Markov chains. *Lecture notes in mathematics*. N.Y.: Springer Verlag. 1155:401–414.
13. Aldous, D. J., and J. Fill. Reversible Markov chains and random walks on graphs. Ch. 8. Available at: <http://www.stat.berkeley.edu/users/aldous/RWG/book.html> (accessed December 25, 2013).
14. Diaconis, P., and D. Stroock. 1991. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* 1:36–61. doi: 10.1214/aoap/1177005980.
15. Mitrophanov, A. Yu. 2003. Stability and exponential convergence of continuous-time Markov chains. *J. Appl. Probab.* 40:970–979. doi: 10.1239/jap/1067436094.
16. Mitrophanov, A. Yu. 2004. The spectral gap and perturbation bounds for reversible continuous-time Markov chains. *J. Appl. Probab.* 41:1219–1222. doi: 10.1239/jap/1101840568.
17. Mitrophanov, A. Yu. 2005. Ergodicity coefficient and perturbation bounds for continuous-time Markov chains. *Math. Inequal. Appl.* 8:159–168. doi: 10.7153/mia-08-15.
18. Mitrophanov, A. Yu. 2006. Stability estimates for finite homogeneous continuous-time Markov chains. *Theory Probab. Appl.* 50(2):319–326. doi: 10.1137/S0040585X97981718.
19. Kartashov, N. V. 1996. *Strong stable Markov chains*. Utrecht, Kiev: VSP, TBiMC. 138 p.
20. Altman, E., K. Avrachenkov, and R. Núñez-Queija. 2004. Perturbation analysis for denumerable Markov chains with application to queueing models. *Adv. Appl. Probab.* 36: 839–853. doi: 10.1239/aap/1093962237.
21. Mouhoubi, Z., and D. Aïssani. 2010. New perturbation bounds for denumerable Markov chains. *Linear Algebra and Its Applications*. 432:1627–1649. doi: 10.1016/j.laa.2009.11.020.
22. Ferre, D., L. Herve, and J. Ledoux. 2013. Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab.* 50:184–194. doi: 10.1239/jap/1363784432.
23. Meyn, S. P., and R. L. Tweedie. 1994. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* 4: 981–1012. doi:10.1214/aoap/1177004900.
24. Yuanyuan, L. 2012. Perturbation bounds for the stationary distributions of Markov chains. *SIAM J. Matrix Anal. Appl.* 33:1057–1074. doi: 10.1137/110838753.
25. Gnedenko, D. B. 1972. On a generalization of Erlang formulae. *Zastosow. Mat.* 12:239–242.
26. Gnedenko, B., and A. Soloviev. 1973. On the conditions of the existence of final probabilities for a Markov process. *Math. Oper. Stat.* 4:379–390.
27. Massey, W. A., and W. Whitt. 1998. Uniform acceleration expansions for Markov chains with time-varying rates. *Ann. Appl. Probab.* 8:1130–1155. doi: 10.1214/aoap/1028903375.
28. Gnedenko, B. V., and I. P. Makarov. 1971. Svoystva resheniy zadachi s poteryami v sluchae periodicheskikh intensivnostey [Properties of the solution to the loss problem for periodic rates]. *Differentsial'nye Uravneniya* [Differential Equations] 7:1696–1698.
29. Zeifman, A. I. 1991. Qualitative properties of inhomogeneous birth and death processes. *J. Math. Sci.* 57:3217–3224. doi: 10.1007/BF01099019.
30. Zeifman, A. I. 1989. Properties of a system with losses in the case of variable rates. *Autom. Rem. Contr.* 50:82–87.
31. Zeifman, A. I. 1991. Some estimates of the rate of convergence for birth and death processes. *J. Appl. Probab.* 28:268–277.

32. Zeifman, A. I. 1995. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. *Stoch. Proc. Appl.* 59:157–173. doi: 10.1016/0304-4149(95)00028-6.
33. Granovsky, B. L., and A. I. Zeifman. 2004. Nonstationary queues: Estimation of the rate of convergence. *Queueing Syst.* 46:363–388. doi: 10.1023/B:QUES.0000027991.19758.b4.
34. Zeifman, A. I., V. E. Bening, and I. A. Sokolov. 2008. *Markovskie tsepi i modeli s nepreryvnym vremenem* [Continuous-time Markov chains and models]. Moscow: Elex-KM. 168 p.
35. Van Doorn, E. A., A. I. Zeifman, and T. L. Panfilova. 2010. Bounds and asymptotics for the rate of convergence of birth–death processes. *Theory Probab. Appl.* 54:97–113. doi: 10.1137/S0040585X97984097.
36. Zeifman, A. I., and D. Isaacson. 1994. On strong ergodicity for nonhomogeneous continuous-time Markov chains. *Stoch. Proc. Appl.* 50:263–273. doi: 10.1016/0304-4149(94)90123-6.
37. Zeifman, A. I. 1998. Stability of birth and death processes. *J. Math. Sci.* 91:3023–3031. doi: 10.1007/BF02432876.
38. Zeifman, A. I., and A. Korotysheva. 2012. Perturbation bounds for $M_t/M_t/N$ queue with catastrophes. *Stochastic models* 28:49–62. doi: 10.1080/15326349.2011.614900.
39. Zeifman, A., V. Korolev, A. Korotysheva, Y. Satin, and V. Bening. 2014 (in press). Perturbation bounds and truncations for a class of Markovian queues. *Queueing Syst.*
40. Zeifman, A. I., S. Leorato, E. Orsingher, Y. Satin, and G. Shilova. 2006. Some universal limits for nonhomogeneous birth and death processes. *Queueing Syst.* 52:139–151. doi: 10.1007/s11134-006-4353-9.
41. Daleckij, Ju, L. and M. G. Krein. 1974. Stability of solutions of differential equations in Banach space. *Am. Math. Soc. Transl.* 43. 386 p.
42. Zeifman, A. I., Ya. Satin, and T. Panfilova. 2013. Limiting characteristics for finite birth–death–catastrophe processes. *Math. Biosci.* 245:96–102. doi: 10.1016/j.mbs.2013.02.009.
43. Thorne, J. L., H. Kishino, and J. Felsenstein. 1991. An evolutionary model for maximum-likelihood alignment of DNA sequences. *J. Mol. Evol.* 33:114–124. doi: 10.1007/BF02193625.
44. Mitrophanov, A. Yu., and M. Borodovsky. 2007. Convergence rate estimation for the TKF91 model of biological sequence length evolution. *Math. Biosci.* 209:470–485. doi: 10.1016/j.mbs.2007.02.011.
45. Satin, Ya. A., A. I. Zeifman, A. V. Korotysheva, and S. Ya. Shorgin. 2011. Ob odnom klasse markovskikh sistem obsluzhivaniya [On a class of Markovian queues]. *Informatika i ee Primeneniya — Inform. Appl.* 5(4):18–24.
46. Satin, Ya. A., A. I. Zeifman, A. V. Korotysheva. 2013. On the rate of convergence and truncations for a class of Markovian queueing systems. *Theory Probab. Appl.* 57:529–539. doi: 10.1137/S0040585X97986151.
47. Zeifman, A. I., A. Korotysheva, Ya. Satin, G. Shilova, and T. Panfilova. 2013. On a queueing model with group services. *Lecture notes in communications in computer and information science.* 356:198–205. doi: 10.1007/978-3-642-35980-4_22.
48. Zeifman, A. I., Y. Satin, G. Shilova, V. Korolev, V. Bening, and S. Shorgin. 2013. On $M_t/M_t/S$ type queue with group services. *ECMS 2013: 27th Conference (European) on Modeling and Simulation Proceedings.* Alesund, Norway. 604–609. doi: 10.7148/2013-0604.
49. Zeifman, A. I., A. V. Korotysheva, Ya. A. Satin, and S. Ya. Shorgin. 2012. Otsenki v nul’-ergodicheskom sluchae dlya nekotorykh sistem obsluzhivaniya [Bounds in null ergodic case for some queueing systems]. *Informatika i ee Primeneniya — Inform. Appl.* 6(4):27–33.

Received August 27, 2013

Contributors

Zeifman Alexander I. (b. 1954) — Doctor of science in physics and mathematics, professor, Head of Department of Applied Mathematics, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; principal scientist, ISEDT, Russian Academy of Sciences, Vologda, Russian Federation; a_zeifman@mail.ru

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vkorolev@cs.msu.su

Korotysheva Anna V. (b. 1988) — senior lecturer, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; a_korotysheva@mail.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; SShorgin@ipiran.ru

ОБ АППРОКСИМАЦИИ И СХОДИМОСТИ ОДНОМЕРНЫХ ПАРАБОЛИЧЕСКИХ ИНТЕГРОДИФФЕРЕНЦИАЛЬНЫХ МНОГОЧЛЕНОВ И СПЛАЙНОВ

В. И. Киреев¹, М. М. Гершкович², Т. К. Бирюкова³

Аннотация: Работа посвящена исследованию методов аппроксимации функций одномерными интегродифференциальными многочленами (ИД-многочленами) второй степени и построенными на их базе консервативными параболическими интегродифференциальными сплайнами (ИД-сплайнами). В большинстве практических вычислительных задач точность исходных данных не превышает точности аппроксимации параболическими многочленами и сплайнами. Традиционные параболические сплайны, основанные на дифференциальных условиях согласования с аппроксимируемой функцией (дифференциальные сплайны), строятся со сдвигом узлов сплайна относительно узлов интерполяции для обеспечения устойчивости процесса аппроксимации, что существенно усложняет расчетные алгоритмы. Кроме того, традиционные дифференциальные сплайны не обладают свойством консервативности, т. е. не сохраняют интегральные свойства аппроксимируемых функций. Авторами разработаны новые параболические ИД-сплайны, основанные на использовании интегральной невязки в качестве условия согласования сплайна и исходной функции. Такие сплайны являются устойчивыми без сдвига их узлов относительно узлов приближаемой функции и консервативными в смысле сохранения площадей под кривыми. В статье доказаны теоремы об аппроксимации функций одномерными параболическими ИД-многочленами и о сходимости построенных на их базе параболических ИД-сплайнов. Предложенные ИД-сплайны рекомендуются для применения при построении математических моделей обработки данных в больших территориально распределенных информационных системах.

Ключевые слова: сплайн; многочлен; параболический; интегродифференциальный; аппроксимация; интерполяция; слаживание; оценка погрешностей; теорема сходимости; математическая модель обработки данных

DOI: 10.14357/19922264140112

1 Введение

Одной из актуальных задач информатики является задача управления большими базами данных, постоянно пополняемыми в реальном времени. Под большой базой данных здесь понимается пространственно распределенная, неоднородная (с отличиями по структуре в узлах сети), иерархически организованная база данных, реализующая следующие функции:

- динамическое пополнение с первичной обработкой и идентификацией данных;
- мониторинг и анализ данных;
- поддержка параллельного осуществления процессов внутренней обработки данных и поиска ответов на внешние запросы, а также подготовки результатов для принятия решений.

Очевидно, что для успешной работы в реальном времени с базами данных, обладающими такими

свойствами, необходимы специальные математические методы, сокращающие трудоемкость вычислительных процессов без заметного ущерба для точности результатов. В ряде случаев для такой цели весьма полезными оказываются сплайны. Это обусловлено тем, что сплайны могут с требуемой точностью восстанавливать кривые и поверхности, сохраняя их специфические особенности, что особенно важно, когда отсутствует строгая математическая модель описания поведения кривой или поверхности.

Применение сплайнов целесообразно при оптимизации функционирования больших баз данных, например, в задачах [1]:

- идентификации информационных объектов, составляющих базы данных, кластеризации данных при выработке суждений о степени их близости между собой;
- поиска зависимостей между информационными объектами;

¹Московский государственный горный университет, Vladimir-Kireyev@mail.ru

²Институт проблем информатики Российской академии наук, makmg@mail.ru

³Институт проблем информатики Российской академии наук, yukonta@mail.ru

- оптимизации процессов передачи данных по каналам связи;
- построения интерполяционных и сглаживающих кривых экспертных оценок, участвующих в подготовке результатов для принятия решений.

С этой целью авторами был разработан и реализован особый вид ИД-сплайнов. В статье приводятся теоретические результаты исследования поведения таких сплайнов, оценивается погрешность аппроксимации, доказываемся сходимость и устойчивость сплайнов.

В настоящее время наиболее развиты и математически обоснованы методы аппроксимации кубическими сплайнами, которые являются устойчивыми, а по способу построения — дифференциальными [2–4]. Однако в большинстве практических вычислительных задач точность исходных данных не превышает точности аппроксимации параболическими многочленами и сплайнами.

Для обеспечения устойчивости традиционных дифференциальных параболических сплайнов приходится осуществлять сдвиг узлов сплайна относительно узлов аппроксимируемой функции [5], что усложняет расчетные алгоритмы. Кроме того, все традиционные дифференциальные сплайны, основанные на дифференциальных условиях согласования с аппроксимируемой функцией, не обладают свойством консервативности, т. е. не сохраняют интегральные свойства аппроксимируемых функций.

В работах [6, 7] и др. авторами предложены ИД-многочлены и ИД-сплайны четных степеней, обладающие свойством консервативности, которое обеспечивается использованием интегрального условия согласования при их построении. Интегродифференциальный способ приближения широко используется при конструировании численных методов [8].

В данной статье, являющейся развитием работ [6, 7], доказана теорема об оценке погрешностей аппроксимации функций различных классов гладкости параболическими ИД-многочленами и теорема сходимости параболических слабосглаживающих ИД-сплайнов $S_{2ID}(x)$ дефекта 1.

Под алгебраическим ИД-многочленом (в общем случае эрмитова типа) $S_{r,i}(x)$ степени r , аппроксимирующим на отрезке $[x_i, x_{i+1}]$ некоторую сеточную функцию $f_i = f(x_i)$, заданную на $[a, b]$ на сетке несовпадающих узлов

$$\Delta_1 : a = x_0 < x_1 < \dots < x_i < x_{i+1} < \dots < x_{n-1} < x_n = b, \quad (1)$$

понимается функция вида

$$S_{r,i}(x) = \sum_{k=0}^r a_{k,i}(x - x_i)^k. \quad (2)$$

Здесь $a_{k,i}$ — коэффициенты, определяемые из совокупности интегральных и дифференциальных условий согласования:

$$\begin{aligned} \delta S_{r,i}^{(-1)}(x_i, x_{i+1}) &= \int_{x_i}^{x_{i+1}} [S_{r,i}(x) - f(x)] dx = 0; \quad (3) \\ \delta S_{r,i}^{(p_1)}(x_k) &= S_{r,i}^{(p_1)}(x_k) - f^{(p_1)}(x_k) = 0 \\ &\quad (k = i, i + 1), \quad (4) \end{aligned}$$

где p_1 — порядки производных, принимающие целые значения.

Количество условий (4) (с различными p_1) определяется степенью многочлена.

Тогда функция $S_r^{[q]}(x) = \bigcup_{i=0}^{n-1} S_{r,i}(x)$, определенная на отрезке $[a, b]$ и принадлежащая классу гладкости $C_{[a,b]}^m$, составленная из звеньев $S_{r,i}(x)$, называется *одномерным алгебраическим ИД-сплайном степени r дефекта q* ($0 \leq m \leq r$, $q = r - m$) с узлами на сетке Δ_1 (1), если каждое звено сплайна при $x \in [x_i, x_{i+1}]$ ($i = 0, \dots, n - 1$) представляется в виде многочлена (2) с коэффициентами $a_{k,i}$, определяемыми из совокупности интегральных и дифференциальных условий согласования (3), (4) (где p_1 — целое, $0 \leq p_1 \leq m$) и условий непрерывности сплайна $S_r^{[q]}(x)$ и его производных во внутренних узлах сетки Δ_1 :

$$S_{r,i-1}^{(p_2)}(x)|_{x=x_i} = S_{r,i}^{(p_2)}(x)|_{x=x_i} \quad (i = 1, \dots, n - 1),$$

где $0 \leq p_2 \leq m$ — порядки производных, такие что $\{p_1\} \cap \{p_2\} = \emptyset$ и $\{p_1\} \cup \{p_2\} = \{0, 1, 2, \dots, m\}$.

Ранее показано [6], что условия согласования

$$\int_{x_i}^{x_{i+1}} S_{2,i}(x) dx = I_i^{i+1},$$

где

$$I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx;$$

$$S_{2,i}(x_k) = f_k;$$

$$f_k = f(x_k) \quad (k = i, i + 1),$$

определяют параболический интерполяционный ИД-многочлен на $[x_i, x_{i+1}]$:

$$S_{2ID,i}(x) = f_i + \left(\frac{6\nabla I_i^{i+1}}{h_{i+1}^2} - \frac{2\Delta f_i}{h_{i+1}} \right) (x - x_i) + \left(-\frac{6\nabla I_i^{i+1}}{h_{i+1}^3} + \frac{3\Delta f_i}{h_{i+1}^2} \right) (x - x_i)^2, \quad (5)$$

где $h_{i+1} = x_{i+1} - x_i$; $\nabla I_i^{i+1} = I_i^{i+1} - f_i h_{i+1}$; $\Delta f_i = f_{i+1} - f_i$.

Далее в разд. 2 проведены исследования погрешностей аппроксимации функций различных классов гладкости с помощью ИД-многочленов $S_{2ID,i}(x)$. В разд. 3 доказана теорема сходимости параболического ИД-сплайна минимального дефекта ($q = 1$). В разд. 4 перечислены основные выводы по данной работе.

2 Оценки погрешностей аппроксимации сеточных функций интегриродифференциальными многочленами

Проведем оценку погрешностей аппроксимации функций $f(x)$ различных классов гладкости m с помощью ИД-многочленов $S_{2ID,i}(x)$ (5) в предположении, что параметры I_i^{i+1} , f_i , f_{i+1} многочлена $S_{2ID,i}(x)$ известны точно или вычислены с точностью не ниже $O(h^{m+2})$ для I_i^{i+1} и $O(h^{m+1})$ для f_i , f_{i+1} (в этом случае порядок аппроксимации максимален).

В пространстве непрерывных на отрезке $[a, b]$ функций используется норма

$$\|g(x)\|_{[a,b]} = \max_{x \in [a,b]} |g(x)|.$$

Обозначим через

$$R^{(p)}(x) = S_r^{(p)}(x) - f^{(p)}(x)$$

($p = 0, 1$ — порядок производной) остаточное слагаемое аппроксимации. Тогда на частичном отрезке $[x_i, x_{i+1}]$ в качестве погрешности принимается норма:

$$\|R^{(p)}(x)\|_{[x_i, x_{i+1}]} = \max_{x \in [x_i, x_{i+1}]} |R^{(p)}(x)|.$$

Многочлен $S_{2ID,i}(x)$ (5) в форме Лагранжа имеет вид:

$$S_{2ID,i}(x) = \frac{6u(1-u)}{h_{i+1}} I_i^{i+1} + (1-u)(1-3u)f_i + u(3u-2)f_{i+1},$$

где

$$u = \frac{x - x_i}{h_{i+1}} \quad (0 \leq u \leq 1).$$

Остаточное слагаемое интерполяции $f(x)$ многочленом $S_{2ID,i}(x)$ на отрезке $[x_i, x_{i+1}]$ выражается разностью

$$R_{2ID}(x) = S_{2ID,i}(x) - f(x) = \frac{\varphi_1(u)}{h_{i+1}} I_i^{i+1} + \varphi_2(u)f_i + \varphi_3(u)f_{i+1} - f(x), \quad (6)$$

где

$$\left. \begin{aligned} \varphi_1(u) &= 6u(1-u); \\ \varphi_2(u) &= (1-u)(1-3u); \\ \varphi_3(u) &= u(3u-2). \end{aligned} \right\} \quad (7)$$

Здесь I_i^{i+1} , f_i и f_{i+1} являются параметрами многочлена.

Рассмотрим случай $f(x) \in C^3_{[x_i, x_{i+1}]}$. Заменяя в (6) параметры $I_i^{i+1} = F_{i+1} - F_i$ и f_i, f_{i+1} (здесь $F_i = F(x_i)$, $F_{i+1} = F(x_{i+1})$, где $F(x)$ — первообразная) их разложениями по формуле Тейлора в точке $x \in (x_i, x_{i+1})$, получим:

$$R_{2ID}(x) = \frac{h_{i+1}^3}{12} u^3(1-u)(-3u^2 + 6u - 2)f'''(\xi) + \frac{h_{i+1}^3}{12} u(1-u)^3(3u^2 - 1)f'''(\eta) \quad (\xi, \eta \in (x_i, x_{i+1})).$$

Здесь учитывается, что точка ξ ($x_i < \xi < x$) для разложения F_i та же, что и для разложения f_i , и точка η ($x < \eta < x_{i+1}$) для разложения F_{i+1} та же, что и для разложения f_{i+1} .

Обозначим

$$\begin{aligned} \psi_1(u) &= (1-u)(-3u^2 + 6u - 2); \\ \psi_2(u) &= u(1-u)^3(3u^2 - 1). \end{aligned}$$

Тогда

$$\begin{aligned} |R_{2ID}(x)| &\leq \\ &\leq \frac{h_{i+1}^3}{12} (|\psi_1(u)| \cdot |f'''(\xi)| + |\psi_2(u)| \cdot |f'''(\eta)|). \end{aligned}$$

По теореме о среднем [2], если $g(x) \in C_{[\xi, \eta]}$ и α и β имеют одинаковые знаки, то

$$\exists \zeta \in [\xi, \eta] : \alpha g(\xi) + \beta g(\eta) = (\alpha + \beta)g(\zeta).$$

Следовательно, если принять в качестве $g(x)$ функцию $|f'''(x)|$ и в качестве коэффициентов α и β значения $|\psi_1(u)|$ и $|\psi_2(u)|$ в данной конкретной точке u ($u \in [0, 1]$), то существует такая точка $\zeta \in [x_i, x_{i+1}]$

Таблица 1 Константы в оценке погрешности интерполяции функции $f(x) \in C^m_{[x_i, x_{i+1}]}$ ($m = 1, 2, 3$) параболическим ИД-многочленом $S_{2ID,i}(x)$

Порядок производной	$T_{3,0}^{(2ID)}$ ($f(x) \in C^3_{[x_i, x_{i+1}]}$)	$T_{2,0}^{(2ID)}$ ($f(x) \in C^2_{[x_i, x_{i+1}]}$)	$T_{1,0}^{(2ID)}$ ($f(x) \in C^1_{[x_i, x_{i+1}]}$)
$p = 0$	$\frac{1}{72\sqrt{3}} \approx 0,0080$	$\frac{1}{25\sqrt{2}} \approx 0,0283$	$\frac{1}{6}$
$p = 1$	$\frac{1}{12}$	$\frac{1}{4}$	2

(поскольку $\xi, \eta \in [x_i, x_{i+1}]$), что $\forall x \in [x_i, x_{i+1}]$ (т.е. $\forall u \in [0, 1]$) выполняется неравенство:

$$\begin{aligned} |R_{2ID}(x)| &\leq \frac{h_{i+1}^3}{12} (|\psi_1(u)| + |\psi_2(u)|) |f'''(\zeta)| \leq \\ &\leq \frac{h_{i+1}^3}{12} (|\psi_1(u)| + |\psi_2(u)|) \|f'''(\zeta)\|. \end{aligned}$$

Правая часть неравенства является мажорантой для модуля остаточного слагаемого, поэтому

$$\begin{aligned} \|R_{2ID}(x)\|_{[x_i, x_{i+1}]} &\leq \\ &\leq \frac{h_{i+1}^3}{12} \max_{u \in [0, 1]} \gamma(u) \cdot \|f'''(x)\|_{[x_i, x_{i+1}]}, \end{aligned}$$

где $\gamma(u) = |\psi_1(u)| + |\psi_2(u)|$.

Максимум функции $\gamma(u)$ при $0 \leq u \leq 1$ достигается в двух точках, симметричных относительно середины отрезка, и равен $1/(6\sqrt{3}) \approx 0,0962$.

Таким образом, при $f(x) \in C^3_{[x_i, x_{i+1}]}$ оценка погрешности интерполяции функции $f(x)$ ИД-многочленом $S_{2ID,i}(x)$ на отрезке $[x_i, x_{i+1}]$ будет иметь вид

$$\begin{aligned} \|R_{2ID}(x)\|_{[x_i, x_{i+1}]} &= \|S_{2ID,i}(x) - f(x)\|_{[x_i, x_{i+1}]} \leq \\ &\leq \frac{h_{i+1}^3}{72\sqrt{3}} \|f'''(x)\|_{[x_i, x_{i+1}]}. \end{aligned}$$

Для производной $f'(x)$ при $f(x) \in C^3_{[x_i, x_{i+1}]}$, а также для функций классов гладкости $C^2_{[x_i, x_{i+1}]}$ и $C^1_{[x_i, x_{i+1}]}$ (и их производных) оценивание погрешностей аппроксимации проводится по тому же алгоритму.

Все полученные результаты обобщаются в виде следующей теоремы.

Теорема 1. Об оценке погрешностей аппроксимации функций ИД-многочленом $S_{2ID,i}(x)$.

Если параболический ИД-многочлен $S_{2ID,i}(x)$ на отрезке $[x_i, x_{i+1}]$ интерполирует функцию $f(x) \in C^m_{[x_i, x_{i+1}]}$ ($m = 1, 2, 3$), причем его параметры $I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx$, $f_k = f(x_k)$ ($k = i, i + 1$) известны

точно или вычислены с точностью не ниже $O(h^{m+2})$, $O(h^{m+1})$ соответственно, то справедливы оценки:

$$\begin{aligned} \|R_{2ID}^{(p)}(x)\|_{[x_i, x_{i+1}]} &= \|S_{2ID,i}^{(p)}(x) - f^{(p)}(x)\|_{[x_i, x_{i+1}]} \leq \\ &\leq T_{m,p}^{(2ID)} h_{i+1}^{m-p} \|f^{(m)}(x)\|_{[x_i, x_{i+1}]}, \end{aligned} \quad (8)$$

где $p = 0, 1$ — порядок производной; $T_{m,p}^{(2ID)}$ — константы, приведенные в табл. 1.

3 Построение и сходимость глобальных параболических интегродифференциальных сплайнов минимального дефекта

Пусть на сетке Δ_1 (1) задана функция $\{f_i = f(x_i) \pm \theta_i\}_{i=0}^n$, где θ_i ($i = 0, \dots, n$) — погрешности измерения или вычисления значений функции, не превышающие $O(H^3)$ ($H = \max_{i=1, \dots, n} h_i$).

Требуется построить глобальный параболический ИД-сплайн $\tilde{S}_{2ID}(x)$ с узлами на сетке Δ_1 , имеющий погрешность аппроксимации

$$\|\tilde{S}_{2ID}(x) - f(x)\|_{[a,b]} = \max_{x \in [a,b]} |\tilde{S}_{2ID}(x) - f(x)|,$$

не превышающую $O(H^3)$ (для $f(x) \in C^m_{[a,b]}$ ($m \geq 3$)), и удовлетворяющий следующим условиям:

- (1) интегральному условию согласования

$$\int_{x_i}^{x_{i+1}} \tilde{S}_{2ID}(x) dx = \hat{I}_i^{i+1}, \quad i = 0, \dots, n - 1; \quad (9)$$

- (2) условию непрерывности сплайна и его первой производной в узлах сетки Δ_1 :

$$\begin{aligned} \tilde{S}_{2ID}^{(p)}(x)|_{x=x_i}^{[x_{i-1}, x_i]} &= \tilde{S}_{2ID}^{(p)}(x)|_{x=x_i}^{[x_i, x_{i+1}]}, \\ p = 0, 1; \quad i &= 1, \dots, n - 1. \end{aligned} \quad (10)$$

Здесь \hat{I}_i^{i+1} — заданные или предварительно вычисленные с точностью не ниже $O(H^4)$ интегралы от функции $f(x)$ на частичных отрезках $[x_i, x_{i+1}]$, образуемых сеткой Δ_1 . Параболический сплайн, удовлетворяющий условиям (10), имеет дефект $q = 1$.

Данный метод аппроксимации будем называть *методом слабого сглаживания*, а соответствующий сплайн — *слабосглаживающим сплайном*.

Как будет показано ниже, для этих сплайнов разность значений сплайна и функции как в узлах сетки Δ_1 , так и на всем рассматриваемом отрезке при аппроксимации функций класса $C_{[a,b]}^m$ ($m \geq 3$) имеет порядок $O(H^3)$ и, таким образом, указанные сплайны близки к интерполяционным. Во многих практических задачах порядок точности исходных данных не превышает $O(H^3)$. В этих случаях слабосглаживающие сплайны в сущности являются интерполяционными, так как обеспечивают выполнение условий интерполяции в узлах сеточной функции с порядком $O(H^3)$.

Однако, в отличие от традиционных интерполяционных параболических дифференциальных сплайнов, устойчивость ИД-сплайнов обеспечивается без сдвига узлов сплайна относительно узлов аппроксимируемой сеточной функции, что обусловлено применением интегрального условия согласования. При этом ИД-сплайны обладают свойством консервативности, а алгоритмы их построения характеризуются простотой реализации и экономичностью.

Для построения слабосглаживающего ИД-сплайна $\tilde{S}_{2ID}(x) = \bigcup_{i=0}^{n-1} \tilde{S}_{2ID,i}(x)$, удовлетворяющего условиям (9) и (10), коэффициенты $a_{k,i}$ ($k = 0, 1, 2$) составляющих его звеньев $\tilde{S}_{2ID,i}(x) = \sum_{k=0}^2 a_{k,i}(x - x_i)^k$ вычисляются из системы уравнений, вытекающей из совокупности соотношений (9) и (10).

Звено $\tilde{S}_{2ID,i}(x)$ на отрезке $[x_i, x_{i+1}]$ с коэффициентами, полученными из (9) и (10) при $p = 0$ (условие интегрального согласования сплайна и аппроксимируемой функции и условие непрерывности сплайна соответственно), имеет вид:

$$\begin{aligned} \tilde{S}_{2ID,i}(x) = & \tilde{f}_i + \left(\frac{6\nabla\hat{I}_i^{i+1}}{h_{i+1}^2} - \frac{2\Delta\tilde{f}_i}{h_{i+1}} \right) (x - x_i) + \\ & + \left(-\frac{6\nabla\hat{I}_i^{i+1}}{h_{i+1}^3} + \frac{3\Delta\tilde{f}_i}{h_{i+1}^2} \right) (x - x_i)^2, \end{aligned}$$

где \tilde{f}_i ($i = 0, \dots, n$) — параметры, равные значениям сплайна $\tilde{S}_{2ID}(x)$ в узлах сетки Δ_1 , $\nabla\hat{I}_i^{i+1} = \hat{I}_i^{i+1} - \hat{f}_i h_{i+1}$, $\Delta\tilde{f}_i = \tilde{f}_{i+1} - \tilde{f}_i$.

Выполнение условия (10) при $p = 1$ (условие непрерывности первой производной сплайна $\tilde{S}_{2ID}(x)$) обеспечивается, если параметры \tilde{f}_i ($i = 0, \dots, n$) удовлетворяют следующим соотношениям (полученным из (10) при $p = 1$ путем алгебраических преобразований):

$$\begin{aligned} \frac{1}{h_i} \tilde{f}_{i-1} + 2 \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \tilde{f}_i + \frac{1}{h_{i+1}} \tilde{f}_{i+1} = \\ = 3 \left(\frac{\hat{I}_i^{i+1}}{h_{i+1}^2} + \frac{\hat{I}_{i-1}^i}{h_i^2} \right), \quad i = 1, \dots, n-1. \end{aligned} \quad (11)$$

Равенства (11) представляют собой трехдиагональную систему линейных алгебраических уравнений с диагональным преобладанием. Эта система в совокупности с краевыми условиями (например, заданными в виде $\tilde{f}_0 = f_0 = f(x_0)$, $\tilde{f}_n = f_n = f(x_n)$) имеет единственное решение, которое можно найти экономичным методом прогонки [9].

Для вычисления интегральных параметров звеньев сплайнов могут быть использованы левосторонние и правосторонние квадратурные формулы, обеспечивающие при $f(x) \in C_{[a,b]}^m$ ($m \geq 3$) порядок точности вычисления интегралов $O(H^4)$ [8]:

$$\hat{I}_{i-1}^i = \frac{h_i^3}{6\bar{H}_i^{i+1}} \left(-\frac{1}{h_{i+1}} f_{i+1} + \frac{\bar{H}_i^{i+1} \bar{H}_i^{3(i+1)}}{h_i^2 h_{i+1}} f_i + \frac{\bar{H}_{2i}^{3(i+1)}}{h_i^2} f_{i-1} \right); \quad (12)$$

$$\hat{I}_i^{i+1} = \frac{h_{i+1}^3}{6\bar{H}_i^{i+1}} \left(\frac{\bar{H}_{3i}^{2(i+1)}}{h_{i+1}^2} f_{i+1} + \frac{\bar{H}_i^{i+1} \bar{H}_{3i}^{i+1}}{h_i h_{i+1}^2} f_i - \frac{1}{h_i} f_{i-1} \right), \quad (13)$$

где $\bar{H}_{ki}^{l(i+1)} = kh_i + lh_{i+1}$ ($k, l > 0$ — натуральные числа).

Для параболического ИД-сплайна $\tilde{S}_{2ID}(x)$ верна следующая теорема сходимости.

Теорема 2. О сходимости одномерного глобального параболического ИД-сплайна.

Пусть функцию $f(x)$ ($x \in [a, b]$), заданную с точностью не ниже $O(H^4)$ ($h_i = x_i - x_{i-1}$, $H = \max_{i=1, \dots, n} h_i$) на сетке Δ_1 (1) с параметром неравномерности сетки

$$Q = \frac{\max_{i=1, \dots, n} h_i}{\min_{i=1, \dots, n} h_i},$$

аппроксимирует слабосглаживающий глобальный параболический ИД-сплайн $\tilde{S}_{2ID}(x)$. Тогда если

$f(x) \in C^3_{[a,b]}$ и параметры \hat{I}_i^{i+1} ($i = 0, \dots, n-1$) определяются по формулам (12), (13), а \tilde{f}_i ($i = 0, \dots, n$) — из трехдиагональной системы линейных алгебраических уравнений (СЛАУ) (11) с краевыми условиями $\tilde{f}_0 = f_0 = f(x_0)$, $\tilde{f}_n = f_n = f(x_n)$, то справедливы оценки

$$\begin{aligned} & \| \tilde{S}_{2ID}^{(p)}(x) - f^{(p)}(x) \|_{[a,b]} \leq \\ & \leq H^{3-p} \left(T_{3,p}^{(2ID)} + K_p Q^{1+p} \right) \| f'''(x) \|_{[a,b]}, \quad (14) \end{aligned}$$

где $p = 0, 1$ — порядок производной, а константы имеют значения:

$$T_{3,0}^{(2ID)} = \frac{1}{72\sqrt{3}}; T_{3,1}^{(2ID)} = \frac{1}{12}; K_0 = \frac{11}{48}; K_1 = \frac{25}{24}.$$

Таким образом, при $f(x) \in C^3_{[a,b]}$ сплайны $\tilde{S}_{2ID}(x)$ равномерно сходятся к функции $f(x)$ на последовательности сеток $\Delta_1^{(n)}$: $a = x_0 < x_1 < \dots < x_i < x_{i+1} < \dots < x_n = b$ по крайней мере со скоростью H^3 , а их производные \tilde{S}'_{2ID} равномерно сходятся к $f'(x)$ по крайней мере со скоростью H^2 с ростом n .

Доказательство. Формула звена сплайна $\tilde{S}_{2ID}(x)$ на отрезке $[x_i, x_{i+1}]$ в форме Лагранжа записывается в виде ИД-многочлена:

$$\begin{aligned} \tilde{S}_{2ID,i}(x) = & \frac{6u(1-u)}{h_{i+1}} \hat{I}_i^{i+1} + (1-u)(1-3u)\tilde{f}_i + \\ & + u(3u-2)\tilde{f}_{i+1}. \end{aligned}$$

Для нахождения погрешности $\| \tilde{S}_{2ID}^{(p)}(x) - f^{(p)}(x) \|_{[a,b]}$ представим ее в виде:

$$\begin{aligned} \| \tilde{S}_{2ID}^{(p)}(x) - f^{(p)}(x) \|_{[a,b]} \leq & \| S_{2ID}^{(p)}(x) - f^{(p)}(x) \|_{[a,b]} + \\ & + \| \tilde{S}_{2ID}^{(p)}(x) - S_{2ID}^{(p)}(x) \|_{[a,b]}, \quad (15) \end{aligned}$$

где $S_{2ID}(x)$ — сплайн, составленный из звеньев (5), параметры которого I_i^{i+1} , f_i , f_{i+1} известны точно (или вычислены с точностью не ниже $O(H^5)$ для I_i^{i+1} и $O(H^4)$ для f_i, f_{i+1}).

Оценки первого слагаемого правой части неравенства (15) вытекают из формулы (8). Поскольку оценки, приведенные в теореме 1, справедливы для любого частичного отрезка $[x_i, x_{i+1}] \subset [a, b]$, $i = 0, \dots, n-1$, то верна формула

$$\begin{aligned} \| S_{2ID,i}^{(p)}(x) - f^{(p)}(x) \|_{[a,b]} \leq \\ \leq T_{m,p}^{(2ID)} H^{m-p} \| f^{(m)}(x) \|_{[a,b]}. \quad (16) \end{aligned}$$

Оценки второго слагаемого получаются следующим образом. На отрезке $[x_i, x_{i+1}]$ разность

$\tilde{S}_{2ID}^{(p)}(x) - S_{2ID}^{(p)}(x)$ при $p = 0, 1$ (соответственно) имеет вид:

$$\begin{aligned} \tilde{S}_{2ID}(x) - S_{2ID}(x) = & \frac{6u(1-u)}{h_{i+1}} \left(\hat{I}_i^{i+1} - I_i^{i+1} \right) + \\ & + (1-u)(1-3u)(\tilde{f}_i - f_i) + \\ & + u(3u-2)(\tilde{f}_{i+1} - f_{i+1}), \quad (17) \end{aligned}$$

$$\begin{aligned} \tilde{S}'_{2ID}(x) - S'_{2ID}(x) = & \frac{6(1-2u)}{h_{i+1}^2} \left(\hat{I}_i^{i+1} - I_i^{i+1} \right) + \\ & + \frac{2(3u-2)}{h_{i+1}} \left(\tilde{f}_i - f_i \right) + \\ & + \frac{2(3u-1)}{h_{i+1}} \left(\tilde{f}_{i+1} - f_{i+1} \right). \quad (18) \end{aligned}$$

Введем функции $\varphi_k(u)$ ($k = 1, 2, 3$) по формулам (7). Дифференцируя $\varphi_k(u)$, получим:

$$\varphi'_1(u) = \frac{d\varphi_1(u)}{du} = 6(1-2u);$$

$$\varphi'_2(u) = \frac{d\varphi_2(u)}{du} = 2(3u-2);$$

$$\varphi'_3(u) = \frac{d\varphi_3(u)}{du} = 2(3u-1).$$

Из (17) и (18) $\forall x \in [a, b]$ вытекает соотношение:

$$\begin{aligned} \| \tilde{S}_{2ID}^{(p)}(x) - S_{2ID}^{(p)}(x) \|_{[a,b]} \leq & \leq \frac{1}{\left(\min_{i=1, \dots, n} h_i \right)^{1+p}} \max_{u \in [0,1]} \left| \varphi_1^{(p)}(u) \right| \times \\ & \times \max_{i=0, \dots, n-1} \left| \hat{I}_i^{i+1} - I_i^{i+1} \right| + \\ & + \frac{1}{\left(\min_{i=1, \dots, n} h_i \right)^p} \max_{u \in [0,1]} \left[\left| \varphi_2^{(p)}(u) \right| + \right. \\ & \left. + \left| \varphi_3^{(p)}(u) \right| \right] \max_{i=0, \dots, n} \left| \tilde{f}_i - f_i \right|. \quad (19) \end{aligned}$$

Можно показать, что

$$\max_{u \in [0,1]} |\varphi_1(u)| = \frac{3}{2} \text{ при } u = \frac{1}{2};$$

$$\max_{u \in [0,1]} |\varphi'(u)| = 1 \text{ при } u = 0;$$

$$\max_{u \in [0,1]} [|\varphi_2(u)| + |\varphi_3(u)|] = 1 \text{ при } u = 0, u = 1;$$

$$\max_{u \in [0,1]} [|\varphi'_2(u)| + |\varphi'_3(u)|] = 6 \text{ при } u = 0, u = 1.$$

Оценки погрешностей вычисления интегралов $\max_{i=0, \dots, n-1} |\hat{I}_i^{i+1} - I_i^{i+1}|$ при нахождении \hat{I}_i^{i+1} ($i = 0, \dots, n-1$) по формулам (12), (13) и погрешностей вычисления значений функции в узлах сетки $\Delta_1 \max_{i=0, \dots, n} |\tilde{f}_i - f_i|$ при нахождении \tilde{f}_i ($i = 0, \dots, n$) из трехдиагональной СЛАУ (11) с

Таблица 2 Оценки погрешностей аппроксимации функции $f(x) = C_{[x_i, x_{i+1}]}^m$ ($m = 1, 2, 3$) и ее первой производной $f'(x)$ слабосглаживающим ИД-сплайном $\tilde{S}_{2ID}(x)$ и его производной $\tilde{S}'_{2ID}(x)$

Погрешность	$f(x) \in C_{[a,b]}^3$	$f(x) \in C_{[a,b]}^2$	$f(x) \in C_{[a,b]}^1$
$\ \tilde{S}_{2ID}^{(p)}(x) - f^{(p)}(x)\ _{[a,b]} \leq$	$H^{3-p}(T_{3,p} + K_{3,p}Q^{1+p})M_3$	$H^{2-p}(T_{2,p} + K_{2,p}Q^{2+p})M_2$	$H^{1-p}(T_{1,p} + (K_{1,p} + K_{1,p}^*Q)Q^{1+p})M_1$
$p = 0$	$T_{3,0} = \frac{1}{72\sqrt{3}}, K_{3,0} = \frac{11}{48}$	$T_{2,0} = \frac{1}{25\sqrt{2}}, K_{2,0} = \frac{9}{24}$	$T_{1,0} = \frac{1}{6}, K_{1,0} = \frac{1}{2}, K_{1,0}^* = \frac{3}{4}$
$p = 1$	$T_{3,1} = \frac{1}{12}, K_{3,1} = \frac{25}{24}$	$T_{2,1} = \frac{1}{8}, K_{2,1} = \frac{19}{12}$	$T_{1,1} = 2, K_{1,1} = 3, K_{1,1}^* = \frac{19}{6}$

Примечания: $M_3 = \|f'''(x)\|_{[a,b]}$, $M_2 = \|f''(x)\|_{[a,b]}$, $M_1 = \|f'(x)\|_{[a,b]}$.

краевыми условиями $\tilde{f}_0 = f_0, \tilde{f}_n = f_n$ найдены в [8] и в случае $f(x) \in C_{[a,b]}^3$ имеют вид:

$$\max_{i=0, \dots, n-1} |\hat{I}_i^{i+1} - I_i^{i+1}| \leq \frac{1}{24} H^4 \|f'''(x)\|_{[a,b]}; \quad (20)$$

$$\max_{i=0, \dots, n} |\tilde{f}_i - f_i| \leq \left(\frac{H^3}{24} + \frac{H^3}{8}\right) Q \|f'''(x)\|_{[a,b]} = \frac{H^3}{6} Q \|f'''(x)\|_{[a,b]}, \quad (21)$$

где $H = \max_{i=1, \dots, n} h_i$.

Итак, из формул (19)–(21) вытекает

$$\|\tilde{S}_{2ID}^{(p)}(x) - S_{2ID}^{(p)}(x)\|_{[a,b]} \leq K_p H^{3-p} Q^{1+p} \|f'''(x)\|_{[a,b]} \quad \text{при } f(x) \in C_{[a,b]}^3, \quad (22)$$

где $p = 0, 1$ — порядок производной; $K_0 = 11/48$, $K_1 = 25/24$.

Таким образом, из формул (15), (16), (22) получается оценка (14).

Теорема 2 доказана.

Аналогичным способом можно найти оценки погрешностей аппроксимации функции $f(x)$ (и ее производной) с помощью сплайна $\tilde{S}_{2ID}(x)$ (и его производной) при $f(x) \in C_{[x_i, x_{i+1}]}^m$, $m = 1, 2$.

Оценки погрешностей аппроксимации функции $f(x) \in C_{[x_i, x_{i+1}]}^m$ ($m = 1, 2, 3$) (и ее первой производной $f'(x)$) с помощью слабосглаживающего ИД-сплайна $\tilde{S}_{2ID}(x)$ (и $\tilde{S}'_{2ID}(x)$) при определении параметров \hat{I}_i^{i+1} ($i = 0, \dots, n-1$) по формулам (12), (13) и нахождении затем значений \tilde{f}_i ($i = 0, \dots, n$) из трехдиагональной СЛАУ (11) с краевыми условиями $\tilde{f}_0 = f_0 = f(x_0), \tilde{f}_n = f_n = f(x_n)$ представлены в табл. 2.

Из оценок, приведенных в табл. 2, следует, что при $f(x) \in C_{[a,b]}^m$ ($m = 1, 2, 3$) ИД-сплайны $\tilde{S}_{2ID}(x)$ равномерно сходятся к функции $f(x)$ на последовательности сеток $\Delta_1^{(n)}$: $a = x_0 < x_1 < \dots < x_i <$

$< x_{i+1} < \dots < x_n = b$ по крайней мере со скоростью H^m с ростом n , а их производные при $f(x) \in C_{[a,b]}^m$ ($m = 2, 3$) равномерно сходятся к $f'(x)$ по крайней мере со скоростью H^{m-1} .

Дальнейшее увеличение степени гладкости функции $f(x)$ не приводит к увеличению порядка приближения относительно H .

Выводы

1. В работе предложены методы аппроксимации функций различных классов гладкости с помощью ИД-многочленов и ИД-сплайнов второй степени, обладающих свойством консервативности.
2. Получены оценки погрешностей аппроксимации функций и их производных параболическими ИД-многочленами, доказана теорема сходимости глобального параболического слабосглаживающего ИД-сплайна дефекта 1.
3. Применение предложенных авторами алгоритмов аппроксимации ИД-сплайнами можно рекомендовать разработчикам больших территориально-распределенных информационных систем при построении математических моделей обработки данных в условиях ограниченных временных и вычислительных ресурсов — реализующих, например, идентификацию информационных объектов, определение степени близости данных, поиск зависимостей [1, 7].

Литература

1. Гершкович М. М., Бирюкова Т. К., Синицын В. И. Проблемы идентификации и распознавания информационных объектов при создании распределенных информационно-телекоммуникационных систем // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и

- символьной информации (Распознавание-2012): Сб. мат-лов X Междунар. науч.-технич. конф. — Курск: ЮЗГУ, 2012. С. 24–26.
2. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. — Новосибирск: Наука, 1980. 350 с.
 3. Завьялов Ю. С., Леус В. А., Скороспелов В. А. Сплайны в инженерной геометрии. — М.: Машиностроение, 1985. 224 с.
 4. Квасов Б. И. Методы изогеометрической аппроксимации сплайнами. — М.: Физматлит, 2006. 360 с.
 5. Стечкин С. Б., Субботин Ю. Н. Сплайны в вычислительной математике. — М.: Наука, 1976. 248 с.
 6. Киреев В. И., Бирюкова Т. К. Интегродифференциальные консервативные сплайны и их применение в интерполяции, численном дифференцировании и интегрировании // Вычислительные технологии, 1995. Т. 4. № 16. С. 233–244.
 7. Бирюкова Т. К., Гершкович М. М., Киреев В. И. Интегро-дифференциальные многочлены и сплайны произвольной четной степени в задачах анализа параметров функционирования распределенных информационных систем // Системы компьютерной математики и их приложения (СКМП-2012): Мат-лы XIII Междунар. науч. конф., посвященной 75-летию профессора Э. И. Зверовича. — Смоленск: СмолГУ, 2012. Вып. 13. С. 67–72.
 8. Киреев В. И., Пантелеев А. В. Численные методы в приборах и задачах. — М.: Высшая школа, 2008. 480 с.
 9. Волков Е. А. Численные методы. — М.: Наука, 1982. 254 с.

Поступила в редакцию 10.12.13

ON APPROXIMATION AND CONVERGENCE OF ONE-DIMENSIONAL PARABOLIC INTEGRODIFFERENTIAL POLYNOMIALS AND SPLINES

V. I. Kireev¹, M. M. Gershkovich², and T. K. Biryukova²

¹Moscow State Mining University, 6 Leninskiy Prosp., Moscow 119991, Russian Federation

²Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The methods for approximation of functions with one-dimensional (1D) integrodifferential polynomials of the 2nd degree and derived conservative parabolic integrodifferential splines are considered. In majority of applied computational tasks, accuracy of source data does not exceed precision of approximation by parabolic polynomials and splines. The nodes of conventional parabolic splines, based on differential matching conditions with approximated function (further named as differential splines), are shifted relatively to interpolation nodes in order to provide stability of approximation process. The shift between spline and approximation nodes complicates computational algorithms drastically. Additionally, traditional differential splines are not conservative, i. e., they do not maintain integral characteristics of approximated functions. The novel integrodifferential parabolic splines that use integral deviation as criteria for matching a spline with a source function are presented. These splines are stable if spline nodes coincide with nodes of approximated functions and conservative with respect to sustaining area under curves. The theorems on approximation of mathematical functions with 1D integrodifferential parabolic polynomials and convergence of parabolic integrodifferential splines are proved. It is suggested to apply the proposed integrodifferential splines for development of mathematical data processing models for large area spread information systems.

Keywords: spline; polynomial; integrodifferential; approximation; interpolation; smoothing; estimation of errors; convergence theorem; mathematical data processing model

DOI: 10.14357/19922264140112

References

1. Gershkovich, M. M., T. K. Biryukova, and V. I. Sinityn. 2012. Problemy identifikatsii i raspoznavaniya informatsionnykh ob"ektov pri sozdanii raspredelennykh informatsionno-telekommunikatsionnykh sistem [Problems of identification and recognition of information's objects in development of information-telecommunication systems]. *Optiko-elektronnye pribory i ustroystva v sistemakh raspoznavaniya obrazov, obrabotki izobrazheniy i simvol'noy informatsii. Raspoznavanie 2012: Sbornik Materialov X Mezhdunarodnoy Nauchno-Tekhnicheskoy Konferentsii*. Kursk: Izd-vo Jugo-Zap. Gos. Un-ta. 24–26.
2. Zav'jalov, Ju. S., B. I. Kvasov, and V. L. Miroshnichenko. 1980. *Metody splayn-funktsiy [Methods of spline functions]*. Novosibirsk: Nauka, 1980. 350 p.

3. Zav'jalov, Ju. S., V. A. Leus, and V. A. Skorospelov. 1985. *Splayny v inzhenernoy geometrii* [Splines in engineering geometry]. Moscow: Mashinostroenie. 224 p.
4. Kvasov, B. I. 2006. *Metody izogeometricheskoy approksimatsii splaynami* [Methods of izogeometric spline approximation]. Moscow: Fizmatlit. 360 p.
5. Stechkin, S. B., and Ju. N. Subbotin. 1976. *Splayny v vychislitel'noy matematike* [Splines in computing mathematics]. Moscow: Nauka. 248 p.
6. Kireev, V. I., and T. K. Biryukova. 1955. Integrodifferentsial'nye konservativnye splayny i ikh primenenie v interpol'yatsii, chislennom differentsirovanii i integrirovanii [Integrodifferential splines and their applications in interpolation, numerical differentiation and quadrature]. *Vychislitel'nye Tekhnologii* 4(16):233–244.
7. Biryukova, T. K., M. M. Gershkovich, and V. I. Kireev. 2012. Integro-differentsial'nye mnogochleny i splayny proizvol'noy chetnoy stepeni v zadachakh analiza parametrov funktsionirovaniya raspredelennykh informatsionnykh sistem [Integrodifferential polynomials and splines of arbitrary even degree in analysis of parameters of functioning of spread information systems]. *Sistemy Komp'yuternoy matematiki i ikh prilozheniya (SKMP-2012): Materialy XIII Mezhdunarodnoy Nauchnoy Konferentsii, posvyashchennoy 75-letiyu Professora E. I. Zverovicha*. Smolensk: Izd-vo SmolGU. 13:67–72.
8. Kireev, V. I., and A. V. Panteleev. 2008. *Chislennyye metody v primerakh i zadachakh* [Numerical methods in examples and problems]. Moscow: Vysshaya shkola. 480 p.
9. Volkov, E. A. 1982. *Chislennyye metody* [Computational methods]. Moscow: Nauka. 254 p.

Received December 10, 2013

Contributors

Kireev Vladimir I. (b. 1938) — Doctor of Science in physics and mathematics, Professor, Moscow State Mining University, 6 Leninskiy Prosp., Moscow 119991, Russian Federation; Vladimir-Kireyev@mail.ru

Gershkovich Maxim M. (b. 1968) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; makmg@mail.ru

Biryukova Tatiana K. (b. 1968) — Candidate of Science (PhD) in physics and mathematics, Institute of Informatics Problems, Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; yukonta@mail.ru

STABILITY ANALYSIS OF AN OPTICAL SYSTEM WITH RANDOM DELAY LINES LENGTHS

E. Morozov^{1,2}, L. Potakhina^{1,2}, and K. De Turck³

Abstract: A new model of an optical buffer system is considered, in which the differences $\{\Delta_n\}$ between the lengths of two adjacent fiber delay lines (FDLs) are *random*. This is an extension of the model considered in [1] where these differences (also referred to as *granularity*) are constant, i. e., $\Delta_n \equiv \text{const}$. The system is modeled by utilizing the random-walk theory and closely-related asymptotic results of the renewal theory, such as the inspection paradox and the Lorden's inequality. A stability analysis is performed based on the regenerative approach. Some numerical results are included as well, showing that the obtained conditions delimit the stability region with high accuracy.

Keywords: optical buffer; stability; stochastic granularity; renewal theory; regeneration; inspection paradox; simulation

DOI: 10.14357/19922264140113

1 Introduction

In networks featuring optical burst switching, data units referred to as bursts travel from source to destination in the form of light (i. e., without conversion to the electrical domain in intermediate switches). As optical random access memory does not as yet exist, a set \mathcal{A} of FDLs is used as a substitute form of buffering. Thus, the set of possible waiting times is denumerable, with each value corresponding to the length of a fixed delay line belonging to \mathcal{A} . As a result, in general even having an infinite set of (different) lines, arriving bursts have to wait for service longer than in the classical case with infinite buffer for awaiting customers. The stability of such a general optical system with a denumerable set \mathcal{A} with deterministic lengths has been analyzed in recent papers [2, 3]. The main idea of the present paper is to consider a denumerable set of FDLs with *random lengths*.

Motivations for this extension are manifold. For heavily-loaded modern large networks, a large number of lines is required. These lines constitute a huge number of possible paths between hosts and users. Moreover, it seems appropriate to describe the differences between their lengths as random variables to reflect variability of the paths. On the other hand, to meet the high quality of service (QoS) requirement and avoid huge difference in transmission times, it seems reasonable to assume that the difference $\Delta > 0$ between the *adjacent increasing* FDL lengths has the same distribution regardless of the lines, reflecting *homogeneity* of the network. In particular, homogeneity may be important for reduction of

reordering in the multipath transmission of a big file by means of a huge number of small packets.

One more argument to support a common distribution of Δ is that the modern networks are very well-connected and contain a huge number of links, such that each path is collected from a number of *elementary (primary)* optical cables with *comparable* lengths. Thus, randomness of Δ reflects variability of FDL lengths, while the common distribution of Δ (homogeneity) reflects the mentioned comparability. Such a setting seems to be natural for a large fiber lines set (infinite in theory) in a local network, where knowledge of the exact FDL lengths is unavailable (or costly) for the some reason. In this case, it is assumed that the most important predetermined information is contained in the distribution of Δ , while other details/info related to the set \mathcal{A} are hardly available or less important.

The final argument for the considered model is that sometimes, it might be easier to assume some randomness in the differences Δ rather than to calculate the exact differences. Moreover, to meet the stability condition, in fact, only knowledge of the first two moments is required (not of the entire distribution); hence, simple sample estimators can be used.

As it will be shown below, randomness of Δ implies considerable changes in the technique of analysis in comparison with the model with the deterministic FDL set \mathcal{A} . As the difference $\Delta = \text{const}$ considered in [1] is called *granularity*, it is natural to refer to the model considered in this paper as the model with *stochastic granularity*.

Further, the term *stochastic inhomogeneity* can be

¹Institute of Applied Mathematical Research, Karelian Research Center, Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Russian Federation

²Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Russian Federation

³Ghent University, TELIN Department, 41 Sint-Pietersnieuwstraat, Gent B-9000, Belgium

used to name a system in which the distribution of the difference Δ depends on the lines. This case will not be further discussed; let only note that inhomogeneity may lead to extra loss of capacity caused by an increase of variance of the void (the idle time following completion of a transmission).

Note that the assumption $\Delta > 0$ reflecting the corresponding ordering of the set of FDL lengths \mathcal{A} seems less motivated for large distributed networks, where the structure of currently available different routes is not as well defined. Nevertheless, the authors believe that such assumption and, hence, the approach developed is suitable for some distributed optical networks. Besides, it may be useful not only for optical systems but for some other bufferless communication networks as well.

In the authors' opinion, this model has intrinsic interest from the mathematical point of view as well.

The structure of the paper is as follows. In section 2, a description of the optical system with random delay lines is given. Then, in Section 3, stability of this system is analyzed. In particular, sufficient stability conditions of such a system are presented using regenerative arguments. The key ingredient of the analysis is application of the so-called inspection paradox from the renewal theory. In section 4, simulation results are presented which verify that the stability conditions found are tight and close to being also necessary.

2 Description of the Model

Based on the assumptions detailed above, define the length of the n th line as $S_n := \Delta_1 + \dots + \Delta_n$, $S_0 := 0$, with independent and identically distributed (i.i.d.) random variables $\{\Delta_n, n \geq 1\}$ distributed as a generic variable Δ . Assume in the following that Δ is nonlattice. (In a more general case, it could be assumed that $S_n = a_0 + \Delta_1 + \dots + \Delta_n$ where a constant (or, possibly, even a random variable (r.v.)) $a_0 > 0$ is a minimal given delay. However, further stability analysis is independent of this assumption.)

Thus, for the n th arriving burst whose waiting time is $W_n = x$, generate a path until the random walk $\{S_i\}$ exceeds the level x , $n \geq 0$. As in previous works [2, 3], this rule is motivated by the requirement to keep the FIFO (first in, first out) discipline. Thus, the following set $\mathcal{A} = \{S_0, S_1, S_2, \dots\}$, in which the origin 0 covers the case of a nonwaiting burst is obtained. Define the number of lines required to guarantee a delay exceeding the threshold $x \geq 0$ as

$$N(x) = \inf\{n \geq 0 : S_n \geq x\}.$$

(Obviously, $\{N(x)\}$ is the renewal process generated by the random walk $\{S_n\}$.) Then, the actual (random) delay of the burst, which meets the workload x , is defined

as

$$S_{N(x)} = \inf\{S_n : S_n - x \geq 0\}.$$

Note that $S_{N(0)} = N(0) = S_0 = 0$. For each $x \geq 0$, the overshoot of the renewal process is defined as

$$S_{N(x)} - x = \Delta(x). \tag{1}$$

By definition, $\Delta(x) \geq 0$ for all $x \geq 0$ and $\Delta(S_k) = 0$ for any $k \geq 0$. In what follows, it is assumed that

$$0 < E\Delta^{2+\varepsilon_0} < \infty \tag{2}$$

for some $\varepsilon_0 > 0$. This assumption is necessary to apply the asymptotic results from the renewal theory detailed below. In particular, (2) implies that for any δ_0 , there are (fixed) constants $D < \infty$, $\delta^* > 0$, such that

$$P(\delta^* < \Delta \leq D) \geq 1 - \delta_0. \tag{3}$$

(Obviously, having $E\Delta < \infty$ is enough to ensure (3).) Let introduce the following quantities, which are crucial for the further analysis:

$$\Delta^* := \frac{E\Delta^2}{E\Delta}; \quad \Delta_0 := \frac{E\Delta^2}{2E\Delta}. \tag{4}$$

Note that by the Lorden's inequality [4] the first term in (4) is the upper bound of the expected overshoot $\Delta(x)$ over all $x \geq 0$, while the second term is the mean stationary overshoot in the random walk $\{S_n\}$ expressing the so-called *inspection paradox*. Further, assume that the input process is renewal with independent and identically distributed (i.i.d.) interarrival times $\{T_k\}$ with a generic time T . Also assume that service (transmission) times $\{B_n\}$ are i.i.d. with the generic element B . Denote $U := B - T$ and introduce the distribution $F_U(x) = P(U \leq x)$, $x \in (-\infty, +\infty)$. Obviously, by the independence between B and T , one can express this distribution in the terms of given distributions of B and T as follows:

$$F_U(x) = \int_{0^-}^{\infty} F_B(x+y)F_T(dy).$$

Let W_k be the remaining workload which the burst k meets upon arrival. In other words, it is the time which is required to transmit the work accumulated in the systems (in the optical buffer and server) prior to arrival of the burst. Define the workload process $W = \{W_k, k \geq 0\}$ and construct regeneration instants $\{\beta_n\}$ for the process W in the following (conventional) way: let $\beta_0 = 0$ and

$$\beta_{n+1} = \inf\{k > \beta_n : W_k = 0\}, \quad n \geq 0 \quad (\inf \emptyset := \infty).$$

In what follows, the focus will be on the zero-delayed regenerative process in which case 0 is the regeneration point, and $\beta := \beta_1$ is the generic regeneration period. Note that the sequence $\{\beta_n\}$ is well-defined since $0 \in \mathcal{A}$.

3 Main Stability Result

The main purpose is to establish conditions ensuring the *positive recurrence* of the workload regenerative process meaning that

$$E\beta < \infty. \quad (5)$$

The importance of positive recurrence for the stability analysis is well-known (see, for instance, [4, 5]). Let use the following characterization of the recurrence property of the renewal process β via the limiting behavior of the forward regeneration time at the instant n , which is defined as

$$\beta(n) = \inf_k (\beta_k - n : \beta_k - n > 0), \quad n \geq 1.$$

It is known [6] that

$$E\beta = \infty \text{ if and only if } \beta(n) \rightarrow \infty \text{ in probability as } n \rightarrow \infty. \quad (6)$$

Thus, to establish (5), it suffices to show that $\beta(n) \not\rightarrow \infty$ (in probability), which is exactly the approach used below. Now, the main stability result can be formulated.

Theorem 1. *Assume that Δ is nonlattice, $EB < \infty$, $ET < \infty$ and that condition (2) holds. Furthermore, assume that the following negative drift assumption*

$$EU + \Delta_0 < 0 \quad (7)$$

and the regeneration assumption

$$P(T > \Delta + B) = \delta_1 > 0 \quad (8)$$

hold. Then $E\beta < \infty$.

Proof. Put $N(x) := 0$ for $x \leq 0$ and define the increments of the workload process as

$$\Delta_W(k) = S_{N((W_k+U_k))} - W_k, \quad k \geq 0.$$

Note that $S_{N((W_k+U_k))}$ is the fiber line length which is selected for the $(k+1)$ th burst meeting the workload $(W_k+B_k-T_k)^+ := (W_k+U_k)^+$, the time when the k th burst brings the transmission (service) equals B_k and the interarrival time between the bursts k and $k+1$ (where the workload decreases) equals T_k . Thus, $\Delta_W(k)$ is the overshoot of the random walk (forming line lengths) across the threshold W_k+U_k . Then as in [2], using independence between U_k and W_k , one can write:

$$E\Delta_W(k) = \int_{y \in R_+} E(S_{N(y+U)} - y)P(W_k \in dy), \quad k \geq 1.$$

If $U \leq -y$, then $E(S_{N(y+U)} - y) = -y$. Thus, as in [2], one has for each fixed y :

$$E(S_{N(y+U)} - y) = -yP(U \leq -y) + \int_{z \geq -y} E\Delta(y+z)F_U(dz) + \int_{z \geq -y} zF_U(dz). \quad (9)$$

Let consider the three terms in the right-hand side of Eq. (9) one by one. Since $EU > -\infty$, one has

$$0 > -yP(U \leq -y) \geq \int_{-\infty}^{-y} xF_U(dx) \uparrow 0 \text{ as } y \rightarrow \infty. \quad (10)$$

For the third term, it holds that (since $y \geq 0$ and $EU < \infty$)

$$\int_{z \geq -y} zF_U(dz) \downarrow EU \text{ as } y \rightarrow \infty. \quad (11)$$

Finally, for the second term, one has

$$\begin{aligned} & \int_{-y}^{\infty} E\Delta(y+z)F_U(dz) \\ &= \int_{-y}^{-y/2} E\Delta(y+z)F_U(dz) + \int_{-y/2}^{\infty} E\Delta(y+z)F_U(dz) \\ &\leq \sup_{0 \leq x \leq y/2} E\Delta(x)P\left(-y < U \leq -\frac{y}{2}\right) \\ &+ \sup_{x \geq y/2} E\Delta(x)P\left(U > -\frac{y}{2}\right) \leq \Delta^*P\left(U \leq -\frac{y}{2}\right) \\ &+ \sup_{x \geq y/2} E\Delta(x)P\left(U > -\frac{y}{2}\right). \end{aligned}$$

Fix now an arbitrary $\varepsilon > 0$ and choose such y_0 that for $y \geq y_0$,

$$P\left(U \leq -\frac{y}{2}\right) \leq \frac{\varepsilon}{4\Delta^*}; \quad (12)$$

$$\int_{z \geq -y} zF_U(dz) \leq EU + \frac{\varepsilon}{4}. \quad (13)$$

Such a choice of y_0 is possible: indeed, for (12), due to (10); for (13), due to (11). As Δ is nonlattice and (2) holds, then $\Delta(x) \Rightarrow \Delta(\infty)$ in distribution, where $E\Delta(\infty) = \Delta_0$. Similarly, the renewal interval $L(x)$ covering the point x converges in distribution, as $x \rightarrow \infty$, to a variable L with the mean $EL = \Delta^*$ [6]. By the modified form of the Lorden's inequality (see, for instance, [7]),

$$\sup_x E[\Delta(x)]^{1+\varepsilon_0} \leq (3 + \varepsilon_0)EL^{1+\varepsilon_0}.$$

It is well-known [6] that the density of L is

$$P(L \in dx) = \frac{1}{E\Delta} xF_{\Delta}(dx)$$

where F_Δ is the distribution of Δ and, hence, $EL^{1+\varepsilon_0} < \infty$ if and only if condition (2) holds. Thus, $\sup_x E[\Delta(x)]^{1+\varepsilon_0} < \infty$, that is, the family $\{\Delta(x), x \geq 0\}$ is uniformly integrable [8]. It then follows from the convergence $\Delta(x) \Rightarrow \Delta(\infty)$ that $E\Delta(x) \rightarrow E\Delta(\infty) = \Delta_0$, see (4). Thus, one can take y_0 in (12) and (13) in such a way that simultaneously,

$$\sup_{x \geq y/2} E\Delta(x) \leq \Delta_0 + \frac{\varepsilon}{4}, \quad y \geq y_0.$$

As ε above is arbitrary, define it now as $\Delta_0 + EU = -2\varepsilon$ (see (7)). Then, picking together the bounds obtained above, one has

$$\begin{aligned} E(\Delta_W(k)|W_k = y) &= E(S_{N(y+U)} - y) \\ &\leq \frac{\varepsilon}{4} + \Delta_0 + \frac{\varepsilon}{4} + EU + \frac{\varepsilon}{4} < -\varepsilon, \quad y \geq y_0. \end{aligned} \quad (14)$$

By the standard Lorden's inequality, $\sup_x E\Delta(x) \leq \Delta^*$ [4] and it then follows that

$$\begin{aligned} E(S_{N(y+B)} - y) &= E(S_{N(y+B)} - y - B) + EB = E[\Delta(y+B)] + EB \\ &\leq \Delta^* + EB \leq \max(\varepsilon, \Delta^* + EB) := C < \infty. \end{aligned}$$

(It is necessary to ensure $\varepsilon/C \leq 1$ in (18) below.) In particular,

$$\begin{aligned} E(\Delta_W(k) | W_k \leq y_0) &= \int_0^{y_0} E(S_{N(y+U)} - y)P(W_k \in dy) \leq C. \end{aligned} \quad (15)$$

Now, for any k , use (14) and (15) to obtain

$$E\Delta_W(k) \leq CP(W_k \leq y_0) - \varepsilon P(W_k > y_0). \quad (16)$$

Now it will be proved, by contradiction, that $W_k \not\rightarrow \infty$ in probability as $k \rightarrow \infty$. Thus, it is assumed that

$$W_k \rightarrow \infty \text{ in probability as } k \rightarrow \infty. \quad (17)$$

Therefore, one can find a value k_0 such that

$$P(W_k \leq y_0) \leq \frac{\varepsilon}{4C}, \quad k \geq k_0,$$

which allows to rewrite (16) as

$$E\Delta_W(k) \leq \frac{\varepsilon}{4} - \varepsilon \left(1 - \frac{\varepsilon}{4C}\right) \leq -\frac{\varepsilon}{2}, \quad k \geq k_0. \quad (18)$$

Using the Lorden's inequality, one obtains (see (1))

$$\begin{aligned} EW_1 &= ES_{N(U_0)} \leq ES_{N(B_0)} = \int_0^\infty ES_{N(z)}F_B(dz) \\ &= \int_0^\infty E\Delta(z)F_B(dz) + \int_0^\infty zB(dz) \leq \Delta^* + EB := C^* \end{aligned}$$

where $F_B(x) := P(B \leq x)$. Continuing this analysis iteratively, one easily obtains $EW_{k_0} < k_0C^* < \infty$ that contradicts (17). Hence, assumption (17) is false, and there exist a nonrandom (sub-)sequence $z_k \rightarrow \infty$ and constants $\delta^* > 0$ and $R < \infty$ such that

$$\inf_k P(W_{z_k} \leq R) \geq \delta^*. \quad (19)$$

Let now proceed by showing that the probability of reaching an empty system is strictly positive. Let discern two cases pertaining to the r.v. T . In the first case, let assume that T is unbounded, while the second case considers bounded T .

1. If it is assumed that the r.v. T is unbounded, then $P(T > x) > 0$ for any $x \geq 0$. It is easy to see that in the event $\{W_{z_k} \leq R, T_{z_k} > R + B_{z_k} + \Delta(W_{z_k} + B)\}$ the burst $z_k + 1$ meets an empty system, and, hence, regeneration occurs. In other words, the event $\{\beta(z_k) = 1\}$ occurs. Now, it will be shown that for a given z_k belonging to the sequence $\{z_k\}$ which satisfies (19), the probability of this event is positive. Indeed,

$$\begin{aligned} P(\beta(z_k) = 1) &\geq P(\{W_{z_k} \leq R, T_{z_k} > R + B_{z_k} + \Delta(W_{z_k} + B)\}) \\ &= \int_{x=0}^R \int_{y=0}^\infty P(W_{z_k} \in dx, B_{z_k} \in dy, \\ &\quad T_{z_k} > R + y + \Delta(x + y)) \\ &= \int_{x=0}^R \int_{y=0}^\infty P(T_{z_k} > R + y + \Delta(x + y))P(W_{z_k} \in dx) \\ &\quad \times P(B_{z_k} \in dy) \\ &\geq \int_{x=0}^R \int_{y=0}^b P(T_{z_k} > R + y + \Delta(x + y)) \\ &\quad \times P(W_{z_k} \in dx)P(B_{z_k} \in dy) \\ &\geq \inf_{x+y \leq R+b} P(T_{z_k} > R + b + \Delta(x + y))\delta^*P(B \leq b) \end{aligned}$$

where inequality (19) and independence between variables W_{z_k}, B_{z_k} , and T_{z_k} are used, and a constant $b > 0$ is such that $P(B \leq b) > 0$. Now, note that

$$\Delta(x + y) \leq R + b + \Delta(R + b), \quad x + y \leq R + b. \quad (20)$$

Moreover, because $E\Delta(R + b) < \infty$, then constants $\varepsilon_0 > 0$ and $D < \infty$ exist such that

$$P(\Delta(R + b) \leq D) \geq 1 - \varepsilon_0. \quad (21)$$

Then, using (20) and (21), one has

$$\begin{aligned} & \inf_{x+y \leq R+b} \mathbb{P}(T_{z_k} > R+b+\Delta(x+y)) \\ & \geq \mathbb{P}(T > 2(R+b)+D, \Delta(R+b) \leq D) \\ & \geq \mathbb{P}(T > 2(R+b)+D)(1-\varepsilon_0) := \hat{\delta} > 0. \end{aligned}$$

Now, one obtains the following lower bound

$$\mathbb{P}(\beta(z_k) = 1) \geq \hat{\delta} \delta^* \mathbb{P}(B \leq b) > 0$$

which is uniform in z_k .

2. Assume now that T is bounded; then, it is possible that

$$\mathbb{P}(T > 2(R+b)+\Delta(R+b)) = 0.$$

In the analysis below, the first moment will be found when the random walk S_n reaches the zero state, starting into the compact set $[0, R+D]$. More exactly, each next burst finds a shorter FDL upon arrival until a nonwaiting burst arrives. Consider the independent events

$$\begin{aligned} \mathcal{D}(z_k+i) & := \{T_{z_k+i} > \Delta_{z_k+i} + B_{z_k+i}, \delta^* \leq \Delta_{z_k} \leq D\} \\ & = \{B_{z_k+i} - T_{z_k+i} < -\Delta, \delta^* \leq \Delta_{z_k} \leq D\}, i \geq 0, \end{aligned}$$

where, by (3) and (8), constants $D < \infty$ and $\delta^* > 0$ are chosen in such a way that (regardless of z_k and i)

$$\mathbb{P}(T_{z_k+i} > B_{z_k+i} + D, \Delta_{z_k+i} \in (\delta^*, D]) \geq \frac{\delta_1}{2}. \quad (22)$$

It is clear that on each event $\mathcal{D}(z_k+i)$, the random walk S_n decreases by at least Δ and thus, each new burst occupies a shorter FDL. Note that $W_{z_k} \leq S_{N(R)}$ in the event $\mathcal{E}(z_k) = \{W_{z_k} \leq R\}$ and that in the event $\mathcal{E}(z_k) \cap \mathcal{D}(z_k)$,

$$\begin{aligned} W_{z_k+1} & = S_{N((W_{z_k}+B_{z_k}-T_{z_k}))} \leq S_{N(R)} - \Delta \\ & = S_{(N(R)-1)+} \end{aligned}$$

where $\mathbb{P}(\mathcal{E}(z_k) \cap \mathcal{D}(z_k)) \geq \delta^* \delta_1 / 2$ in view of (19) and (22). Continuing in such a way, one finds that in the event

$$\mathcal{B}(z_k) := \mathcal{E}(z_k) \cap \bigcap_{i=0}^{N(R)} \mathcal{D}(z_k+i),$$

a burst arrives, among the bursts with the numbers $\{z_k, \dots, z_k+N(R)\}$, which finds the transmission channel free, and thus regeneration occurs. Moreover, it is easy to see that in the event $\mathcal{B}(z_k)$,

$$N(R) \leq \left\lceil \frac{R+D}{\delta^*} \right\rceil := C_0 < \infty.$$

Now, it follows that

$$\mathbb{P}(\beta(z_k) \leq C_0) \geq \mathbb{P}(\mathcal{B}(z_k)) \geq \delta^* \left[\frac{\delta}{2} \right]^{C_0} > 0.$$

As the lower bound holds for all z_k satisfying (19) and the sequence $\{z_k\}$ is deterministic, then $\beta(n) \not\rightarrow \infty$, and it follows from (6) that $\mathbb{E}\beta < \infty$. ■

4 Verification of the Stability Region by Simulation

In this section, the simulation results, which illustrate the stability/instability domain of the described optical system and the accuracy of the stability conditions found above, are presented. The simulation has been carried out by means of the “*R*” *toolbox*: the language and the environment for statistical computing [9]. The conditions of Theorem 1 imply the positive recurrence of the workload process W .

First, let check the main negative drift condition (7) which includes the stationary overshoot Δ_0 of the FDL length. As simulations below suggest, condition (7) is, in fact, a *stability criterion*, meaning that it appears to sharply delineate the regions of stability and instability. On the other hand, a less tight condition has been suggested in previous related works [2, 3]. This condition, in adaptation to the model described, can be written as

$$\mathbb{E}U + \Delta^* < 0 \quad (23)$$

where Δ^* from (4) is used. Indeed, condition (23) is suggested in [2, 3] for the model with the different *deterministic* distances $\{\Delta_n\}$, and in this case, the appearance of the bound $\Delta^* \geq \sup_x \mathbb{E}\Delta(x)$ is motivated by the Lorden’s inequality. However, simulations of the model with deterministic granularity $\Delta = \text{const}$ have shown that condition (23) is indeed overly restrictive. The inspection paradox presented above has revealed the source of this redundancy and has allowed to replace Δ^* by the smaller quantity Δ_0 in (7). It explains why below, the accuracy of both conditions (7) and (23) is verified. In particular, simulations confirm that in the intermediate case when condition (23) is violated but condition (7) holds, the system remains stable. Visually, absence of a clear tendency in the workload process indicates that the system is stable, while instability is characterized by an increase of W_n as $n \rightarrow \infty$. These evident observations are used below to determine stability/instability of the system.

In the numerical experiments, 5000 bursts have been simulated following a Poisson input with rate λ and with the i.i.d. exponential service times with rate μ . Moreover, the following distributions of the difference Δ have been considered: (i) exponential; (ii) uniform; and (iii) Pare-

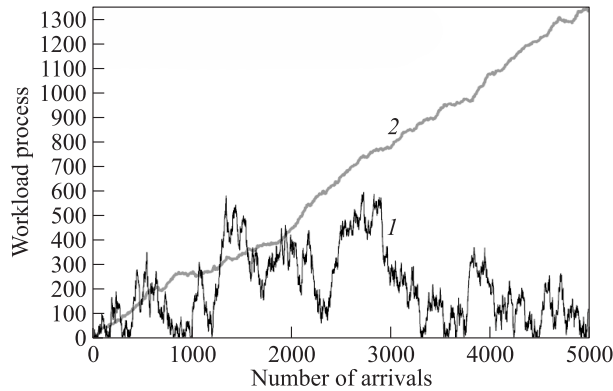


Figure 1 The system with exponential $\Delta \sim \exp(3)$, $\Delta_0 = 0.333$, $\Delta^* = 0.667$, and $\rho = 0.94$: 1 – $EU + \Delta_0 = -0.268 < 0$, $\lambda = 0.1$, and $\mu = 0.1064$; and 2 – $EU + \Delta_0 = 0.273 > 0$, $\lambda = 1$, and $\mu = 1.0638$

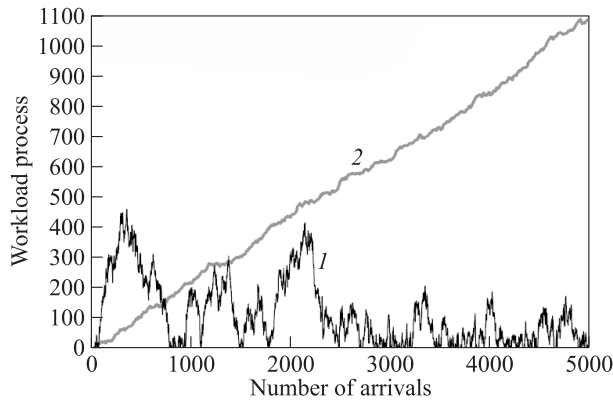


Figure 2 The system with uniform $\Delta \in [0, 0.8]$, $\Delta_0 = 0.267$, $\Delta^* = 0.533$, and $\rho = 0.94$: 1 – $EU + \Delta_0 = -0.230 < 0$, $\lambda = 0.12$, and $\mu = 0.1276$; and 2 – $EU + \Delta_0 = 0.217 > 0$, $\lambda = 1.2$, and $\mu = 1.2766$

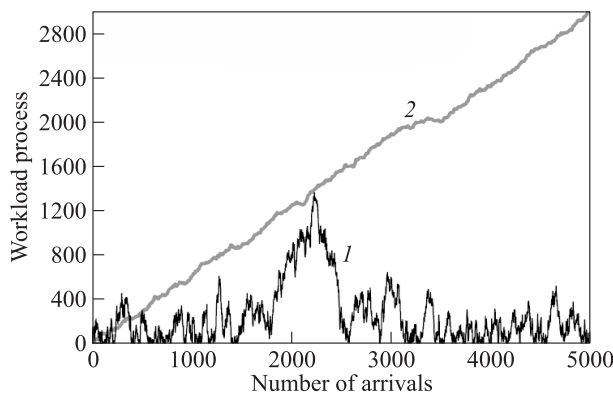


Figure 3 The system with Δ that has Pareto distribution with parameter $\alpha = 4$, $\Delta_0 = 0.75$, $\Delta^* = 1.5$, and $\rho = 0.94$: 1 – $EU + \Delta_0 = -0.527 < 0$, $\lambda = 0.047$, and $\mu = 0.05$; and 2 – $EU + \Delta_0 = 0.622 > 0$, $\lambda = 0.47$, and $\mu = 0.5$

to, $F_\Delta(x) = 1 - x^{-\alpha}$, $x \geq 1$ ($F_\Delta(x) = 0$, $x < 1$). The simulations have shown similar results for a range of traffic intensity $\rho := \lambda/\mu$, but below, the present results for the most illustrative value $\rho = 0.94$ are presented, in which case, an increase of the workload process is more easy to observe if condition (7) is violated. (For smaller values of ρ , this tendency requires much more time to be detected.)

Figure 1 describes the system with exponential Δ with parameter 3. As one can see, satisfying condition (7) implies stability. At the same time, in this case, $EU + \Delta^* = 0.065$, and a definite decision about stability could not be made using condition (23). Otherwise, if condition (7) is violated, the system is unstable. In this case, the value $EU + \Delta^* = 0.607$ also indicates instability.

Figure 2 describes the system with Δ uniformly distributed in the interval $[0, 0.8]$. Satisfying condition (7) implies stability; however, in this case, $EU + \Delta^* = 0.037$. Otherwise, if condition (7) is violated, visual observation confirms instability and it agrees with the positivity of the sum $EU + \Delta^* = 0.483$.

Finally, Fig. 3 corresponds to the system where Δ has Pareto distribution with parameter $\alpha = 4$. One can see that satisfying condition (7) implies stability, while in this case, $EU + \Delta^* = 0.223$. Otherwise, if condition (7) is violated, the system is unstable, and again, it is consistent with the value $EU + \Delta^* = 1.372$.

Thus, in all considered cases the simulations confirm that condition (7) allows to delimit the stability/instability region with high accuracy and that condition (7) is indeed overly strict.

5 Concluding Remarks

In this work, a new model of an optical buffer system is suggested. The main assumption is that the distances $\{\Delta_n\}$ between the lengths of the adjacent increasing FDLs are random and have the same distribution for all lines. This is a natural extension of the model with the *deterministic granularity* $\Delta_n \equiv const$ considered in [1]. For this reason, the new model considered above can be referred to as an optical system with *stochastic granularity*. The key new ingredient of the stability analysis is the renewal theory, including such asymptotic results as the *inspection paradox* and the *Lorden's inequality*. The regenerative approach is applied to obtain sufficient stability conditions of the system.

As simulations show, the stability condition found allows to delimit the stability domain with high accuracy suggesting that it is in fact the stability criterion of the model.

As an extension of the model, one may consider the case when distributions F_n of $\{\Delta_n\}$ are different for $n = 1, \dots, n_0$, being the same, $F_n \equiv F$, for $n > n_0$ with a deterministic n_0 . Another modification may assume that $F_n \rightarrow F$ (weakly) as $n \rightarrow \infty$. The authors believe that the analysis developed above can be applied in both cases. Finally, if we were to weaken the requirement $\Delta > 0$, which may be problematic for distributed optical networks, allowing Δ to have arbitrary sign, when the fundamental rule FIFO is violated. An analysis of such an extension is probably possible within the framework of the extended renewal theory (where the renewal “interval” can be negative); however, this analysis is beyond the scope this paper.

Acknowledgments

The work of the first and second authors is partially supported by the Program of Strategy development of the Petrozavodsk State University within the framework of the research activity. The third author is a post-doctoral fellow with the Research Foundation-Flanders (FWO-Vlaanderen).

Contributors

Morozov Evsei V. (b. 1947) — Doctor of Science in physics and mathematics, professor, leading scientist, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; professor, Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; emorozov@karelia.ru

Potakhina Lyubov V. (b. 1989) — PhD student, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences, 11 Pushkinskaya Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; engineer, Petrozavodsk State University, 33 Lenin Str., Petrozavodsk 185910, Republic of Karelia, Russian Federation; lpotakhina@gmail.com

Koen De Turck (b. 1981) — postdoctoral fellow with the Research Foundation Flanders (FWO-Vlaanderen), SMACS Research Group, Department of Telecommunications and Information Processing (TELIN), Ghent University, 41 Sint-Pietersnieuwstraat, Gent B-9000, Belgium; kdeturck@telin.ugent.be

References

1. Callegati, F. 2000. Optical buffers for variable length packets. *IEEE Comm. Lett.* 4(9):292–294.
2. Rogiest, W., E. Morozov, D. Fiems, K. Laevens, and H. Bruneel. 2010. Stability of single-wavelength optical buffers. *Eur. Trans. Telecomm.* 21(3):202–212.
3. Morozov, E., W. Rogiest, K. De Turck, and D. Fiems. 2012. Stability of multiwavelength optical buffers with delay-oriented scheduling. *Trans. Emerging Telecomm. Technol.* 23(3):217–226.
4. Asmussen, S. 2003. *Applied probability and queues*. 2nd ed. NY: Springer-Verlag. 440 p.
5. Morozov, E., and R. Delgado. 2009. Stability analysis of regenerative queueing systems. *Automation Remote Control* 70(12):1977–1991.
6. Feller, W. 1971. *An introduction to probability theory and its applications*. Vol. II. New York: John Wiley & Sons. 704 p.
7. Chang, J. 1994. Inequalities for the overshoot. *Ann. Appl. Probab.* 4(4):1223–1233.
8. Billingsley, P. 1968. *Convergence of probability measures*. Wiley. 296 p.
9. *R Foundation for Statistical Computing*. Vienna, Austria. <http://www.R-project.org/>.

Received November 8, 2013

АНАЛИЗ УСТОЙЧИВОСТИ СИСТЕМЫ ПЕРЕДАЧИ ДАННЫХ С ОПТИЧЕСКИМИ ЛИНИЯМИ ЗАДЕРЖКИ СЛУЧАЙНОЙ ДЛИНЫ

Е. В. Морозов¹, Л. В. Потахина², К. Де Турк³

¹Институт прикладных математических исследований КарНЦ РАН, Россия, Республика Карелия, г. Петрозаводск 185910, ул. Пушкинская 11; Петрозаводский государственный университет, Россия, Республика Карелия, г. Петрозаводск 185910, пр. Ленина 33; emorozov@karelia.ru

²Институт прикладных математических исследований КарНЦ РАН, Россия, Республика Карелия, г. Петрозаводск 185910, ул. Пушкинская 11; Петрозаводский государственный университет, Россия, Республика Карелия, г. Петрозаводск 185910, пр. Ленина 33; lpotahina@gmail.com

³Университет Гента, Sint-Pietersnieuwstraat 41, Гент В-9000, Бельгия; kdeturck@telin.ugent.be

Аннотация: Рассмотрена новая модель оптической системы, в которой разности упорядоченных (по возрастанию) длин двух соседних оптических кабелей являются независимыми одинаково распределенными случайными величинами. Это является обобщением рассмотренной ранее модели, где эти разности были детерминированными величинами. Эта система моделируется с использованием теории случайного блуждания и тесно связанных с ней асимптотических результатов теории восстановления, таких как парадокс инспекции и неравенство Лордена. Развита теория устойчивости, основанная на регенеративном подходе. Включены также некоторые численные результаты, которые показывают, что полученные условия позволяют определить область устойчивости с высокой точностью.

Ключевые слова: оптический буфер; устойчивость; теория восстановления; регенерация; парадокс инспекции; неравенство Лордена; имитационное моделирование

DOI: 10.14357/19922264140113

Работа первых двух авторов частично поддержана Программой стратегического развития Петрозаводского государственного университета. Работа третьего автора поддержана Research Foundation-Flanders (FWO-Vlaanderen).

Литература

1. Callegati F. Optical buffers for variable length packets // IEEE Comm. Lett., 2000. Vol. 4. No. 9. P. 292–294.
2. Rogiest W., Morozov E., Fiems D., Laevens K., Bruneel H. Stability of single-wavelength optical buffers // Eur. Trans. Telecomm., 2010. Vol. 21. No. 3. P. 202–212.
3. Morozov E., Rogiest W., De Turck K., Fiems D. Stability of multiwavelength optical buffers with delay-oriented scheduling // Trans. Emerging Telecomm. Technol., 2012. Vol. 23. No. 3. P. 217–226.
4. Asmussen S. Applied probability and queues. — 2nd ed. — NY: Springer-Verlag, 2003. 440 p.
5. Morozov E., Delgado R. Stability analysis of regenerative queueing systems // Automation Remote Control, 2009. Vol. 70. No. 12. P. 1977–1991.
6. Feller W. An introduction to probability theory and its applications. Vol. II. — 1971. New York: John Wiley & Sons. 704 p.
7. Chang J. Inequalities for the overshoot // Ann. Appl. Probab., 1994. Vol. 4. No. 4. P. 1223–1233.
8. Billingsley P. Convergence of probability measures. — Wiley, 1968. 296 p.
9. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org/>.

Поступила в редакцию 08.11.2013

Архипов Олег Петрович (р. 1948) — кандидат технических наук, директор Орловского филиала Института проблем информатики Российской академии наук

Бирюкова Татьяна Константиновна (р. 1968) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Бобков Сергей Геннадьевич (р. 1955) — доктор технических наук, заведующий отделением Научно-исследовательского института системных исследований Российской академии наук

Васильев Николай Семенович (р. 1952) — доктор физико-математических наук, профессор, МГТУ им. Н. Э. Баумана

Гершкович Максим Михайлович (р. 1968) — старший научный сотрудник Института проблем информатики Российской академии наук

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник Института проблем информатики Российской академии наук

Ерошенко Александр Андреевич (р. 1989) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Института проблем информатики Российской академии наук

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой Вологодского государственного университета; старший научный сотрудник Института проблем информатики Российской академии наук; главный научный сотрудник ИСЭРТ Российской академии наук

Зыкин Сергей Владимирович (р. 1959) — доктор технических наук, профессор, заведующий лабора-

торией Института математики им. С. Л. Соболева Сибирского отделения Российской академии наук, Новосибирск

Киреев Владимир Иванович (р. 1938) — доктор физико-математических наук, профессор Московского государственного горного университета

Козеренко Елена Борисовна (р. 1959) — кандидат филологических наук, заведующая лабораторией Института проблем информатики Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета; ведущий научный сотрудник Института проблем информатики Российской академии наук

Коротышева Анна Владимировна (р. 1988) — старший преподаватель Вологодского государственного университета

Кун Де Турк (р. 1981) — научный сотрудник исследовательской группы SMACS факультета телекоммуникаций и обработки информации Университета Гента, Бельгия

Лупенцов Олег Сергеевич (р. 1986) — аспирант Омского государственного института сервиса

Лучко Олег Николаевич (р. 1961) — кандидат педагогических наук, профессор, заведующий кафедрой Омского государственного института сервиса

Малашенко Юрий Евгеньевич (р. 1946) — доктор физико-математических наук, заведующий сектором Вычислительного центра им. А. А. Дородницына Российской академии наук

Маньяков Юрий Анатольевич (р. 1984) — кандидат технических наук, научный сотрудник Орловского филиала Института проблем информатики Российской академии наук

Маренко Валентина Афанасьевна (р. 1951) — кандидат технических наук, доцент, старший научный

сотрудник Института математики им. С. Л. Соболева Сибирского отделения Российской академии наук

Морозов Евсей Викторович (р. 1947) — доктор физико-математических, профессор, ведущий научный сотрудник Института прикладных математических исследований Карельского научного центра Российской академии наук; профессор Петрозаводского государственного университета, Петрозаводск

Назарова Ирина Александровна (р. 1966) — кандидат физико-математических наук, научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Павлов Игорь Валерианович (р. 1945) — доктор физико-математических наук, профессор МГТУ им. Н. Э. Баумана

Потахина Любовь Викторовна (р. 1989) — аспирантка Института прикладных математических исследований Карельского научного центра Российской академии наук; инженер Петрозаводского государственного университета, Петрозаводск

Рождественский Юрий Владимирович (р. 1952) — кандидат технических наук, заведующий сектором Института проблем информатики Российской академии наук

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель

науки РФ, заведующий отделом Института проблем информатики Российской академии наук

Сиротинин Денис Олегович (р. 1984) — кандидат технических наук, научный сотрудник Орловского филиала Института проблем информатики Российской академии наук

Соколов Игорь Анатольевич (р. 1954) — академик (действительный член) Российской академии наук, доктор технических наук, директор Института проблем информатики Российской академии наук

Степченко Юрий Афанасьевич (р. 1951) — кандидат технических наук, заведующий отделом Института проблем информатики Российской академии наук

Сурков Алексей Викторович (р. 1978) — старший научный сотрудник Научно-исследовательского института системных исследований Российской академии наук

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Российской академии наук

Шоргин Сергей Яковлевич (р. 1952.) — доктор физико-математических наук, профессор, заместитель директора Института проблем информатики Российской академии наук

ОБЪЯВЛЕНИЯ О КОНФЕРЕНЦИЯХ

<http://www.icnaam.org/>



Специальная сессия «Прикладные проблемы теории вероятностей и математической статистики, относящиеся к моделированию информационных систем»

22–28 сентября 2014 г.

Родос, Греция

В рамках 12-й Международной конференции по численному анализу и прикладной математике (12th International Conference of Numerical Analysis and Applied Mathematics — ICNAAM 2014), которая состоится на острове Родос (Греция) с 22 по 28 сентября 2014 г., будет проведена специальная сессия «Прикладные проблемы теории вероятностей и математической статистики, относящиеся к моделированию информационных систем» (Applied Problems in Probability Theory and Mathematical Statistics Related to Modeling of Information Systems — APPT + MS), организуемая с участием Института проблем информатики Российской академии наук (ИПИ РАН).

Сопредседатели сессии

А. И. Зейфман (Вологодский государственный университет, ИПИ РАН)
А. К. Горшенин (ИПИ РАН)

Члены программного комитета

А. А. Грушо (ИПИ РАН, МГУ им. М. В. Ломоносова)
В. Ю. Королев (МГУ им. М. В. Ломоносова, ИПИ РАН)
А. В. Печинкин (ИПИ РАН, РУДН)
К. Е. Самуйлов (РУДН)
С. Я. Шоргин (ИПИ РАН)

Тематика специальной сессии

- Применение теории вероятностей в моделировании информационных систем
- Применение случайных процессов в моделировании информационных систем
- Применение математической статистики в моделировании информационных систем
- Теория массового обслуживания
- Дискретные вероятностные модели
- Модели информационной безопасности
- Современные задачи теории телетрафика

Труды конференции публикуются в изданиях, индексируемых в системах **Web of Sciences** и **Scopus**.

Важные даты

- 20 апреля 2014 г.** Ранняя регистрация с оплатой льготного регистрационного взноса
20 июня 2014 г. Обычная регистрация с оплатой регистрационного взноса
31 июля 2014 г. Обычная регистрация с оплатой повышенного регистрационного взноса
20 июля 2014 г. Представление расширенных тезисов
29 июля 2014 г. Извещение о принятии доклада
1 августа 2014 г. Представление окончательного варианта расширенных тезисов для публикации в издании AIP Conference Proceedings
1 августа 2014 г. Представление полных статей для публикации в журналах

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.
Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Рекомендуется для составления списка “References” пользоваться размещенной на сайте <http://translit.ru/> (опция BGN) бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaistvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’—kompozit” [Experimental study of the strength of joints “steel—composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publ. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)

<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;

- indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.

Important! Keywords must not be sentences.

9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.ru>, option BGN).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Article from the continuing publication (collection of works, proceedings):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy "stal'-kompozit" [Experimental study of the strength of joints "steel-composite"]. *Trudy MGTU "Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem"* [Bauman MSTU "Mathematical Modeling of Complex Technical Systems" Proceedings]. 593:125–130.

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publ. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author's name in the "Subject" field;
- an article and additional materials have to be attached using the "attach" function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. "Informatics and Applications" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

IPI RAN, Vavilova Str., 44, block 2, Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)

<http://www.ipiran.ru/english/journal.asp>