ISSN 1997-3217

# Труды
## КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РОССИЙСКОЙ АКАДЕМИИ НАУК

№ 5, 2011

transactions.krc.karelia.ru

## Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Выпуск 2

### СОДЕРЖАНИЕ



ТРУДЫ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН. № 5, 2011

Подписной индекс в каталоге «Роспечать» **57397**

Карельский научный центр
Российской академии наук

# ТРУДЫ
## КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РОССИЙСКОЙ АКАДЕМИИ НАУК

Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Выпуск 2

Петрозаводск
2011

**Труды Карельского научного центра**
**Российской академии наук**
№ 5, 2011. Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, вып. 2



ISSN 1997-3217

Зав. редакцией Н. В. Михайлова
Адрес редакции: 185910 Петрозаводск, ул. Пушкинская, 11
тел. (8-8142)780109; (8-8142)769600
E-mail: trudy@krc.karelia.ru
Электронная полнотекстовая версия: http://transactions.krc.karelia.ru

# ПРЕДИСЛОВИЕ

Данный выпуск Трудов Карельского научного центра РАН является продолжением издававшихся ранее сборников Института прикладных математических исследований и первого выпуска Трудов КарНЦ РАН в серии «Математическое моделирование и информационные технологии». Включение этих трудов в Перечень ВАК заметно подняло интерес к изданию, что сразу сказалось на количестве статей, полученных редколлегией, на расширении тематики и «географии мест работы» авторов.

Настоящий выпуск содержит работы, посвященные решению различных задач теории вероятностей, математической и прикладной статистики, теории игр и вычислительной математики. Хотя такие статьи содержат достаточно глубокие теоретические результаты, они направлены на решение актуальных прикладных задач. Это отражено в ряде публикаций, непосредственно направленных на создание и использование математических моделей конкретных объектов.

Некоторые из статей содержат новые и интересные постановки задач, в них предлагаются и обсуждаются оригинальные методы моделирования, исследование которых еще далеко от завершения. Однако редколлегия сочла возможной публикацию этих статей для привлечения внимания к предлагаемым методам и к активизации дискуссии о перспективах их использования.

Одним из важных событий в научной жизни КарНЦ РАН стало введение в эксплуатацию высокопроизводительного вычислительного кластера. Результаты использования и перспективы применения высокопроизводительных вычислений освещаются в разделе «Хроника». Также в разделе «Хроника» есть информация о проведенных конференциях.

Труды Карельского научного центра РАН
№ 5. 2011. С. 4–9

*УДК 519.2*

# О ЧИСЛЕ ДЕРЕВЬЕВ ЗАДАННОГО ОБЪЕМА В СЛУЧАЙНОМ НЕПОМЕЧЕННОМ НЕКОРНЕВОМ ЛЕСЕ

**Е. С. Берникович**

*Институт прикладных математических исследований Карельского научного центра РАН*

Рассматривается множество $F_{N,n}$ всех случайных лесов, состоящих из $N$ упорядоченных некорневых деревьев и $n$ незанумерованных вершин, на котором задано равномерное распределение вероятностей. Для таких лесов доказаны предельные теоремы о числе деревьев заданного объема при $N, n \to \infty$ так, что $1 < n/N \leqslant L = 2.0512772...$

К л ю ч е в ы е  с л о в а: случайные леса, число деревьев заданного объема, обобщенная схема размещения, предельные теоремы.

## E. S. Bernikovich. ON THE NUMBER OF TREES OF A GIVEN SIZE IN A RANDOM UNLABELLED UNROOTED FOREST

We consider the set $F_{N,n}$ of all random forests consisting of $N$ ordered unrooted trees and $n$ unlabelled vertices with uniform probability distribution on this set. We prove limit theorems for the number of trees of a given size in such forest as $N, n \to \infty$ so that $1 < n/N \leqslant L = 2.0512772...$

K e y  w o r d s: random forests, number of trees of a given size, generalized allocation scheme, limit theorems.

## Введение

Нахождение предельных распределений таких важных характеристик как максимальный объем дерева и число деревьев заданного объема является традиционным направлением исследований различных типов случайных лесов (см. [Pavlov, 2000; Павлов, Лосева, 2002; Хворостянская, 2002; Колчин, 2004]). Для изучавшихся ранее типов случайных лесов было показано, что такие распределения близки и часто отличаются друг от друга только значениями параметров. Основными методами получения результатов были методы теории ветвящихся процессов (в тех случаях, когда лес генерируется процессом Гальтона-Ватсона ([Pavlov, 2000]) и обобщенная схема размещения частиц по ячейкам, введенная и подробно изученная В. Ф. Колчиным (см., например, [Колчин, 2004]). В настоящей статье доказываются предельные теоремы, описывающие распределение числовой характеристики $\mu_r$, равной числу деревьев, содержащих $r$ вершин, в случае, когда $N$, $n \to \infty$ так, что $1 < n/N \leqslant L = 2.0512772...$, для непомеченных некорневых лесов. Заметим, что для таких лесов в работе [Берникович, Павлов, 2011] получено полное описание предельного поведения другой характеристики – максимального объема дерева, где под объемом понимается число вершин, содержащихся в этом дереве.

## Основные результаты

Обозначим $F_{N,n}$ – множество случайных непомеченных лесов, состоящих из $N$ некорневых деревьев, упорядоченных одним из $N!$ возможных способов, с общим числом вершин $n$. Зададим на этом объекте равномерное распределение вероятностей.

Пусть независимые одинаково распределенные целочисленные случайные величины $\xi_1, ..., \xi_N$ имеют распределение следующего вида:

$$p_k = \mathbf{P}\{\xi_1 = k\} = t_k \lambda^k t^{-1}(\lambda), \quad (1)$$

$$k = 1, 2, ..., \quad 0 < \lambda \leqslant R,$$

где $t_k$ – число непомеченных некорневых деревьев, содержащих $k$ вершин. Рассмотрим производящую функцию

$$t(x) = \sum_{k=1}^{\infty} t_k x^k. \quad (2)$$

Свойства функции (2) изучены в [Харари, Палмер, 1977]. Показано, что начало этой суммы вглядит следующим образом:

$$t(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 6x^6 + 11x^7 + ...,$$

и радиус сходимости ряда $R = 0.3383219...$ Кроме того, функцию $t(x)$ можно представить в виде

$$t(x) = a_0 - a_1(R - x) + a_2(R - x)^{3/2} + ..., \quad (3)$$

где $a_0 = t(R) = 0.5628769..., a_1 = t'(R) = 3.4127749..., a_2 = 6.4243753....$ При $k \to \infty$

$$t_k = \alpha \left(R^k k^{5/2}\right)^{-1} + O\left(\left(R^k k^{7/2}\right)^{-1}\right), \quad (4)$$

где $\alpha = (3a_2/4\sqrt{\pi})\, R^{3/2} = 0.5349485...$

В работе [Берникович, Павлов, 2011] показано, что справедливо равенство:

$$\mathbf{P}\{\eta_1 = k_1, ..., \eta_N = k_N\} =$$

$$\mathbf{P}\{\xi_1 = k_1, ..., \xi_N = k_N | \xi_1 + ... + \xi_N = n\}, \quad (5)$$

где $\eta_1, ..., \eta_N$ – случайные величины, равные объемам деревьев леса из $F_{N,n}$. Это равенство означает, что для рассматриваемых случайных лесов выполнены условия обобщенной схемы размещения частиц по ячейкам (см., например, [Колчин, 2004]).

Пусть

$$L = a_1 R / a_0 = 2,0512772... \quad (6)$$

Следуя [Берникович, Павлов, 2011], в качестве $\lambda$ берем решение уравнения

$$\lambda t'(\lambda)/t(\lambda) = n/N. \quad (7)$$

Обозначим $m = \mathbf{E}\,\xi_1, \quad \sigma^2 = \mathbf{D}\,\xi_1$. Из (1), (2) и (7) получаем, что

$$m = \frac{\lambda t'(\lambda)}{t(\lambda)} = \frac{n}{N}, \quad \sigma^2 = \frac{\lambda^2 t''(\lambda)}{t(\lambda)} + m - m^2. \quad (8)$$

Положим

$$\sigma_{rr}^2 = p_r \left(1 - p_r - \frac{(m-r)^2}{\sigma^2} p_r\right). \quad (9)$$

Справедливы следующие утверждения о предельном поведении числа вершин заданного объема $\mu_r$.

**Теорема 1.** *Пусть $N, n \to \infty, \quad n/N \to 1, n - N \to \infty$. Тогда для любого $r \geqslant 2$ равномерно относительно $u_r = (k - Np_r)/\sqrt{Np_r}$ в любом фиксированном конечном интервале для целых неотрицательных $k$*

$$\mathbf{P}\{\mu_r = k\} = \frac{1}{k!}(Np_r)^k exp\{-Np_r\}(1 + o(1)).$$

**Теорема 2.** *Пусть $N, n \to \infty$ так, что $1 < C_1 \leqslant n/N \leqslant C_2 < L$. Тогда равномерно относительно $u_r = (k - Np_r)/(\sigma_{rr}\sqrt{N})$ в любом фиксированном конечном интервале для целых неотрицательных $k$*

$$\mathbf{P}\{\mu_r = k\} = (\sigma_{rr}\sqrt{2\pi N})^{-1} e^{-{u_r}^2/2}(1 + o(1)).$$

**Теорема 3.** *Пусть $N, n \to \infty$ так, что $n/N \to L, \quad N\,(L - n/N)^3 \to \infty$. Тогда для целых неотрицательных $k$:*
*1. при фиксированном $r$ равномерно относительно $u_r = (k - Np_r) / (\sigma_{rr}\sqrt{N})$ в любом фиксированном конечном интервале*

$$\mathbf{P}\{\mu_r = k\} = (\sigma_{rr}\sqrt{2\pi N})^{-1} e^{-{u_r}^2/2}(1 + o(1));$$

*2. при $r \to \infty$ равномерно относительно $u_r = (k - Np_r)/\sqrt{Np_r}$ в любом фиксированном конечном интервале*

$$\mathbf{P}\{\mu_r = k\} = \frac{1}{k!}(Np_r)^k exp\{-Np_r\}(1 + o(1)).$$

Ниже приводятся леммы 1–4, с помощью которых будут доказаны сформулированные теоремы.

## Вспомогательные утверждения

Пусть $r$ – натуральное число. Введем независимые вспомогательные случайные величины $\xi_i^{(r)}$, $i = 1, ..., N$, такие, что

$$\mathbf{P}\{\xi_i^{(r)} = k\} = \mathbf{P}\{\xi_i = k | \xi_i \neq r\}. \quad (10)$$

Положим

$$\zeta_N = \xi_1 + ... + \xi_N, \quad \zeta_N^{(r)} = \xi_1^{(r)} + ... + \xi_N^{(r)}.$$

Обозначим также $m_r = \mathbf{E}\,\xi_1^{(r)}$, $\sigma_r^2 = \mathbf{D}\,\xi_1^{(r)}$. Из (1), (8) и (10) вытекают следующие соотношения:

$$m_r = \frac{m - rp_r}{1 - p_r},$$

$$\sigma_r^2 = \frac{\sigma^2}{(1-p_r)^2}\left(1 - p_r - p_r\frac{(m-r)^2}{\sigma^2}\right). \quad (11)$$

Пусть $\varphi(u), \varphi^{(r)}(u)$, $\varphi_S^{(r)}(u)$ – характеристические функции случайных величин $\xi_1$, $\xi_1^{(r)}$, $(\zeta_S^{(r)} - Sm_r)/(\sigma_r\sqrt{S})$ соответственно. Используя (1), (2) и (10), нетрудно получить, что

$$\varphi(u) = \frac{t(\lambda R^{iu})}{t(\lambda)}, \varphi^{(r)}(u) = \frac{\varphi(u) - p_r e^{iur}}{1 - p_r}. \quad (12)$$

В условиях обобщенной схемы размещения, как показано в [Колчин, 2004], удобно использовать следующую лемму, которая является следствием соотношения (5).

**Лемма 1.** *Справедливо равенство*

$$\mathbf{P}\{\mu_r = k\} =$$

$$\left(\begin{array}{c} N \\ k \end{array}\right) p_r^k (1-p_r)^{N-k} \frac{\mathbf{P}\{\zeta_{N-k}^{(r)} = n - kr\}}{\mathbf{P}\{\zeta_N = n\}}.$$

Лемма 1 показывает, что для изучения предельного поведения $\mu_r$ достаточно изучить предельное поведение сумм независимых случайных величин $\zeta_N$ и $\zeta_N^{(r)}$ и биномиальных вероятностей $\left(\begin{array}{c} N \\ k \end{array}\right) p_r^k (1-p_r)^{N-k}$.

Из соотношений (1) – (4) и (6) – (8) нетрудно получить утверждения следующей леммы.

**Лемма 2.** *Пусть* $N, n \to \infty$.
*1. Если* $n/N \to 1$, *то*

$$\lambda = (n/N - 1)(1 + o(1)), \quad \sigma^2 = 2\lambda(1 + o(1)).$$

*2. Если* $1 < C_1 \leqslant n/N \leqslant C_2 < L$, *то*

$$0 < C_1 \leqslant \lambda \leqslant C_2 < R, \quad \sigma^2 = 2\lambda(1 + o(1)).$$

*3. Если* $n/N \to L$, *то*

$$\lambda = R - \frac{4}{9a_2^2}\left(1 - \frac{n}{LN}\right)^2 (1 + o(1)),$$

$$\sigma^2 = (3a_2)/(4a_0)R^2(R-L)^{-1/2}(1 + o(1)).$$

При выполнении условий теорем 1–3 из (8) и (11) видим, что

$$\sigma_r = \sigma(1 + o(1)). \quad (13)$$

Леммы 3 и 4 будем доказывать для случая $r \to \infty$, поскольку для фиксированного $r$ их утверждения следуют из результатов работ [Колчин, 2003; Павлов, 2005].

**Лемма 3.** *Пусть* $N, n \to \infty$ *так, что выполнено одно из условий:*
*1.* $n/N \to 1, \quad n - N \to \infty, \quad r \geqslant 2$;
*2.* $1 < C_1 \leqslant n/N \leqslant C_2 < L$;
*3.* $n/N \to L, \quad N(L - n/N)^3 \to \infty$.
*Тогда при* $S = N(1-p_r)(1+o(1))$ *равномерно по* $u$ *в любом конечном интервале*

$$\varphi_S^{(r)}(u) \to e^{-u^2/2}.$$

*Доказательство.* В дальнейшем нам потребуется явный вид третьей производной $(\ln\varphi^{(r)}(u))'''_u$. Для удобства обозначим $\rho(u) = \varphi(u) - p_r e^{iur}$. Из (12) получаем, что

$$(\ln\varphi^{(r)}(u))''' = (\varphi'''(u) + ir^3 p_r e^{iur})\rho^{-1}(u)$$
$$- 3(\varphi'(u) - irp_r e^{iur})(\varphi''(u) + r^2 p_r e^{iur})\rho^{-2}(u)$$
$$+2(\varphi'(u) - irp_r e^{iur})^3 \rho^{-3}(u). \quad (14)$$

Используя (1), (2), (4), (8), (11), (13) и лемму 2, можно убедиться в выполнении соотношения

$$\sigma_r\sqrt{S} \to \infty \quad (15)$$

при каждом из условий леммы 2.

Поскольку

$$\varphi_S^{(r)}(u) = exp\left\{-\frac{iSm_r u}{\sigma_r\sqrt{S}}\right\}\left(\varphi^{(r)}\left(\frac{u}{\sigma_r\sqrt{S}}\right)\right)^S,$$

справедливо равенство

$$\ln\varphi_S^{(r)}(u) = -\frac{iSm_r u}{\sigma_r\sqrt{S}} + S\ln\varphi^{(r)}\left(\frac{u}{\sigma_r\sqrt{S}}\right). \quad (16)$$

При достаточно малых $u$ имеет место разложение

$$\ln\varphi^{(r)}(u) = u\left(\frac{\partial\ln\varphi^{(r)}(u)}{\partial u}\right)_{u=0}$$
$$+\frac{u^2}{2}\left(\frac{\partial^2\ln\varphi^{(r)}(u)}{\partial u^2}\right)_{u=0} + \frac{u^3}{3!}Q(u)$$
$$= ium_r - \frac{u^2\sigma_r^2}{2} + \frac{u^3}{6}Q(u), \quad (17)$$

где

$$|Q(u)| \leqslant 2\max_{|\tau|\leqslant|u|}|\ln\varphi^{(r)}(\tau)'''_\tau|. \quad (18)$$

Отсюда и из (15) – (18) следует, что при выполнении условий леммы при любом фиксированном $u$ справедливо соотношение

$$\ln \varphi_S^{(r)}(u) = -\frac{u^2}{2} + \frac{u^3}{6\sigma_r^3\sqrt{S}} Q\left(\frac{u}{\sigma_r\sqrt{S}}\right). \quad (19)$$

Для доказательства леммы достаточно показать, что второе слагаемое равенства (19) стремится к нулю.

Из (18) следует, что

$$\left|\frac{u^3}{6\sigma_r^3\sqrt{S}} Q\left(\frac{u}{\sigma_r\sqrt{S}}\right)\right| \leqslant \frac{|u|^3}{3} Q_1(u), \quad (20)$$

где

$$Q_1(u) = \max_{|\tau|\leqslant|u/(\sigma_r\sqrt{S})|} |\ln \varphi^{(r)}(\tau)'''_\tau| / (\sigma_r^3\sqrt{S}). \quad (21)$$

Рассмотрим случай, когда $1 < C_1 \leqslant n/N \leqslant C_2 < L$. Из (1), (2) и (12) следует, что производные $\varphi'(u), \varphi''(u), \varphi'''(u)$ ограничены, а $r^3 p_r \to 0$ при $r \to \infty$. Отсюда, из (11), (13) – (15) и (19) – (21) получаем, что

$$\ln \varphi_S^{(r)}(u) = (-u^2/2)(1 + o(1)). \quad (22)$$

Пусть теперь

$$n/N \to 1, \quad n - N \to \infty, \quad r \geqslant 2.$$

Из (1) и (11), применяя лемму 2, нетрудно вывести, что

$$\sigma_r^3\sqrt{S} = \sqrt{8}(n-N)^{3/2} N^{-1}(1 + o(1)). \quad (23)$$

Из (2) и (12) получаем выражения:

$$\varphi(u) = 1 + O(\lambda), \varphi'(u) = i + O(\lambda),$$
$$\varphi''(u) = -1 + O(\lambda), \varphi'''(u) = -i + O(\lambda).$$

Ясно, что при $r \geqslant 2$ из (1) следует, что $p_r \to 0$. Таким образом, из (14) получаем, что $\ln \varphi^{(r)}(u)'''_u = O(\lambda)$. Откуда, применяя (23), получаем соотношение

$$\ln \varphi^{(r)}(u)'''_u / (\sigma_r^3\sqrt{S}) = (n-N)^{-1/2}(1 + o(1)),$$

а значит из (19) – (21) следует (22).

Осталось рассмотреть случай

$$n/N \to L, \quad N\,(L - n/N)^3 \to \infty.$$

Аналогично (23), делаем вывод, что в этом случае

$$\sigma_r^3\sqrt{S} = (3a_2)/(4a_0)^{3/2} R^3 (R-L)^{-3/4}(1 + o(1)). \quad (24)$$

Из (1), (2) и (12) следует, что $\varphi'(u) \to \text{const}, \varphi''(u) \to \infty, \varphi'''(u) \to \infty$, а

$$t'''(\lambda)/(\sigma^3\sqrt{N}) \asymp ((L - n/N)N)^{-1/2}.$$

Тогда из (19) – (21) и (24) получаем (22). □

Ниже приводится лемма 4, в которой устанавливается локальная сходимость распределения суммы $\zeta_S^{(r)}$ к нормальному закону.

**Лемма 4.** *Пусть $N, n \to \infty$ так, что выполнено одно из условий леммы 3. Тогда при $S = N(1-p_r)(1+o(1))$*

$$\mathsf{P}\left\{\zeta_S^{(r)} = l\right\} = \frac{1}{\sigma_r\sqrt{2\pi S}} e^{-z^2/2}\,(1 + o(1))$$

*равномерно по $l$, для которых $z = (l - Sm_r)/\sigma_r\sqrt{S}$ находится в любом фиксированном конечном интервале.*

*Доказательство.* Следуя классическому доказательству локальных предельных теорем, представим рассматриваемую вероятность по формуле обращения в виде интеграла

$$\mathsf{P}\left\{\zeta_S^{(r)} = l\right\} = \frac{1}{2\pi\sigma_r\sqrt{S}} \int\limits_{-\pi\sigma_r\sqrt{S}}^{\pi\sigma_r\sqrt{S}} e^{-izt}\varphi_S^{(r)}(t)dt. \quad (25)$$

Поскольку

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-izt} e^{-t^2/2} dt, \quad (26)$$

разность

$$R_s = \sqrt{2\pi}\left(\sigma_r\sqrt{2\pi S}\,\mathsf{P}\left\{\zeta_S^{(r)} = l\right\} - e^{-z^2/2}\right)$$

можно представить в виде суммы

$$R_s = I_1 + I_2 + I_3 + I_4, \quad (27)$$

где

$$I_1 = \int\limits_{-A}^{A} e^{-izt}\left(\varphi_S^{(r)}(t) - e^{-t^2/2}\right) dt,$$

$$I_2 = \int\limits_{A<|t|\leqslant\varepsilon\sigma_r\sqrt{S}} e^{-izt}\varphi_S^{(r)}(t)dt,$$

$$I_3 = \int\limits_{\varepsilon\sigma_r\sqrt{S}<|t|\leqslant\pi\sigma_r\sqrt{S}} e^{-izt}\varphi_S^{(r)}(t)dt,$$

$$I_4 = -\int\limits_{A<|t|} e^{-izt-t^2/2} dt,$$

положительные постоянные $A, \varepsilon$ будут выбраны позднее. Для доказательства леммы достаточно показать, что разность $R_s$ можно сделать сколь угодно малой при больших $n, N$.

Из (26) следует, что $|I_4| \leqslant \int\limits_{A<|t|} e^{-t^2/2}dt$, и, выбрав достаточно большое $A$, мы можем сделать $I_4$ сколь угодно малым.

Оценим $I_3$. Пусть $1 < C_1 \leqslant n/N \leqslant C_2 < L$. Поскольку максимальный шаг распределения $\xi_1^{(r)}$ равен 1, $0 < C_3 \leqslant \lambda \leqslant C_4 < R$, при $\varepsilon < |t| \leqslant \pi$ выполняется неравенство

$$\left|\varphi^{(r)}(t)\right| \leqslant e^{-C_5}. \qquad (28)$$

Здесь и далее $C_5, C_6, \ldots$ означают некоторые положительные постоянные. Поэтому

$$|I_3| \leqslant 2 \int\limits_{\varepsilon\sigma_r\sqrt{S}}^{\pi\sigma_r\sqrt{s}} e^{-C_5 S}dt \leqslant C_6\sigma_r\sqrt{S}e^{-C_5 S}$$

и $I_3 \to 0$ при $S \to \infty$ и постоянном $\varepsilon > 0$. При условии $n/N \to L, \quad N\left(L - n/N\right)^3 \to \infty$, выполнено (28), поэтому, проводя аналогичные выкладки, находим, что $|I_3| \to 0$ и в этом случае.

Пусть теперь $n/N \to 1, \quad n - N \to \infty$. Из (2), (3), (12) и того, что $p_r \to 0$, получаем соотношение $|\varphi^{(r)}(t)|^S \leqslant e^{-C_8\lambda S}$. Тогда по лемме 2

$$|I_3| \leqslant C_9\sigma_r\sqrt{S}e^{-C_8\lambda S}$$

и $I_3 \to 0$ при $S \to \infty$ и постоянном $\varepsilon > 0$.

Оценим теперь $I_1$ и $I_2$ при фиксированных $\varepsilon$ и $A$. По лемме 3 слабая сходимость к нормальному распределению равномерна на любом конечном интервале и, следовательно, интеграл $I_1 \to 0$ при $S \to \infty$. Для интеграла $I_2$ справедлива оценка

$$|I_2| \leqslant \int\limits_{A<|t|\leqslant\varepsilon\sigma_r\sqrt{S}} |\varphi_S^{(r)}(t)|dt.$$

При каждом из условий леммы 3 и $t < \varepsilon\sigma_r\sqrt{S}$ из леммы 2, (20) и (21) следует выполнение следующего неравенства:

$$\left|\frac{t}{3\sigma_r^3\sqrt{S}}Q\left(\frac{t}{\sigma_r\sqrt{S}}\right)\right| \leqslant C_{10}\varepsilon,$$

тогда из (19):

$$|\varphi_S^{(r)}(t)| < e^{-C_7 t^2}.$$

Отсюда делаем вывод, что интеграл $I_2$ может быть сделан сколь угодно малым при достаточно большом $A$.

□

## Доказательства теорем

В работе [Берникович, Павлов, 2011] было показано, что для суммы $\zeta_N$ при выполнении условий теорем 1 – 3 выполняется соотношение

$$\mathsf{P}\left\{\zeta_N = n\right\} = (\sigma\sqrt{2\pi N})^{-1}(1 + o(1)) \qquad (29)$$

равномерно относительно $n$.

При выполнении условий теоремы 1 $p_r \to 0$. Поэтому, как известно, для целых положительных $k$

$$\binom{N}{k} p_r^k(1 - p_r)^{N-k} = \frac{(Np_r)^k}{k!}e^{-Np_r}(1 + o(1)) \qquad (30)$$

равномерно относительно $(k - Np_r)/\sqrt{Np_r}$ в любом конечном интервале. Для оценки вероятности $\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n - kr\right\}$ используем лемму 4, полагая $S = N - k$ и $l = n - kr$ :

$$\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n - kr\right\} = \left(\sigma_r\sqrt{2\pi(N-k)}\right)^{-1}$$

$$\times \exp\left\{-\frac{\left(n - kr - (N-k)m_r\right)^2}{2\sigma_r^2\left(N-k\right)}\right\}(1 + o(1)).$$

Заметим, что $N - k = N(1 - p_r - u\sqrt{p_r/N})$. Отсюда, из леммы 2, (8), (9), (11), (13), вытекает, что

$$\frac{\left(n - kr - (N-k)m_r\right)^2}{2\sigma_r^2\left(N-k\right)} \to 0,$$

следовательно,

$$\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n - kr\right\} =$$

$$\left(\sigma\sqrt{2\pi(N-k)}\right)^{-1}(1 + o(1)).$$

Отсюда и из (29) получаем, что

$$\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n - kr\right\} / \mathsf{P}\left\{\zeta_N = n\right\} \to 1. \qquad (31)$$

Для завершения доказательства осталось применить лемму 1 и (30).

Для доказательства теоремы 2 заметим, что

$$k = Np_r + u_r\sigma_{rr}\sqrt{N}, \quad N - k = N(1 - p_r)(1 + o(1)), \qquad (32)$$

где $u_r \leqslant C < \infty$. Используя нормальное приближение для биномиального распределения при $Np_r(1 - p_r) \to \infty$, находим, что

$$\binom{N}{k} p_r^k(1 - p_r)^{N-k} = \frac{1 + o(1)}{\sqrt{2\pi Np_r(1 - p_r)}}$$

$$\times \exp\left\{-\frac{(k-Np_r)^2}{2Np_r(1-p_r)}\right\} \qquad (33)$$

равномерно относительно

$$(k-Np_r)/\sqrt{Np_r(1-p_r)}$$

в любом фиксированном конечном интервале. Поскольку из (8) и (9) следует, что $\sigma_{rr}^2 \leqslant p_r(1-p_r)$, то равенство (33) справедливо равномерно и по $(k-Np_r)/(\sigma_{rr}\sqrt{N})$.

По лемме 4 при $S = N-k$ и $l = n-kr$, из (8), (9), (11) и (32) можно установить, что

$$\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n-kr\right\} = \left(\sigma_r\sqrt{2\pi N(1-p_r)}\right)^{-1}$$

$$\times \exp\left\{-\frac{(m-r)^2 p_r u_r^2}{2\sigma^2\left(1-p_r+(u_r\sigma_{rr})/\sqrt{N}\right)}\right\}(1+o(1)). \qquad (34)$$

Нетрудно получить, что

$$\frac{(k-Np_r)^2}{2Np_r(1-p_r)} + \frac{(m-r)^2 p_r u_r^2}{2\sigma^2(1-p_r)} = \frac{u_r^2}{2}.$$

Отсюда, из (29), (33) и (34), применяя лемму 1, получаем утверждение теоремы 2.

Утверждения теоремы 3 очевидным образом следуют из лемм 1 и 4, (29) – (31), (33).

## Литература


*Берникович Е. С., Павлов Ю. Л.* О максимальном объеме дерева случайного непомеченного некорневого леса // Дискретная математика. 2011. Т. 23, № 1. С. 3–20.

*Колчин А. В.* Предельные теоремы для обобщенной схемы размещения // Дискретная математика. 2003. Т. 15, № 4. С. 148–157.

*Колчин В. Ф.* Случайные графы. М.: Физматлит, 2004. 256 с.

*Павлов Ю. Л., Лосева Е. А.* Предельные распределения максимального объема дерева в случайном рекурсивном лесе // Дискретная математика. 2002. Т. 14, № 1. С. 60–74.

*Павлов Ю. Л.* Предельные теоремы для объемов деревьев в случайном непомеченном лесе // Дискретная математика. 2005. Т. 17, № 2. С. 70–86.

*Харари Ф., Палмер Э.* Перечисление графов. М.: Мир, 1977. 326 с.

*Хворостянская Е. В.* Об условии возникновения гигантского дерева в случайном непомеченном лесе // Дискретная математика. 2002. Т. 19, № 3. С. 35–50.

*Pavlov Yu.* Random forests. Utrecht: VSP, 2000.


## СВЕДЕНИЯ ОБ АВТОРЕ:


**Берникович Елена Сергеевна**
аспирант лаб. теории вероятностей и компьютерной статистики
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: bern@krc.karelia.ru
тел.: (8142) 761218

**Bernikovich, Elena**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: bern@krc.karelia.ru
tel.: (8142) 761218

Труды Карельского научного центра РАН
№ 5. 2011. С. 10–17

*УДК 338.45:621.31(470.22)*

# ИССЛЕДОВАНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ЕДИНИЧНОЙ ТЕХНОЛОГИЧЕСКОЙ ЛИНИИ ТОПЛИВНО-ЭНЕРГЕТИЧЕСКОГО ХОЗЯЙСТВА РЕГИОНА

**Г. А. Борисов, Т. П. Тихомирова**

*Институт прикладных математических исследований Карельского научного центра РАН*

В статье обосновывается математическая модель топливно-энергетического хозяйства региона в виде системы рекурсивных уравнений и изложены результаты аналитических и численных исследований ее, показывающие определяющую роль неравномерности графика конечной мощности в увеличении потерь мощности и энергии.

К л ю ч е в ы е  с л о в а: технологические переделы, потери мощности, потери энергии, расходная характеристика, неравномерность графика нагрузки, стабилизация нагрузки.

## G. A. Borisov, T. P. Tikhomirova. STUDY OF THE MATHEMATICAL MODEL OF A SINGLE PROCESSING LINE IN THE REGION'S FUEL-&-ENERGY SECTOR

The paper substantiates the mathematical model of the region's fuel-&-energy sector through a system of recursive equations, and reports the results of its analytical and numerical studies, which demonstrate the irregularity of the output curve plays a decisive role in the increase in power and energy losses.

K e y  w o r d s: process transitions, power losses, energy losses, metering characteristic, irregularity of the load curve, load stabilization.

---

Одним из наиболее общих и эффективных способов исследования сложных систем большой размерности является переход от исходной сложной системы к исследованию более простых систем, по свойствам которых можно восстановить точно или приближенно свойства исходной [Крон, 1972; Первозванский, Гайцгори, 1979]. Таким переходом от сложной системы к более простой является разделение системы на изолированные части, основанное на пренебрежении слабыми связями между ними.

Другим способом снижения размерности сложных систем является агрегирование однотипных подсистем, используя малость различий между ними и допустимую возможность суммирования параметров таких подсистем.

Объединяющим эти два способа приемом в исследовании сложных систем является моделирование единичной нити или единичной технологической линии [Гуд, Макол, 1962], когда сложная система моделируется линейной последовательностью разнотипных подсистем технологических переделов, каждый из которых моделирует свойства агрегированных однотипных элементов.

Топливно-энергетическое хозяйство региона является сложной системой, поскольку оно обладает всеми ее чертами – целостностью, большой размерностью, сложностью прямых и обратных связей, нерегулярным распределением нагрузок во времени, наличием состязательных, конкурирующих сторон [Гуд, Макол, 1962]. Вследствие сложных связей и большой размерности топливно-энергетическое хозяйство необозримо [Макаров, Мелентьев, 1973], поэтому представим его в виде единичной технологической линии, образованной технологическими переделами.

Технологический передел образуется из отдельных однотипных составляющих, выполняющих одинаковые функции с близкими технико-экономическими характеристиками. Он описывается в установившемся режиме расходной характеристикой – функциональной зависимостью входной мощности от выходной.

Расходная характеристика $i$-го передела топливно-энергетического хозяйства, состоящего из множества однотипных параллельно работающих агрегатов, представляет собой нелинейную функцию мощности $P_{i-1}$, поступающей на элемент системы от его выходной мощности $P_i$, т. е. $P_{i-1} = f(P_i)$. Будем считать, что функция во всем диапазоне допустимого изменения нагрузки является дифференцируемой, выпуклой, монотонно возрастающей [Болотов, 1947].

Ввиду отсутствия идеальных элементов и обязательного наличия потерь мощности и ее расхода на собственные нужды всегда

$$P_{i-1} > P_i, \tag{1}$$

$$P_{i-1} = P_i + \Delta P_i, \tag{2}$$

где $\Delta P_i$ – потери мощности и ее расход на собственные нужды в $i$-м элементе системы.

Величины $\Delta P_i$ могут быть разделены на две составляющие – постоянную и переменную, зависящую от величины выходной мощности (нагрузки) передела $P_i$. Постоянная составляющая образуется так называемыми потерями холостого хода $\Delta P_{0i}$, идущими у энергетических и технологических машин на различные виды трения, у электрических аппаратов переменного тока – на потери мощности холостого хода, обусловленные вихревыми токами магнитопроводов, у линий электропередач высокого напряжения переменного тока – тлеющим разрядом проводов, утечками тока по изоляторам.

Переменная составляющая потерь мощности обычно представляется нелинейной функцией с положительной второй производной, характеризующей увеличенное возрастание потерь в сравнении с ростом нагрузки. В электрических машинах, аппаратах и ЛЭП это потери «в меди» ($\Delta P$), изменяющиеся квадратично относительно нагрузки. Также квадратично относительно нагрузки (подачи) изменяются потери напора в газодинамических и гидровлических процессах и в третьей степени – потери мощности.

В соответствии с располагаемой информацией [Борисов, Тихомирова, 2010], получаемой с помощью действующей системы коммерческого учета, в единичной технологической линии регионального энергетического хозяйства Карелии (рис. 1) можно выделить следующие переделы с однородными технологическими преобразованиями энергии:

- добыча и транспортировка природного газа;
- производство тепла паровыми котлами;
- производство электроэнергии паровыми турбогенераторами;
- транспорт электроэнергии по линиям электропередачи;
- преобразование электроэнергии в конечные виды или услуги центробежными агрегатами с частотно-регулируемыми электроприводами.

Источником первичной энергии выбран природный газ, превалирующий на ТЭЦ Карелии. Выбор в качестве последнего передела центробежных электронасосных, вентиляторных и компрессорных агрегатов обусловлен значительной долей (до 25 %) вырабатываемой электроэнергии, потребляемой ими.

Исходя из данных коммерческого учета [Борисов, Тихомирова, 2010] и гидродинамических процессов на газопроцессорных станциях и газопроводах, расходная характеристика первого передела описывается уравнением

$$P_0 = 0,09P_{H1} + P_1 + 0,1803P_{H1}\left(\frac{P_1}{P_{H1}}\right)^3, \tag{3}$$

где $P_{H1}$ – номинальная мощность первого передела. Для второго передела расходная характеристика паровых котлов (аналогичных котлам БКЗ-420) с паспортным к.п.д. 90,5 % [Горшков, 1949]

$$P_1 = 0,0203P_{H2} + P_2 + 0,41P_{H2}\left(\frac{P_2}{P_{H2}}\right)^3, \tag{4}$$

где $P_{H2}$ – номинальная мощность второго передела.

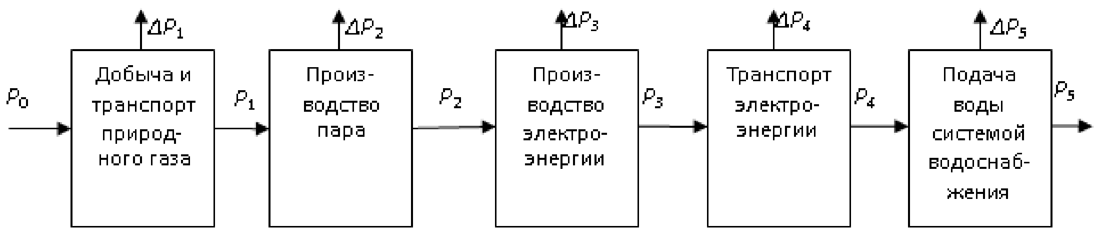



*Рис. 1.* Структурная схема единичной технологической линии энергетического хозяйства региона

Для теплофикационных паротурбогенераторов с регулируемым отбором пара учетом собственных нужд расходная характеристика [Горшков, 1949] составится из двух прямолинейных отрезков, описываемых двумя линейными уравнениями:

$$P_2 = 0,068P_{H3} + 2,065P_3 \quad \text{при } P_3 \leqslant 0,78P_{H3}, \tag{5}$$

$$P_2 = 0,068P_{H3} + 2,065P_3 + (2,95 - 2,065)(P_3 - 0,78P_{H3}) \quad \text{при } 0,78P_{H3} < P_3 \leqslant P_{H3}, \tag{6}$$

где $P_{H3}$ – номинальная мощность третьего передела.

Электрическую сеть со среднегодовым уровнем потерь энергии в 11,8 % с учетом заполнения графика нагрузки 0,68, удельного времени потерь $\tau = 0{,}498$, потерь холостого хода можно представить в виде эквивалента с расходной характеристикой [Железко и др., 2006]

$$P_3 = 0,0205P_{H4} + P_4 + 0,157P_{H4}\left(\frac{P_4}{P_{H4}}\right)^2, \tag{7}$$

где $P_{H4}$ – номинальная мощность четвертого передела.

Последний передел — центробежный электронасос с водопроводной сетью может быть описан уравнением

$$P_4 = 0,8P_{H5} + P_5 + 4,6P_{H5}\left(\frac{P_5}{P_{H5}}\right)^3, \tag{8}$$

где $P_{H5}$ – номинальная мощность пятого передела.

Уравнение (8) получено по экспериментальным данным публикации [Сарач, Бастунский, 1995]. В уравнении (8) $P_{H5} = 2{,}35$ кВт.

В целом единичная технологическая линия топливно-энергетического хозяйства описывается системой рекурсивных уравнений

$$\begin{cases} P_4 = 0,8P_{H5} + P_5 + 4,6P_{H5}\left(\frac{P_5}{P_{H5}}\right)^3; \\ P_3 = 0,0205P_{H4} + P_4 + 0,157P_{H4}\left(\frac{P_4}{P_{H4}}\right)^2; \\ P_2 = 0,068P_{H3} + 2,065P_3 \text{ при } 0 < P_3 \leqslant 0,78P_{H3}; \\ P_2 = 0,068P_{H3} + 2,065P_3 + (2,95 - 2,065)(P_3 - 0,78P_{H3}) \\ \qquad \text{при } 0,78P_{H3} < P_3 \leqslant P_{H3}; \\ P_1 = 0,0203P_{H2} + P_2 + 0,41P_{H2}\left(\frac{P_2}{P_{H2}}\right)^3; \\ P_0 = 0,09P_{H1} + P_1 + 0,1803P_{H1}\left(\frac{P_1}{P_{H1}}\right)^3. \end{cases} \tag{9}$$

При $P_{H5}$= 2,35 кВт, $P_{H4}$= 15,04 кВт, $P_{H3}$ = 17,71 кВт, $P_{H2}$= 41,22 кВт, $P_{H1}$= 58,96 кВт в соответствии с системой уравнений (9) мгновенное значение мощности каждого $P_i$ передела нелинейно зависит от мощности системы на ее выходе $P_5$ (рис. 2). Потери мощности на каждом $i$-м переделе равны разности мощностей на его входе и выходе $\Delta P_i = P_i(P_5) - P_{i-1}(P_5)$ и зависят однозначно от величины конечной мощности.

Эффективность каждого передела принято оценивать по максимальному значению коэффициента полезного использования (к.п.и.), равного максимальному значению отношения выходной мощности к входной, т. е.

$$\eta_i = \frac{P_i(P_5)}{P_{i-1}(P_5)}. \tag{10}$$

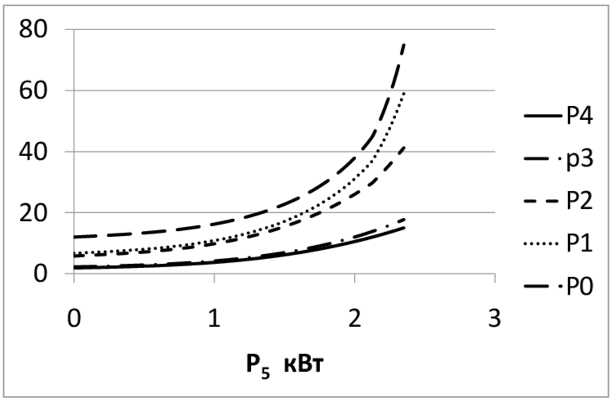



*Рис. 2.* Зависимость мощностей переделов от конечной мощности $P_5$

Зависимость к.п.и. каждого передела и всей технологической линии иллюстрирует рис. 3.

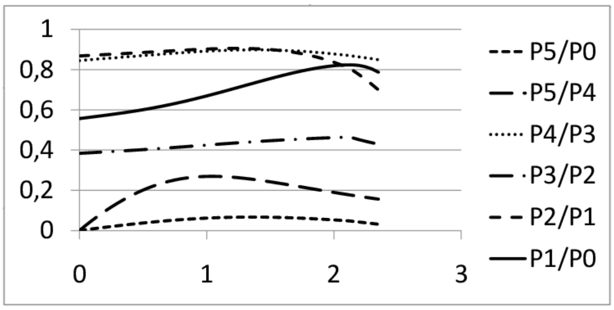



*Рис. 3.* Зависимости к.п.и. от конечной мощности

Численные значения максимальных к.п.и. и значения конечной мощности $P_5$, при которых они достигаются, приведены в табл. 1.

*Таблица 1.* Значения максимальных к.п.и.

| к.п.и. | $P_5$, кВт |
|---|---|
| $\eta = 6,67$ | 1,35 |
| $\eta_5 = 27$ | 1,05 |
| $\eta_4 = 89,8$ | 1,4 |
| $\eta_3 = 46,5$ | 2,12 |
| $\eta_2 = 90,5$ | 1,3 |
| $\eta_1 = 82,3$ | 2,13 |

Из таблицы следует, что максимальные к.п.и. переделов и всей технологической линии достигаются при различных значениях конечной мощности, определяющей однозначно потери мощности на каждом переделе.

Известно, что в связи с изменением во времени передаваемой мощности в линиях электропередач изменяется величина нагрузочных потерь энергии в зависимости от характеристик графика нагрузки – коэффициента формы, числа часов наибольших потерь [Железко и др., 2006]. Влияние изменчивости графика нагрузки, тем более конечной нагрузки, на других переделах и в целом по энергетическому хозяйству неизвестно. Кроме того, графики нагрузок имеют значительную случайную составляющую, обусловленную непредсказуемыми изменениями влияющих на энергопотребление природных и антропогенных факторов. Ввиду этого рассмотрим на $i$-м переделе с расходной характеристикой, ограничившись полиномом третьей степени

$$P_{i-1} = a_{0i} + a_{1i}P_i + a_{2i}P_i^2 + a_{3i}P_i^3 \tag{11}$$

(здесь $a_{ki}$, k = 0,1,2,3 – положительные коэффициенты)
влияние простейшего циклического процесса изменения графика нагрузки $P_i$ с двумя равновеликими отклонениями во времени $\pm\Delta P_i$ от среднего значения $P_{i\,\text{ср}}$ за период T (рис. 4).

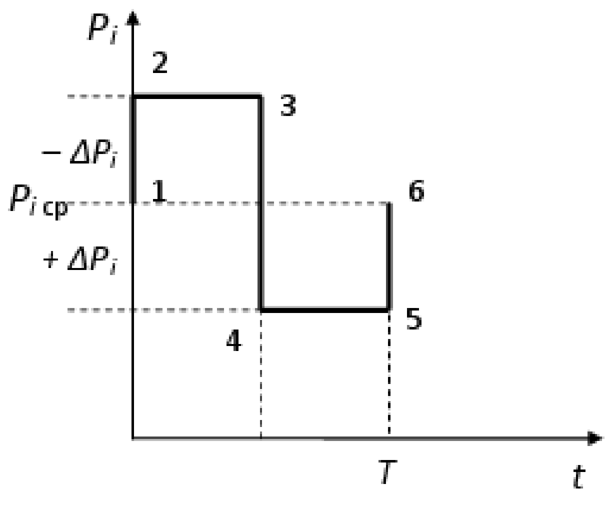



*Рис. 4.* Простейший график изменения нагрузки $P_i = f(t)$

Тогда в первый полупериод потребляемая переделом энергия $W_1$ составит:

$$W_1 = \frac{T}{2} \cdot P_{i-1}\left(P_{i\,\text{ср}} + \Delta P_i\right) = $$
$$= \frac{T}{2}\Big(a_{0i} + \sum_{k=1}^{3} a_{ki}\left(P_{i\,\text{ср}} + \Delta P_i\right)^k\Big). \quad (12)$$

Во второй полупериод потребляемая переделом энергия $W_2$ составит:

$$W_2 = \frac{T}{2} \cdot P_{i-1}\left(P_{i\,\text{ср}} - \Delta P_i\right) = $$
$$= \frac{T}{2}\Big(a_{0i} + \sum_{k=1}^{3} a_{ki}\left(P_{i\,\text{ср}} - \Delta P_i\right)^k\Big). \quad (13)$$

Потребление энергии $i$-м переделом за весь период $T$:

$$W = W_1 + W_2 = T\Big[\,a_{0i} + a_{1i}P_{i\,\text{ср}} + $$
$$+ a_{2i}\left(P_{i\,\text{ср}}^2 + \Delta P_i^2\right) + \quad (14)$$
$$+ a_{3i}\left(P_{i\,\text{ср}}^3 + P_{i\,\text{ср}} \cdot \Delta P_i^2 + P_{i\,\text{ср}}^2 \Delta P_i^2\right)\Big].$$

В сравнении с равномерным графиком нагрузки, когда $P_i = P_{i\,\text{ср}}$, а

$$W_p = P_{i\,\text{ср}} \cdot T = $$
$$= T\left(a_{0i} + a_{1i}P_{i\,\text{ср}} + a_{2i}P_{i\,\text{ср}}^2 + a_{3i}P_{i\,\text{ср}}^3\right) \quad (15)$$

потребление энергии $i$-м переделом увеличилось на величину

$$\Delta W = W - W_p = a_{2i}\Delta P_{i\,\text{ср}}^2 + $$
$$+ a_{3i}P_{i\,\text{ср}} \cdot \Delta P_i^2 + a_{3i}P_{i\,\text{ср}}^2 \cdot \Delta P_i. \quad (16)$$

Таким образом, отклонение на переделе мощности от средней стабильной приводит к увеличению потерь мощности на величину, зависящую от коэффициентов расходной характеристики при нелинейных членах и от величины отклонения от среднего. От величины постоянного и линейного члена расходной характеристики потери мощности при неравномерности нагрузки не зависят.

Рассмотрим единичную технологическую линию с расходными характеристиками технологических переделов, описываемых системой (9). Каждая расходная характеристика является полиномом второй или третьей степени с положительными коэффициентами у свободного члена (условно-постоянными потерями, не зависящими от нагрузки) и положительными коэффициентами у членов первой, второй и третьей степени. Ввиду этого расходные характеристики являются непрерывными выпуклыми функциями с положительными производными.

В технологической линии управляющим параметром является мгновенное значение мощности последнего передела $P_5$. В однозначном соответствии с этим параметром определяются мощности всех предшествующих переделов $P_4$, $P_3$, $P_2$, $P_1$ и $P_0$ по уравнениям (9). Разница между входной и выходной мощностью передела определяет потери мощности $i$-го передела, зависящих от мощности последующего и в конце концов зависящих от конечной мощности потребителя.

Известно, что потребление конечной мощности потребителей энергосистемы в большинстве случаев изменяется во времени с характерными суточными, недельными, годовыми циклами и случайными составляющими, прогнозирование которых имеет высокую степень неопределенности. Оценим влияние на потери энергии неравномерности графиков нагрузки. Для технологической линии переделов проведен расчет потребляемой мощности, энергии и потерь энергии при различных формах графиков, приведенных на рис. 5. Все эти суточные графики имеют одинаковую конечную энергию, равную 28,2 кВт·ч, и одинаковый коэффициент заполнения [Железко и др., 2006]

$$k_з = \frac{P_{\max}}{P} = \frac{P_{\max} \cdot 24}{W}, \quad (17)$$

равный 0,5, но существенно отличаются формой (рис. 5). В табл. 2 приведены результаты расчета энергетических характеристик переделов и всей технологической линии.

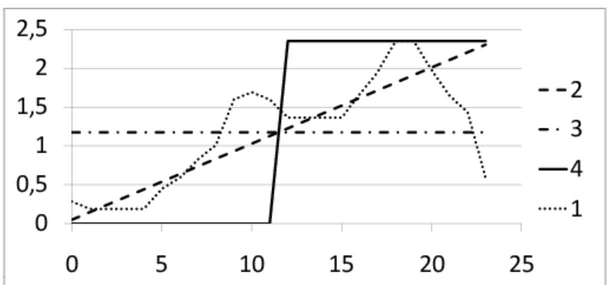



*Рис. 5.* Суточные графики потребления конечной энергии

*Таблица 2.* Результаты расчетов

| № графика | Вид графика | Потребление первичной энергии $W_0$, кВт | Увеличение $W_0$ из-за неравномерности графика, % | Максимум первичной мощности, кВт | Потери первичной энергии | Максимум потерь первичной мощности |
|---|---|---|---|---|---|---|
| 1 | Коммунальная нагрузка | 588,7 | 136,6 | 74,9 | 560,5 | 72,55 |
| 2 | Линейный рост нагрузки до $P_5$ | 571,4 | 132,5 | 74,9 | 543,2 | 72,55 |
| 3 | Равномерная нагрузка | 431,0 | 100 | 12,5 | 402,8 | 10,15 |
| 4 | Постоянная нагрузка 12 час. | 1042,8 | 241 | 74,9 | 1014,6 | 72,55 |

Анализ табл. 2 показывает, что при одном и том же количестве потребляемой конечной энергии и конечной максимальной мощности:

- потери энергии каждого передела определяются однозначно графиком конечного потребления, увеличиваясь от конечного передела к первому, и зависят от неравномерности графика;
- на потери энергии оказывает сильное влияние неравномерность графика нагрузки потребителя, увеличивая их в системе на 32,5–141 %;
- при неравномерном графике нагрузки требующаяся пиковая первичная мощность может возрастать в 6 раз.

Ввиду сильного влияния неравномерности потребляемой конечной энергии необходимо определить возможность снижения потерь энергии от неравномерности потребления конечной энергии путем сглаживания неравномерности.

Для снижения потерь энергии, вызванных неравномерностью графика потребления мощности, определим закон регулирования во времени выходной мощности $i$-го передела единичной технологической линии, расходная характеристика которого является в общем случае полиномом $m$-й степени:

$$P_{i-1} = a_{0i} + a_{1i}P_i + a_{2i}P_i^2 + ... + a_{mi}P_i^m, \quad (18)$$

где $P_{i-1}$ – мощность на входе $i$-го передела; $P_i$ – мощность на выходе передела. Мощность на выходе передела изменяется во времени по графику (рис. 6).

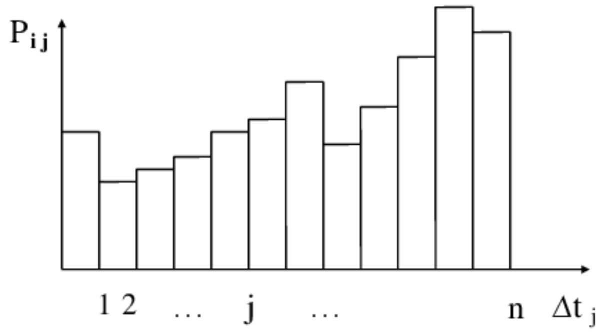



*Рис. 6.* График мощности на выходе $i$-го передела $P_i = f(\Delta t_j)$

Энергия на выходе передела за промежуток времени $T = n \cdot \Delta t_j$ составит величину:

$$W_i = \sum_{j=1}^{n} \Delta t_j \cdot P_{ij}. \quad (19)$$

Требуется найти множество таких $\{P_{i1},\ P_{i2}, \ldots P_{ij}, \ldots P_{in}\}$, чтобы сумма потерь энергии на $i$-м переделе была минимальна.

При заданной суммарной энергии на выходе передела $W_i$, минимум суммы потерь энергии на $i$-м переделе будет соответствовать минимальной сумме энергии на его входе, т. е.

$$\min W_{i-1} = \sum_{j=1}^{n} \Delta t_j P_{i-1,\,j}(P_{ij}). \quad (20)$$

При выборе одинаковых интервалов $\Delta t_j = \Delta t$

$$\min W_{i-1} = \min \Delta t \sum_{j=1}^{n} P_{i-1,\,j}(P_{ij}). \quad (21)$$

С учетом условия (19) сформулирована задача нахождения условного минимума методом неопределенных множителей Лагранжа, при котором условием достижения минимума является равенство производных:

$$\frac{\partial P_{i-1,1}}{\partial P_{i,1}} = \frac{\partial P_{i-1,2}}{\partial P_{i,2}} = ... = \frac{\partial P_{i-1,n}}{\partial P_{i,n}}, \qquad (22)$$

или:

$$\begin{aligned} & a_{1i} + 2a_{2i}P_{i1} + ... + ma_{mi}P_{i1}^{m-1} = \\ & = a_{1i} + 2a_{21}P_{i2} + ... + ma_{mi}P_{i2}^{m-1} = \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & = a_{1i} + 2a_{21}P_{ij} + ... + ma_{mi}P_{ij}^{m-1} = \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & = a_{1i} + 2a_{21}P_{in} + ... + ma_{mi}P_{in}^{m-1}. \end{aligned} \qquad (23)$$

Решение системы уравнений (23) дает

$$P_{i1} = P_{i2} = \ldots = P_{ij} = \ldots = P_{in},$$

т. е. минимум потребляемой энергии на $i$-м переделе и, следовательно, минимум потерь энергии на нем получается в случае постоянства мощности на выходе передела, равной среднему значению мощности за рассматриваемый период. Действуя аналогичным образом, получаем при нелинейных расходных характеристиках такой же вывод для всех переделов технологической линии, а также и для единичной технологической линии в целом.

Таким образом, нелинейность расходных характеристик технологических переделов единичной линии и неравномерность графика конечной мощности являются причинами значительного увеличения потерь первичной мощности и энергии, которые можно снизить при том же количестве конечной энергии выравниванием графика потребляемой конечной мощности.

Выводы

Анализ математической модели единичной технологической линии топливно-энергетического хозяйства региона показал:

1. Мощность переделов технологической линии однозначно определяется конечной потребляемой мощностью.
2. Максимальные значения к.п.и. всей технологической линии и ее переделов достигаются при различных значениях конечной мощности.
3. Неравномерность графика мощности передела, вызванная отклонением ее от среднего стабильного значения, приводит к увеличению потерь мощности на величину, зависящую от коэффициентов расходной характеристики переделов при нелинейных членах и от величины отклонения мощности от ее среднего значения.
4. Потери энергии на переделах и технологической линии зависят не только от количества конечной энергии, коэффициента заполнения графика ее получения, но и от формы графика конечной мощности.
5. Условный минимум потерь энергии на отдельном переделе и технологической линии достигается при ее потреблении с постоянной мощностью.
6. Численное моделирование коммунально-бытовых нагрузок топливно-энергетического хозяйства региона показало, что наименьшие суммарные потери мощности технологической линии составляют свыше 93 % от первичной мощности, неравномерность графика конечной нагрузки приводит к увеличению потерь энергии на 36 % и потерь первичной мощности в 6 раз в сравнении со стабильным режимом потребления.
7. Потери мощности и энергии, вызванные неравномерностью графика конечной нагрузки, могут быть снижены путем его выравнивания.

## Литература


*Болотов В. В.* Теоретические основы выбора экономического режима электроэнергетической системы. М.; Л.: Изд-во АН СССР, 1947. 278 с.

*Борисов Г. А., Тихомирова Т. П.* Характеристики и свойства потерь энергии и мощности на переделах энергетического хозяйства региона // Труды Карельского научного центра Российской академии наук, 2010. № 3. С. 4–10.

*Гориков А. С.* Технико-экономические показатели электрических станций М.; Л.: ГЭИ, 1949. 287 с.

*Гуд Г. Х., Макол Р. Э.* Системотехника. Введение в проектирование больших систем. М.: Сов. Радио, 1962. 383 с.

*Железко Ю. С., Артемьев А. В., Савченко О. В.* Расчет, анализ и нормирование потерь энергии в электрических сетях. М.: Издательство НЦ ЭНАС, 2006. 280 с.

*Крон Г.* Исследование сложных систем по частям (диакоптика). М.: Наука, Главная редакция физико-математической литературы, 1972. 544 с.

*Макаров А. А., Мелентьев Л. А.* Методы исследования и оптимизации энергетического хозяйства. Новосибирск: Наука, 1973. 274 с.

*Первозванский А. А., Гайцгори В. Г.* Декомпозиция, агрегирование и приближенная оптимизация. М.: Наука, Главная редакция физико-математической литературы, 1979. 344 с.

*Сарач Б. М., Бастунский А. М.* Заводские и натурные испытания насосных агрегатов с преобразователями частоты // Электротехника. 1995. № 7. С. 19–20.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Борисов Георгий Александрович**
старший научный сотрудник, к. т. н.,
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: borisov@krc.karelia.ru
тел.: (8142) 766312

**Borisov, George**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: borisov@krc.karelia.ru
tel.: (8142) 766312

**Тихомирова Тамара Петровна**
ученый секретарь, к. т. н.,
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: tihomiro@krc.karelia.ru
тел.: (8142) 785520

**Tikhomirova, Tamara**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
e-mail: tihomiro@krc.karelia.ru
tel.: (8142) 785520

Труды Карельского научного центра РАН
№ 5. 2011. С. 18–27

*УДК 517.977.1*

# УСТОЙЧИВОСТЬ ИНТЕГРАЛЬНЫХ ОПЕРАТОРОВ НАБЛЮДЕНИЯ АНАЛИТИЧЕСКИХ СИСТЕМ

**Ю. В. Заика**

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассматривается задача определения фазового состояния нелинейной динамической системы по известной траекторной информации. В качестве операторов обработки измерений приняты линейные интегральные функционалы. Показано, что в аналитическом случае можно обеспечить устойчивость решения обратной задачи по отношению к малым вариациям весовых функций.

Ключевые слова: наблюдаемость, устойчивость, аналитические системы.

## Yu. V. Zaika. STABILITY OF INTEGRAL OBSERVABILITY OPERATORS OF ANALYTICAL SYSTEMS

The observability problem of determining the phase state of a nonlinear dynamic system from known information about the path is considered. Linear integral functionals are taken as operators of measurements processing. We show that in the analytic case one can ensure the stability of the inverse problem solution with respect to small variations of weight functions.

Key words: observability, stability, analytic systems.

---

## Критерии наблюдаемости нелинейных систем

### Постановка задачи

Рассмотрим в области $U \subseteq \mathbb{R}^n$ нелинейную систему наблюдения

$$\dot{x} = f(x),\ y = g(x), \quad g : U \to \mathbb{R}^m,\ m < n, \quad (1)$$

моделирующую закон движения и доступную информацию о движении. Вектор-функции $f$, $g$ считаем гладкими. Для доказательства основных результатов потребуется вещественная аналитичность: $f, g \in C^\omega(U)$. Задан промежуток наблюдения $[0,T]$ и область возможных конечных состояний $U_T = \{x(T)\} \subseteq U$. Решения $x(\,\cdot\,;x,T)$ $\big(x(T\,;x,T) = x \in U_T\big)$ продолжимы на отрезок $[0,T]$. Задача наблюдения состоит в определении по информации

$$y(\,\cdot\,;x,T) = g\big(x(\,\cdot\,;x,T)\big) : [0,T] \to \mathbb{R}^m$$

фазового вектора $x = x(T) \in U_T$. Запись $y(\,\cdot\,;x,T)$ означает, что известная на отрезке времени $[0,T]$ вектор-функция измерений $y(\cdot)$ однозначно определяется искомым неизвестным состоянием $x$ в момент $T$. Предполагается, что задачу необходимо решать систематически. Поэтому нас интересуют операции вычисления по любой возможной реализации $y(\cdot)$ соответствующего $x(T)$ из области $U_T$. Можно ставить задачу в терминах неизвестных начальных данных $x^0 = x(0) \in U_0$. Но обычно интересуются фазовым состоянием к моменту окончания наблюдения. В рамках модели по данным $x(T)$ $(x^0)$ можно численно восстановить решение и траекторию движения.

Установить наблюдаемость пары $(f,g)$ (биекцию $y(\cdot) \leftrightarrow x(T) \in U_{\scriptscriptstyle T})$ непосредственно по соответствию $x \mapsto y(\,\cdot\,;x,T)$ затруднительно, поскольку речь идет об обращении отображения в пространство вектор-функций. Поэтому обычно переходят к исследованию так называемого отображения наблюдаемости

$$H : x \mapsto y(\,\cdot\,;x,T) \mapsto z \in \mathbb{R}^p,$$

вычисляя значения $p$ функционалов на $y(\cdot)$. Наблюдаемость $(f,g)$ в множестве $V \subseteq U_{\scriptscriptstyle T}$ на отрезке времени $[0,T]$ означает биекцию $y(\,\cdot\,;x,T) \leftrightarrow x \in V$. Если $H$ инъективно на множестве $V \subseteq U_{\scriptscriptstyle T}$ $\big(H(x) \leftrightarrow x \in V\big)$, то $(f,g)$ наблюдаема в $V$ и вектор $x = x(T) \in V$ однозначно определяется по $z = H(x) \in H(V)$. Здесь значения $z$ известны после обработки измерений $y(\cdot)$ (вместе с дополнительной информацией $x_{\scriptscriptstyle T} \in V$, если строго $V \subset U_{\scriptscriptstyle T}$). Возможные способы построения $H$ $(p = \ell\, m)$:

$a)$ $x \mapsto \big(y'(T),\, \dot{y}'(T),\, \ldots,\, y^{(\ell-1)\prime}(T)\big)' \in \mathbb{R}^p;$

$b)$ $x \mapsto \big(y'(t_1),\, \ldots,\, y'(t_\ell)\big)',\ t_i \in [0,T];$

$c)$ $x \mapsto \big(\langle k_1, y\rangle,\, \ldots,\, \langle k_p, y\rangle\big)',\ \langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{L_2};$

$d)$ $x \mapsto \Big(\int_0^T k_1(\tau,y)\,d\tau, \ldots, \int_0^T k_p(\tau,y)\,d\tau\Big)'.$

Достаточные условия наблюдаемости обычно получают на основе анализа инъективности отображения $H \in C^1\big(U_{\scriptscriptstyle T}, \mathbb{R}^p\big)$ в подобласти $V \subseteq U_{\scriptscriptstyle T}$ при $p = n$ (см. [Никайдо, 1972; Ортега, Рейнболдт, 1975; Кирин, 1993]). При этом не только на $H$, но и на $V$ накладываются ограничения. При построении $H$ можно использовать как способы a) — d), так и их комбинации ($p$ функционалов вида $y_i^{(r)}(s)$, $y_j(t_k), \ldots$). При этом критерии 1)—4) можно применять не только к $H$, но и к $MH$, $\det M \neq 0$.

С целью упрощения обозначений и без существенного для дальнейшего изложения ограничения общности можно считать $m = 1$.

Приведем некоторые результаты аналитической теории наблюдения. Ее развитие можно проследить в серии статей К. Е. Старкова в журнале «Автоматика и телемеханика» (80—90-е гг.). Для полиномиальной пары $(f,g)$ при определении $x_{\scriptscriptstyle T} = x(T) \in U_{\scriptscriptstyle T}$ вместо $y(\cdot)$ достаточно ограничиться вычислением конечного числа производных $y^{(i)}(t_*)$, $t_* \in [0,T]$ [Inoye, 1977]. Но их количество $p = p(f,g,U_{\scriptscriptstyle T})$ может оказаться сколь угодно большим, хотя семейство $\big\{y(\,\cdot\,;x,T)\big|\, x \in U_{\scriptscriptstyle T}\big\}$ «всего лишь» $n$-параметрическое. Для стационарной наблюдаемой вещественной аналитической пары $(f,g)$ без потери информации об искомом $x_{\scriptscriptstyle T}$ вместо $y(\cdot)$ можно ограничиться набором $2n+1$ значений $y(t_j)$ [Козеренко, 1987]. Моменты времени $t_j$ фиксируются и не зависят от $y(\cdot)$. Но в общем случае множество «удачных» программ наблюдений $\{t_1, \ldots, t_{2n+1}\}$ не открыто в $[0,T]^{2n+1}$. С учетом погрешностей задания $t_j$ это может привести к потере наблюдаемости. Устойчивые к возмущениям дискретные программы рассмотрены в [Заика, 1999]. Интегральные операторы наблюдения исследуются в [Zaika, 2003]. Применение аналитической теории к простейшей модели движения центра масс летательного аппарата в вертикальной плоскости изложено в [Заика, 1988].

Если на измерение значений функции $y(t)$ существенное влияние оказывают различного рода неконтролируемые помехи, то предпочтительнее использовать интегральные операции обработки информации $y(t)$. Основы соответствующего математического аппарата в линейном случае изложены в книге [Красовский, 1968]. Напомним известный результат, который будет взят за основу обобщения. Пусть $f = Fx$, $g = Gx$, где $F$, $G$ — матрицы $n \times n$, $m \times n$. Если в сопряженной системе

$$\dot{V}(t) = -F'V(t) + G'k(t),\ V(0) = 0, \qquad (2)$$

построить управление $k(\cdot)$ из условия $V(T) = h$, то по $y(\cdot)$ вычисляется проекция неизвестного $x_{\scriptscriptstyle T} = x(T)$ на вектор $h$: $h'x_{\scriptscriptstyle T} = \langle k, y\rangle_{L_2}$ $\forall\, x_{\scriptscriptstyle T} \in \mathbb{R}^n$. Совокупность всех $h \in \mathbb{R}^n$, для которых по любой возможной реализации $y(\cdot)$ однозначно восстанавливается проекция $h'x_{\scriptscriptstyle T}$, описывается множеством достижимости $\mathcal{D}_{\scriptscriptstyle T} = \{V(T)\}$. Этот подход Н. Е. Кириным [Иванов, Кирин, 1988; Кирин, 1993] обобщен на нелинейный случай. Построение оператора восстановления по $y(\cdot)$ значений данной функции $\varphi : U_{\scriptscriptstyle T} \to \mathbb{R}$ в интегральной форме

$$\varphi(x_{\scriptscriptstyle T}) = \int_0^T k\big(\tau, y(\tau)\big)d\tau\ \ \forall\, x_{\scriptscriptstyle T} = x(T) \in U_{\scriptscriptstyle T} \qquad (3)$$

сводится к следующей задаче управления. В сопряженной системе

$$v_t(t,x) + v_x(t,x)\cdot f(x) = k\big(t, g(x)\big),\ v(0,x) = 0,$$

требуется выбрать функцию $k(\,\cdot\,,\cdot)$ из условия $v(T,x) = \varphi(x)$, $x \in U_{\scriptscriptstyle T}$. В линейном случае $(f,g) = (F,G)$, $k(t,y) = k'(t)y$ получаем $v(t,x) = V'(t)x$, где вектор-функция $V(t)$ удовлетворяет соотношениям (2). Нелинейная задача построения операции наблюдения для области фазового пространства является распределенной. Важно, что сопряженное уравнение линейное по паре функций $(k,v)$ и возможно применение теории управления и методов решения линейных граничных задач.

**Сведения из комплексного анализа**

*1. Аналитические подмножества.* Пусть $\Omega$ — область в $\mathbb{C}^n$. Множество $A \subseteq \Omega$ называется (комплексным) аналитическим подмножеством $\Omega$ [Чирка, 1985], если для каждой точки $a \in \Omega$ найдутся ее окрестность $U$ и голоморфные в ней функции $f_1, \ldots, f_N$, такие, что $A \cap U = \{z \in U | \, f_1(z) = \ldots = f_N(z) = 0\}$. Локально множество $A$ определяется общими нулями конечных наборов голоморфных функций и замкнуто в $\Omega$. В изложении [Эрве, 1965] такие $A$ называются аналитическими множествами в $\Omega$ (без приставки «под»). Можно оставить лишь требование открытости $\Omega$, не меняя определения. В [Чирка, 1985] понятие аналитического множества «занято» несколько иным объектом $(a \in A)$.

**Теорема 1** ([Чирка, 1985]). *Если $\{A_\alpha\}_{\alpha \in I}$ — произвольное семейство аналитических подмножеств $\Omega$, то $A = \cap_{\alpha \in I} A_\alpha$ — тоже аналитическое подмножество $\Omega$, причем для любого $K$ с компактным замыканием в $\Omega$ ($K \Subset \Omega$) найдется конечное подмножество $J \subseteq I$, такое, что $A \cap K = \left(\cap_{\alpha \in J} A_\alpha\right) \cap K$.*

*2. Ростки голоморфных функций.* Рассмотрим кольцо $I_n$ степенных рядов от $n$ комплексных переменных, сходящихся в заданной открытой окрестности $U$ фиксированной точки $z_0 \in \mathbb{C}^n$. Идеалом $J$ в $I_n$ называется всякая аддитивная подгруппа $I_n$, удовлетворяющая условию $I_n J = J$ $(a \in I_n,\ b \in J \Rightarrow ab \in J)$. Кольцо $I_n$, как известно, нётерово. Это означает, что в произвольном идеале $J \subseteq I_n$ найдутся такие элементы $b_1, \ldots, b_p$ (конечный базис $J$), что любой элемент $b \in J$ представим в виде линейной комбинации $b = \sum_{i=1}^{p} a_i b_i$, $a_i \in I_n$.

Если окрестность $U$ не фиксировать, то приходим к понятию кольца $\mathcal{H}_n$ ростков голоморфных функций в точке $z_0 \in \mathbb{C}^n$ [Эрве, 1965]. Приведем определения. Заданные и голоморфные в открытых окрестностях $U_1$, $U_2$ точки $z_0 \in \mathbb{C}^n$ функции $f_1$, $f_2$ эквивалентны, если в некоторой открытой окрестности $U_3 \subseteq U_1 \cap U_2$ точки $z_0$ они тождественны. Ростками голоморфных функций в точке $z_0$ называются классы эквивалентности функций, определенных и голоморфных в открытых множествах, содержащих $z_0$. Две функции, голоморфные в открытых окрестностях $z_0$, совпадают в некотором поликруге $P = \{z : |z_k - z_{0k}| < r_k \in \mathbb{R}_+ \backslash \{0\},\ 1 \leqslant k \leqslant n\}$ тогда и только тогда, когда они имеют одно и то же тейлоровское разложение в $z_0$. Поэтому $\mathcal{H}_n$ изоморфно кольцу сходящихся степенных рядов. Сходимость ряда означает, что существует открытый поликруг $P$, в котором ряд сходится. Росток, порожденный функцией $f$, обозначаем $\widehat{f}$. Сложение и умножение ростков определяются их представителями: $\widehat{f}_1 + \widehat{f}_2$ — росток, порожденный функцией $f_1 + f_2$, а росток $\widehat{f}_1 \cdot \widehat{f}_2$ порожден функцией $f_1 \cdot f_2$. Кольцо $\mathcal{H}_n$ нётерово. Справедливо следующее утверждение.

**Теорема 2** ([Эрве, 1965], с. 44). *Для любого идеала $J$ кольца $\mathcal{H}_n$ можно указать:*

(I) *в некоторой открытой окрестности $U$ точки $z_0$ конечное множество голоморфных функций $h_1, \ldots, h_r$ с ростками в $z_0$ из $J$;*

(II) (*базис векторов в пространстве $\mathbb{C}^n$ и*) *последовательность открытых поликругов $\{P_i, i \geqslant 1\}$ с центром в точке $z_0$ и радиусами, монотонно стремящимися к нулю $\left(\overline{P}_1 \subset U,\ r_k = r_k(i) > r_k(i+1) > 0\right)$;*

(III) *числа $\varrho_i > 0$, $i \geqslant 1$, со следующим свойством: для каждой голоморфной в поликруге $P_i$ функции $h$ с ростком $\widehat{h} \in J$ существуют такие голоморфные в $P_i$ функции $\alpha_1, \ldots, \alpha_r$, что в $P_i$ имеет место разложение*

$$h(z) = \sum\nolimits_{j=1}^{r} \alpha_j(z) h_j(z), \tag{4}$$

$$\|\alpha_j\|_{P_i} = \sup\nolimits_{z \in P_i} |\alpha_j(z)| \leqslant \varrho_i \, \|h\|_{P_i}\ \forall j.$$

При переходе к первоначальному базису $\mathbb{C}^n$ вместо поликругов $P_i$ получим последовательность окрестностей $U_i$ точки $z_0$.

**Следствие.** *Пусть $\mathcal{F}$ — семейства функций, голоморфных в открытой окрестности $U$ точки $z_0 \in \mathbb{C}^n$. Тогда можно указать (открытый) поликруг $P \subseteq U$ с центром в $z_0$ и набор функций $f_1, \ldots, f_N \in \mathcal{F}$, обладающие следующим свойством: для каждой $f \in \mathcal{F}$ существуют такие голоморфные в поликруге $P$ функции $\alpha_1, \ldots, \alpha_N$, что в $P$ справедливо представление $f = \alpha_1 f_1 + \ldots + \alpha_N f_N$.*

В поликруге $P$ множество общих нулей функций $f \in \mathcal{F}$ совпадет с множеством общих нулей конечного подмножества $\mathcal{F}$.

*3. Число определяющих функций.* Пусть $\{f_\alpha\}_{\alpha \in I}$ — семейство функций, голоморфных в открытом множестве $\Omega \subseteq \mathbb{C}^n$. Тогда множество их общих нулей $Z$ есть аналитическое подмножество в $\Omega$, причем существуют такие голоморфные в $\Omega$ функции $g_0, \ldots, g_n \in \mathcal{O}(\Omega)$, что множество их общих нулей совпадает с множеством $Z$ [Чирка, 1985](с. 54).

Схема доказательства следующая. Пусть $\Omega_i$ — связные компоненты $\Omega$, не принадлежащие $Z$, и $a_i \in \Omega_i \setminus Z$ — произвольно выбранные точки. Для каждого $i$ найдется функция $f_{\alpha_i}$, такая, что $f_{\alpha_i}(a_i) \neq 0$. Представим $\Omega$ в виде такого счетного объединения $\Omega = \cup D_j$, что $D_j$ — ограниченные открытые множества, $D_j \subset D_{j+1}$, $K_j = \overline{D}_j \subset \Omega$, и для каждого компакта $K \subset \Omega$ найдется $s$ из условия $K \subset K_s$. Подберем индукцией по $j$ числа $c_j$ так, чтобы выполнялись следующие два неравенства:

$$\left|c_j f_{\alpha_j}(z)\right| < 2^{-j} \;\forall\, z \in K_j,\; \forall\, i \leqslant j,$$

$$\left|\sum_{k=1}^{j} c_k f_{\alpha_k}(a_i)\right| > 2^{-1}\left|c_i f_{\alpha_i}(a_i)\right|. \qquad (5)$$

Ряд $\sum c_i f_{\alpha_i}$ равномерно на компактах сходится в $\Omega$ к голоморфной функции, которую обозначим $g_n$. По построению выполнено

$$g_n(a_i) \neq 0 \;\forall\, i \;\Rightarrow\; \dim(Z_{g_n} \cap \Omega_i) < n,$$

где множество $Z_{g_n}$ — нули $g_n$ в $\Omega$, $\Omega \not\subset Z$. Остальные функции $g_{n-1}, \ldots, g_0$ строятся в [Чирка, 1985] по индукции в форме (счетных) линейных комбинаций некоторых функций $f_\alpha$ аналогичным образом: $g_s\big|_Z \equiv 0$ и все неприводимые компоненты множества $Z_{g_n} \cap \ldots \cap Z_{g_s}$ размерности $\geqslant s$ в $\Omega$ принадлежат $Z$. Множество общих нулей $g_0, \ldots, g_n$ совпадает с $Z$.

**Наблюдение по проекциям в $L_2$**

Остановимся на линейных операторах (3):

$$\varphi(x_{\scriptscriptstyle T}) = \langle k, y\rangle_{L_2},\; x_{\scriptscriptstyle T} = x(T) \in U_{\scriptscriptstyle T},$$

где $L_2 = L_2[0,T]$, $m = 1$. Допустимые весовые функции $k(\cdot)$ обработки измерений $y(\cdot)$ считаем кусочно непрерывными на отрезке времени $[0,T]$. Функционалы $y(\cdot) \mapsto \langle k, y\rangle$ и сами числа (моменты) $\langle k, y\rangle$ будем называть проекциями. С вычислительной точки зрения важно иметь конечномерное представление функции $y(\cdot)$. Можно ли подобрать такие $k_1(\cdot), \ldots, k_p(\cdot)$, чтобы сужение $y(\cdot)$ до значений конечного числа функционалов $J_i[y(\cdot)] = \langle k_i, y\rangle_{L_2}$ не приводило к потере информации об искомом $x(T)$? Имеется в виду биективное соответствие

$$y(\,\cdot\,; x_{\scriptscriptstyle T}, T) \leftrightarrow \big(J_1[y(\cdot)], \ldots, J_p[y(\cdot)]\big),\; x_{\scriptscriptstyle T} \in U_{\scriptscriptstyle T}.$$

В случае успеха «запоминание» $y(\cdot)$ сводится к интегрированию функций $k_i(t)y(t)$ по мере поступления измерений $y(t)$, что сравнительно легко осуществляется техническими средствами. Иной акцент вопроса: возможна ли ситуация, когда пара $(f,g)$ наблюдаема $\big($инъективно отображение $x_{\scriptscriptstyle T} \mapsto y(\cdot)\big)$, но по конечному числу проекций $\langle k_i, y\rangle$ однозначно определять $x_{\scriptscriptstyle T}$ невозможно? Здесь функции $k_i(\cdot)$, $1 \leqslant i \leqslant p$, фиксируются одни и те же для всех возможных $y(\cdot)$, $x_{\scriptscriptstyle T} \in U_{\scriptscriptstyle T}$. Если указанные наборы $k_i(\cdot)$ существуют, как выбрать по возможности минимальным $p$? Пусть $k(\cdot)$ фиксирована. Цепочка $x_{\scriptscriptstyle T} \mapsto y(\cdot) \mapsto \langle k, y\rangle$ порождает функцию $\varphi(x_{\scriptscriptstyle T}) = \langle k, y\rangle$. Как дать аналитическое описание $\varphi(\cdot)$? Важен и в определенном смысле обратный вопрос. Обычно измеряется часть фазовых координат, требуется лишь восстанавливать оставшиеся или, более общо, значения заданных функций $\varphi(x_{\scriptscriptstyle T})$. Как для заданной функции $\varphi$ подобрать $k$, чтобы выполнялось представление (3) с требуемой точностью? Изложим некоторые результаты в случае аналитичности $f, g$ по $x$.

**Определение 1.** Функцию $\varphi\colon U_{\scriptscriptstyle T} \to \mathbb{R}$ назовем наблюдаемой в множестве $M \subseteq U_{\scriptscriptstyle T}$, если существует функционал $\Lambda$ из условия $\varphi(x) = \Lambda[y(\cdot)]$, где $x = x_{\scriptscriptstyle T} \in M$, $y(\cdot) = y(\,\cdot\,; x_{\scriptscriptstyle T}, T)$.

Такие функции $\varphi$ будем также называть *наблюдаемыми компонентами* пары $(f,g)$. Наблюдаемость $\varphi$ в $M$ означает, что ее значения $\varphi(x)$ на неизвестном априори фазовом векторе $x = x(T)$ однозначно восстанавливаются по доступной в результате измерений информации $y(\,\cdot\,; x, T)$, если дополнительно известно включение $x \in M$. Наблюдаемость пары $(f,g)$ эквивалентна наблюдаемости всех координат $\varphi(x) = x_i$, $1 \leqslant i \leqslant n$, в области $U_{\scriptscriptstyle T}$. Когда исследуется наблюдаемость функции $\varphi$ в $M$ и $\varphi$ задана лишь на подмножестве $N$ $\big(M \subseteq N \subseteq U_{\scriptscriptstyle T}\big)$, то считаем ее доопределенной в $U_{\scriptscriptstyle T} \setminus N$ произвольно. Обозначим через $\Phi(M)$ множество всех наблюдаемых в $M$ функций $\varphi$. Очевидно, $\Phi(N) \subseteq \Phi(M)$ при $M \subseteq N$.

**Определение 2.** Базисом множества $\Phi(M)$ наблюдаемых в множестве $M$ функций назовем такую конечную совокупность $\varphi_i \in \Phi(M)$, $1 \leqslant i \leqslant p$, что имеет место функциональная зависимость: $\forall\, \varphi \in \Phi(M)$, $\forall\, x \in M$,

$$\varphi(x) = F_\varphi[\varphi_1(x), \ldots, \varphi_p(x)].$$

В записи $F_\varphi$ индекс отражает зависимость функции $F$ от $\varphi$. Множество $\Phi(M)$ является нелинейной (функциональной) оболочкой базисных наблюдаемых функций. Вычислив по измерениям $y(\,\cdot\,; x, T)$ $(x \in M)$ значения $\varphi_1(x), \ldots, \varphi_p(x)$, дополнительной информации о неизвестном векторе $x = x(T)$ из $y(\cdot)$ уже извлечь невозможно. Наблюдаемость $(f,g)$ в $M \subseteq U_{\scriptscriptstyle T}$ означает биекцию $\big(\varphi_1(x), \ldots, \varphi_p(x)\big) \leftrightarrow x \in M$. Если последним свойством обладает один из базисов, то это же справедливо и для любого другого (при условии их существования). Действительно, пусть $\Lambda_i$ — функционалы, соответствующие базисным $\varphi_i \in \Phi(M)$ согласно определению 1, $\{k_i,\, i \geqslant 1\}$ — полная в $L_2 = L_2[0,T]$ система, т. е. $\bigl\{\langle\varphi, k_i\rangle,\, i \geqslant 1\bigr\} \leftrightarrow \varphi(\cdot) \in L_2$. Тогда

$$\psi_i \in \Phi(U_{\scriptscriptstyle T}) \subseteq \Phi(M),\ \ \psi_i(x) \equiv \bigl\langle k_i,\, y(\,\cdot\,; x, T)\bigr\rangle.$$

Знак тождества $\equiv$ используем также в смысле равенства по определению в зависимости от контекста. Базисность означает, что

$$\psi_i(x) = F_i[\varphi_1(x), \ldots, \varphi_p(x)]\ \forall\, i \geqslant 1,\, \forall\, x \in M.$$

Поэтому по значениям $\varphi_i(x) = \Lambda_i[y(\,\cdot\,; x, T)]$, $1 \leqslant i \leqslant p$, числа $\psi_i(x)$, $i \geqslant 1$, определяются однозначно. В силу полноты $\{k_i,\, i \geqslant 1\}$ имеем

$$\bigl(\varphi_1(x), \ldots, \varphi_p(x)\bigr) \leftrightarrow y(\,\cdot\,; x, T),\ \ x \in M,$$

и вместо функций $y(\cdot)$ можно оперировать векторами $\bigl(\varphi_1(x), \ldots, \varphi_p(x)\bigr)$, $x \in M$. Эти же функции $\varphi_i$ образуют базис $\Phi(N)\ \forall\, N \subseteq M$.

Функционалы $\Lambda_i$ в определении 1 могут быть различной природы, в частности, $\Lambda_i[y(\cdot)] = y(t_i)$, $\Lambda_i[y(\cdot)] = y^{(i)}(t_*)$. Ограничимся классом линейных интегральных операций обработки измерений $\Lambda[y(\cdot)] = \langle k, y\rangle$. Очевидно, функции $\psi(x) = \bigl\langle k, y(\,\cdot\,; x, T)\bigr\rangle$ наблюдаемы в любом подмножестве области $U_{\scriptscriptstyle T}$, т. е. $\psi \in \Phi(M)\ \forall M \subseteq U_{\scriptscriptstyle T}$. При необходимости класс допустимых весовых функций $k(\cdot)$ можно расширить до пространства $L_2$.

**Теорема 3.** *Пусть система наблюдения $(f, g)$ вещественная аналитическая: $f \in C^{\omega}(U, \mathbb{R}^n)$, $g \in C^{\omega}(U, \mathbb{R})$. Тогда для любого множества $M$ с компактным замыканием в области $U_{\scriptscriptstyle T}$ ($M \Subset U_{\scriptscriptstyle T}$) из произвольной полной в $L_2[0,T]$ системы допустимых весовых функций $\bigl\{k_i,\, i \geqslant 1\bigr\}$ можно выделить такие $k_{i_\nu}(\cdot)$ ($1 \leqslant \nu \leqslant p$), что базис множества $\Phi(M)$ образуют компоненты*

$$\varphi_\nu\colon U_{\scriptscriptstyle T} \to \mathbb{R},\ \ \varphi_\nu(x) \equiv \bigl\langle k_{i_\nu}, y(\,\cdot\,; x, T)\bigr\rangle.$$

*Доказательство.* Определим в $U_{\scriptscriptstyle T} \times U_{\scriptscriptstyle T}$

$$\Delta\psi_i\bigl(x^1, x^2\bigr) = \psi_i\bigl(x^1\bigr) - \psi_i\bigl(x^2\bigr)$$
$$= \bigl\langle k_i, y\bigl(\cdot\,; x^1, T\bigr) - y\bigl(\cdot\,; x^2, T\bigr)\bigr\rangle,\ \ x^j \in U_{\scriptscriptstyle T}.$$

Включение $\psi \in C^{\omega}(U, \mathbb{R})$ означает, что функция $\psi$ в некоторой окрестности каждой точки области $U$ представима сходящимся степенным рядом. По теореме Пуанкаре решение $x(t; x, T)$ аналитически зависит от начальных данных $x = x(T)$. Поэтому (см. [Эрве, 1965], с. 14) функции $\Delta\psi_i$ голоморфны в $U_{\scriptscriptstyle T} \times U_{\scriptscriptstyle T}$. В силу леммы Абеля о сходимости степенных рядов можем считать, что $\Delta\psi_i$ заданы и голоморфны в $W = U^c_{\scriptscriptstyle T} \times U^c_{\scriptscriptstyle T} \subseteq \mathbb{C}^{2n}$, где комплексная область $U^c_{\scriptscriptstyle T}$ является достаточно малой окрестностью области $U_{\scriptscriptstyle T}$ в $\mathbb{C}^n$. Это продолжение можно задать формулой

$$\Delta\psi_i\bigl(z^1, z^2\bigr) = \bigl\langle k_i, y\bigl(\cdot\,; z^1, T\bigr) - y\bigl(\cdot\,; z^2, T\bigr)\bigr\rangle,$$

$z^j \in U^c_{\scriptscriptstyle T}$. Смысл записи $y(\,\cdot\,; z, T)$ ($z \in U^c_{\scriptscriptstyle T}$) сохраняется, поскольку решения дифференциального уравнения $\dot{x} = f(x)$ можно рассматривать и при комплексных условиях Коши $x(T) = z \in U^c_{\scriptscriptstyle T} \subseteq \mathbb{C}^n$. Продолжимость таких решений на отрезок времени $[0, T]$ гарантируется для достаточно малой комплексной окрестности $U^c_{\scriptscriptstyle T}$ исходной области $U_{\scriptscriptstyle T} \subseteq \mathbb{R}^n$.

Обозначим через $Z_i$ множество нулей функции $\Delta\psi_i$ в области $W$. Тогда общие нули $\Delta\psi_i$ образуют аналитическое подмножество $Z = \cap_{j=1}^{\infty} Z_j$ области $W$ и существуют такие номера $i_1, \ldots, i_p$, что

$$Z \cap (M \times M) = \bigl(\cap_{\nu=1}^{p} Z_{i_\nu}\bigr) \cap (M \times M).$$

Из $\Delta\psi_{i_\nu}\bigl(x^1, x^2\bigr) = 0$ для $1 \leqslant \nu \leqslant p$, $x^j \in M$ следует $\Delta\psi_i\bigl(x^1, x^2\bigr) = 0$, $i \geqslant 1$, и в силу полноты системы $\bigl\{k_i,\, i \geqslant 1\bigr\}$ получаем равенство $y\bigl(\cdot\,; x^1, T\bigr) = y\bigl(\cdot\,; x^2, T\bigr)$. Отсюда $\forall\, x \in M$

$$\bigl(\varphi_1(x), \ldots, \varphi_p(x)\bigr) =$$
$$= \bigl(\langle k_{i_1}, y\rangle, \ldots, \langle k_{i_p}, y\rangle\bigr) \leftrightarrow y(\,\cdot\,; x, T),$$
$$\varphi(x) = \Lambda[y(\,\cdot\,; x, T)] = F_\varphi[\varphi_1(x), \ldots, \varphi_p(x)],$$

$\forall\, \varphi \in \Phi(M)$. Согласно определению функции $\varphi_i$ образуют конечный базис $\Phi(M)$. $\square$

Проблему поиска базиса наблюдаемых компонент пары $(f, g)$ можно сформулировать в алгебраических терминах. Рассмотрим в кольце $\mathcal{O}(W)$ голоморфных в $W = U^c_{\scriptscriptstyle T} \times U^c_{\scriptscriptstyle T} \subseteq \mathbb{C}^n$ функций идеал, порожденный $\{\Delta\psi_i,\, i \geqslant 1\}$. Элементы этого идеала — конечные линейные комбинации функций $\Delta\psi_i$ с коэффициентами из кольца $\mathcal{O}(W)$. Конечный базис этого идеала (когда он существует) и определяет номера базисных проекций $\langle k_i, y\rangle$ для множества $M = U_{\scriptscriptstyle T}$ ($\Rightarrow \forall M \subseteq U_{\scriptscriptstyle T}$). В частности ([Эрве, 1965], с. 50), $\forall\, \bar{x} \in U_{\scriptscriptstyle T}$ существует окрестность

$$P_\varepsilon = \bigl\{z \in \mathbb{C}^n\colon \|z - \bar{x}\| = \max_i |z_i - \bar{x}_i| < \varepsilon\bigr\} \subseteq U^c_{\scriptscriptstyle T}$$

$\bigl(P_\varepsilon \cap \mathbb{R}^n \subseteq U_{\scriptscriptstyle T}\bigr)$ и конечный набор функций $\Delta\psi_{i_1}, \ldots, \Delta\psi_{i_q}$ из условий: $\bigl(z^1, z^2\bigr) \in P_\varepsilon \times P_\varepsilon$,

$$\Delta\psi_j\bigl(z^1, z^2\bigr) = \sum\nolimits_{\nu=1}^{q} \alpha_{j\nu}\bigl(z^1, z^2\bigr)\Delta\psi_{i_\nu}\bigl(z^1, z^2\bigr),$$

$j \geqslant 1$, $\alpha_{j\nu} \in \mathcal{O}\big(P_\varepsilon \times P_\varepsilon\big)$. Тогда справедливо

$$\Delta\psi_{i_\nu}\big(x^1, x^2\big) = 0,\ 1 \leqslant \nu \leqslant q,\ x^j \in M = P_\varepsilon \cap U_{\scriptscriptstyle T}$$

$$\Rightarrow \Delta\psi_i\big(x^1, x^2\big) = 0,\ i \geqslant 1 \Rightarrow y(\cdot\,; x^1) = y(\cdot\,; x^2).$$

Базисом множества $\Phi(M)$ будут функции $\varphi_\nu(x) = \psi_{i_\nu}(x) = \big\langle k_{i_\nu}, y(\,\cdot\,; x, T)\big\rangle$:

$$\big(\varphi_1(x), \ldots, \varphi_q(x)\big) \leftrightarrow y(\,\cdot\,; x, T), \quad x \in M.$$

Если нет ограничений на весовые функции $\big($в смысле включения $k(\cdot) \in \{k_i\}\big)$, то результат можно усилить $\big(M = U_{\scriptscriptstyle T},\ p = 2n + 1\big)$.

**Теорема 4.** *Пусть* $\big\{k_i,\ i \geqslant 1\big\}$ *— полная в* $L_2[0, T]$ *система непрерывных функций,* $f \in C^\omega(U, \mathbb{R}^n)$, $g \in C^\omega(U, \mathbb{R})$. *Тогда существует семейство наборов из* $2n + 1$ *функций* $\{r_i(\cdot)\}$, *для которых наблюдаемые компоненты*

$$\varphi_i(x) = \big\langle r_i, y(\,\cdot\,; x, T)\big\rangle,\ 0 \leqslant i \leqslant 2n,$$

*образуют базис* $\Phi(U_{\scriptscriptstyle T})$ $\big($*и* $\Phi(M)\,\forall M \subseteq U_{\scriptscriptstyle T}\big)$. *Каждая* $r_j(\cdot)$ *представима равномерно сходящимся на* $[0, T]$ *рядом по элементам* $\{k_i\}$.

*Доказательство.* Рассмотрим голоморфные в области $\Omega = W = U^c_{\scriptscriptstyle T} \times U^c_{\scriptscriptstyle T} \subseteq \mathbb{C}^{2n}$ функции

$$\Delta\psi_i : W \to \mathbb{C}, \quad \Delta\psi_i\big(z^1, z^2\big) = \psi_i\big(z^1\big) - \psi_i\big(z^2\big).$$

Они определяются аналитическим продолжением функций $\psi_i(x)$ из области $U_{\scriptscriptstyle T}$ в достаточно малую окрестность (область) $U^c_{\scriptscriptstyle T} \supset U_{\scriptscriptstyle T}$ в $\mathbb{C}^n$:

$$\psi_i(z) = \big\langle k_i,\, y(\,\cdot\,; z, T)\big\rangle,\ z = x(T) \in U^c_{\scriptscriptstyle T}.$$

Коррекция составления линейных комбинаций состоит в том, что коэффициенты $c_j$ (см. неравенства (5)) будем подбирать из условия

$$\begin{aligned}
&\big|\, c_j\, \Delta\psi_{\alpha_j}\big(z^1, z^2\big)\big| \\
&= \big|\big\langle\, c_j\, k_{\alpha_j},\, y\big(\cdot\,; z^1, T\big) - y\big(\cdot\,; z^2, T\big)\big\rangle\big| \\
&\leqslant \big\|\, c_j\, k_{\alpha_j}\big\|_{\scriptscriptstyle C} \cdot \big\|\, y\big(\cdot\,; z^1, T\big) - y\big(\cdot\,; z^2, T\big)\big\|_{L_1} \\
&< 2^{-j}\ \forall\, \big(z^1, z^2\big) \in K_j,
\end{aligned}$$

сохраняя при этом второе для $c_j$ неравенство в (5). Эта модификация обеспечит не только сходимость ряда $\sum c_i \Delta\psi_{\alpha_i}$ в области $W$ к голоморфной функции, но и сходимость ряда $\sum c_i k_{\alpha_i}$ в $C[0, T]$. Используя такие построения по индукции и обозначая суммы рядов через $r_{2n}, \ldots, r_0$, приходим к следующему утверждению. Множество общих нулей функций

$$q_i\big(z^1, z^2\big) = \big\langle r_i,\, y\big(\cdot\,; z^1, T\big) - y\big(\cdot\,; z^2, T\big)\big\rangle,$$

$0 \leqslant i \leqslant 2n$, в области $W$ совпадает с $Z = \cap_{j=1}^{\infty} Z_j$ $\big(Z_j$ — нули $\Delta\psi_j$ в $W\big)$. В силу полноты системы $\big\{k_i,\ i \geqslant 1\big\}$ любые неравные на отрезке $[0, T]$ функции $y\big(\cdot\,; x^1, T\big) \neq y\big(\cdot\,; x^2, T\big)$, $x^j \in U_{\scriptscriptstyle T}$, имеют различный набор проекций:

$$\big\{\big\langle r_i,\, y\big(\cdot\,; x^1, T\big)\big\rangle\big\} \neq \big\{\big\langle r_i,\, y\big(\cdot\,; x^2, T\big)\big\rangle\big\}.$$

Из взаимно однозначного соответствия

$$y(\,\cdot\,; x, T) \leftrightarrow \big(\varphi_0(x), \ldots, \varphi_{2n}(x)\big),\ x \in U_{\scriptscriptstyle T},$$

$\varphi_i(x) = \big\langle r_i,\, y(\,\cdot\,; x, T)\big\rangle$, следует базисность набора функций $\varphi_i$ в множестве $\Phi(U_{\scriptscriptstyle T})$. Удачных наборов $\{r_j\}$ бесконечно много: имеется определенный произвол в выборе систем $\{k_i\}$ и коэффициентов рядов для функций $r_j$. □

Если брать $k_i(t) = t^i$, то можно построить $r_j(t)$ вещественными аналитическими в форме степенного ряда. Допустимо использование и разрывных $k_i(t)$, если доказывать сходимость рядов для $r_j(t)$ в пространстве $L_2$.

Наблюдаемость пары $(f, g)$ в $M \subseteq U_{\scriptscriptstyle T}$ $\big(y(\,\cdot\,; x, T) \leftrightarrow x \in M\big)$ характеризуется тем, что для полной в $L_2$ $\big($или хотя бы в $\mathcal{Y} = \{\, y(\cdot)\}\big)$ системы $\{k_i,\ i \geqslant 1\}$ множество общих нулей функций $\Delta\psi_i\big(x^1, x^2\big)$ в $M \times M$ совпадает с диагональю $\{(x, x) |\, x \in M\}$. Для базисных $r_j(\cdot)$ вектор $x_{\scriptscriptstyle T} \in M$ однозначно определяется по набору $2n + 1$ проекций $\mu_i = \langle r_i, y\rangle$. Если $(f, g)$ не является полностью наблюдаемой, поиск базиса $\Phi(M)$ $(M \subseteq U_{\scriptscriptstyle T})$ сохраняет прикладное значение, поскольку для «отслеживания» значений $u(x(t))$ заданной функции $u(x)$ по предыстории измерений на $[t - T, t]$ нет необходимости в промежуточном восстановлении полного фазового вектора $x(t)$. Если какая-либо из базисных весовых функций $k(\cdot)$ не удовлетворяет ограничениям реализации $|k(t)| \leqslant \overline{k} = \mathrm{const}$, то вместо нее следует взять $\alpha k(\cdot)$ с малым множителем $\alpha$. Важны лишь «проекции на направления».

## Устойчивость интегральных операторов наблюдения

Прежде чем переходить к уточнению понятия устойчивости операторов наблюдения, исследуем аналитическую структуру элементов множества достижимости $\mathcal{D}_{\scriptscriptstyle T} = \{v(T, \cdot)\}$ линейной сопряженной системы управления. В стационарном линейном случае, когда $\dot{x} = Fx$, $y = Gx$, имеем $\mathcal{D}_{\scriptscriptstyle T} = \{V(T)\} = \mathcal{L}(\mathcal{K})$, где $\mathcal{L}$ — линейная оболочка столбцов матрицы управляемости $\mathcal{K} = \big(G', F'G', \ldots, F'^{n-1}G'\big)$. Попытаемся найти аналог такого конечного описания в общем (бесконечномерном) случае.

**Локальное представление элементов $\mathcal{D}_{T}$**

Рассмотрим вещественную аналитическую систему наблюдения $(f,g)$: $x \in U \subseteq \mathbb{R}^n$,

$$\dot{x} = f(x),\ y = g(x), \quad y(\cdot)\colon [0,T] \to \mathbb{R}.$$

С целью упрощения обозначений пишем $f, g \in C^{\omega}(U)$ и считаем $m = 1$, т. е. остановимся на анализе наблюдаемости по скалярным измерениям на отрезке $[0,T]$. Это сужение задачи в дальнейшем непринципиально. Кроме того, ограничимся линейными весовыми функциями $k(t,y) = k(t)y$. Допустимы $k(\cdot) \in C[0,T]$ (при необходимости $k(\cdot) \in KC$).

Перейдем к задаче представления элементов множества достижимости

$$\mathcal{D}_{T} = \big\{v(T,\cdot)\colon U_{T} \to \mathbb{R},$$
$$k(t,y) = k(t)y,\ v(T,x) = \big\langle k, y(\,\cdot\,; x, T)\big\rangle\big\}$$

в форме линейных комбинаций конечного числа функций $L_f^i g$, где $L_f^0 g(x) = g(x)$,

$$L_f^{i+1} g(x) = \partial_x\big(L_f^i g(x)\big)\cdot f(x),\ x \in U.$$

В приложениях обычно компоненты вектор-функций $f$, $g$ являются суперпозициями элементарных функций, тогда и $L_f^i g$ таковые. В операторных терминах производные

$$L_f^i g = \mathcal{A}^i B\ \big(B = g,\ \mathcal{A} = \partial_x(\cdot) f\big)$$

представляют аналог столбцов матрицы управляемости: $(f,g) = (F,G) \Rightarrow L_f^i g(x) = GF^i x$, $F'^{j-1}G'$ – $j$-й столбец $\mathcal{K}$. Производные выхода $y^{(i)}(t)$ равны $L_f^i g\big(x(t)\big)$. Теоретически удобно исследовать разрешимость системы уравнений $L_f^i g(x) = y^{(i)}(T)$ , $0 \leqslant i \leqslant n-1$, относительно $x = x_{T}$ в области $U_{T}$. Но последовательное дифференцирование измерений $y(t)$ практически неприемлемо. В этом контексте интегральные операторы корректны: каждая операция производится независимо от другой и происходит сглаживание измерений.

Последующие построения носят локальный характер. Фиксируем произвольную точку $\bar{x} \in U_{T}$ и достаточно малый куб $\overline{\Pi} \subset U_{T}$,

$$\Pi = \big\{x \in \mathbb{R}^n\colon \|x-\bar{x}\| = \max|x_i-\bar{x}_i| < \delta\big\} \subset U_{T}.$$

В силу теоремы единственности для вещественных аналитических функций выполнено $y|_{[0,T]} \leftrightarrow y|_{[t_1,t_2]}$, $0 \leqslant t_1 < t_2 \leqslant T$. Поэтому для теоретического анализа наблюдаемости без существенного ограничения общности можно считать отрезок наблюдения таким, что функция $y(t;x,T)$ на множестве $\{(t,x)\} = (-\varepsilon, T+\varepsilon) \times \Pi$, $\varepsilon > 0$, разлагается в ряд по степеням $(t-T)$ и компонент вектора $x - \bar{x}$. Степенные ряды по $x - \bar{x}$ в кубе $\Pi$ для производных и элементов множества достижимости

$$L_i(x) = L_f^i g(x) = y^{(i)}(T; x, T),\ w(x) = v(T,x)$$

определяют такие голоморфные функции $L_i^c : P \to \mathbb{C}$, $w^c : P \to \mathbb{C}$, что

$$L_i^c\big|_{\Pi} = L_i\big|_{\Pi},\ w^c\big|_{\Pi} = w\big|_{\Pi},$$

$P = \big\{z \in \mathbb{C}^n\colon \|z - \bar{x}\| = \max|z_i - \bar{x}_i| < \delta\big\}$. По непрерывности $L_i^c$, $w^c$ продолжаются на замыкание $\overline{P}$ (иначе уменьшим значение $\delta > 0$).

Рассмотрим идеал $J$ в кольце $\mathcal{H}_n$ ростков голоморфных функций в точке $\bar{x}$, порожденный $\{L_i^c : P \to \mathbb{C},\ i \geqslant 0\}$. Элементы $J$ — конечные линейные комбинации ростков $\hat{L}_j^c$ с коэффициентами из $\mathcal{H}_n$. Пусть окрестности $Q_i$ точки $z_0 = \bar{x}$ (поликруги $P_i$ в подходящем базисе $\mathbb{C}^n$), базисные функции $h_1, \ldots, h_r$, константы $\varrho_i$, $i \geqslant 1$, выбраны согласно теореме 2.

Фиксируем номера $s \geqslant 1$, $p \geqslant 1$ из условий $\overline{Q}_s \subset P$ и справедливости в достаточно малой окрестности $Q_s$ $\forall j \geqslant p$ представлений

$$L_j^c(z) = \sum\nolimits_{\nu=0}^{p-1} \beta_{j\nu}(z) L_\nu^c(z), \tag{6}$$
$$\|\beta_{j\nu}\|_{Q_s} \leqslant \varrho\,\|L_j^c\|_{Q_s},\ \beta_{j\nu} \in \mathcal{O}\big(Q_s, \mathbb{C}\big),$$

$0 \leqslant \nu \leqslant p-1$, $j \geqslant p$, $\varrho > 0$. Это возможно, поскольку все функции $L_j^c$ в окрестностях $Q_i$, $i \geqslant 1$, являются линейными комбинациями голоморфных функций $h_1, \ldots, h_r$. Последние, в свою очередь, в некоторой окрестности точки $\bar{x}$ представимы комбинациями конечного числа $L_\nu^c$ по определению идеала $J$ (коэффициенты — голоморфные функции). Существование константы $\varrho$, независящей от номеров $\nu$ и $j$, следует из оценок в теореме 2.

Итак, без существенного ограничения общности полагаем, что на множестве $\{(t,x)\} = (-\varepsilon, T+\varepsilon) \times \Pi$ $(\varepsilon > 0)$ $y(t;x,T) =$

$$= \sum_{j=0}^{\infty} \frac{(t-T)^j}{j\,!}\, y^{(j)}(T) = \sum_{j=0}^{\infty} \frac{(t-T)^j}{j\,!}\, L_j(x)$$
$$= \sum_{j,i_1,\ldots,i_n \geqslant 0} b_{j,i_1,\ldots,i_n}\cdot(t-T)^j \prod_{\nu=1}^{n}(x_\nu - \bar{x}_\nu)^{i_\nu}.$$

Тогда для $w(x) = v(T,x) = \big\langle k, y(\,\cdot\,; x, T)\big\rangle$ при $x \in \Pi$ справедливо представление

$$w(x) = \sum\nolimits_{j=0}^{\infty} c_j L_j(x), \quad c_j = \big\langle k, (\tau - T)^j\big\rangle / j\,!.$$

По теореме Абеля о сходимости степенных рядов из полученных разложений следует, что в открытом поликруге $P$ определены функции

$$w^c(z) = \sum\nolimits_{j=0}^{\infty} c_j L_j^c(z),\ z \in P,$$

$w^c\big|_\Pi = w\big|_\Pi,\ L_j^c\big|_\Pi = L_j\big|_\Pi$. Используя представление (6), получим в окрестности $Q_s$

$$w^c(z) = \sum_{j=0}^{p-1} c_j L_j^c + c_p \sum_{\nu=0}^{p-1} \beta_{p\nu} L_\nu^c + \ldots, \qquad (7)$$

а при вещественных значениях аргумента:

$$\eta_{i\nu}(x) \equiv \operatorname{Re} \beta_{i\nu}(x),\ w(x) = \sum_{j=0}^{p-1} c_j L_j(x)$$

$$+ c_p \sum_{\nu=0}^{p-1} \eta_{p\nu}(x) L_\nu + c_{p+1} \sum_{\nu=0}^{p-1} \eta_{p+1,\nu}(x) L_\nu + \ldots.$$

Вещественная часть голоморфной функции $\beta_{i\nu}(x)$ является вещественной аналитической в пересечении $\mathcal{O}_s \cap \mathbb{R}^n \subset \Pi$.

**Лемма 1.** *Ряд в представлении* (7) *сходится абсолютно и равномерно в* $Q_s$.

*Доказательство.* Достаточно доказать абсолютную и равномерную в области $Q_s$ сходимость рядов $c_\nu + c_p \beta_{p\nu}(z) + c_{p+1}\beta_{p+1,\nu}(z) + \ldots$, конечной линейной комбинацией которых и получается разложение (7). Сходящийся степенной ряд для $y(t;x,T)$ определяет голоморфную функцию $\eta(\zeta, z)$, где $|\zeta - T| < T + \varepsilon$, $z \in P$, $\eta\big|_A = y\big|_A$, $A \equiv (-\varepsilon, 2T + \varepsilon) \times \Pi$. Для каждого $z \in P$ в силу неравенств Коши

$$\frac{\left|L_j^c(z)\right|}{j!} \leqslant \max_\zeta \frac{|\,\eta(\zeta,z)|}{\widetilde{T}^j},\ \ |\zeta - T| = \widetilde{T},\ T < \widetilde{T},$$

$$\widetilde{T} < T + \varepsilon,\ \left\|L_j^c\right\|_P \leqslant j! L \widetilde{T}^{-j},\ L \equiv \sup_{\zeta,z} |\,\eta(\zeta,z)|,$$

$|\zeta - T| = \widetilde{T}$, $z \in P$, где $L < +\infty$ в силу $\overline{\Pi} \subset U_{\scriptscriptstyle T}$, $\overline{P} \subset U_{\scriptscriptstyle T}^c$ ($\delta$ достаточно мало). С учетом

$$|c_i|\, i! = \left| \int_0^T k(\tau)(\tau - T)^i\, d\tau \right| \leqslant T^i \ell,\ \ell = \mathrm{const},$$

в окрестности $Q_s$ получаем последовательность оценок: $q = T\widetilde{T}^{-1} < 1$,

$$|c_\nu| + |c_p|\,|\beta_{p\nu}| + |c_{p+1}|\,|\beta_{p+1,\nu}(z)| + \ldots$$

$$\leqslant |c_\nu| + |c_p|\,\|\beta_{p\nu}\|_{Q_s} + |c_{p+1}|\,\|\beta_{p+1,\nu}\|_{Q_s} + \ldots$$

$$\leqslant |c_\nu| + \varrho\,|c_p|\,\left\|L_p^c\right\|_P + \varrho\,|c_{p+1}|\,\left\|L_{p+1}^c\right\|_P + \ldots$$

$$\leqslant |c_\nu| + \varrho \ell L q^p + \varrho \ell L q^{p+1} + \ldots.$$

□

Заменим при $t \in (-\varepsilon, T + \varepsilon)$, $x \in Q_s \cap \mathbb{R}^n$ в представлении $y(t;x,T) =$

$$= L_0(x) + (t - T)L_1(x) + 0.5(t - T)^2 L_2(x) + \ldots$$

производные $L_j$, $j \geqslant p$, линейными комбинациями согласно разложениям (6) и «соберем коэффициенты» при функциях $L_0, \ldots, L_{p-1}$:

$$y(t;x,T) = \sum_{i=0}^{p-1} \gamma_i(t,x) L_i(x),\ L_i \equiv L_f^i g, \qquad (8)$$

$$v(T,x) = \sum_{i=0}^{p-1} \sigma_i(x) L_i(x),\ \sigma_i(x) \equiv \big\langle k, \gamma_i(\cdot, x)\big\rangle,$$

$x \in Q_s \cap \mathbb{R}^n$, $t \in (-\varepsilon_0, T + \varepsilon_0)$, $\varepsilon_0 > 0$, $T + \varepsilon_0 < \widetilde{T} < T + \varepsilon$. В силу леммы абсолютная и равномерная сходимость комплексных функциональных рядов обеспечивает возможность перегруппировки слагаемых и вещественную аналитичность функций $\gamma_i(t,x)$, $\sigma_i(x)$. Результат является следствием неравенств Коши и оценок в окрестности $Q_s$ коэффициентов $\beta_{j\nu}$ из представлений (6) с константой $\varrho \neq \varrho(\nu, j)$.

В локальной постановке задачи наблюдения неопределенность в начальных данных $x(T)$ мала и требуется исследовать наблюдаемость в окрестности опорного движения с $x(T) = \bar{x}$. Поэтому сформулируем итог проведенных рассуждений в следующей форме.

**Теорема 5.** *Пусть* $f, g \in C^\omega(U)$, *отрезок времени наблюдения* $[0,T]$ *и область* $U_{\scriptscriptstyle T} = \{x(T)\}$ *достаточно малы* ($U_{\scriptscriptstyle T} : \|x - \bar{x}\| < \delta$). *Тогда в* $U_{\scriptscriptstyle T}$ *элементы множества достижимости* $\mathcal{D}_{\scriptscriptstyle T} = \{v(T,\cdot)\,|\,k(t,y) = k(t)y\}$ *имеют представление* (8), *где функции* $\gamma_i(t,x)$ *являются вещественными аналитическими в области* $(t', t'') \times U_{\scriptscriptstyle T} \supset [0,T] \times U_{\scriptscriptstyle T}$, *а функции* $\sigma_i(x)$ — *в области* $U_{\scriptscriptstyle T}$ $\big(\gamma_i \neq \gamma_i(k)\big)$.

В отличие от линейного случая, в полученном конечном разложении элементов множества достижимости по «столбцам матрицы управляемости» $\mathcal{A}^i B = L_f^i\, g$ коэффициенты $\sigma_i$ являются функциями фазового состояния. Если у набора $L_0, \ldots, L_{p-1}$ имеются два различных общих нуля в области $U_{\scriptscriptstyle T}$, то пара $(f, g)$ заведомо неполностью наблюдаема в $U_{\scriptscriptstyle T}$. При $m > 1$ имеем $\langle k, y\rangle_{L_2^m} = \langle k_1, y_1\rangle + \ldots + \langle k_m, y_m\rangle$ и представление вида (8) останется в силе, только $\sigma_i$ — строки $(\sigma_{i1}, \ldots, \sigma_{im})$.

Результат обобщает следующие построения. При $f = Fx$, $g = Gx$ имеем $y(t;x,T) =$

$$= G \exp\big\{(t - T)F\big\} x = \sum_{i=0}^{\infty} \frac{(t - T)^i}{i!}\, GF^j x.$$

Для номеров $j \geqslant p$, $p = \mathrm{rank}(G', \ldots, F'^{n-1}G')$, можно строки $GF^j$ выразить как линейные комбинации $p$ строк $G, GF, \ldots, GF^{p-1}$. Меняя порядок суммирования, получаем

$$y(t;x,T) = \sum_{j=0}^{p-1} \gamma_j(t) GF^j x = \sum_{j=0}^{p-1} \gamma_j(t) L_j(x),$$

$$v(T,x) = \langle k, y \rangle = \sum_{j=0}^{p-1} \sigma_j L_j(x),\ \sigma_j = \langle k, \gamma_j \rangle_{L_2}.$$

Коэффициенты $\gamma_j(t)$ обладают свойством $\gamma(T) = \big(\gamma_0(T), \ldots, \gamma_{p-1}(T)\big)' = e_1$, $\dot{\gamma}(T) = e_2$, $\ldots$, $\gamma^{(p-1)}(T) = e_p$, как и $\gamma_j(t,x)$, построенные в (8): $\gamma(T,x) = e_1$, $\gamma_t(T,x) = e_2, \ldots$, $\{e_i\}$ — канонический базис $\mathbb{R}^n$. Впрочем, суммированием рядов заниматься необязательно. Достаточно воспользоваться представлением матричной экспоненты $\exp\{Ft\} = \alpha_0(t)E + \ldots + \alpha_{p-1}(t)F^{p-1}$. Здесь (как и при $m > 1$) в качестве $p$ можно взять степень характеристического или минимального аннулирующего полинома матрицы $F$. Поиск $\alpha_j(t)$ сводится к решению линейного однородного скалярного дифференциального уравнения $p$-го порядка.

### Вариации весовых функций

Перейдем к вопросу об устойчивости локального базиса наблюдаемых компонент к малым вариациям весовых функций $k(\cdot)$, что существенно с вычислительной точки зрения. Для этого установим важное свойство коэффициентов $\gamma_i(t,x)$. Само представление вида (8) неединственно: можно формально увеличить значение $p$ (полагая соответствующие $\gamma_j = 0$), изменить $\gamma_i$ добавлением нетривиальной тождественной нулю комбинации производных $L_j$ и т. п. Фиксируем в разложении (8) именно те коэффициенты $\gamma_j(t,x)$, которые построены выше ($T$, $U_{\scriptscriptstyle T}$ достаточно малы):

$$\gamma_j(t,x) = \frac{(t-T)^j}{j!} + \sum_{\nu=p}^{\infty} \frac{(t-T)^\nu}{\nu!}\, \eta_{\nu j}(x),$$
$$0 \leqslant j \leqslant p-1,\ \eta_{\nu j}(x) = \mathrm{Re}\,\beta_{\nu j}(x),$$

$t \in (-\varepsilon_0, T+\varepsilon_0)$, $\varepsilon_0 > 0$, $x \in U_{\scriptscriptstyle T}$, $m = 1$. Фиксируем произвольную полную в $L_2[0,T]$ систему $\{k_i, i \geqslant 1\}$. Тогда для элементов $w_i(x) = v_i(T,x) = \big\langle k_i, y(\,\cdot\,;x,T)\big\rangle$, $x \in U_{\scriptscriptstyle T}$, получим $(w_1(x), w_2(x), \ldots)' =$

$$= \begin{pmatrix} \langle k_1, \gamma_0\rangle & ,\ldots, & \langle k_1, \gamma_{p-1}\rangle \\ \langle k_2, \gamma_0\rangle & ,\ldots, & \langle k_2, \gamma_{p-1}\rangle \\ \ldots & ,\ldots, & \ldots \end{pmatrix}$$
$$\times\, (L_0, \ldots, L_{p-1})'. \qquad (9)$$

**Лемма 2.** *В представлении* (9) *элементов множества достижимости* $w_i \in \mathcal{D}_{\scriptscriptstyle T}$, $i \geqslant 1$, *среди строк матрицы* $\Gamma = \big\{\langle k_i, \gamma_j(\cdot,x)\rangle\big\}$ *при любом фиксированном* $x \in U_{\scriptscriptstyle T}$ *можно найти* $p$ *линейно независимых строк.*

Без ограничения общности предполагаем, что базис строк матрицы находится среди $p$ первых строк. В матричной записи получаем представление $\Gamma' = (M, MN)$, $M = M_{p\times p}$. Предположим противное:

$$\mathrm{rank}\, M < p \Rightarrow \Gamma c = 0,\ c = (c_0, \ldots, c_{p-1})' \neq 0,$$

$x = \hat{x} \in U_{\scriptscriptstyle T}$. Тогда выполняется $\langle k_i, c'\gamma\rangle = 0$, $\gamma = (\gamma_0, \ldots, \gamma_{p-1})'$. В силу полноты системы $\{k_i, i \geqslant 1\}$ в $L_2$ имеем $\forall\, t \in [0,T]$ $c'\gamma(t,\hat{x}) =$

$$\sum_{j=0}^{p-1} \frac{(t-T)^j}{j!}\, c_j + \frac{(t-T)^p}{p!} \sum_{j=0}^{p-1} \eta_{pj}(\hat{x}) c_j + \ldots = 0.$$

Отсюда все $c_j = 0$ и получаем противоречие.

Пусть пара $(f,g)$ вещественно аналитична в $U \subseteq \mathbb{R}^n$. Фиксируем $T$ и окрестность $Q$ некоторой опорной точки $\bar{x} \in U$, для которых справедливо конечное представление (8), а также любую полную в $L_2[0,T]$ систему $\{k_i, i \geqslant 1\}$ допустимых весовых функций.

**Теорема 6.** *Если область неопределенности* $U_{\scriptscriptstyle T}$ $(\bar{x} \in U_{\scriptscriptstyle T} \subseteq Q)$ *достаточно мала, то можно выделить такие* $k_{i_1}, \ldots, k_{i_q}$, *что:*

1) *элементы* $w_{i_\nu}(x) = v_{i_\nu}(T,x) = \langle k_{i_\nu}, y\rangle$ $\big(x \in U_{\scriptscriptstyle T},\ 1 \leqslant \nu \leqslant q\big)$ *образуют базис множества достижимости* $\mathcal{D}_{\scriptscriptstyle T}$ $\big($*и* $\Phi(U_{\scriptscriptstyle T})\big)$;

2) *базис множества* $\mathcal{D}_{\scriptscriptstyle T}$ *образуют также функции* $\widetilde{w}_{i_\nu}(x) = \langle k_{i_\nu} + \xi_{i_\nu}, y\rangle$ *при достаточно малых возмущениях* $\|\xi_{i_\nu}\|_{L_1} < \tilde{\varepsilon}$.

*Доказательство.* Считаем $U_{\scriptscriptstyle T}$ окрестностью опорной точки $\bar{x} \in U$ (требование малости $U_{\scriptscriptstyle T}$ далее уточним). Рассмотрим идеал $\tilde{J} \subset \mathcal{H}_{2n}$ ростков голоморфных функций в точке $(\bar{x}, \bar{x}) \in U_{\scriptscriptstyle T} \times U_{\scriptscriptstyle T}$, порожденный множеством $\big\{\Delta L_i^c : P \times P \to \mathbb{C},\ i \geqslant 0\big\}$, $\Delta L_i^c\big(x^1, x^2\big) = L_i^c\big(x^1\big) - L_i^c\big(x^2\big)$. Используем построения из доказательства теоремы 5. Представления

$$\Delta L_j^c\big(z^1, z^2\big) = \textstyle\sum_{\nu=0}^{q-1} \widetilde{\beta}_{j\nu}\big(z^1, z^2\big) \Delta L_\nu^c, \qquad (10)$$
$$\|\widetilde{\beta}_{j\nu}\|_{\widetilde{Q}_s} \leqslant \widetilde{\varrho}\, \|\Delta L_j^c\|_{\widetilde{Q}_s}, \quad \widetilde{\varrho} \neq \widetilde{\varrho}(\nu, j), \quad j \geqslant q,$$

справедливые в некоторой комплексной окрестности $\widetilde{Q}_s$ точки $(\bar{x}, \bar{x})$ в произведении $P \times P$, доказывается аналогично разложениям (6). Считаем, что $U_{\scriptscriptstyle T} \times U_{\scriptscriptstyle T} \subset \widetilde{Q}_s \cap \mathbb{R}^{2n}$. Точно так же (только удваиваем размерность) приходим к выражениям (9) с заменой элементов $w_i(x)$ на разности $\Delta w_i\big(x^1,x^2\big)=w_i\big(x^1\big)-w_i\big(x^2\big)=\big\langle k_i, y\big(\cdot\,;x^1,T\big)-y\big(\cdot\,;x^2,T\big)\big\rangle$, заменой функций $L_\nu(x)$ на разности $\Delta L_\nu\big(x^1,x^2\big)=L_\nu\big(x^1\big)-L_\nu\big(x^2\big)$ и номера $p$ на $q$. Для этого в разложении в функциональный ряд

$$\Delta y = y\big(t;x^1,T\big)-y\big(t;x^2,T\big)$$
$$=\Delta L_0+(t-T)\Delta L_1+0.5(t-T)^2\Delta L_2+\dots$$

следует заменить $\Delta L_j$, $j\geqslant q$, линейными комбинациями функций $\Delta L_0,\dots,\Delta L_{q-1}$ по формуле (10) и поменять порядок суммирования.

Фиксируем теперь номера $i_1,\dots,i_q$ линейно независимых в точке $(\bar{x},\bar{x})$ строк соответствующей матрицы $\Gamma$: $\big(\Delta w_{i_1},\dots,\Delta w_{i_q}\big)=$

$$=\big(\Delta L_0,\dots,\Delta L_{q-1}\big)R,\quad x^{1,2}\in U_{\scriptscriptstyle T}.$$

При этом $\det R(\bar{x},\bar{x})\neq 0$. Элементы матрицы $R$ имеют вид скалярных произведений $\big\langle k_{i_\nu},\gamma_j\big(\cdot\,,x^1,x^2\big)\big\rangle$ $\big(0\leqslant j\leqslant q-1,\ 1\leqslant\nu\leqslant q\big)$. Поэтому при достаточно малых допустимых возмущениях $\|\xi_{i_\nu}\|_{L_1[0,T]}<\tilde{\varepsilon}$ матрица $\widetilde{R}$ с элементами $\big\langle k_{i_\nu}+\xi_{i_\nu},\gamma_j\big\rangle$ останется невырожденной в точке $(\bar{x},\bar{x})$ и ее окрестности $U_{\scriptscriptstyle T}\times U_{\scriptscriptstyle T}$. Здесь, если необходимо, снова уменьшаем область $U_{\scriptscriptstyle T}$. Окончательно получаем

$$\big(\Delta\widetilde{w}_{i_1},\dots,\Delta\widetilde{w}_{i_q}\big)=\big(\Delta L_0,\dots,\Delta L_{q-1}\big)\widetilde{R},\quad(11)$$
$$\big(x^1,x^2\big)\in U_{\scriptscriptstyle T}\times U_{\scriptscriptstyle T},\ \det\widetilde{R}\neq 0,\ \|\xi_{i_\nu}\|_{L_1}<\tilde{\varepsilon},$$
$$\Delta\widetilde{w}_{i_\nu}\big(x^1,x^2\big)=\big\langle k_{i_\nu}+\xi_{i_\nu},\Delta y\big\rangle.$$

Из $\Delta\widetilde{w}_{i_\nu}\big(x^1,x^2\big)=0$, $\nu\leqslant q$, следует $\Delta L_j\big(x^1,x^2\big)=0$, $j\leqslant q-1$, и $\Delta L_j=0$, $j\geqslant 0$, $y\big(\cdot\,;x^1,T\big)=y\big(\cdot\,;x^2,T\big)$. Поэтому

$$\big(\widetilde{w}_{i_1}(x),\dots,\widetilde{w}_{i_q}(x)\big)\leftrightarrow y\big(\cdot\,;x,T\big),\quad x\in U_{\scriptscriptstyle T}.$$

Функции $\widetilde{w}_{i_\nu}$ образуют базис множества достижимости $\mathcal{D}_{\scriptscriptstyle T}$ сопряженной системы и множества $\Phi(U_{\scriptscriptstyle T})$ наблюдаемых функций $\varphi$. $\square$

**Следствие.** *Если дополнительно пара $(f,g)$ наблюдаема в области $U_{\scriptscriptstyle T}$, то фазовый вектор $x(T)\in U_{\scriptscriptstyle T}$ однозначно определяется по $q$ проекциям $\mu_\nu=\langle k_{i_\nu},y\rangle$ из системы уравнений $v_{i_\nu}(T,x)=\mu_\nu$, $1\leqslant\nu\leqslant q$. Однозначность восстановления $x(T)$ останется при малом возмущении весовых функций $k_{i_\nu}(\cdot)$.*

## Литература


*Заика Ю. В.* Вычисление фазовых переменных ЛА по результатам траекторных измерений // Методы восстановления и анализа динамики управляемых процессов. М.: МО, 1988. Вып. 1. С. 57–71.

*Заика Ю. В.* Устойчивые дискретные программы наблюдений в аналитических динамических системах // Математические заметки. 1999. Т. 66, № 2. С. 194–201.

*Иванов А. П., Кирин Н. Е.* Сопряженные задачи теории управления. Л.: ЛГУ, 1988. 88 с.

*Кирин Н. Е.* Методы оценивания и управления в динамических системах. СПб.: СПбГУ, 1993. 308 с.

*Козеренко К. В.* О числе замеров // ДАН СССР. 1987. Т. 296, № 5. С. 1069–1071.

*Красовский Н. Н.* Теория управления движением. Линейные системы. М.: Наука, 1968. 476 с.

*Никайдо Х.* Выпуклые структуры и математическая эконимика. М.: Мир, 1972. 523 с.

*Ортега Дж., Рейнболдт В.* Итерационные методы решения нелинейных систем уравнений. М.: Мир, 1975. 560 с.

*Чирка Е. М.* Комплексные аналитические множества. М.: Наука, 1985. 270 с.

*Эрве М.* Функции многих комплексных переменных. М.: Мир, 1965. 265 с.

*Inoye Y.* On the observability of autonomous nonlinear systems // Journal of Math. Analysis and Applications. 1977. Vol. 60, N. 1. P. 236–247.

*Zaika Yu.* Integral observability operators of nonlinear dynamical systems // International Journal of Mathematics and Mathematical Sciences. 2003. N. 55. P. 3519–3538.


## СВЕДЕНИЯ ОБ АВТОРЕ:


**Заика Юрий Васильевич**
зав. лаб. моделирования природно-технических систем, д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: zaika@krc.karelia.ru
тел.: (8142) 766312

**Zaika, Yury**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: zaika@krc.karelia.ru
tel.: (8142) 766312

Труды Карельского научного центра РАН
№ 5. 2011. С. 28–32

*УДК 519.833*

# ИГРА $N$ ЛИЦ С ОПТИМАЛЬНОЙ ОСТАНОВКОЙ

**А. А. Ивашко, Е. Е. Ивашко**

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассматривается следующая игра $n$ лиц с оптимальной остановкой. Каждый из $n$ игроков независимо от остальных получает в качестве очков значение случайной величины, равномерно распределенной на $[0, b]$, $b \geqslant 1$. Игрок должен принять решение: остановиться после первого шага или продолжить и получить дополнительно значение следующей независимой случайной величины, которое прибавляется к полученным ранее очкам. Игроки не знают ни значений наблюдений, ни решений, принятых другими игроками. Побеждает тот игрок, чья сумма очков окажется наиболее близкой, но не превышающей 1. Если сумма очков каждого из игроков превысила 1, то побеждает игрок, получивший наименьшее число очков. Каждый игрок стремится максимизировать вероятность своего выигрыша. Найдены оптимальные стратегии игроков в данной задаче. Рассмотрено обобщение игры на случай разладки.

Ключевые слова: игры с оптимальной остановкой, пороговая стратегия, разладка.

### A. A. Ivashko, E. E. Ivashko. $N$-PERSON OPTIMAL STOPPING GAME

The following $n$-person optimal stopping game is considered. Each of $n$ players gets a random number from $[0, b]$, $b \geqslant 1$ as a score. A player has to decide whether to stop after the first draw or to continue to a second draw, in which case the value of a second random variable is added to their score. Players do not know the scores or the decisions of other players. The player with the largest total score not exceeding 1 wins. In the case the total scores of each player exceeds 1, the player with the lowest score closest to 1 wins. The objective of each player is to maximize the probability of one's win. Optimal strategies of the players are presented. The disorder version of the game is also investigated.

Key words: optimal stopping games, threshold strategy, disorder.

## Введение

Рассматривается следующая игра $n$ лиц с оптимальной остановкой. Игрок $i$ независимо от остальных участников получает в качестве очков значение случайной величины $Y_1^i$, $i = 1,..,n$. Игрок должен принять решение: *остановиться* после первого шага или *продолжить* и получить дополнительно значение случайной величины $Y_2^i$, $i = 1,..,n$, которое прибавляется к полученным ранее очкам. $Y_1^i$ и $Y_2^i$ — независимые равномерно распределенные на отрезке $[0, b]$, $b \geqslant 1$, случайные величины. Побеждает тот игрок, чья сумма очков окажется наиболее близкой, но не превышающей 1. Если сумма очков каждого из игроков превысила 1, то побеждает игрок, получивший наименьшее число очков. Игроки не знают ни значений наблюдений, ни решений, принятых другими игроками. Каждый игрок стремится максимизировать вероятность своего выигрыша.

Данная задача является моделью популярного в США телевизионного игрового шоу «The Price is Right». В работе [Kaynar, 2009] было представлено решение этой задачи при $b = 1$, основанное на перечислении всех вероятностей выигрыша игрока. Такой метод приводит к необходимости перебора большого числа вариантов и построению объемных формул при большом числе игроков. Как следствие, в указанной статье представлено решение только для случаев двух и трех игроков. В данной статье решение получено методом динамического программирования в общем виде для случая $n$ игроков и при различных значениях параметра равномерного распределения наблюдаемых случайных величин. Также рассмотрено обобщение задачи на случай разладки, в котором распределение поступающих случайных величин может измениться в случайный момент времени.

Другие задачи с оптимальной остановкой были исследованы в работах [Coe, Butterworth, 1995; Tijms, 2007]. Метод динамического программирования для решения задач с оптимальной остановкой применялся в работах [Mazalov, 1996; Sakaguchi, 2005, 2007]. В работе [Mazalov, Ivashko, 2010] рассмотрена многошаговая игра с оптимальной остановкой.

## Равномерное распределение наблюдений на отрезке $[0,1]$

Рассмотрим случай $b = 1$, т. е. наблюдаемые игроками случайные величины независимы и равномерно распределены на отрезке $[0,1]$.

Предположим, что каждый из игроков $i$, $i = 1, ..., n$ для принятия решения об остановке/продолжении использует следующую однопороговую стратегию: остановиться на первом шаге, если полученное значение случайной величины $Y_1^i$ больше или равно пороговому значению, и продолжить в противном случае. Пусть Игроки $1, ..., n-1$ используют пороговую стратегию $0 < t < 1$, а Игрок $n$ ищет свою наилучшую ответную стратегию $t < v < 1$.

Обозначим $X_i$ — сумма очков, полученных игроком $i$ $(i = 1, ..., n-1)$ в результате использования стратегии $t$. $X_i$ равно $Y_1^i$ при остановке на первом шаге или $Y_1^i + Y_2^i$ при продолжении.

Для вычисления ожидаемого выигрыша Игрока $n$ рассмотрим вероятности следующих событий:

$$P(X_i \leqslant y) = y - t + \int_0^t (y-x)dx, \; t \leqslant y \leqslant 1;$$

$$P(X_i > 1) = \int_0^t x dx;$$

$$P_{y>1}(X_i > y) = \int_{y-1}^t (1-(y-x))dx, \; y > 1,$$

где $i = 1, ..., n-1$. Так как игроки независимы друг от друга, то получим, что рассмотренные вероятности одинаковы для всех игроков.

Обозначим $H(v,t)$ ожидаемый выигрыш Игрока $n$ при остановке на первом шаге в случае, когда общее число его очков равно $v$. Тогда

$$H(v,t) = \sum_{k=0}^{n-1} C_{n-1}^k \Big(P(X_1 \leqslant v)\Big)^{n-1-k} \Big(P(X_1 > 1)\Big)^k = (v - t + vt)^{n-1},$$

так как $P(X_1 \leqslant v) = P(X_2 \leqslant v) = ... = P(X_{n-1} \leqslant v)$, и $P(X_1 > 1) = P(X_2 > 1) = ... = P(X_{n-1} > 1)$.

$H'(v,t)$ — ожидаемый выигрыш Игрока $n$, если он продолжит игру, и число его очков на первом шаге равно $v$. Получим

$$H'(v,t) = \int_v^1 H(y,t)dy + \int_1^{v+1} (P_{y>1}(X_1 > y))^{n-1} dy = \frac{1-(v-t+vt)^n}{n(t+1)} + \frac{t^{2n-1}-(t-v)^{2n-1}}{2^{n-1}(2n-1)},$$

т.к. $P_{y>1}(X_1 > y) = P_{y>1}(X_2 > y) = ... = P_{y>1}(X_{n-1} > y)$.

Нетрудно показать, что для любого $0 < t < 1$ функция $H(v,t)$ возрастает, а функция $H'(v,t)$ убывает по $v$. Значение оптимального порога $t$ таково, что ожидаемый выигрыш Игрока $n$ при продолжении равен ожидаемому выигрышу при остановке, т. е. $H(v,t) = H'(v,t)$.

В силу симметрии задачи положим $v = t$. Окончательно получим, что значением оптимального порога $t$ для принятия первого наблюдения является корень уравнения:

$$t^{2(n-1)} = \frac{1-t^{2n}}{n(t+1)} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}.$$

При $n = 2$ эта формула дает то же решение, что и представленное в работе [Kaynar, 2009]: $2t^3 + 3t^2 + 3t - 3 = 0$.

Значения оптимальных порогов для различных значений $n$ представлены в табл. 1.

*Таблица 1.* Оптимальные пороги для различных значений $n$

| $n$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 100 | 500 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | 0,563 | 0,661 | 0,718 | 0,757 | 0,785 | 0,806 | 0,823 | 0,837 | 0,849 | 0,974 | 0,993 |

## Равномерное распределение наблюдений на отрезке $[0, b]$, где $b > 1$

Рассмотрим случай, когда наблюдаемые случайные величины распределены на отрезке $[0, b]$, где $b > 1$. Отметим, что изменение параметра распределения уменьшит значение оптимального порога $t$ по сравнению с вариантом задачи, рассмотренной в предыдущем пункте. Это является следствием того, что при принятии решения о продолжении увеличивается вероятность превышения 1.

Действуя методом, использованным ранее, получаем, что выигрыши игроков $H(v,t)$ и $H'(v,t)$ при остановке и при продолжении, соответственно, имеют следующий вид:

$$H(v,t)=\sum_{k=0}^{n-1} C_{n-1}^{k}\left(P(X_1\leqslant v)^{n-k-1}\right)\left(P(X_1>1)\right)^k$$
$$=(P(X_1\leqslant v)+P(X_1>1))^{n-1}$$
$$=\left[\frac{bv-t+tv+b(b-1)}{b^2}\right]^{n-1},$$

так как $P(X_1 \leqslant v) = P(X_2 \leqslant v) = ... = P(X_{n-1} \leqslant v)$, $P(X_1 > 1) = P(X_2 > 1) = ... = P(X_{n-1} > 1)$,
и

$$H'(v,t) = \frac{1}{b}\left[\int\limits_v^1 H(y,t)dy + \int\limits_1^{v+b} (P_{y>1}(X_1>y))^{n-1}\,dy\right]$$
$$= \frac{1}{b}\left[\int\limits_v^1 H(y,t)dy + \int\limits_1^{b} (P_{b>y>1}(X_1>y))^{n-1}\,dy + \int\limits_b^{v+b} (P_{y>b}(X_1>y))^{n-1}\,dy\right]$$
$$= \frac{1}{b}\left[\frac{(b^2)^n-(bv-t+tv+b(b-1))^n}{n(b+t)b^{2(n-1)}} + \frac{(t(b-1+\frac{t}{2})+b(b-1))^n-(\frac{t^2}{2})^n}{b^{2(n-1)}(t+b)n} + \frac{t^{2n-1}-(t-v)^{2n-1}}{2^{n-1}(2n-1)b^{2(n-1)}}\right],$$

так как $P_{y>1}(X_1 > y) = P_{y>1}(X_2 > y) = ... = P_{y>1}(X_{n-1} > y)$, $P_{b>y>1}(X_1 > y) = P_{b>y>1}(X_2 > y) = ... = P_{b>y>1}(X_{n-1} > y)$ и $P_{y>b}(X_1 > y) = P_{y>b}(X_2 > y) = ... = P_{y>b}(X_{n-1} > y)$.

Решая уравнение $H(v,t) = H'(v,t)$, а затем полагая $v = t$, получим:

$$\left[bt-t+t^2+b(b-1)\right]^{n-1} = \frac{1}{b}\left[\frac{b^{2n}-(bt-t+t^2+b(b-1))^n}{n(b+t)} + \frac{(t(b-1+\frac{t}{2})+b(b-1))^n-(\frac{t^2}{2})^n}{(t+b)n} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}\right].$$

На рис. 1 изображены значения оптимальных порогов $t(b)$ как функции от $b$ при $n = 2$.

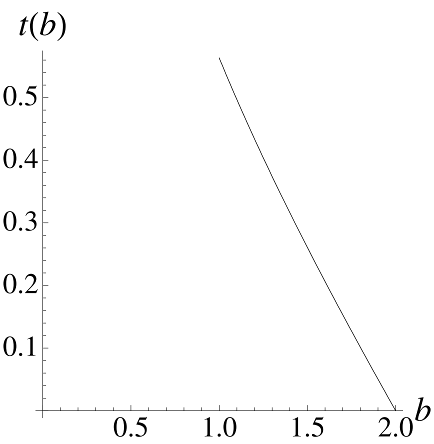



*Рис. 1.* Значения оптимальных порогов $t(b)$ при $n = 2$

Заметим, что оптимальный порог для $b \geqslant 2$ равен 0. Это означает, что игрок должен всегда останавливаться на первом шаге независимо от полученного количества очков.

## Комбинированный вариант

Интерес представляет следующий комбинированный вариант исходной задачи. Рассмотрим задачу, в которой распределение на первом шаге является равномерным на отрезке $[0, 1]$, а на втором — на отрезке $[0, b]$, $b > 1$. Тогда, учитывая различные распределения получаемых значений, ожидаемые выигрыши игроков примут вид:

$$H(v,t) = \left[v - t + \frac{t}{b}(v - \frac{t}{2}) + \frac{t(b-1+\frac{t}{2})}{b}\right]^{n-1}$$
$$= \left[v + \frac{tv}{b} - \frac{t}{b}\right]^{n-1},$$

и

$$H'(v,t) = \frac{1}{b}\left[\frac{b^n-(vb+tv-t)^n}{nb^{n-1}(b+t)} + \left(\frac{t}{b}\right)^{n-1}\frac{(b-1+\frac{t}{2})^n-(\frac{t}{2})^n}{n} + \frac{(t)^{2n-1}-(t-v)^{2n-1}}{2^{n-1}(2n-1)b^{n-1}}\right].$$

Следовательно, значение оптимального порога будет удовлетворять уравнению

$$\left[tb + t^2 - t\right]^{n-1} = \frac{1}{b}\left[\frac{b^n-(tb+t^2-t)^n}{n(b+t)} + t^{n-1}\frac{(b-1+\frac{t}{2})^n-(\frac{t}{2})^n}{n} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}\right].$$

На рис. 2 представлены значения оптимальных порогов $t(b)$ для $n = 2$.

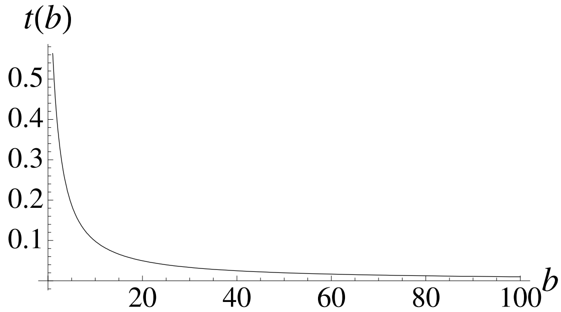



*Рис. 2.* Значения оптимальных порогов $t(b)$ при $n = 2$

При $b \to \infty$ значение оптимального порога стремится к нулю.

## Вариант задачи с разладкой

Рассмотрим обобщение исходной задачи на случай «разладки». Пусть изначально распределение обоих наблюдений является равномерным на отрезке $[0, 1]$. В случайный момент времени (перед первым или вторым шагом) одновременно для всех наблюдателей может произойти разладка, и распределение случайных величин изменится на равномерное на отрезке $[0, b]$, где $b > 1$. Момент разладки — это случайная величина, имеющая геометрическое распределение с параметром $1 - \alpha$ ($0 \leqslant \alpha \leqslant 1$). Игрокам известны параметры $\alpha$ и $b$, но они не обладают информацией, в какой момент произошла разладка. Будем искать однопороговую стратегию, которая устанавливается игроками в начале игры и не меняется в дальнейшем.

Решение данной задачи строится на основе комбинации трех предыдущих игр. Ожидаемый выигрыш игрока зависит от того, на каком шаге произойдет разладка. С вероятностью $\alpha^2$ распределение случайных величин не изменится ни перед первым шагом, ни перед вторым. В этом случае игрок, использующий стратегию $t$, получает выигрыш $t^{2(n-1)}$. С вероятностью $\alpha(1-\alpha)$ разладка произойдет только перед вторым шагом, тогда игрок получит выигрыш $\left[tb + t^2 - t\right]^{n-1}$, соответствующий комбинированному варианту задачи. Наконец, с вероятностью $1-\alpha$ разладка произойдет перед первым шагом, и тогда каждый игрок получит в качестве выигрыша величину $\left[bt - t + t^2 + b(b-1)\right]^{n-1}$, которая равна выигрышу игрока в задаче с равномерным распределением обеих случайных величин на отрезке $[0, b]$. Аналогично можно провести рассуждения для выигрыша при продолжении. Далее, приравнивая выигрыш при остановке и выигрыш при продолжении, получаем, что оптимальные пороги находятся из следующего уравнения:

$$
\begin{aligned}
&\alpha^2 t^{2(n-1)} + \alpha(1-\alpha)\left[tb + t^2 - t\right]^{n-1} \\
&+(1-\alpha)\left[bt - t + t^2 + b(b-1)\right]^{n-1} \\
&= \alpha^2\left[\frac{1-t^{2n}}{n(t+1)} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}\right] + \alpha(1-\alpha)\frac{1}{b}\left[\frac{b^n-(tb+t^2-t)^n}{n(b+t)}\right. \\
&\left.+\, t^{n-1}\frac{(b-1+\frac{t}{2})^n-(\frac{t}{2})^n}{n} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}\right] \\
&+(1-\alpha)\frac{1}{b}\left[\frac{b^{2n}-(bt-t+t^2+b(b-1))^n}{n(b+t)}\right. \\
&\left.+\frac{(t(b-1+\frac{t}{2})+b(b-1))^n-(\frac{t^2}{2})^n}{(t+b)n} + \frac{t^{2n-1}}{2^{n-1}(2n-1)}\right]
\end{aligned}
$$

или

$$
\begin{aligned}
&\alpha^2 t^{2(n-1)} + \alpha(1-\alpha)\left[tb + t^2 - t\right]^{n-1} \\
&+(1-\alpha)\left[bt - t + t^2 + b(b-1)\right]^{n-1} \\
&= \alpha^2\frac{1-t^{2n}}{n(t+1)} + \alpha(1-\alpha)\frac{1}{bn}\left[\frac{b^n-(tb+t^2-t)^n}{(b+t)}\right. \\
&\left.+t^{n-1}\left((b-1+\tfrac{t}{2})^n-(\tfrac{t}{2})^n\right)\right] \\
&+(1-\alpha)\frac{1}{b(t+b)n}\left[b^{2n}-(bt-t+t^2+b(b-1))^n\right. \\
&\left.+(t(b-1+\tfrac{t}{2})+b(b-1))^n-(\tfrac{t^2}{2})^n\right] \\
&+\left(\alpha^2-\frac{\alpha^2}{b}+\frac{1}{b}\right)\frac{t^{2n-1}}{2^{n-1}(2n-1)}.
\end{aligned}
$$

Значения оптимальных порогов для различных $n$, $\alpha = 0,1, b = 1,5$ представлены в табл. 2.

*Таблица 2.* Значения оптимальных порогов $t$ для различных $n$ ($\alpha = 0,1$, $b = 1,5$)

| $n$ | 2 | 3 | 4 | 5 | 10 | 20 | 50 |
|---|---|---|---|---|---|---|---|
| $t$ | 0,277 | 0,424 | 0,516 | 0,579 | 0,735 | 0,839 | 0,920 |

## Заключение

В представленной статье рассмотрена игровая задача оптимальной остановки для $n$ игроков. Для данной задачи рассмотрены три варианта с различными параметрами равномерного распределения наблюдаемых игроками случайных величин. Представлено обобщение задачи на случай разладки. Для иллюстрации полученных решений приведены результаты численного моделирования значений оптимальных порогов.

Авторы выражают благодарность проф. В. В. Мазалову за его помощь в постановке задачи и обсуждении полученных результатов. Исследования поддержаны грантом РФФИ, проект 10-01-00089-а и Отделением математических наук РАН.

## Литература

*Coe P. R., Butterworth W.* Optimal stopping in «The Showcase Showdow» // The American Statistician. 1995. N. 49. P. 271–275.

*Kaynar B.* Optimal stopping in a stochastic game // Probability in the Engeneering and Information Sciences. 2009. N. 23. P. 51–60.

*Mazalov V. V.* A Game related to optimal stopping of two sequences of independent random variables having different distributions // Mathematica Japonica. 1996. 43. N. 1. P. 121–128.

*Sakaguchi M.* Equilibrium in two-player games of «Showcase Showdown» // Scientiae Mathematicae Japonicae. 2005. N. 61. P. 145–151.

*Sakaguchi M.* Players' information in two-player games of «Score Showdown» // Game Theory and Applications. 2007. N. 11 / Edit. Petrosjan and Mazalov N.Y. Nova Science Publ. P. 111–124.

*Tijms H. C.* Understanding probability, 2nd ed. Cambridge: Cambridge Univ. Press. 2007.

*Mazalov V. V., Ivashko A. A.* Equilibrium in $n$-person game of Showcase-Showdown // Probability in the Engineering and Informational Sciences, Cambridge Univ. Press, 2010. N. 24. P. 397–403.


## СВЕДЕНИЯ ОБ АВТОРАХ:


**Ивашко Анна Антоновна**
научный сотрудник
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: aivashko@krc.karelia.ru
тел.: (8142) 766312

**Ivashko, Anna**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: aivashko@krc.karelia.ru
tel.: (8142) 766312

**Ивашко Евгений Евгеньевич**
научный сотрудник
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: ivashko@krc.karelia.ru
тел.: (8142) 766312

**Ivashko, Evgeny**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: ivashko@krc.karelia.ru
tel.: (8142) 766312

Труды Карельского научного центра РАН
№ 5. 2011. С. 33–37

*УДК 517.977*

# ИНВАРИАНТНЫЕ МНОЖЕСТВА СИСТЕМЫ УПРАВЛЕНИЯ ПРОЦЕССОМ БИОЛОГИЧЕСКОЙ ОЧИСТКИ

**А. Н. Кириллов**

*Институт прикладных математических исследований*
*Карельского научного центра РАН*

Рассмотрена многомерная нелинейная динамическая система, описывающая процесс биологической очистки. Построены инвариантные множества системы. Предложен метод стабилизации процесса, основанный на приведении траекторий в эти множества.

К л ю ч е в ы е с л о в а: инвариантное множество, динамическая система, управление, биологическая очистка.

## A. N. Kirillov. INVARIANT SETS OF A BIOLOGICAL TREATMENT PROCESS CONTROL SYSTEM

A multidimensional nonlinear dynamic system used to describe the biological treatment process is considered. Invariant sets of the system are built. A process stabilization method based on routing the trajectories into these sets is proposed.

K e y w o r d s: invariant sets, dynamic system, control, biological treatment.

## Введение

Задача стабилизации процесса биологической очистки сточных вод играет важную роль в проблеме охраны окружающей среды. От ее успешного решения зависит качество питьевой воды, количество которой в расчете на одного человека в период с 1970 по 2002 гг., по данным Центра экологической политики России, уменьшилось вдвое [Данилов-Данильян, 2009]. Сложность процесса биоочистки, ограниченная возможность проведения натурных экспериментов [Евилевич, Брагинский, 1979] приводят к необходимости активно использовать математическое моделирование для управления этим процессом. В 1987 г. была создана так называемая модель ASM1 очистки сточных вод с помощью активного ила [Henze et al., 1987], которая стимулировала использование математического моделирования в практике инженерных расчетов. Модель ASM1 представляет собой систему нелинейных обыкновенных дифференциальных уравнений, аналитическое и качественное исследование которой практически невозможно. Имеется достаточно большое количество работ, связанных с упрощением данной модели. При этом используют линеаризацию, декомпозицию, уменьшение размерности [Steffens et al., 1997; Chachuat et al., 2003; Smets et al., 2003]. Следует заметить, что несмотря на сложность, в модели ASM1 не учтена многовидовость состава микроорганизмов активного ила.

Процессы, происходящие в системе биоочистки, до сих пор мало изучены, поэтому построение адекватной модели, учитывающей различные стороны этого процесса, в настоящее время невозможно. Видимо, следует сосредоточиться на разработке комплекса достаточно простых и приемлемых с инженерной точки зрения моделей, применение которых для повышения эффективности процесса биоочистки не вызывало бы больших затруднений.

В настоящей работе предлагается метод стабилизации процесса биоочистки, основанный на построении инвариантного множества в динамической системе управления. В качестве трофической функции используется функция Моно [Brune, 1985; Вавилин, 1986]. При этом учтен многовидовой состав сообщества микроорганизмов. Управляющим воздействием является скорость возвратного потока ила. Допустимые управления – кусочно-постоянны и принимают всего два значения, что упрощает практическую реализацию алгоритма стабилизации.

## Модель и постановка задачи

Будем рассматривать процесс биологической очистки, выделив из его технологической схемы аэротенк и вторичный отстойник. Рассмотрим сообщество микроорганизмов, находящееся в аэротенке. Пусть происходит параллельное потребление субстратов [Вавилин, 1986], т. е. $i$-й микроорганизм потребляет $i$-й субстрат, $i = 1, ..., n$. При этом используем трофическую функцию Моно. Таким образом, получаем динамическую систему, описывающую процесс биологической очистки

$$\dot{x}_i = ua_{1i} + f_i(x_i, s_i) - (b + u)x_i, \qquad (1)$$

$$\dot{s}_i = ba_{2i} - d_i f_i(x_i, s_i) - (b + u)s_i, \qquad (2)$$

$$i = 1, ..., n,$$

где

$$f_i(x_i, s_i) = \frac{\mu_i x_i s_i}{k_i + s_i},$$

$x_i, s_i$ – концентрация $i$-х видов микроорганизма и субстрата, соответственно; $d_i = \frac{1}{Y_i}$, $Y_i$ – коэффициент утилизации $i$-го вида субстрата в биомассу $i$-го вида микроорганизмов; $k_i, \mu_i$ – константа полунасыщения и максимальная удельная скорость роста, соответственно, микроорганизмов $i$-го вида; $u, a_{1i}$ – скорость и концентрация, соответственно, $i$-го вида микроорганизмов в возвратном потоке; $b, a_{2i}$ – скорость и концентрация, соответственно, $i$-го вида субстрата на входе. Параметры системы, если особо не оговорено, считаем постоянными. Будем полагать, что управлением является скорость возвратного потока $u$ как наиболее технически просто регулируемый параметр. При этом вводим ограничения

$$u \in [u_1, u_2],$$

где $u_1, u_2$ – граничные значения скорости $u$, $0 \leqslant u_1 < u_2$. Пусть допустимое управление является кусочно-постоянной функцией $u = u(x, s)$, $u = (u_1, ..., u_n)$, $s = (s_1, ..., s_n)$.

Задача стабилизации процесса биологической очистки сточных вод состоит в нахождении такого допустимого управления $u$, при котором, начиная с некоторого момента времени $t^*$, выполняются условия

$$x_i(t) \in [0, x_{im}], \quad s_i(t) \in [0, s_{im}], \qquad (3)$$

$$t \geqslant t^*, \quad i = 1, ..., n,$$

где $x_{im}, s_{im}$ – предельно допустимые концентрации $i$-го вида микроорганизмов и субстратов.

Множество $M$ является инвариантным для системы управления $\dot{z} = f(z, u)$, если найдется допустимое управление $u$ такое, что соответствующая траектория $x(t, u)$ принадлежит $M$ при всех $t \geqslant t_0$, если она начинается в точке $x(t_0, u) = x_0 \in M$. Таким образом, получаем задачу нахождения допустимого управления, при котором множество

$$M = \{(x, s) : x_i \in [0, x_{im}], s_i \in [0, s_{im}], i = \overline{1, n}\}$$

будет инвариантным для системы (1), (2). При этом также возникает задача перевода начальных точек в $M$ за конечное время. Дальнейшее изложение посвящено решению этих задач.

**Замечание 1.** Будем считать, что допустимое управление не только кусочно-постоянно, но принимает лишь граничные значения. Во-первых, такое управление технически проще реализовать, а во-вторых, добавление к граничным значениям точек из интервала $(u_1, u_2)$ не расширит возможностей стабилизации системы, что будет видно из дальнейшего изложения. Таким образом, метод управления обладает свойством релейности, т. е. является управлением типа «bang-bang control».

## Инвариантные множества

В работе [Кириллов, 1994] рассматривалась система типа (1), (2) для случая одного вида микроорганизма и субстрата и, в частности, решалась задача стабилизации положений равновесия. Следует отметить, что требование асимптотической устойчивости отдельных состояний сложной экологической системы излишне жестко и в реальности не осуществимо. Достаточно требовать устойчивость по Лагранжу концентраций субстратов и микроорганизмов. Для природных экологических систем, это впервые отмечено в работе [Свирежев, Логофет, 1978], где исследовалась динамика взаимодействия популяций. Таким образом, приходим к задаче построения инвариантных множеств системы (1), (2).

Следующая лемма 1 дает достаточное условие инвариантности множества $M$.

**Лемма 1.** *Множество*

$$M = \{(x, s) : x_i \in [0, x_{im}], s_i \in [0, s_{im}]\},$$

*где* $i = \{1, ..., n\}$, *является инвариантным для системы (1), (2) при достаточно больших значений* $x_{im}, s_{im}$.

*Доказательство.* Для доказательства достаточно рассмотреть значения знаков $\dot{x}_i, \dot{s}_i$ из системы (1), (2) для конкретного значения $i$ на границах множества (3), откуда получаем утверждение для конкретного значения $i$. Осталось заметить, что $M = M_1 \times ... \times M_n$. $\square$

Пусть $z_i = d_i x_i + s_i$, $D_i = d_i a_{1i} u + b a_{2i}$, $z = z_1 + ... + z_n$, $D = D_1 + ... + D_n$. Рассмотрим систему

$$\dot{z_i} = D_i - (b + u) z_i, \quad i = 1, ..., n. \tag{4}$$

**Лемма 2.** *Множества*

$$M_{iz} = \{(x, s) : z_i = \tfrac{D_i}{b+u}\}, \forall i \in \{1, ..., n\},$$

$$M_z = \{(x, s) : z = \tfrac{D}{b+u}\}$$

*являются инвариантными и асимптотически устойчивыми для системы (4).*

*Доказательство.* Понятие асимптотической устойчивости инвариантных множеств рассматривалось в [Зубов, 1982]. Умножив уравнение (1) на $d_i$ и прибавив его к уравнению (2), получим линейную систему (4), из вида которой следует утверждение леммы для множества $M_{iz}$. Сложив уравнения системы (4), получим уравнение $\dot{z} = D - (b + u) z$, откуда следует утверждение леммы для множества $M_z$. $\square$

Пусть $A(u_j) = (z_1(u_j), ..., z_n(u_j))$ – положение равновесия системы (1), (2), соответствующее управлению $u_j, j = 1, 2$. Рассмотрим параллелепипед

$$P = \{z : z_i \in [c_{1i}, c_{2i}], i = 1, ..., n\} \in R^n,$$

где $c_{1i}, c_{2i}$ – заданные постоянные, $P^0 = \{z : z_i \in (c_{1i}, c_{2i}), i = 1, ..., n\}$ – соответствующий открытый параллелепипед. Пусть $[A(u_1), A(u_2)] = \{z : z = \alpha A(u_1) + (1 - \alpha) A(u_2), \alpha \in [0, 1]\}$ – отрезок в $R^n$, с концами $A(u_1), A(u_2)$, $(A(u_1), A(u_2)) = [A(u_1), A(u_2)] \setminus \{A(u_1, A(u_2)\}$ – внутренность отрезка $[A(u_1), A(u_2)]$. Пусть $R^n_+$ – множество точек $z \in R^n$ с неотрицательными координатами.

**Теорема 1.** *Пусть*

$$P^0 \cap [A(u_1), A(u_2)] = [B_1, B_2],$$

*где* $B_1 \neq B_2, [B_1, B_2] \in (A(u_1), A(u_2))$.

*Тогда любую точку* $z_0 \in R^n_+$ *с помощью допустимого управления можно за конечное время перевести вдоль траекторий системы (4) в параллелепипед* $P$, *который является инвариантным множеством системы (4). При этом отрезок* $[B_1, B_2]$ *будет инвариантным асимптотически устойчивым множеством.*

*Доказательство.* Обозначим $A(u_1) \equiv A_1$, $A(u_2) \equiv A_2$. Для любой начальной точки $z_0$ положим $u = u_1$ (или $u = u_2$) до тех пор, пока траектория, начинающаяся в этой точке, не попадет в достаточно малую окрестность $U(A_1)$ точки $A_1$ (или $A_2$). Последнее очевидно в силу глобальной асимптотической устойчивости положения равновесия $A_1$ системы (4) [Кириллов, 1994]. Траектории системы (4) – прямые. Действительно, из (4) следует, что

$$z_i(t) = w_{i0} \exp(-(b + u)(t - t_0)) + \frac{D_i}{b + u},$$

где $w_{i0} = z_i(t_0) - \frac{D_i}{b+u}$. Тогда любая траектория является прямой (лучом) $z = w_0 + \mathbf{D}\tau$, где $w_0 = (w_{10}, ..., w_{n0}), \mathbf{D} = (\frac{D_1}{b+u_j}, ..., \frac{D_n}{b+u_j})$, $\tau = \exp(-(b + u_j)(t - t_0))$. Очевидно, отрезок $[A_1, A_2]$ принадлежит траекториям, проходящим через точку $A_1$ при $u = u_2$ и через точку $A_2$ при $u = u_1$, соответственно. Пусть $l$ – прямая, которой принадлежит отрезок $[A_1, A_2]$. Расстояние от точи $z$ до прямой $l$ обозначим через $\rho(z, l) = min||z - y||$, где минимум нормы берется по всем точкам $y \in l$. Рассмотрим цилиндр $\bar{C}(\varepsilon) = \{z : \rho(z, l) \leqslant \varepsilon\}$ с осью $l$. Поскольку отрезок $[A_1, A_2]$ пересекает внутренность $P^0$ параллелепипеда $P$, то цилиндр $\bar{C}(\varepsilon)$ при достаточно малом $\varepsilon$ также пересекает $P^0$. Пусть окрестность $U(A_1)$ настолько мала, что $U(A_1) \subset \bar{C}(\varepsilon)$.

После попадания траектории в описанную окрестность $U(A_1)$, что возможно в силу асимптотической устойчивости положения равновесия $A_1$ при $u = u_1$, переключаем управление на значение $u = u_2$. Тогда траектория, оставаясь в цилиндре $\bar{C}(\varepsilon)$, также пересечет $P^0$. Пусть $\hat{P} \equiv \bar{C}(\varepsilon) \cap P$, $\bar{B} \equiv \hat{P} \setminus P^0$. Траектория входит в параллелепипед $P$, точнее в область $\hat{P}$, пересекая ее границу $\bar{B}$. Переключим управление на значение $u = u_1$ в момент выхода траектории из области $\hat{P}$ через границу $\bar{B}$. Далее вторично переключаем управление при попадании траектории на границу $\bar{B}$. Получаем последовательность управлений и соответствующих отрезков траекторий $[R_k, R_{k+1}]$, где $R_k \in \bar{B}$. При этом $[R_k, R_{k+1}] \to [B_1, B_2]$ в том смысле, что $\rho([R_k, R_{k+1}], [B_1, B_2]) \to 0$ при $k \to \infty$, где $\rho([R_k, R_{k+1}], [B_1, B_2]) \equiv min||z-y||$ по всем $z \in [R_k, R_{k+1}]$, $y \in [B_1, B_2]$, $k = 0, 1, 2, ...,$. Действительно, если $z(t, t_0, z_0, u_j) \equiv z(t)$, $\tilde{z}(t, t_0, \tilde{z}_0, u_j) \equiv \tilde{z}(t)$, то

$$\begin{gathered}\rho([R_k, R_{k+1}], [B_1, B_2]) \\ \leqslant ||z(t) - \tilde{z}(t)|| = ||z(t_{0k}) - \tilde{z}(t_{0k})|| \\ \times \exp(-(b + u_j)(t - t_{0k}),\end{gathered}$$

где $t \in [t_{0k}, t_{0(k+1)}]$, $t_{0k}$ – момент переключения управления, $z(t, t_0, z_0, u_j)$ – траектория системы (4), соответствующая управлению $u_j$ и удовлетворяющая начальному условию $z(t_0, t_0, z_0, u_j) = z_0$. Таким образом, отрезок $[B_1, B_2]$ – инвариантное асимптотически устойчивое множество системы (4).

□

**Замечание 2.** Как следует из доказательства теоремы 1, конкретный вид функций $f_i(x_i, s_i)$ не влияет на метод стабилизации, поэтому можно сформулировать утверждение, аналогичное теореме 1, но без конкретизации этих функций.

**Замечание 3.** В силу линейности системы (4), несложно получить оценку времени попадания траектории из любой начальной точки в $\hat{P}$.

**Замечание 4.** Построенный алгоритм стабилизации очевидным образом применим в случае, когда концентрации субстратов $a_{2i}$ на входе в очистную систему точно неизвестны, но известны промежутки, которым они принадлежат $a_{2i} \in [a_{2imin}, a_{2imax}]$. Для этого требуется соответствующим образом скорректировать построение цилиндра $\bar{C}(\varepsilon)$. При этом при небольших изменениях полученный результат можно использовать и в случае кусочно-постоянных $a_{2i}$.

**Замечание 5.** В теореме 1 доказана асимптотическая устойчивость множества $[B_1, B_2]$. Это означает, что при соответствующем управлении в системе (4), а вместе с тем и в системе (1), (2), возникает асимптотический периодический устойчивый режим. В работе [Матрос, 1987] исследовался вопрос о повышении эффективности химико-технологических процессов в результате искусственного введения периодических режимов их функционирования. Предложенный метод управления позволяет исследовать этот вопрос для процессов биологической очистки.

## Стабилизация

Вернемся к задаче стабилизации процесса биологической очистки, состоящей в приведении траектории системы (1), (2) за конечное время в область (3). Нижеледующее утверждение дает достаточные условия разрешимости этой задачи.

**Теорема 2.** *Пусть для всех $i = 1, ..., n$, $j = 1, 2$*

$$a_{1i} - Y_i a_{2i} \neq 0,$$

$$\frac{D_i}{(b + u_j)d_i} \leqslant x_{im}, \quad \frac{D_i}{b + u_j} \leqslant s_{im}.$$

*Тогда для любого начального состояния $(x_0, s_0) \in R^{2n}_+$ через конечное время будут выполняться условия (3).*

*Доказательство.* Рассмотрим какую-либо одну подсистему (1),(2) при конкретном значении $i$, $(x_i, s_i) \in R^2_+$. Если выполняется первое условие теоремы, то две прямые $d_i x_i + s_i = \frac{D_i}{b+u_j}$, соответствующие $j = 1, 2$, не совпадают [Кириллов, 1994]. Второе и третье условия гарантируют, что отрезки обеих прямых, принадлежащие $R^2_+$, принадлежат прямоугольнику $\{(x_i, s_i) : x_i \in [0, x_{im}], s_i \in [0, s_{im}]\}$. Отсюда с использованием теоремы 1 следует заключение теоремы 2. □

**Замечание 6.** Условие теоремы 2 слишком жесткое. Его можно заменить условием, вытекающим из того факта, что траектории системы (4) с течением времени локализуются в сколь угодно малой окрестности отрезка $[B_1, B_2]$.

## Заключение

Получен метод решения задачи стабилизации процесса биологической очистки сточных вод. В его основе лежит приведение траекторий системы в инвариантное множество с помощью кусочно-постоянного управления, принимающего только два граничных значения. При этом структура системы (1), (2) позволяет свести решение задачи стабилизации к ее решению для линейной относительно $z \in R^n_+$ системы (4). Представляется перспективным применение предложенного метода для системы (1), (2) с переменной размерностью, что соответствует переменному видовому составу сообщества микроорганизмов. Соответствующая математическая модель разработана в работе [Кириллов, 2010], где размерность системы зависит от значения управляющего воздействия. Ранее такая зависимость была использована для управления системой «хищник-жертва» [Kirillov, 1997].

## Литература


*Вавилин В. А.* Время оборота биомассы и деструкция биомассы органического вещества в системах биологической очистки. М.: Наука, 1986. 144 с.

*Данилов-Данильян В. И.* Водные ресурсы мира и перспективы водохозяйственного комплекса России. М.: Институт устойчивого развития; Центр экологической политики России, 2009. 88 с.

*Евилевич М. А., Брагинский Л. Н.* Оптимизация биохимической очистки сточных вод. Л.: Стройиздат, 1979. 159 с.

*Зубов В. И.* Динамика управляемых систем. М.: Наука, 1982. 286 с.

*Кириллов А. Н.* Задача стабилизации экологических систем // Обозрение прикладной и промышленной математики. 1994. № 6. С. 883–892.

*Кириллов А. Н.* Метод динамической декомпозиции в моделировании процесса биологической очистки сточных вод // Обозрение прикладной и промышленной математики. 2010. № 4. С. 496–505.

*Матрос Ю. Ш.* Каталитические процессы в нестационарных условиях. Новосибирск: Наука, 1987. 232 с.

*Свирежев Ю. М., Логофет Д. О.* Устойчивость биологических сообществ. М.: Наука, 1978. 352 с.

*Brune D.* Optimal control of the complete-mix activated sludge process // Environtmental Technology Letters. 1985. Vol. 6. P. 467–476.

*Chachuat B., Roche N., Latifi M. A.* Reduction of the asm1 model for optimal control of small-size activated sludge treatment plants // Journal of water science. 2003. Vol. 16. P. 5–26.

*Henze M., Grady C., Gujer W. et al.* A general model for single-sludge activated sludge wastewater treatment systems // Water research. 1987. Vol. 6. P. 505–515.

*Kirillov A. N.* The stabilization problem for certain class of ecological systems // International Journal of Software Engineering and Knowledge Engineering. 1997. Vol. 7. N. 2. P. 247–251.

*Smets I. Y., Hagebart J. V., Carrette R., Van Impe J. F.* Linearization of the activated sludge model ASM1 for fast and reliable predictions // Water research. 2003. Vol. 37. P. 1831–1851.

*Steffens M. A., Lant P. A., Newell R. B.* A systematic approach for reducing complex biological wastewater treatment model // Water science and technology. 1997. Vol. 31. P. 590–606.


## СВЕДЕНИЯ ОБ АВТОРЕ:


**Кириллов Александр Николаевич**
ведущий научный сотрудник, д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: kirillov@krc.karelia.ru
тел.: (8142) 763370

**Kirillov, Alexandr**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: kirillov@krc.karelia.ru
tel.: (8142) 763370

Труды Карельского научного центра РАН
№ 5. 2011. С. 38–44

*УДК 517.93: 574.34*

# МЕТОДЫ ИССЛЕДОВАНИЯ УСТОЙЧИВОСТИ ПОЛОЖЕНИЙ РАВНОВЕСИЯ НЕАВТОНОМНЫХ СИСТЕМ И НЕКОТОРЫЕ ПРИМЕРЫ ИХ ПРИМЕНЕНИЯ

**А. В. Ласунский**

*Новгородский государственный университет имени Ярослава Мудрого*

Рассмотрено влияние линейной схемы введения пассивных переменных на устойчивость положений равновесия некоторых неавтономных систем дифференциальных уравнений. Полученные признаки устойчивости иллюстрируются на примере неавтономной модели изолированной популяции и модифицированной модели хищник-жертва

$$\dot{x} = \alpha(t)x(1 - yK^{-1}),\ \dot{y} = \beta(t)y(xL^{-1} - 1).$$

К л ю ч е в ы е с л о в а: пассивные стадии жизнедеятельности, устойчивость положения равновесия, первый метод Ляпунова.

## A. V. Lasunsky. METHODS OF INVESTIGATING THE STABILITY OF EQUILIBRIUM POSITIONS IN THE NONAUTONOMOUS SYSTEMS, AND SOME APPLICATIONS THEREOF

We consider the effect of the linear scheme of introduction of passive variables on the stability of the equilibrium of some nonautonomous systems of differential equations. The detected signs of stability are illustrated by the example of a nonautonomous model of an isolated population and a modified predator-prey model.

K e y w o r d s: passive stages of live, stability of equilibrium position, Liapunov's first method.

## Введение

Для решения задачи об устойчивости по первому приближению А. М. Ляпуновым [Ляпунов, 1956] был предложен метод характеристических показателей (первый метод Ляпунова). Одной из основных задач первого метода Ляпунова является оценка изменения характеристических (и других показателей) линейной системы

$$\dot{x} = A(t)x, \quad x \in \mathbb{R}^n,$$

$$A(t) \in \mathbb{C}[t_0, +\infty), \quad \sup_{t \in R^+} \|A(t)\| < M \qquad (1)$$

при различных возмущениях. В теории линейных систем (1) большую роль играют введенные и изученные А. М. Ляпуновым правильные системы дифференциальных уравнений. Эти системы включают в себя приводимые и почти приводимые системы [Былов, 1962] и играют ведущую роль в теории устойчивости по первому приближению. Определение правильности по Ляпунову дается в терминах характеристических показателей решений рассматриваемой системы. С практической точки зрения проверка правильности вызывает определенные затруднения, так как характеристические показатели системы в общем случае не известны. Неслучайно получение достаточных признаков правильности системы остается актуальным и в настоящее время. Интерес представляют результаты, которые удается сформулировать в терминах коэффициентов системы. Поведение траекторий в окрестности положений равновесия исследуется проще с помощью теорем об устойчивости и неустойчивости по первому приближению в автономном случае. Система первого приближения в этом случае всегда правильна, а характеристические показатели совпадают с вещественными частями собственных чисел матрицы коэффициентов. Для неавтономных систем применение аналогичных теорем вызывает затруднение в связи с отсутствием общих методов определения характеристических чисел системы первого приближения, а также установления правильности этой системы. Если мы обратимся к моделированию динамики численности биологических популяций, то ясно, что в реальных биологических сообществах коэффициенты рождаемости и смертности не постоянны. Интерес представляет случай периодического изменения мальтузианских коэффициентов, что соответствует сезонным изменениям в природе. В результате эволюции возникают различные биологические механизмы адаптации, которые позволяют повысить живучесть данного вида. Одним из таких процессов является способность биологических особей, как простейших, так и высокоразвитых переходить из активного состояния в пассивное при наступлении неблагоприятных условий (смена времени года, резкое уменьшение рациона питания и т. д.). Простейшая формализация этого явления может быть осуществлена введением пассивных переменных в математическую модель динамики численности биологических популяций. Взаимодействие активное состояние – пассивное состояние и наоборот можно смоделировать с помощью линейной надстройки к исходной нелинейной модели популяции. В работе [Ильичев, 1992] для автономных непрерывных моделей рассмотрена линейная схема описания механизма образования пассивных стадий (ПС-механизм), проведено исследование устойчивости положений равновесия изолированной и взаимодействующих популяций (конкуренция, хищничество и т. д.) с учетом данного фактора. Разумеется, переход к пассивным стадиям жизнедеятельности – это сложный биологический процесс. Линейная надстройка с постоянными коэффициентами к исходной нелинейной модели динамики популяции является одной из попыток формализовать этот процесс. В настоящей работе получены признаки устойчивости положений равновесия неавтономных систем и рассмотрены их приложение с точки зрения влияния линейной схемы введения пассивных стадий жизнедеятельности на устойчивость положений равновесия некоторых неавтономных моделей биологических сообществ.

## Изолированные популяции и пассивные стадии жизнедеятельности

Пусть численность изолированной популяции $x(t)$ некоторого биологического вида описывается уравнением

$$\dot{x} = f(t, x). \tag{2}$$

Следуя работе [Ильичев, 1992], предположим, что данные организмы наряду с активным состоянием обладают еще и пассивным состоянием, находясь в котором они развиваются и не испытывают влияние других особей. Также предположим, что этот переход обладает свойством линейности. Модель (2) преобразуется к виду

$$\dot{x} = f(t, x) - qx + ps, \ \dot{s} = qx - ps. \tag{3}$$

Здесь через $s(t)$ обозначена часть популяции, находящаяся в пассивном состоянии; $p, q$ – положительные параметры, характеризующие переходы $s \to x$ и $x \to s$ соответственно. Отметим, что если $x_0$ – положение равновесия модели (2), то $(x_0,\ s_0)$ – положение равновесия модели (3), где $s_0 = qp^{-1}x_0$. Обозначим $\partial f(t, x_0)/\partial x = \lambda(t)$. Это коэффициент перед $x - x_0$ в разложении $f(t, x)$ в ряд Тейлора. Какой экологический смысл имеет этот коэффициент? Одной из основных моделей роста популяций изолированных видов является логистическое уравнение [Coleman, 1979],

$$\dot{x} = r(t)x(t)(1 - x(t)K^{-1}(t)), \quad t \geqq 0,$$

где $r(t)$ и $K(t)$ – положительные функции на $[0; +\infty)$. Коэффициент $r(t)$ характеризует скорость роста (размножения) популяции. Если $K(t) = K$, то логистическое уравнение имеет положительное положение равновесия $x_0 = K$. Для этого уравнения $\partial f(t, K)/\partial x = -r(t)$. Ясно, что коэффициент $\lambda(t)$ связан с мальтузианским коэффициентом роста популяции.

Исследуем положение равновесия $(x_0,\ s_0)$ системы (3) на устойчивость с помощью теоремы Ляпунова об устойчивости по первому приближению [Демидович, 1967]. Система первого приближения для (3) в окрестности положения равновесия $(x_0,\ s_0)$ имеет вид

$$\dot{u} = \begin{pmatrix} \lambda(t) - q & p \\ q & -p \end{pmatrix} u = A(t)u. \qquad (4)$$

Если существует предел $\lim\limits_{t\to+\infty} \lambda(t) = \lambda_0$ , то система (4) почти приводима, а значит правильна [Адрианова, 1992]. Правильность системы (4) в этом случае можно проверить и по определению. Если $\lambda_0 < 0$, то характеристические показатели предельной системы $\dot{u} = A_0 u$ отрицательны, так как совпадают с действительными частями корней характеристического уравнения [Былов и др., 1966] $z^2 + (p + q - \lambda_0)z - p\lambda_0 = 0$. Если $\lambda_0 > 0$, то среди характеристических показателей предельной системы есть положительный и можно применить теорему Четаева о неустойчивости по первому приближению [Четаев, 1946]. Автономные линейные системы имеют устойчивые характеристические показатели [Адрианова, 1992].Так как характеристические показатели линейной системы в случае их устойчивости инвариантны относительно бесконечно малых линейных возмущений [Адрианова, 1992], то убеждаемся в справедливости следующих теорем.

**Теорема 1.** *Если существует* $\lim\limits_{t\to+\infty} \lambda(t) = \lambda_0$ *и* $\lambda_0 < 0$, *то положение равновесия* $(x_0,\ s_0)$ *системы* (3) *асимптотически устойчиво при любых положительных* $p$ *и* $q$.

Иными словами, введение ПС-механизма в модель (2) сохраняет асимптотическую устойчивость соответствующего положения равновесия.

**Теорема 2.** *Если существует* $\lim\limits_{t\to+\infty} \lambda(t) = \lambda_0$ *и* $\lambda_0 > 0$, *то положение равновесия* $(x_0,\ s_0)$ *системы* (3) *неустойчиво при любых положительных* $p$ *и* $q$.

Таким образом, неустойчивость положения равновесия неустранима за счет введения линейной модели ПС-механизма.

Применение метода вариации произвольных постоянных и неравенства Гронуолла-Беллмана позволяет убедиться в справедливости следующего утверждения.

**Теорема 3.** *Пусть система* (4) *правильна и* $\lambda(t) = \lambda_0 + \varepsilon(t)$. *Если* $\lambda_0 < 0$ *и верхнее интегральное среднее функции* $|\varepsilon(t)|$ *достаточно мало, то положение равновесия* $(x_0,\ s_0)$ *асимптотически устойчиво.*

**Замечание 1.** Теорема 1 является частным случаем теоремы 3. Если $\varepsilon \to 0$, то система (4) правильна, а интегральное среднее функции $|\varepsilon(t)|$ равно нулю.

*Доказательство.* Собственные числа матрицы

$$A_0 = \begin{pmatrix} \lambda_0 - q & p \\ q & -p \end{pmatrix},$$

$\lambda_0 < 0, \quad p > 0, \quad q > 0$ – различные отрицательные числа. Воспользуемся оценкой матричной экспоненты

$$\|\exp A_0(t-\tau)\| \leqslant \|S\| \|S^{-1}\| \|\exp(J(t-\tau))\| \leqslant$$

$M \exp \Lambda(t-\tau), \quad t \geqslant \tau$, где в данном случае $\Lambda < 0$ – наибольшее собственное число матрицы $A_0$ [Адрианова, 1992]. Здесь $S$ – матрица преобразования $A_0$ к канонической форме Жордана $J$, которая в нашем случае диагональная. Разумеется, значение постоянной $M$ зависит от выбора матричной нормы. Систему (4) запишем в виде $\dot{u} = A_0 u + B(t)u$, где $B(t) = \begin{pmatrix} \varepsilon(t) & 0 \\ 0 & 0 \end{pmatrix}$. По методу вариации для любого решения этой системы имеем

$$u(t) = \exp(A_0(t-t_0))u(t_0)$$
$$+ \int_{t_0}^{t} \exp(A_0(t-\tau))B(\tau)u(\tau)d\tau,$$

откуда

$$\|u(t)\| \leqslant M\exp(\Lambda(t-t_0))\|u(t_0)\|$$
$$+ \int_{t_0}^{t} M\exp(\Lambda(t-\tau))|\varepsilon(\tau)|\, \|u(\tau)\| d\tau,$$
$$\|u(t)\| \exp(-\Lambda t) \leqslant M\exp(-\Lambda t_0)\|u(t_0)\|$$
$$+ \int_{t_0}^{t} M\exp(-\Lambda\tau))|\varepsilon(\tau)|\, \|u(\tau)\| d\tau.$$

По лемме Гронуолла-Беллмана [Адрианова, 1992] имеем

$$\|u(t)\| \exp(-\Lambda t)$$
$$\leqslant M\exp(-\Lambda t_0)\|u(t_0)\| \exp\int_{t_0}^{t} M|\varepsilon(\tau)|d\tau.$$

Для характеристического показателя любого решения $u(t)$ системы (4) справедлива оценка

$$\chi[u(t)] \leqslant \Lambda + M \varlimsup_{t\to\infty} t^{-1} \int_{t_0}^{t} |\varepsilon(\tau)|\, d\tau.$$

Так как $\Lambda < 0$, то при достаточно малом верхнем интегральном среднем функции $|\varepsilon(t)|$ показатели системы (4) отрицательны. В частности, если $\lim\limits_{t\to+\infty} t^{-1}\int_{t_0}^{t}|\varepsilon(\tau)|\,d\tau = 0$, то $\chi[u(t)] \leqslant \Lambda < 0$.

Посмотрим, какие достаточные условия отрицательности характеристических показателей дает применение метода замораживания [Адрианова, 1992] к системе (4). Воспользуемся следующей леммой из работы [Ласунский, 2009].

**Лемма 1.** *Пусть матрица коэффициентов $A(t)$ системы* (1) *второго порядка такова, что*

$$\det A(t) \geqslant \gamma_1 > 0, \quad \mathrm{Sp\,A(t)} \leqslant -\gamma_2 < 0,$$

$$\|A(t_2) - A(t_1)\| \leqslant \delta|t_2 - t_1|$$

*с достаточно малой постоянной Липшица $\delta$, тогда характеристические показатели этой системы отрицательны.*

Применительно к системе (4) приходим к следующему утверждению.

**Теорема 4.** *Пусть система* (4) *правильна, $-M \leqslant \lambda(t) \leqslant -\varepsilon < 0$ и дополнительно функция $\lambda(t)$ удовлетворяет условию Липшица с достаточно малой постоянной Липшица, тогда положение равновесия $(x_0, s_0)$ системы* (3) *асимптотически устойчиво.*

Дадим экологическую интерпретацию полученных выше результатов. Насколько реальны для жизнедеятельности популяций условия предыдущих теорем. Условие $\lim\limits_{t\to+\infty} \lambda(t) = \lambda_0$ теорем 1 и 2 означает, что мальтузианские коэффициенты с течением времени стабилизируются. В теоремах 3 и 4 требуется правильность линейной системы. Так как линейные системы с периодической матрицей коэффициентов правильны, то это условие заведомо выполняется для случая периодического мальтузианского коэффициента $\lambda(t)$, что соответствует сезонным изменениям в природе. На первый взгляд может показаться (см. условия теоремы 4), что если функция $\lambda(t)$ удовлетворяет условию Липшица с достаточно малой постоянной Липшица, то эта функция близка к постоянной функции. Это не так. Периодическая функция $\lambda(t) = -2 + sin(\delta t)$ отделена от нуля и из теоремы Лагранжа следует, что она удовлетворяет условию Липшица с постоянной $\delta$.

В заключение этого раздела отметим, что по коэффициентному критерию [Изобов, 1972] система (4) интегрально разделена, а следовательно, характеристические показатели этой системы различны и устойчивы [Былов, 1965]. Необходимым условием правильности системы (4) является существование строгого интегрального среднего у функции $\lambda(t)$ [Демидович, 1967].

## Взаимодействующие популяции и пассивные стадии жизнедеятельности

Действие ПС-механизма в случае двух популяций более сложно, чем в случае изолированной популяции. Даже в автономном случае ПС-механизм может «испортить» устойчивую динамику взаимодействующих популяций. Устойчивая матрица

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, Sp\ A < \ 0, det\ A\ >\ 0$$

при расширении $B = \begin{pmatrix} a_{11}-q & a_{12} & p \\ a_{21} & a_{22} & 0 \\ q & 0 & -p \end{pmatrix}$ с параметрами $p, q > 0$ может стать неустойчивой. Действительно, характеристический многочлен для матрицы $B$ имеет вид

$$\lambda^3 + (p + q - SpA)\lambda^2$$
$$+(detA - pSpA - qa_{22})\lambda + pdetA = 0.$$

Нетрудно найти элементы матрицы A и значения параметров p и q, для которых $detA - pSpA - qa_{22} < 0$ . Например, можно взять $p = q = 0,5$, $a_{11} = -18$, $a_{12} = 24$, $a_{21} = 13$, $a_{22} = 17$. Так как не все коэффициенты характеристического многочлена для матрицы $B$ одного знака, то матрица $B$ неустойчива.

Рассмотрим действие ПС-механизма на примере модели хищник-жертва вида

$$\dot{x} = \alpha(t)x(1-yK^{-1}),\ \dot{y} = \beta(t)y(xL^{-1}-1). \quad (5)$$

Модель отличается от классической модели Вольтерры [Свирежев, Логофет, 1978] тем, что мальтузианские коэффициенты $\alpha(t)$ и $\beta(t)$ зависят от времени. Эта система исследована в работе [Ласунский, 2008]. Для нетривиального положения равновесия $(L, K)$ теорема Ляпунова об устойчивости по первому приближению не применима, так как среди характеристических показателей решений системы первого приближения есть хотя бы один неотрицательный, что вытекает из неравенства Ляпунова для суммы характеристических показателей.

Если отношение мальтузианских коэффициентов $\alpha(t)/\beta(t) = m$ постоянно, то нетривиальное положение равновесия системы (5) можно исследовать на устойчивость вторым (прямым) методом Ляпунова. Заменой $x - L =$

$u$, $y-K=v$ переведем положение равновесия в начало координат для системы

$$\dot u = -m\beta(t)(u+L)vK^{-1}, \quad \dot v = \beta(t)(v+K)uL^{-1}.$$

Можно убедиться, что функция $F(u,v) = (u+L)(v+K)^m \exp(-uL^{-1} - mvK^{-1})$ является интегралом этой системы. Рассмотрим функцию $V(u,v) = F(u,v) - F(0,0)$. Ясно, что $V(0,0) = 0$. Производная от этой функции в силу системы тождественно равна нулю. Функция $V$ отрицательно определена в некоторой окрестности начала координат, что следует из вида первых членов разложения функции $V$ в ряд Тейлора в окрестности $(0;0)$: $V(u,v) = -0,5K^mL^{-1}u^2 - 0,5mLK^{m-2}v^2 + \dots$ Положение равновесия $(L,\, K)$ системы (5) устойчиво по Ляпунову [Меркин, 1976]. Отметим, что если интеграл $\int_0^t \alpha(u)\,du$ неограничен, то тривиальное положение равновесия неустойчиво, так как при $y$, тождественно равном нулю, мы имеем неограниченный рост $x$.

Исследуем введение ПС-механизма у жертв $x$ и хищников $y$ на характер положений равновесия. Введение ПС-механизма для популяции жертв в модели (5) приводит к системе

$$\begin{cases} \dot x = \alpha(t)x(1-yK^{-1}) - qx + ps, \\ \dot y = \beta(t)y(xL^{-1}-1), \\ \dot s = qx - ps. \end{cases} \tag{6}$$

В случае постоянных коэффициентов $\alpha$ и $\beta$ классическая модель Вольтерры хищник-жертва (5) имеет устойчивое (типа центр) положение равновесия $(L,K)$. Введение ПС-механизма позволяет положение равновесия сделать асимптотически устойчивым. Отметим, что В. Г. Ильичевым рассматривалась задача о возможности стабилизации неустойчивой динамики взаимодействующих популяций в автономном случае при подходящем выборе параметров $p$ и $q$ в ПС-механизме.

**Теорема 5.** *Если существуют положительные пределы $\lim\limits_{t\to+\infty}\alpha(t) = \alpha$, $\lim\limits_{t\to+\infty}\beta(t) = \beta$, то нетривиальное положение равновесия $(L, K, qp^{-1}L)$ системы (6) асимптотически устойчиво при любых положительных $p$ и $q$, тривиальное же положение равновесия неустойчиво при любых положительных $p$ и $q$.*

*Доказательство.* Система первого приближения $\dot u = A(t)u$ в окрестности нетривиального положения равновесия имеет вид

$$\dot u = \begin{pmatrix} -q & -LK^{-1}\alpha(t) & p \\ KL^{-1}\beta(t) & 0 & 0 \\ q & 0 & -p \end{pmatrix} u. \tag{7}$$

Обозначим $\lim\limits_{t\to+\infty} A(t) = A_0$. Матрица $A_0$ имеет следующее характеристическое уравнение $\lambda^3 + (p+q)\lambda^2 + \alpha\beta\lambda + p\alpha\beta = 0$. Применение критерия Гурвица приводит к системе неравенств $p+q > 0$, $\quad q\alpha\beta > 0$, $\quad p\alpha\beta > 0$, следовательно, характеристические показатели предельной системы $\dot u = A_0 u$ отрицательны. Система (7) почти приводима, а значит, правильна [Адрианова, 1992]. Показатели Ляпунова в случае их устойчивости инвариантны относительно линейных возмущений, стремящихся к нулю на $+\infty$, поэтому показатели системы (7) совпадают с отрицательными показателями предельной системы. По теореме Ляпунова положение равновесия асимптотически устойчиво.

Для тривиального положения равновесия системы (7) система первого приближения имеет вид

$$\dot u = \begin{pmatrix} \alpha(t)-q & 0 & p \\ 0 & -\beta(t) & 0 \\ q & 0 & -p \end{pmatrix} u. \tag{8}$$

Характеристическое уравнение матрицы коэффициентов этой системы следующее

$$(\lambda + \beta(t))(\lambda^2 + (p+q-\alpha(t))\lambda - p\alpha(t)) = 0.$$

Если функции $\alpha(t)$ и $\beta(t)$ имеют положительные пределы при $t \to +\infty$ , то система (8) правильна, причем среди ее характеристических показателей, которые совпадают с характеристическими показателями предельной системы, есть положительный показатель. По теореме Четаева тривиальное положение равновесия системы (6) неустойчиво при любых положительных $p$ и $q$.

**Теорема 6.** *Пусть система (7) правильна, коэффициенты $\alpha(t), \beta(t) \in [\varepsilon, M]$, $\quad \varepsilon > 0$ и удовлетворяют условию Липшица с достаточно малой постоянной Липшица, тогда нетривиальное положение равновесия системы (6) асимптотически устойчиво при любых положительных $p$ и $q$.*

*Доказательство.* Для применения теоремы Ляпунова нужно убедиться, что характеристические показатели системы (7) отрицательны. Для этого воспользуемся оценкой сверху характеристических показателей, которую можно получить с помощью метода замораживания. Сначала убедимся, что величина $\gamma = \sup\limits_{t\in R_+} \max\limits_n \mathrm{Re}\lambda_n(t)$ отрицательна, где $\lambda_n(t)$, $n = 1,2,3$ – собственные числа матрицы $A(t)$. Характеристическое уравнение для матрицы $A(t)$ имеет вид

$$\lambda^3 + (p+q)\lambda^2 + \alpha(t)\beta(t)\lambda + p\alpha(t)\beta(t) = 0. \tag{9}$$

Из критерия Гурвица следует, что положительность $p, q, \alpha(t), \beta(t)$ влечет отрицательность вещественных частей собственных чисел матрицы $A(t)$ для всех $t \geqslant 0$, откуда следует лишь, что $\gamma \leqslant 0$. Но в случае $\gamma = 0$ оценка метода замораживания не проходит. Покажем, что если положительные функции $\alpha(t)$ и $\beta(t)$ ограничены и отделены от нуля, то $\mathrm{Re}\lambda_n(t) \leqslant -\varepsilon < 0, n = 1, 2, 3$, а следовательно, $\gamma < 0$.

Из ограниченности коэффициентов $\alpha(t)$ и $\beta(t)$ следует ограниченность модуля любого корня $\lambda(t)$ уравнения (9). Так как свободный член уравнения (9) отделен от нуля и все его корни ограничены, то модуль любого корня отделен от нуля. Если все корни уравнения (9) действительны, то они отрицательны по критерию Гурвица и отделены от нуля. Рассмотрим случай $\lambda_1(t) \in \mathbb{R}, \quad \lambda_{2,3}(t) = \sigma(t) \pm i\omega(t)$. Модули корней отделены от нуля, поэтому $\lambda_1(t) \leqslant -\varepsilon < 0$. Покажем, что $\sigma(t) \leqslant -\varepsilon < 0$. С одной стороны, верхний угловой минор второго порядка матрицы Гурвица для многочлена (9) равен

$$\Delta_2 = \begin{vmatrix} p+q & p\alpha(t)\beta(t) \\ 1 & \alpha(t)\beta(t) \end{vmatrix} = q\alpha(t)\beta(t)$$

и, следовательно, положителен и отделен от нуля. С другой стороны, выражая коэффициенты многочлена (9) через его корни $\lambda_1(t)$ и $\lambda_{2,3} = \sigma \pm i\omega$, для определителя $\Delta_2$ получаем

$$\Delta_2 = \begin{vmatrix} -\lambda_1 - 2\sigma & -\lambda_1(\sigma^2+\omega^2) \\ 1 & 2\sigma\lambda_1 + \sigma^2 + \omega^2 \end{vmatrix}$$

$$= -2\sigma((\lambda_1+\sigma)^2 + \omega^2) \geqslant \varepsilon > 0.$$

Так как сомножитель $(\lambda_1+\sigma)^2 + \omega^2$ ограничен, то функция $\sigma(t)$ отрицательна и отделена от нуля. Итак, постоянная $\gamma < 0$ . Так как функции $\alpha(t)$ и $\beta(t)$ удовлетворяют условию Липшица с достаточно малой постоянной Липшица, то $\|A(t_2) - A(t_1)\| \leqslant \delta|t_2 - t_1|$ с достаточно малой постоянной $\delta > 0$ . Из метода замораживания следует, что для характеристического показателя $\chi$ любого нетривиального решения системы (7) справедлива оценка $\chi \leqslant \gamma + C\delta^{1/3}$, где $C$ – некоторая постоянная. Если постоянная $\delta$ достаточно мала, то $\chi < 0$ и нетривиальное положение равновесия системы (6) асимптотически устойчиво по теореме Ляпунова для неавтономных систем.

Введение ПС – механизма для популяции хищников в модели (5) приводит к системе

$$\begin{cases} \dot{x} = \alpha(t)x(1 - yK^{-1}), \\ \dot{y} = \beta(t)y(xL^{-1} - 1) - qy + ps, \\ \dot{s} = qy - ps, \end{cases} \tag{10}$$

которая имеет нетривиальное положение равновесия $(L, K, qp^{-1}K)$. Нетрудно проверить, что характеристическое уравнение для матрицы системы первого приближения в окрестности этого положения равновесия совпадает с уравнением (9). Для тривиального положения равновесия системы (10) соответствующее характеристическое уравнение имеет вид $(\lambda - \alpha(t))(\lambda^2 + (p + q + \beta(t))\lambda + p\beta(t)) = 0$. Ясно, что справедливы утверждения, аналогичные теоремам 5 и 6 при линейной схеме перехода в пассивные стадии для популяции хищников. Условиям теорем 5 и 6 с экологической точки зрения можно придать интерпретацию, которая была дана для условий предыдущих теорем.

## Литература

*Адрианова Л. Я.* Введение в теорию линейных систем дифференциальных уравнений. СПб.: Изд-во С.-Петербургского университета, 1992. 240 с.

*Былов Б. Ф.* Почти приводимые системы дифференциальных уравнений // Сиб. мат. журн. 1962. Т. 3, № 3. С. 333–359.

*Былов Б. Ф.* О приведении системы линейных уравнений к диагональному виду // Мат. сб. 1965. Т. 67, № 3. С. 338–344.

*Былов Б. Ф. и др.* Теория показателей Ляпунова. М.: Наука, 1966. 576 с.

*Демидович Б. П.* Лекции по математической теории устойчивости. М.: Наука, 1967. 472 с.

*Изобов Н. А.* Коэффициентный признак устойчивости показателей Ляпунова двумерной линейной системы // Укр. мат. журн. 1972. Т. 24, № 3. С. 306–315.

*Ильичев В. Г.* Пассивные стадии – стабилизирующий фактор в динамических системах (на примере экологических систем) // Автоматика и телемеханика. 1992. № 12. С. 88–95.

*Ласунский А. В.* Устойчивость стационарных состояний некоторых популяционных моделей с переменными коэффициентами // Математическое моделирование. 2008. Т. 20, № 5. С. 69–77.

*Ласунский А. В.* Состояния равновесия неавтономной модели Лотки-Вольтерры при наличии убежища для жертвы // Дифференциальные уравнения. 2009. Т. 45, № 3. С. 445–448.

*Ляпунов А. М.* Собр. соч. в 6 томах. Т. 2. М.; Л.: Изд-во АН СССР, 1956. 473 с.

*Меркин Д. Р.* Введение в теорию устойчивости движения. М.: Наука, 1976. 320 с.

*Свирежев Ю. М., Логофет Д. О.* Устойчивость биологических сообществ. М.: Наука, 1978. 352 с.

*Четаев Н. Г.* Устойчивость движения. М.; Л.: Гостехиздат, 1946. 204 с.

*Coleman B. D.* Nonautonomous logistic equations as models of the adjustment of populations to environmental changes // Mathematical Biosciences. 1979. Vol. 45. P. 159–173.

## СВЕДЕНИЯ ОБ АВТОРЕ:

**Ласунский Александр Васильевич**
доцент кафедры высшей математики
Новгородский государственный университет имени Ярослава Мудрого
ул. Большая Санкт-Петербургская, 41, Великий Новгород, Россия, 173003
эл. почта: Alexandr.Lasunsky@novsu.ru
lasunskim@mail.ru
тел.: (8162) 629 968

**Lasunsky, Alexandr**
Novgorod State University
41, B.Saint Petersburgskaya St., 173003,Veliky Novgorod, Russia
e-mail: Alexandr.Lasunsky@novsu.ru lasunskim@mail.ru
tel.: (8162) 629 968

Труды Карельского научного центра РАН
№ 5. 2011. С. 45–54

*УДК 528.8+681.3:574.4*

# КЛАССИФИКАЦИЯ СКАНЕРНЫХ СНИМКОВ МЕТОДОМ МОДЕЛИРОВАНИЯ СПЕКТРАЛЬНОГО ПРОСТРАНСТВА

**П. Ю. Литинский**

*Институт леса Карельского научного центра РАН*

Описывается метод классификации снимков космических сканеров (Landsat TM/ETM+, IRS и аналогичных), основанный на моделировании спектрального пространства в 3D-осях xyz, где x и y – соответственно первая и вторая главные компоненты логарифмированной матрицы снимка (каналы G, R, NIR, SWIR2), а z – индекс стресса влажности MSI (отношение каналов SWIR1 и NIR). Декомпозиция неоднозначных спектральных классов производится по геоморфометрической модели и по разновременным данным. Метод оптимален для моделирования структуры и динамики таежных экосистем.

К л ю ч е в ы е с л о в а: дистанционное зондирование, классификация многоспектральных сканерных снимков, геоинформационное моделирование.

## P. Yu. Litinsky. MULTISPECTRAL IMAGERY CLASSIFICATION METHOD BASED ON SPECTRAL SPACE MODELING

A method for satellite imagery (Landsat TM/ETM+, IRS, etc.) classification based on spectral space modeling is described. The 3D model is built in $xyz$-*axes*, where $x$ and $y$ are the first two principal components of the image matrix in logarithmic form (bands G, R, NIR, SWIR2), and $z$ is the moisture stress index MSI (SWIR1/NIR). Ambiguous spectral classes are decomposed using the geomorphometric model and time series imagery. The method is optimal for taiga ecosystem structure and dynamics modeling.

K e y w o r d s: remote sensing, multispectral imagery classification, geoinformation modeling.

## Введение

Первый этап построения математической модели структурно-функциональной организации биосферы – создание структурной (геоинформационной) модели экосистем на основе данных дистанционного зондирования (ДЗ). Наиболее распространенный в настоящее время тип систем ДЗ – космические сканеры высокого пространственного разрешения (размер пиксела 10–30 м), такие как Landsat TM/ETM+, Spot, IRS, регистрирующие кривую спектрального отражения (сигнатуру) в нескольких диапазонах для каждого пиксела снимка (рис. 1).

Наиболее характерная особенность сигнатур всех растительных категорий земной поверхности – минимальное отражение в красном диапазоне и максимальное – в ближнем инфракрасном. Первое объясняется поглощением хлорофиллом энергетически насыщенных участков видимого спектра, второе –

изменением коэффициента преломления в губчатой паренхиме, между воздухом в межклеточных пространствах и гидратированной целлюлозой клеточных стенок [Горышина, 1979].

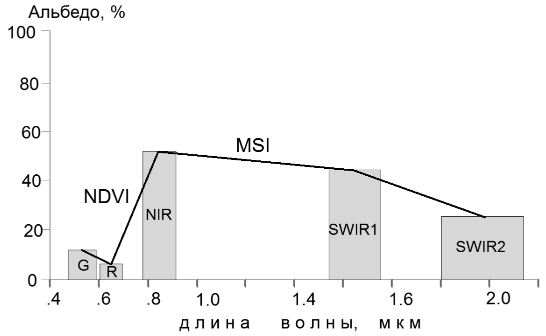



*Рис. 1.* Спектр отражения зеленой растительности и спектральные диапазоны каналов сканера Landsat-TM. Ширина столбцов диаграммы соответствует диапазонам: G – зеленый, R – красный, NIR – ближний инфракрасный, SWIR1 и SWIR2 – средневолновые инфракрасные. Названия индексов NDVI и MSI расположены над отрезками сигнатуры, определяющими величину соответствующих индексов

Поэтому наиболее общая характеристика растительного покрова на сканерном снимке выражается через отношение каналов NIR и R – так называемый вегетационный индекс NDVI, Normalized Difference Vegetation Index [Tucker, 1979; Sabins, Floyd, 1987]. Индекс рассчитывается как разность значений отражения в ближней инфракрасной и красной областях спектра, поделенная на их сумму. Для зеленой растительности отражение в красной области всегда меньше, чем в ближней инфракрасной, поэтому значения NDVI для растительности всегда больше нуля, а максимальное значение индекса при прочих равных условиях наблюдается в пикселах с максимальным количеством фотосинтезирующей биомассы. Чем ближе к вертикали отрезок сигнатуры на участке R-NIR, тем выше значение индекса.

Различия в водном режиме растений наиболее сильно проявляются в средневолновом инфракрасном диапазоне SWIR1, поэтому для выявления участков с различными по влажности условиям произрастания используется спектральный индекс стресса влажности – MSI, moisture stress index, равен отношению каналов SWIR1 и NIR [Rock et al., 1985]. Чем выше значение индекса (ближе к горизонтали соответствующий участок сигнатуры), тем больше дефицит влаги в листьях (хвое). Таким образом, наиболее информативны для классификации растительных экосистем два отрезка сигнатуры – на участке R-NIR-SWIR1 (данное «геометрическое» пояснение способствует лучшему пониманию изложенного ниже материала).

## Состояние вопроса (традиционный подход)

При создании геоинформационных моделей на основе сканерных снимков обычно используется так называемая классификация «с обучением» (иногда употребляется термин «управляемая классификация»), основанная на математической экстраполяции данных наземных ключевых участков на всю охваченную снимком территорию. В целом процесс включает следующие стадии:

– составление схемы классификации (списка целевых категорий земной поверхности);

– закладка ключевых участков в выбранных категориях;

– обработка снимка (собственно классификация);

– оценка достоверности классификации.

Проблема в составлении схемы классификации состоит в том, что реальная информация, содержащаяся в снимках, далеко не всегда согласуется с представлениями, заложенными в типологии растительности [Пузаченко и др., 1998]. На исследуемой территории всегда могут обнаружиться категории с неизвестными ранее характеристиками.

Методы обработки снимков, применяемые для управлемой классификации, основаны на стандартных методах многомерной статистики – сигнатуры сканерного снимка рассматриваются как объекты многомерного спектрального пространства (СП). Наиболее достоверным принято считать метод максимального подобия [Richards, Xiuping, 1999]. Классификация проводится по максимуму плотности вероятности принадлежности сигнатуры к той или иной категории, т. е. метод является вариантом дискриминантного анализа и, соответственно, для него необходимо статистически достоверное количество сигнатур каждой категории. Поэтому, как отмечают многие исследователи, приемлемая достоверность классификации достигается только при весьма значительном количестве правильно подобранных ключевых участков [Замятин, 2006; Шаталов и др., 2007]. Применяется также метод гиперпараллелепипедов, метод наименьшего евклидова расстояния (или расстояния Махаланобиса при достаточном количестве ключевых участков), метод множественной регрессии и другие.

Для оценки достоверности полученных результатов проводится выборочная проверка на местности, которая по сути дела показывает, какой процент *площади* каждой категории соответствует категории ключевого участка (каппа-статистика). Насколько при этом соответствует действительности реальная *форма* объектов, остается неизвестным. Назначение геоинформационной модели – быть структурной основой для создания функциональной модели, в которой важнейшее значение приобретает характер территориальной сопряженности, определяющий межэкосистемные связи, и, следовательно, необходимо точное отображение структуры.

Нужно отметить, что при наличии ключевых участков, представляющих *все* категории земной поверхности: леса различного типа, состава, возраста, полноты и т. д., типы болот, вырубок, сельхозугодий и т. д., плотно заполняющие все СП, и в количестве, обеспечивающем статистическую достоверность для каждой категории, с практически одинаковым результатом может использоваться любой из вышеуказанных методов обработки, однако на практике трудно обеспечить наличие такого количества ключевых участков.

Проблема заключается также в том, что вследствие технического ограничения приборов ДЗ (регистрируется интегральная информация участка поверхности Земли площадью несколько сотен квадратных метров) одна и та же сигнатура может соответствовать нескольким принципиально различным категориям, в связи с чем делается вывод о недостаточно высокой точности подобных методов [Ершов, Барталев, 1998], а некоторые авторы заявляют об их несостоятельности [Пузаченко и др., 1998].

В ходе наших работ по применению дистанционных методов для изучения структуры и динамики таежных экосистем [Литинский, 1996, 2003, 2007] сформировалось понимание основного недостатка традиционного подхода: при любом методе обработки и даже при наличии большого количества ключевых участков СП остается «черным ящиком». Исследователь не может оценить реальные размеры и раположение категорий в СП, их пространственную сопряженность, т. е. по сути дела действует вслепую. То есть основная проблема состоит не в выборе метода обработки снимка, а в подборе ключевых участков.

## Описание метода

При создании геоинформационной модели, включающей все категории земной поверхности, необходимо принципиально иначе подойти к классификации – визуализировать все СП (т. е. создать его модель), исследовать закономерности его организации, чтобы целенаправленно подойти к выбору ключевых участков. Самый простой способ это сделать – трансформировать многомерную матрицу сканерного снимка методом главных компонент [Richards, Xiuping, 1999]. Величины яркости различных каналов в высокой степени взаимно коррелируют, поэтому на снимках Landsat 7 различных типов таежных ландшафтов суммарная доля первых двух компонент составляет 96–98 % вариации [Литинский, 2007]. Доля третьей и последующих компонент сопоставима с измерительной погрешностью сканера. Теоретически это позволяет предположить, что с минимальными потерями информации абстрактное многомерное СП можно представить в наиболее удобном для анализа виде – в двухмерной прямоугольной системе координат (ось х – первая компонента, у – вторая).

При размещении сигнатур в осях компонент происходит их автоматическая ординация – сходные по форме сигнатуры располагаются рядом друг с другом, а наиболее контрастные – по различным краям поля распределения, образуя проекцию спектрального гиперэллипсоида, визуальный образ СП (рис. 2). Размерность осей компонент на рисунке не указана, поскольку в данном случае их конкретные величины значения не имеют.

Суммарная величина формирующих сигнатуру значений (общая яркость) возрастает слева направо, вдоль первой главной компоненты, вторая компонента характеризует качественные различия категорий. Конфигурация спектрального гиперэллипсоида несколько меняется в зависимости от типа сканера, сезона съемки и особенностей территории, однако существуют и четкие общие закономерности. Водные поверхности находятся в его левой, наиболее «темной» части. Области лишенных растительности категорий (с отрицательным значением индекса NDVI, т. е. с превышением уровня красного канала над ближним инфракрасным) располагаются вдоль верхней части эллипсоида, а по направлению вправо и вниз значения индекса возрастают, индицируя увеличение фотосинтезирующей биомассы.

Несмотря на удовлетворительную, на первый взгляд, ординацию, результаты классификации снимка показали, что в некоторых случаях в одной и той же области СП находятся принципиально различные категории – например, сельскохозяйственные земли (луга) и некоторые участки верховых болот, причем различия в оттенках этих категорий заметны даже на RGB-растре, синтезированном из инфракрасных каналов. Это свидетельствует о том, что фактически СП не плоское, а имеет некое третье измерение.

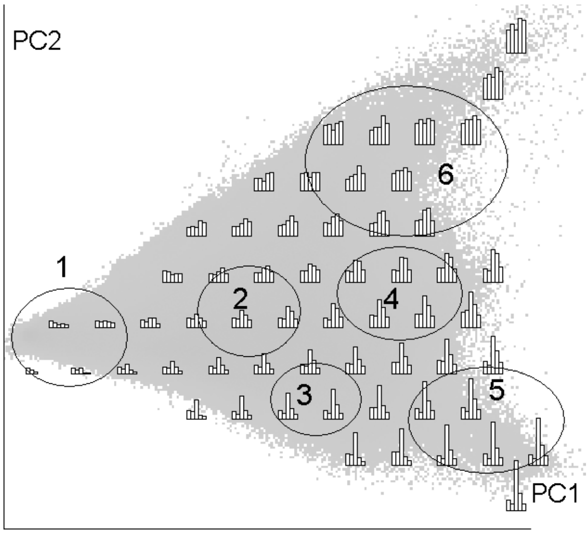



*Рис. 2.* Схема ординации сигнатур снимка таежной территории в двухмерном пространстве главных компонент PC1 и PC2. Цифрами обозначены области: 1 – водная поверхность, 2 – хвойные леса, 3 – хвойно-лиственные леса, 4 – кустарники, 5 – травянистая растительность, 6 – лишенные растительности категории. Диаграммы сигнатур показывают относительный уровень яркости каналов G, R, NIR, SWIR1, SWIR2 слева направо

Для выяснения физического смысла этого измерения были выведены в виде графика все сигнатуры снимка, соответствующие нескольким определенным точкам СП (рис. 3, b, c, d). Оказалось, что в каждой точке находятся существенно различные сигнатуры, особенно четко это выражено в точке, которой соответствует график d, где явно просматриваются две категории – вышеупомянутые луга и болота.

Отрезки графиков на участке R-NIR-SWIR1 сходятся в точке NIR – т. е. в каждой точке СП находятся сигнатуры с существенно различными значениями индексов как NDVI, так и MSI, и поэтому ординацию нельзя признать удовлетворительной. Разброс сигнатур по всем каналам, кроме NIR, показывает, что для решения задачи классификации требуется ввести третье измерение, но физический смысл этого измерения не поддается однозначной интерпретации.

Причина такого характера организации СП в том, что растительные категории занимают подавляющую часть любого снимка таежных территорий, а уровень ближнего инфракрасного канала в данных категориях значительно больше, чем у остальных каналов. Компонентный анализ «считает» эту особенность наиболее важной характеристикой сигнатуры и помещает объекты с принципиально различными биологическими характеристиками в одну область пространства.

Уменьшить влияние высокого уровня ближнего инфракрасного канала на положение сигнатуры в СП, максимально сохранив при этом индивидуальные особенности сигнатур, и повысить таким образом дифференциацию категорий, можно логарифмированием исходных значений матрицы снимка. При этом СП в целом несколько «сжимается» по вертикали, левая («водная») его часть, наоборот, становится более объемной, в средней части существенных изменений не происходит (рис. 4, a).

Более всего трансформируется зона высоких уровней инфракрасного канала – перекрывавшиеся в исходном пространстве категории не только отделяются друг от друга, но и их положение в СП становится более компактным (рис. 4, b, c, d). Отрезки сигнатур на участке R-NIR почти параллельны, следовательно, СП логарифмированной матрицы удовлетворительно упорядочено по значениям индекса NDVI. На участке NIR-SWIR1 наблюдается максимальная вариация значений сигнатур, т. е. третья размерность СП становится четко определенной – она соответствует индексу влажности MSI, и таким образом 3D-пространство в осях PC1-PC2-MSI может быть использовано для создания модели СП сканерного снимка.

## Программная реализация метода

Формирование модели СП, ее анализ и последующая классификация снимка могут быть реализованы в профессиональных пакетах обработки ДЗ (например, ERDAS), включающих классификацию методом feature space, который позволяет создавать полигоны в произвольной форме в двухканальном поле распределения сигнатур. Затем производится разделение перекрывающихся сегментов по значению MSI с использованием растровой алгебры, поэтому весь процесс представляет собой достаточно сложную последовательность операций. Для его упрощения нами разработан программный модуль, с помощью которого все операции осуществляются автоматически, и создается модель СП в наглядном и удобном для анализа и корректировки виде.

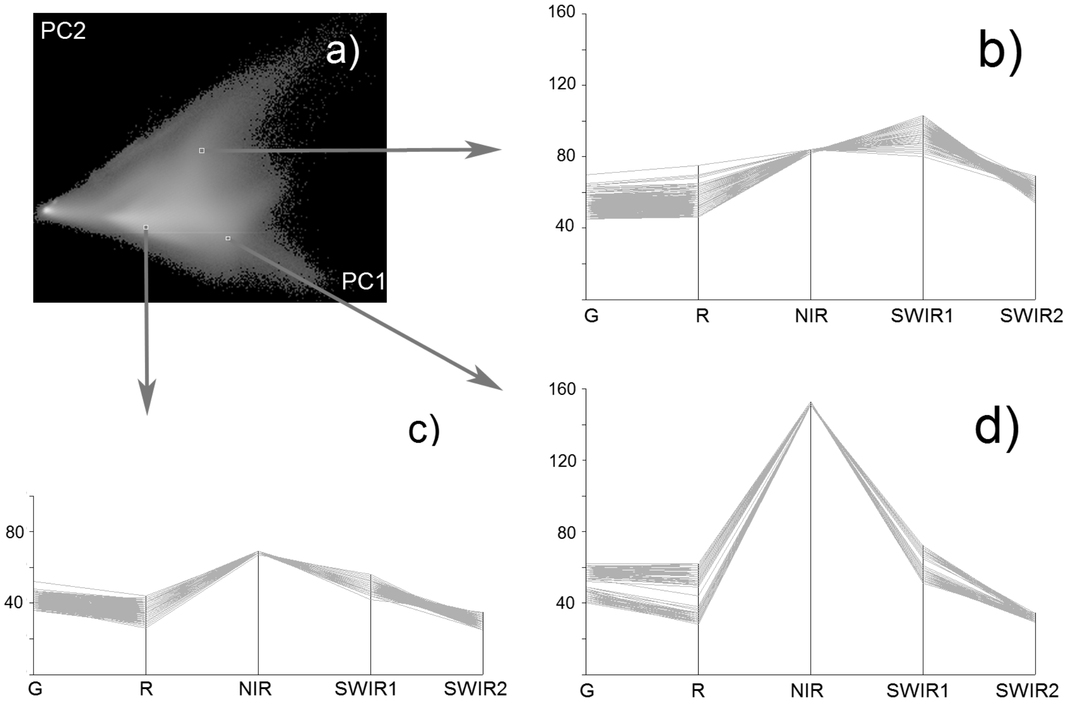



*Рис. 3.* Спектральное пространство в осях компонент (a) и сигнатуры отдельных точек (b, c, d)

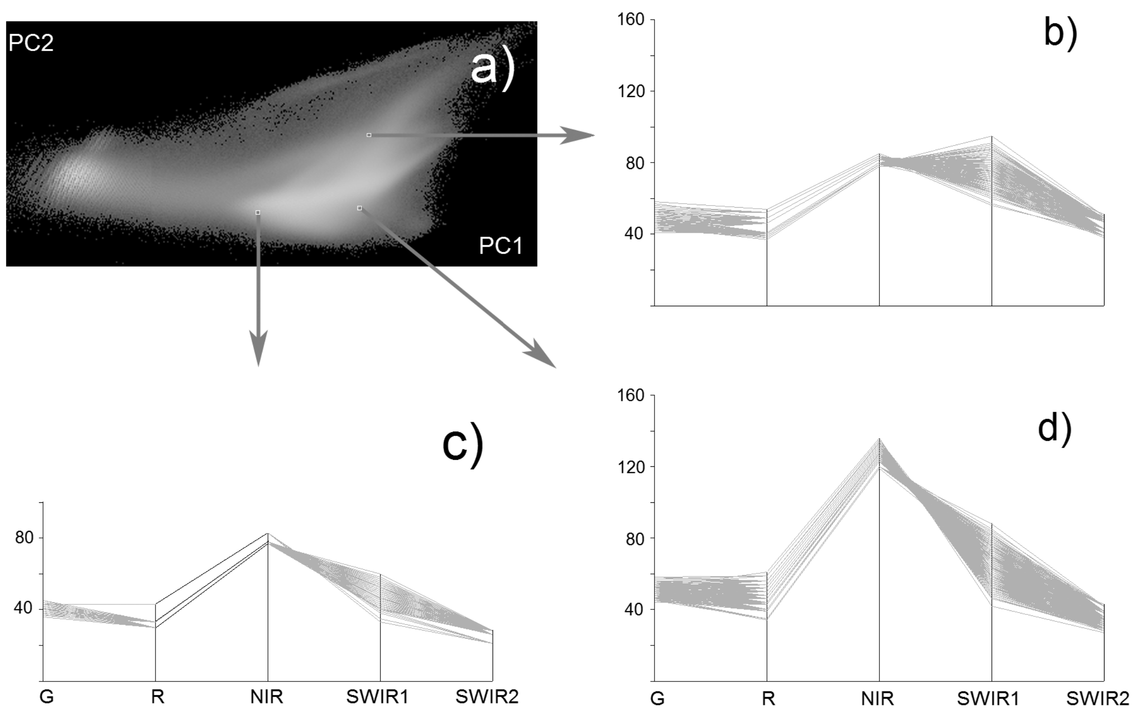



*Рис. 4.* Логарифмированное СП (a) и сигнатуры отдельных точек (b, c, d)

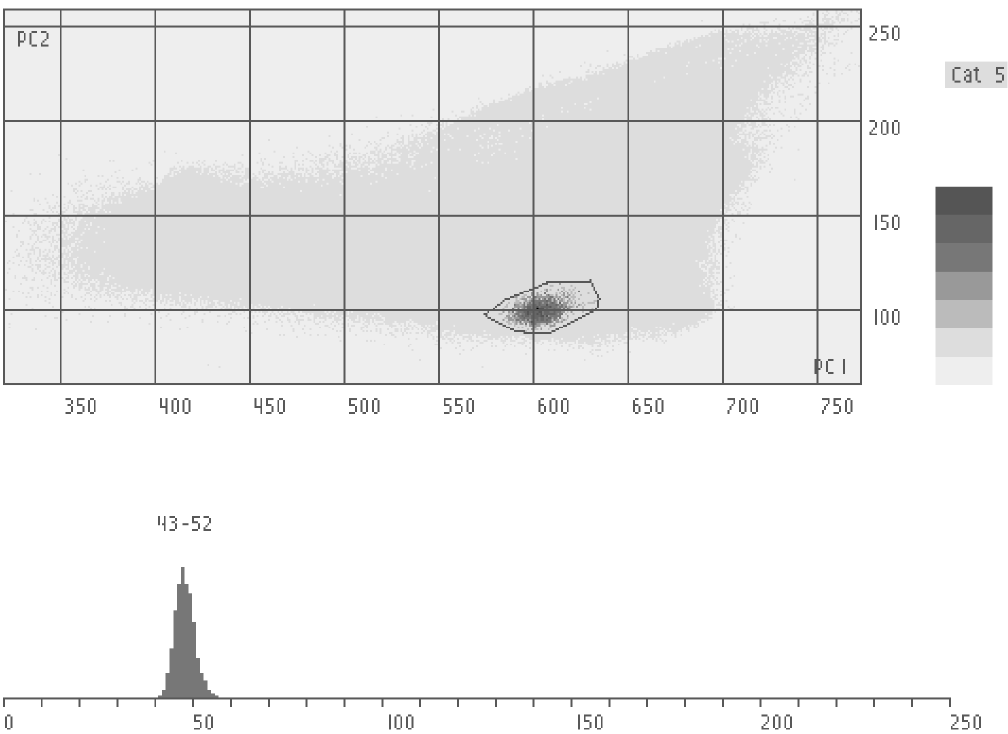



*Рис. 5.* Вид растрового файла, формируемого модулем создания модели СП. Вверху – поле распределения сигнатур категории в осях PC1 и PC2 и полигон основания призмы, внизу – гистограмма MSI (призма категории 5 на рис. 6)

Модуль работает в двух режимах: 1) создания модели СП и 2) собственно классификации. Для работы модуля необходимы растровые файлы компонент PC1 и PC2, файл индекса MSI и файл ключевых участков.

В первом режиме работы модуля при считывании значений PC1, PC2 и MSI для каждой категории ключевого участка создается поле распределения сигнатур в осях PC1 и PC2, и гистограмма значений MSI, которые сохраняются в виде растрового файла *catN.bmp*, где N – номер категории ключевого участка (рис. 5).

Растр позиционирован в координатах PC1-PC2, привязка записывается в соответствующий файл bpw (координаты верхнего левого угла растра равны минимальному значению 1-й компоненты (x) и максимальному – второй (y), размер пиксела равен единице). «Ширина» и «высота» поля распределения равны интервалам значений компонент от минимума до максимума. Так как логарифмированные значения компонент обычно не превышают 10, для зрительного увеличения пространства (размера поля распределения в пикселах) применяется коэффициент (50–150). Создается также растр поля распределения всех сигнатур снимка *cat0.bmp*, на растрах категорий его очертания показаны серым фоном.

Одновременно создается общий для всех категорий ключевых участков векторный файл полигонов (*shape*-файл ESRI), оконтуривающих точки растра СП с частотой встречаемости выше заданной (ГИС-оператор *convex hull*), посредством чего отсекаются внешние пологие участки колокола распределения. Атрибутивная информация полигонов – номер категории ключевого участка и интервал значений индекса MSI – сохраняется в файле *dbf*, создается также соответствующий индексный файл *shx*. Данный набор файлов и представляет собой векторное представление модели СП снимка, он может вводиться в любой ГИС-пакет для анализа и редактирования, или же сразу использоваться для классификации. Таким образом, каждая категория земной поверхности представлена в модели в виде призматической области СП, в которой основания призмы описываются полигонами векторного файла, а ее высота и положение в оси MSI – атрибутивными значениями полигона (рис. 6).

В режиме классификации сначала векторный файл преобразуется в 8-битовый 3D-растр модели СП, в котором область (призма) каждой категории образована байтами, значения которых соответствуют номеру категории. Затем при считывании из файлов значений PC1, PC2 и MSI (x, y, z) пикселу растра результата

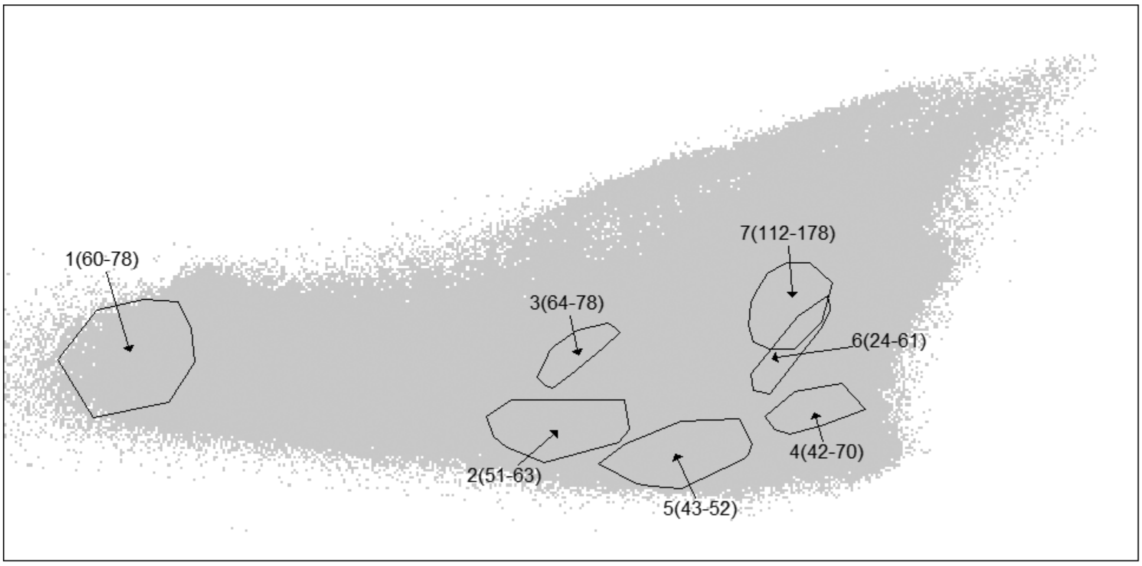



*Рис. 6.* Отображение векторного файла модели СП. Цифрами обозначены номера категорий и в скобках – интервалы значений MSI. Категории, пересекающиеся в проекции на плоскость осей компонент, разделяются по положению в оси MSI

классификации (формат 8-битовый RAW) присваивается значение байта модели СП, находящемуся в соответствующих координатах 3D-растра. Таким образом, данный способ классификации близок к известному методу параллелепипедов, но призма более точно отграничивает положение категории в СП.

Растровые файлы, необходимые для работы модуля, могут быть созданы любым из вышеупомянутых пакетов обработки данных ДЗ. Модуль тестировался в комплексе со свободно распространяемым пакетом Quantum GIS (http://qgis.org), который включает векторную ГИС (упрощенный аналог ArcMap ESRI), а также растровый пакет GRASS, содержащий широкий набор модулей анализа геоинформации (но лишь один модуль для управляемой классификации – методом максимального подобия). В векторной ГИС создается *shape*-файл полигонов ключевых участков и производится анализ и при необходимости редактирование файла модели СП. Файлы компонент создаются модулем GRASS **i.pca**, файл MSI – модулем **r.mapcalc**, растр ключевых участков – модулем **v.to.rast** из *shape*-файла их полигонов.

Модуль имеет 10 параметров запуска. Первый – режим работы: 1 – создание модели СП, 2 – классификация. Второй параметр зависит от режима, в первом – это имя файла ключевых участков, во втором – имя shape-файла модели СП. Остальные восемь параметров одинаковы в обоих режимах. Это имена файлов компонент с соответствующими интервалами их значений (min-max), имя файла MSI, число рядов и колонок растра. Интервалы компонент и размер растра можно получить, используя модуль **r.info**. Все файлы должны иметь одинаковую размерность по числу рядов и колонок растра.

## Обсуждение

Рассматриваемый комплексный метод классификации сканерных снимков, основанный на моделировании СП, включает три основных составляющих:

- способ трансформации СП;
- способ выделения областей СП;
- принцип работы с моделью СП (методика создания геоинформационной модели).

Применение для построения модели СП стандартного метода сокращения размерностей – компонентного анализа – не дает удовлетворительных результатов. Визуальный анализ структуры гиперэллипсоида снимка, проецированного на плоскость осей двух первых компонент, показал, что сигнатуры упорядочены, в первую очередь, по уровню ближнего инфракрасного канала, что не дает возможности четко определить физический смысл явно выраженного третьего измерения СП. Кроме того, более «яркие» категории (болота, луга, вырубки) занимают большее место, чем «темные» лесные, что не соответствует степени их реальной гетерогенности.

Для выравнивания размеров различных по яркости областей СП было применено логарифмирование значений сигнатур, и оказалось, что при этом пространство структурируется таким образом, что его третья размерность становится четко определенной – это индекс стресса влажности. Таким образом, трансформация СП для построения модели заключается в расчете двух первых главных компонент логарифмированной матрицы снимка и индекса влажности в качестве третьей размерности.

Физический смысл осей данной трансформации, таким образом, близок к таковому у классической трансформации Tasseled Cap – brightness, greeness, wetness (яркость, «зеленость», влажность) [Kauth, Thomas, 1976]. Точнее, в логарифмированном пространстве компонент «зеленость» (относительное количество фотосинтезирующей биомассы) увеличивается примерно по диагонали, от его верхнего левого угла к нижнему правому, однако для разделения сигнатур в пространстве это принципиального значения не имеет. СП в трансформации Tasseled Cap также упорядочено по значениям и NDVI, и MSI, однако она была разработана для классификации сельскохозяйственных культур и поэтому, напротив, еще более увеличивает размер соответствующей области (категория 5 на рис. 2), что совершенно нецелесообразно при классификации таежных экосистем, где, наоборот, в первую очередь, нужно максимально «развернуть» лесную область.

При данной трансформации визуализация областей СП в виде поля распределения в осях PC1 и PC2 и гистограммы MSI ( см. рис. 5) дает возможность оценивать качество подбора ключевых участков – большой разброс точек поля и растянутая гистограмма свидетельствуют о том, что ключевой участок в действительности включает несколько различных категорий.

Построение модели СП в виде призм, описываемых shape-файлом, в автоматическом или ручном режиме, обеспечивает достаточно удобное, наглядное и свободное управление классификацией. Она перестает быть однонаправленной – от ключевых участков ко всему снимку. Принципиальное отличие от традиционного подхода заключается в том, что по модели СП можно оценить степень представленности категорий ключевых участков (свободные места на рис. 6) и, сформировав соответствующую призму, узнать, где территориально расположены категории, для которых на данном этапе нет наземных данных. Этим обеспечивается эффективный механизм обратной связи при подборе ключевых участков.

Становится также возможной интерполяция промежуточных характеристик: логично предположить, что в промежутке СП между сегментами с известными свойствами будут располагаться области переходных зон (экотонов). В стандартных пакетах анализа ландшафтной структуры (например, модули пакета GRASS с общим префиксом **r.le** экотон рассматривается как линейный объект, а при использовании модели появляется возможность исследования экотонов как полигональных объектов без закладки соответствующих ключевых участков. Таким образом, классификация становится в полном смысле управляемой.

Данные возможности определяют и методику создания геоинформационной модели. Вместо разовой акции она превращается в постепенный исследовательский процесс выяснения соответствия между областями СП и наземными данными. Схема классификации перестает быть наперед заданной, а формируется и уточняется в процессе моделирования СП, иными словами, модель и представляет собой схему классификации. Меняется также роль выборочной оценки достоверности – она становится не конечным, а промежуточным этапом постепенного формирования геоинформационной модели.

Так, нередко в процессе создания модели достаточно обширной территории анализ наземных данных показывает, что одну и ту же область СП вследствие ограниченной разрешающей способности сканера занимают весьма различные по некоторым параметрам категории. Количество и структура фотосинтезирующей биомассы, ее влагообеспеченность – основные факторы, определяющие форму сигнатуры – в данных категориях практически одинаковы. Соответственно, исчерпаны и возможности для математической дифференциации категорий по положению в СП.

Практически идентичны по форме сигнатуры некоторых участков болот и старых вырубок с кустарниковой растительностью. Следует отметить, что для некоторых задач – например, при оценке запаса фитомассы в данный момент времени, эти категории могут совпадать, но с точки зрения структурно-функциональной организации это принципиально различные категории. Дифференциацию этих категорий в данном случае можно осуществить известным методом использования разновременных снимков. Если в прошлом в данном месте был лес, то данная категория является вырубкой, если же она остается неопределенной, следует рассмотреть траекторию ее движения в СП. Изменения сигнатур болот в основном имеют сезонный характер, в зависимости от температурного режима и осадков. Динамика сигнатур вырубок принципиально иная. После рубки исходная форма сигнатуры (см. рис. 1) резко изменяется – значение красного канала становится больше, чем зеленого, так как фотосинтезирующая биомасса практически отсутствует, возрастают и уровни всех каналов, поскольку нет поглощения лесным пологом. С появлением растительности форма сигнатуры начинает приближаться к исходной, но уровни всех каналов остаются высокими, с течением времени постепенно уменьшаясь. Эти изменения обуславливают и соответствующий дрейф вырубки в СП – сигнатура сначала выходит за пределы «растительной» области (категория 2 на рис. 2), оказывается в области номер 6, и затем медленно возвращается обратно, проходя при этом на стадии «старая вырубка» область, занимаемую также некоторыми участками болот (категория 4 на рис. 2). Следовательно, если на разновременных снимках наблюдаются небольшие колебания положения участка «неопределенной» категории в СП – то это болото, если же сигнатура медленно перемещается по описанной выше траектории – участок является вырубкой. Таким образом одновременно и выявляется динамика растительного покрова, и детализируется его структура.

Одинаковы также сигнатуры некоторых типов низкополнотных сосняков и ельников. Объясняется это тем, что сканер регистрирует интегральное отражение пиксела площадью несколько сотен квадратных метров, и в случае редкого древостоя большую роль играет отражение и от напочвенного покрова, нивелирующего отражение от крон. В этом случае разновременные данные бесполезны, декомпозиция может проводиться путем привлечения данных радарной съемки и цифровых моделей высот (используется экологическая приуроченность различных типов леса к формам рельефа). В перспективе представляется очень эффективным использование данных лидарной съемки [Данилин и др., 2005], позволяющей получить не только детальную модель рельефа, но и данные об архитектонике древесного полога, однако рассмотрение этих вопросов выходит за рамки данной статьи.

Таким образом, при создании геоинформационной модели необходимо исследование технических возможностей конкретного сканера для выявления различных категорий на конкретной территории, и метод моделирования СП является инструментом решения этой задачи.

И последнее – по порядку, но не по значимости. Назначение геоинформационной модели – служить структурной основой для построения комплексной структурно-функциональной newline модели биосферы, и метод классификации, основанный на моделировании СП, предоставляет необходимую для моделирования возможность простого математического описания качественных характеристик категорий посредством локализации их положения по координатам спектрального пространства [Литинский, 2007].

## Заключение

Описанный метод предназначен для создания геоинформационной модели таежных территорий на основе информации наиболее распространенных в настоящее время космических сканеров Landsat TM/ETM+ и аналогичных, и ориентирован на извлечение максимума информации о структуре и динамике наземных экосистем с учетом ограничений, накладывемых пространственным и спектральным разрешением этих приборов. В основе метода – построение 3D-модели СП сканерного снимка, которая позволяет целенаправленно подойти к подбору ключевых участков (составлению схемы классификации) и выявить области СП с неоднозначным биологическим содержимым.

Для создания модели предлагается способ трансформации СП методом главных компонент для логарифмированной матрицы снимка. Третья размерность – индекс влажности MSI. Данная трансформация более оптимальна для таежных экосистем по сравнению с классической трансформацией Tasseled Cap.

Процесс классификации включает следующие стадии:

1. логарифмирование матрицы снимка;
2. создание растровых файлов двух главных компонент, индекса MSI и ключевых участков;
3. формирование векторного файла модели СП, его анализ и редактирование;
4. классификация по модели СП.

Для реализации стадий 3 и 4, не включенных в стандартные пакеты обработки данных ДЗ, разработана соответствующая программа (свид. о гос. регистрации № 2011613929 от 20.05.2011).

Однако классификация снимка – лишь первый, хотя и самый важный этап создания геоинформационной модели. Информация сканирования дает наиболее генерализованные классы растительного покрова, которые невозможно получить с помощью любых других систем ДЗ. Но в различных типах ландшафта одну и ту же область СП могут занимать категории с весьма различными экологическими характеристиками. Декомпозиция неоднозначных сегментов может проводиться путем анализа траекторий категорий на разновременных снимках, с использованием модели рельефа и данных других систем ДЗ (радаров, лидаров), анализа текстуры и контекста, что и обозначает направления дальнейших исследований.

## Литература


*Горышина Т. К.* Экология растений. М.: Высш. школа, 1979. 368 с.

*Данилин И. М., Медведев Е. М., Мельников С. Р.* Лазерная локация Земли и леса.: Учеб. пособие. Красноярск.: Институт леса им. Сукачева. 2005. 182 с.

*Ершов Д. В., Барталев С. А.* Использование метода декомпозиции спектральных смесей при обработке многоспектральных спутниковых изображений в решении задач мониторинга лесов // Материалы второго Всерос. совещ. «Аэрокосмические методы и геоинформационные системы в лесоведении и лесном хозяйстве». М., 1998. С. 118–119.

*Замятин А. В.* Анализ динамики ландшафтного покрова на основе данных дистанционного зондирования Земли // Исследование Земли из космоса. 2006. № 6 С. 50–64.

*Литинский П. Ю.* Оценка динамики деградации лесов в зоне воздействия выбросов Костомукшского ГОКа дистанционными методами // Проблемы антропогенной трансформации лесных биогеоценозов Карелии. Петрозаводск: КарНЦ РАН, 1996. С. 182–192.

*Литинский П. Ю.* Ландшафтно-экологическая ГИС Восточной Фенноскандии // Труды Карельского научного центра РАН. Вып. 5. Петрозаводск: КарНЦ РАН, 2003. С. 100–107.

*Литинский П. Ю.* Трехмерное моделирование структуры и динамики таежных ландшафтов. Петрозаводск: КарНЦ РАН, 2007. 107 с.

*Пузаченко Ю. Г., Алещенко Г. М., Молчанов Г. С., Пузаченко А. Ю.* Анализ аэрофотоизображения для выделения типов территориальных структур // Материалы второго Всерос. совещ. «Аэрокосмические методы и геоинформационные системы в лесоведении и лесном хозяйстве». М., 1998. С. 156–159.

*Шаталов А. В., Жирин В. М., Сухих В. И. и др.* Анализ информативности космических снимков высокого разрешения QuickBird // Междунар. конф. «Аэрокосмические методы и геоинформационные технологии в лесоведении и лесном хозяйстве». М., 2007. С. 168–174.

*Kauth R. J., Thomas G. S.* The Tasseled Cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat // Proceedings on the Symposium on Machine Processing of Remotely Sensed Data, 4b: 41-51, 6 June – 2 July 1976 (West Lafayette, Indiana: LARS, Purdue University).

*Richards J. A., Xiuping Jia.* Remote Sensing Digital Image Analysis. Berlin, Springer, 1999. 400 p.

*Rock B., Williams D., Vogelman J.* Field and airborn spectral characterization of suspected acid deposition damage in red spruce (Picea rubens) from Vermont. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Indiana, 1985. P. 71–81.

*Sabins Floyd F., Jr.* Remote Sensing Principles and Interpretation. New York: W. H. Freeman & Co, 1987. 449 p.

*Tucker C. J.* Red and Photographic Infrared Linear Combination for Monitoring Vegetation. Remote Sensing of Environment, 1979. Vol. 8. P. 127–150.


## СВЕДЕНИЯ ОБ АВТОРЕ:


**Литинский Петр Юрьевич**
старший научный сотрудник
Институт леса КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: litinsky@krc.karelia.ru
тел.: (8142) 768160

**Litinsky, Peter**
Forest Research Institute, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: litinsky@krc.karelia.ru
tel.: (8142) 768160

Труды Карельского научного центра РАН
№ 5. 2011. С. 55–62

*УДК 519.872.6*

# СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГАУССОВСКОЙ ОЧЕРЕДИ

**О. В. Лукашенко[1], Е. В. Морозов[1], М. Пагано[2]**

[1] *Институт прикладных математических исследований Карельского научного центра РАН*
[2] *Факультет информационного инжиниринга, Университет Пизы, Италия*

Статья посвящена оцениванию некоторых характеристик гауссовских очередей (систем с гауссовским входным потоком). Основное внимание уделено системе с фрактальным броуновским входным процессом (фрактальное броуновское движение, ФБД). Исследуется аппроксимация вероятности большого уклонения процесса загрузки. Методом имитационного моделирования анализируется гауссовская очередь в режиме большого числа входящих потоков. Работа выполнена при поддержке РФФИ, проект 10-07-00017.

Ключевые слова: гауссовская очередь, структура зависимости, фрактальное броуновское движение.

**O. V. Lukashenko, E. V. Morozov, M. Pagano. SIMULATION OF THE GAUSSIAN QUEUE**

The paper is devoted to the dependence structure of Gaussian queues (queues fed by Gaussian input). The focus is on the system with fractional Brownian input (fractional Brownian motion, fBm). An approximation of the probability of large deviation is investigated. A Gaussian queue in the multiple-source mode is analyzed by simulation. The study was supported by RFBR, project 10-07-00017.

Key words: Gaussian queue, dependence structure, fractional Brownian motion.

---

## Введение

В настоящее время гауссовские модели систем обслуживания достаточно широко используются, поскольку позволяют адекватно описать входной поток для широкого класса сетевых узлов [Norros, 1994; Kilpi, Norros, 2002; Mandjes, 2007]. В частности, они позволяют включить в моделирование самоподобие и долговременную зависимость, которые играют важную роль в описании динамики современных телекоммуникационных систем. Напомним, что самоподобие означает, что поведение процесса качественно не меняется при изменении масштаба времени, а долговременная зависимость означает столь медленное убывание автокорреляционной функции, что автокорреляционный ряд оказывается несуммируемым. Самым известным и наиболее хорошо изученным гауссовским процессом является фрактальное броуновское движение (ФБД), которое обладает свойствами самоподобия и долговременной зависимости. В частности, ФБД возникает при суперпозиции большого числа независимых так называемых on/off-источников с тяжелыми хвостами на больших масштабах времени. (Более подробно см. [Taqqu et al., 1995].)

Следует отметить, что гауссовские очереди (очереди с гауссовским входным потоком) обладают очень сложной структурой зависимости. Этот факт не позволяет в явном виде получить выражения для различных стационарных характеристик такой очереди, в частности, для вероятности переполнения, т. е. вероятности превышения некоторого уровня $b$. Этот уровень принято ассоциировать с конечным размером буфера, хотя физически буфер отсутствует и поступающая информация не теряется. Отсутствие точных аналитических результатов вызывает необходимость исследования асимптотик соответствующих характеристик. Применительно к очередям обычно выделяют два типа асимптотик: асимптотики при растущем буфере $b$, а также асимптотики в режиме многих источников: число гауссовских источников растет и пропорционально растут размер буфера и скорость обслуживания. Например, для вероятности переполнения результаты вида

$$\begin{gathered} \mathsf{P}\left(Q > b\right) \sim f_1(b), \;\; b \to \infty \\ \mathsf{P}\left(Q^n > nb\right) \sim f_2(n), \;\; n \to \infty, \end{gathered}$$

где $a \sim b$ означает, что $a/b \to 1$, называются точными асимптотиками, здесь $f_1, f_2$ – некоторые явно заданные функции, $Q$ – процесс загрузки. Иногда удается получить лишь логарифмические асимптотики:

$$\begin{gathered} \ln \mathsf{P}\left(Q > b\right) \sim f_3(b), \;\; b \to \infty \\ \ln \mathsf{P}\left(Q^n > nb\right) \sim f_4(n), \;\; n \to \infty, \end{gathered}$$

дающие менее полную информацию о вероятности переполнения. Отметим наиболее важные работы [Duffield, O'Connell, 1995; Narayan, 1998; Hüsler, Piterbarg, 1999; Debicki, Mandjes, 2002], посвященные этой проблематике. В работе [Es-Saghouani, Mandjes, 2009] предпринята попытка исследовать структуру зависимости гауссовской очереди. Поскольку выражение для корреляционной функции гауссовской очереди в явном виде получить пока не удалось, в [Es-Saghouani, Mandjes, 2009] введена и исследована иная (но близкая) мера зависимости, также основанная на двумерном распределении процесса загрузки. Для этой функции найдена асимптотика для случая большого числа источников. В данной статье мы постарались найти такое количество источников (приемлемый уровень агрегирования), когда указанные асимптотические результаты имеют достаточную точность. В статье также приведены результаты имитационного моделирования по оценке вероятности превышения заданного уровня в гауссовской очереди с ФБД в качестве входного потока.

## Гауссовские очереди

Пусть $A(t)$ есть объем трафика, поступающего в систему на интервале времени $[0, t]$, и пусть $A(s,t) = A(t) - A(s)$, $t > s > 0$. Заметим, что $A(-t) =_d -A(t)$, где $=_d$ означает равенство по распределению. Будем предполагать, что входной поток (входной трафик) имеет следующий вид [Norros, 1994]:

$$A(t) = mt + \sqrt{am} B_H(t), \tag{1}$$

где $m > 0$ – средняя интенсивность трафика, $B_H(t)$ – ФБД, описывающее флуктуации трафика вокруг линейно возрастающего среднего, $1/2 < H < 1$ – так называемый параметр Херста, $a > 0$ – некоторая константа. Предполагается, что система имеет одно обслуживающее устройство с постоянной скоростью обслуживания $C > m$. Процесс загрузки (в стационарном режиме) для такой системы определяется следующим образом (см., например, [Mandjes, 2007]):

$$Q_t =_d \sup_{s \leqslant t} \left( A(s,t) - C(t-s) \right),$$

Ввиду стационарности $Q_t =_d Q_s$ для любых $t$, $s$. Для удобства иногда зависимость $Q_t$ от $t$ будем опускать. Далее предполагаем, что имеется $n$ независимых, одинаково распределенных гауссовских источников описанного выше вида, а скорость обслуживания пропорциональна количеству источников, т. е. $C = nc$. (Это обычное предположение при асимптотическом анализе системы с растущим числом источников.) Тогда с учетом (1) стационарный процесс загрузки в момент времени $t$ удовлетворяет следующим стохастическим равенствам:

$$\begin{aligned} Q_t^n \;\; &=_d \;\; \sup_{s \leqslant t} \left( \sum_{i=1}^n A_i(s,t) - nc(t-s) \right) \\ &=_d \;\; \sup_{s \leqslant t} \left( \sum_{i=1}^n \sqrt{am} B_H^i(s,t) - nr(t-s) \right) \\ &=_d \;\; \sup_{s \leqslant t} \left( \sqrt{amn} B_H(s,t) - nr(t-s) \right) \\ &=_d \;\; \sup_{s \geqslant 0} \left( \sqrt{amn} B_H(s) - nrs \right), \end{aligned}$$

здесь $B_H^i(u,v) = B_H^i(v) - B_H^i(u)$ – нагрузка, поступающая от источника $i$ на интервале $[u, v]$, $u \leqslant v$, $B(u,v)$ – нагрузка от типичного источника, а параметр $r = c - m$. Отметим, что предполагается выполненным условие $r > 0$, обеспечивающее существование стационарного режима для процесса загрузки.

Для произвольного заданного $b$ стационарная вероятность переполнения (превышения уровня $B = nb$) $\pi_n$ определяется как

$$\pi_n(b) = \mathsf{P}\left(Q^n > nb\right). \quad (2)$$

Точные выражения для (2) получены лишь в некоторых простейших случаях. Так, при $H = 1/2$, т. е. для стандартного броуновского движения, в работе [Такач, 1971] доказано, что

$$\pi_n(b) = \exp\left(-\frac{2nrb}{am}\right).$$

Также точное выражение можно получить и при $H = 1$. В этом случае процесс $B_1(t) =_d t \cdot N(0,1)$ является случайной прямой (лучом в случайном направлении из начала координат), где $N(0,1)$ – стандартная нормальная случайная величина. В этом случае имеем

$$\begin{aligned} Q^n &=_d \sup_{s\geqslant 0}\left(\sqrt{amn}B_1(s) - nrs\right) \\ &=_d \sup_{s\geqslant 0}\left(\left(\sqrt{amn}N(0,1) - nr\right)s\right). \end{aligned}$$

Из этого следует, что

$$Q^n = \begin{cases} \infty, & \text{если } N(0,1) > \frac{r\sqrt{n}}{\sqrt{am}}, \\ 0, & \text{иначе} \end{cases} \quad (3)$$

Выражение (3) имеет понятную физическую интерпретацию. В систему в каждый момент времени поступает одинаковая нагрузка, поэтому возможны два случая: либо система справляется с этой работой, а значит, очереди не образуется, либо работа неограниченно накапливается. Тогда для вероятности переполнения $\pi_n(b)$ при любых $b$ справедливо следующее выражение:

$$\begin{aligned} \pi_n(b) &= \mathsf{P}\left(N(0,1) > \frac{r\sqrt{n}}{\sqrt{am}}\right) \\ &= \frac{1}{\sqrt{2\pi}}\int_{\frac{r\sqrt{n}}{\sqrt{am}}}^{\infty} e^{-x^2/2}dx. \end{aligned} \quad (4)$$

В общем случае ($1/2 < H < 1$) аналогичный точный результат пока не найден. Однако имеет место ряд асимптотических результатов. В частности, было доказано [Botvich, Duffield, 1995], что справедливы следующие предельные соотношения:

$$-\lim_{n\to\infty}\frac{1}{n}\ln\pi_n(b) = \inf_{t\in\mathbb{R}_+}\frac{V^2(t)}{2am}, \quad (5)$$

где

$$V(t) = \frac{(b+rt)^2}{t^H}.$$

Значение $t_*$, при котором достигается минимум в (5) – это так называемое наиболее вероятное время переполнения. Для ФБД можно получить, что

$$t_* = \frac{H}{1-H}\cdot\frac{b}{r}, \quad (6)$$

$$V(t_*) = \left(\frac{b}{1-H}\right)^{1-H}\left(\frac{r}{H}\right)^H.$$

Выражение (5) означает, что при достаточно большом $n$

$$\pi_n(b) \approx \exp\left(-\frac{n}{2am}V^2(t_*)\right). \quad (7)$$

## Структура зависимости процесса загрузки

Для изучения структуры зависимости проанализируем двумерное распределение процесса загрузки. Естественный подход к изучению такой структуры – это анализ корреляционной функции:

$$\delta_n(T) = \frac{\mathsf{E}Q_0^nQ_T^n - (\mathsf{E}Q_0^n)(\mathsf{E}Q_T^n)}{\sqrt{Var(Q_0^n)Var(Q_T^n)}}.$$

Очевидно, что $\delta_n(T) \to 0$ при $T \to \infty$. Однако остается неясной скорость убывания, так как для ФБД точная формула даже для средней стационарной нагрузки $\mathsf{E}Q_0^n$ не найдена.

Поэтому вводятся в рассмотрение иные меры зависимости. Мы будем анализировать следующую характеристику меры зависимости, введенную в работе [Es-Saghouani, Mandjes, 2009]:

$$K_n(T) = \frac{\mathsf{P}(Q_0^n > np, Q_T^n > nq)}{\mathsf{P}(Q_0^n > np)\mathsf{P}(Q_T^n > nq)}, \quad (8)$$

где параметры $p, q > 0$. В [Es-Saghouani, Mandjes, 2009] также показано, что для $K_n(T)$ справедлива асимптотика:

$$\lim_{n\to\infty}\frac{1}{n}\ln K_n(T) = k_0T^{2H-2} + o(T^{2H-2}), \quad (9)$$

где параметры

$$k_0 = \frac{(2H-1)r^2}{Hm^2}s_*^{2-2H}t_*^{2-2H},$$

$$s_* = \frac{p}{r}\cdot\frac{H}{1-H} \qquad t_* = \frac{q}{r}\cdot\frac{H}{1-H},$$

(см. (6)). Это означает, что при достаточно больших $n$

$$K_n(T) \approx \exp\left(nk_0T^{2H-2} + n \cdot o\left(T^{2H-2}\right)\right). \quad (10)$$

На первый взгляд, эта аппроксимация может быть использована для оценки зависимости процесса загрузки. Однако формулу (10) нужно применять с осторожностью, ввиду присутствия множителя $n$ при неизвестной функции $o\left(T^{2H-2}\right)$.

Следует отметить, что процесс загрузки наследует структуру зависимости входного потока в смысле соотношения (8). Действительно, пусть

$$\lambda(T) = \frac{\mathsf{P}\left(W_0^n > np, W_T^n > nq\right)}{\mathsf{P}\left(W_0^n > np\right)\mathsf{P}\left(W_T^n > nq\right)},$$

где

$$W_0^n = \sum_{i=1}^{n} A_i(0,1),\ W_T^n = \sum_{i=1}^{n} A_i(T, T+1),$$

т. е. суммарная работа на интервалах $[0,1]$ и $[T, T+1]$ соответственно, причем в силу стационарности приращений ФБД, $W_0^n =_d W_T^n$. Обозначим

$$\Lambda(T) = \lim_{n\to\infty} \frac{1}{n} \ln \lambda(T).$$

По теореме Крамера (см. [Mandjes, 2007; Es-Saghouani, Mandjes, 2009])

$$\Lambda(T) \sim const \cdot T^{2H-2},\ T \to \infty,$$

что согласуется с (9).

Также можно говорить о наличии долговременной зависимости (долгой памяти) для процесса загрузки при достаточно больших $n$ и $T$. Действительно, обозначим

$$R_n(T) = Corr\left(I(Q_0^n > np), I(Q_T^n > nq)\right),$$

где $I$ означает индикатор. Принимая во внимание соотношения (7), (10), можно получить (см. [Mandjes, 2007; Es-Saghouani, Mandjes, 2009]), что при достаточно больших значениях параметров $n$ и $T$

$$\begin{aligned} R_n(T) &\approx (e^{nk_0T^{2H-2}} - 1)\psi(n) \\ &\approx \psi(n)T^{2H-2}, \end{aligned} \quad (11)$$

где $\psi(n)$ – некоторая (независящая от $T$) функция. Выражение (11) показывает, что функция $R_n(T)$ несуммируема при $H > 1/2$. Это позволяет говорить о том, что долговременная зависимость входного потока ФБД влечет за собой долговременную зависимость процесса загрузки.

## Оценивание стационарных характеристик

В данном разделе рассматриваются оценки величин $\pi_n(b), K_n(T)$. Для этого генерируется $N$ реализаций ФБД, каждая из которых содержит $M$ наблюдений (т. е. $M$ – длина траектории). Для построения этих реализаций использован RMD-метод (Random Midpoint Displacement) [Lau et al., 1995], временная сложность которого составляет $O(M)$. Для оценивания стационарных характеристик системы, длина $M$ генерируемой траектории должна быть достаточно большой, чтобы устранить (или по крайней мере уменьшить) влияние начального, так называемого переходного периода. (Отметим, что подробно вопрос моделирования ФБД обсуждается в [Coeurjolly, 2000].)

Обозначим $k$-ю последовательность (траекторию) значений ФБД $X_0^k, ..., X_M^k$, где $k = 1, ..., N$. Обозначим также

$$\hat{Q}_u^k = \max_{s\in S}\left[\sqrt{anm}X^k(s,u) - nr(u-s)\right], \quad (12)$$

где $S = \{0, 1, ..., u\}$. Стандартная выборочная оценка для $\pi_n(b)$ имеет вид:

$$\hat{\pi}_n(b) \triangleq \frac{1}{N}\sum_{i=1}^{N} I\left(\hat{Q}_u^k > nb\right),$$

где момент $u$ выбран достаточно большим.

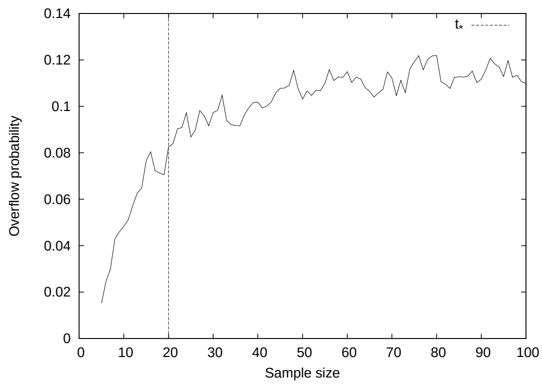



*Рис. 1.* Стабилизация оценки $\hat{\pi}_{10}(1)$ с ростом числа наблюдений

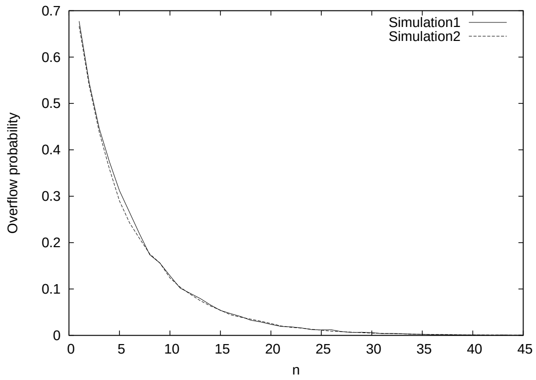



*Рис. 2.* Сравнение двух методов оценки стационарной вероятности переполнения $\pi_n(1)$

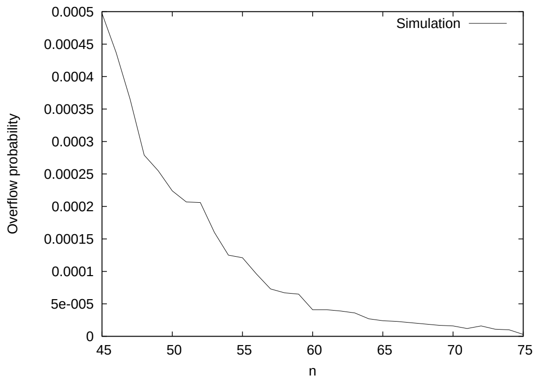



*Рис. 3.* Зависимость вероятности переполнения $\pi_n(1)$ от количества источников

Рис. 1 показывает зависимость оценки вероятности переполнения $\pi_{10}(1)$ от длины траектории. При моделировании использованы следующие значения параметров: $n = 10; H = 0,8; c = 1; m = 0,8; b = 1; N = 10^4$. В соответствии с формулой (6) получаем $t_* = 20$. На рис. 1 видна определенная стабилизация значений оценки после наиболее вероятного времени переполнения $t_*$. Отметим вновь, что для оценивания стационарных характеристик начальная часть траектории должна быть пропущена, т. е. общая длина траектории должна быть достаточно большой.

Стационарную вероятность переполнения (2) можно оценивать и как долю времени нахождения процесса загрузки выше уровня $B = nb$, т. е.

$$\frac{1}{M}\sum_{k=1}^{M} I\left(Q(kh) > nb\right), \qquad (13)$$

где процесс загрузки в дискретном времени определяется известной рекурсией Линдли:

$$Q(t+h) = \left(Q(t) - nhr + \sqrt{nam}B_H(t,t+h)\right)^+,$$

где $h > 0$ – так называемый шаг дискретизации (выбор $h$ обсуждается ниже), $Q(0) = 0$.

Значение (13) вычисляется для каждой траектории. В качестве оценки вероятности переполнения берется среднее полученных значений. Заметим, что совпадение в пределе этих оценок теоретически доказано лишь на основе так называемого свойства PASTA, которое применимо лишь в случае пуассоновского входного потока. Однако рис. 2 показывает, что результаты, полученные по схеме (12) (Simulation1) и схеме (13) (Simulation2), не отличаются значительно и для рассматриваемой модели.

Следует отметить, что схема (13) предпочтительнее с точки зрения вычислительных затрат. Так, например, используя $N = 1000$ траекторий ФБД, этот метод позволяет оценить вероятности порядка $10^{-5}$ (см. рис. 3), в то время как метод (12) позволяет достичь такой же точности при генерации как минимум $N = 10^5$ траекторий, что естественным образом сказывается на времени работы алгоритма.

Теперь сравним результаты численного моделирования вероятности $\pi_n(b)$ с аппроксимацией (7). Чтобы определить, при каком $n$ эта аппроксимация имеет заданную точность, рассмотрим так называемые невязки

$$\Delta_n = \left|\pi_n(b) - \hat{\pi}_n(b)\right|, \qquad (14)$$

при некоторых фиксированных значениях $b$, $r$. Требуется найти такое минимальное значение $n$, при котором $\Delta_n < \varepsilon$, для некоторого заранее заданного значения $\varepsilon > 0$. (Таким образом, $\varepsilon$ – это точность аппроксимации.)

На рис. 4 представлен график зависимости $\Delta_n$ от числа источников $n$ (логарифмический масштаб) при следующих параметрах: $b = 1; c = 1; m = 0,8; h = 1; N = 1000; M = 10000$. Этот график наглядно показывает, каким должен быть уровень агрегирования $n$, чтобы обеспечить необходимую точность аппроксимации (7).

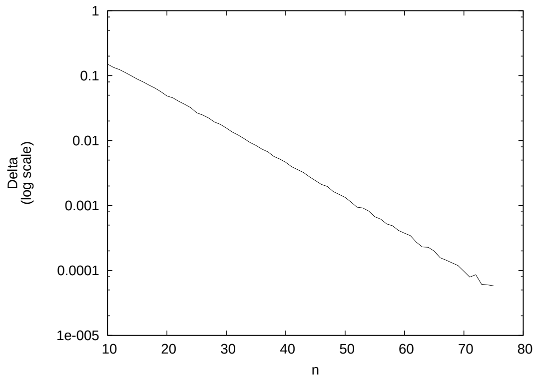



*Рис. 4.* Зависимость $\Delta_n$ от числа источников

Зависимость $\Delta_n$ от шага дискретизации

| № | $\Delta_n$, $h = 1$ | $\Delta_n$, $h = 1/16$ |
|---|---|---|
| 1 | 0,151152 | 0,150463 |
| 2 | 0,148348 | 0,149770 |
| 3 | 0,148442 | 0,146604 |
| 4 | 0,150842 | 0,146778 |
| 5 | 0,151446 | 0,151623 |
| 6 | 0,151327 | 0,147263 |
| 7 | 0,148494 | 0,144033 |
| 8 | 0,152131 | 0,147650 |
| 9 | 0,147023 | 0,148453 |
| 10 | 0,151670 | 0,149856 |
| 11 | 0,148143 | 0,150300 |
| 12 | 0,149802 | 0,145306 |
| 13 | 0,149432 | 0,148748 |
| 14 | 0,155012 | 0,145321 |
| 15 | 0,147368 | 0,148984 |
| 16 | 0,152423 | 0,151441 |
| 17 | 0,147264 | 0,149285 |
| 18 | 0,148313 | 0,150049 |
| 19 | 0,147407 | 0,150392 |
| 20 | 0,150885 | 0,152498 |
| Среднее: | 0,149846 | 0,148741 |

Для изучения влияния шага дискретизации $h$ на результаты оценивания, для двух значений $h = 1/16, 1$ проведена серия из 20 испытаний. Каждое испытание состоит в подсчете значения выражения (14) по схеме (13) при фиксированных $n = 10$ и $b = 1$. Поскольку аппроксимация (7) получена для процесса загрузки в непрерывном времени, то можно ожидать, что с уменьшением величины шага $h$ значения $\Delta_n$ также должны уменьшаться. Отметим, что при численном моделировании исходный процесс с непрерывным временем заменяется на процесс с дискретным временем, что неизбежно приводит к ошибке дискретизации. Фактически траектория процесса представляет собой линейный интерполяционный сплайн, построенный по точкам $X_0, X_h, ..., X_{Mh}$. В этой связи уместно привести аналитический результат о величине ошибки такой дискретизации в случае стандартного броуновского движения $B_{1/2}(t)$. Пусть $h = h_N = 1/L$, где $L > 0$, $B^h(t)$ – реализация стандартного броуновского движения с шагом $h$. Тогда средняя ошибка линейной интерполяции удовлетворяет соотношению (см. [Asmussen, Glynn, 2007]):

$$\mathsf{E}\int_0^1 \left|B^h(t) - B_{1/2}(t)\right| dt = \sqrt{\pi/32}/L^{1/2}.$$

Результаты эксперимента представлены в таблице, где видна тенденция к уменьшению $\Delta_n$ с уменьшением $h$. Таким образом, чтобы точнее оценивать характеристики случайного процесса с непрерывным временем, необходимо уменьшать шаг дискретизации $h$. С другой стороны, уменьшение $h$ ведет к увеличению объема вычислений, что ставит проблему баланса между величиной шага $h$ и объемом вычислений.

На рис. 5 представлен график зависимости (логарифмический масштаб) оценки вероятности переполнения от величины параметра Херста $H$ при следующих значениях остальных параметров: $b = 1; c = 1; m = 0,8; n = 10; M = 10^4; N = 10^4$.

На рис. 6 представлен график зависимости вероятности переполнения от размера буфера $b$ при $H = 1$ со следующим параметрами: $n = 10; c = 1; m = 0,8; M = 10^4; N = 10^4$. В силу (4) $\pi_{10}(b) \approx 0,23975$ для любого $b$. Как видно, результаты численного моделирования согласуются с теоретическими.

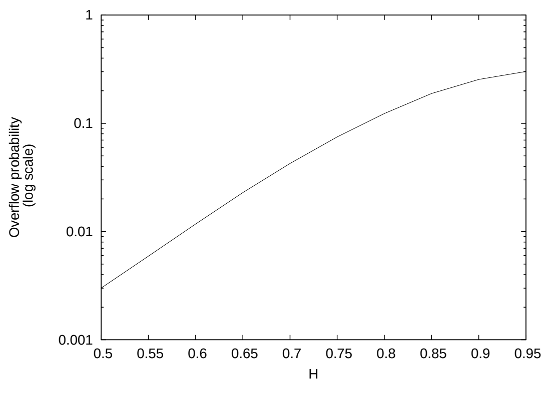



*Рис. 5.* Зависимость $\hat{\pi}_{10}(1)$ от параметра Херста $H$

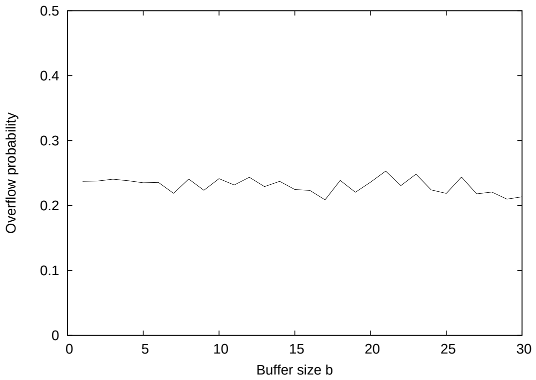



*Рис. 6.* Оценка вероятности $\pi_{10}(b)$ при $H = 1$

Чтобы оценить величину $K_n(T)$, нужно оценить вероятности, входящие в выражение (8). Напомним, что величины $\hat{Q}_0^k$, $\hat{Q}_T^k$ определяются в (12). Тогда оценка двумерного распределения $\mathsf{P}\left(Q_0^n > np,\, Q_T^n > nq\right)$ определяется как выборочное среднее:

$$\hat{\pi}_{n,T}(p,g) = \frac{1}{N}\sum_{k=1}^{N} I\left(\hat{Q}_0^k > np, \hat{Q}_T^k > nq\right).$$

С учетом этого оценка для $K_n(T)$ запишется в виде отношения:

$$\hat{K}_n(T) = \frac{\hat{\pi}_{n,T}(p,g)}{\hat{\pi}_n(p)\hat{\pi}_n(g)}.$$

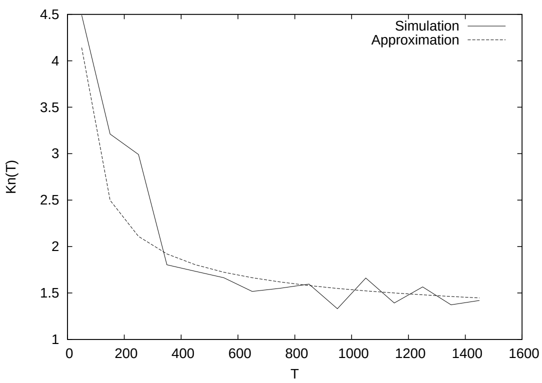



*Рис. 7.* Оценка $K_{10}(T)$

На рис. 7 представлен результат численного моделирования величины $\hat{K}_{10}(T)$ и его сравнение с аппроксимацией (10). В моделировании использованы следующие значения параметров: $H = 0,8; c = 1; m = 0,8; n = 10; p = 1; q = 2; N = 10^4$.

Поскольку получение большого количества траекторий ФБД большой длины за приемлемое время моделирования является критически важным для надежного оценивания параметров гауссовских очередей, в дальнейшем исследовании предполагается применение ускоренных методов моделирования, а также использование высокопроизводительных технологий.

## Заключение

В статье рассмотрена проблема оценивания характеристик гауссовской очереди в случае нескольких независимых источников. Основное внимание уделено оценке так называемой вероятности переполнения. Приведен краткий обзор основных теоретических результатов для процесса загрузки. Представлены результаты численных экспериментов, которые дают хорошее согласие с известными аналитическими результатами.

## Литература


*Такач Л.* Комбинаторные методы в теории случайных процессов. М.: Мир, 1971. 264 с.

*Asmussen S., Glynn P.* Stochastic Simulation: algorithms and analysis. Springer. 2007. P. 488.

*Botvich D., Duffield N.* Large deviations, the shape of the loss curve, and economies of scale in large multiplexers // Queueing Systems. 1995. Vol. 20. P. 293–320.

*Coeurjolly J. F.* Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study // Journal of Stat. Software. 2000. Vol. 5.

*Debicki K., Mandjes M.* Exact overflow asymtotics for queues with many Gaussian inputs. Report PNA-R0209 March 31, 2002.

*Duffield N., O'Connell N.* Large deviations and overflow probabilities for the general single server queue, with applications // Proceedings of the Cambridge Philosophical Society. 1995. Vol. 118. P. 363–374.

*Es-Saghouani A., Mandjes M.* On the dependence structure of Gaussian queues // Stochastic Models. 2009. Vol. 25. P. 221–247.

*Hüsler J., Piterbarg V. I.* Extremes of a certain class of Gaussian processes // Stochastic Processes and their Applications. 1999. Vol. 83. P. 257–271.

*Kilpi J., Norros I.* Testing the Gaussian approximation of aggregate traffic, Proceedings of the second ACM SIGCOMM Workshop, Marseille, France. 2002. P. 49–61.

*Lau W-C., Erramilli A., Wang J. L., Willinger W.* Self-similar traffic generation: the random midpoint displacement algorithm and its properties. Proceedings of ICC '95. 1995.

*Mandjes M.* Large Deviations of Gaussian Queues. Chichester: Wiley, 2007. P. 339.

*Narayan O.* Exact asymptotic queue length distribution for fractional Brownian traffic // Advances in Performance Analysis. 1998. Vol. 1. P. 39–63.

*Norros I.* A storage model with self-similar input // Queueing Systems. 1994. Vol. 16. P. 387–396.

*Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling // Computer communication review. 1997. Vol. 27. P. 5–23.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Лукашенко Олег Викторович**
аспирант
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: lukashenko-oleg@mail.ru
тел.: (8142) 763370

**Lukashenko, Oleg**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: lukashenko-oleg@mail.ru
tel.: (8142) 763370

**Морозов Евсей Викторович**
ведущий научный сотрудник д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: emorozov@karelia.ru
тел.: (8142) 763370

**Morozov, Evsey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: emorozov@karelia.ru
tel.: (8142) 763370

**Пагано Микеле**
преподаватель
факультет информационной инженерии Университета г. Пизы, Италия
эл. почта: m.pagano@iet.unipi.it
тел.: +39 050 2217575

**Pagano, Michele**
Associated Professor
Department of Information Engineering,
University of Pisa, Italy
e-mail: m.pagano@iet.unipi.it
tel.: +39 050 2217575

Труды Карельского научного центра РАН
№ 5. 2011. С. 63–74

*УДК 519.872.1*

# ОЦЕНИВАНИЕ ВЕРОЯТНОСТИ БЛОКИРОВКИ В СИСТЕМЕ С ПОВТОРНЫМИ ВЫЗОВАМИ И ПОСТОЯННОЙ СКОРОСТЬЮ ВОЗВРАЩЕНИЯ ЗАЯВОК С ОРБИТЫ

**Е. В. Морозов, Р. С. Некрасова**

*Институт прикладных математических исследований Карельского научного центра РАН*

В статье исследуется вероятность блокировки заявки в системе с повторными вызовами и постоянной интенсивностью повторных заявок, идущих с орбиты на сервер. Изучаются теоретические свойства, а также свойства различных статистических оценок этой вероятности. Анализ проводится с помощью регенеративного метода. Исследован эффект уменьшения дисперсии оценки. Результаты имитационного моделирования хорошо согласуются с известными аналитическими результатами, а также с соответствующими результатами для системы с потерями.

Ключевые слова: система с повторными вызовами, вероятность блокировки, зона стационарности/нестационарности, (квази-)регенерации, уменьшение дисперсии.

**E. V. Morozov, R. S. Nekrasova. ESTIMATION OF BLOCKING PROBABILITY IN A RETRIAL QUEUING SYSTEM WITH CONSTANT RETRIAL RATE**

The paper investigates the blocking probability in a retrial system with constant retrial rate. Both theoretical properties and the properties of various estimates of the probability are studied. The analysis is based on the regenerative method. The variance reduction effect is studied. Simulation results are consistent with known analytical results and with corresponding results for a lossy system.

Key words: retrial queuing system, blocking probability, stability/instability region, (quasi-) regenerations, variance reduction.

## Введение

Данная статья посвящена применению регенеративного метода для доверительного оценивания стационарной вероятности блокировки (ухода на орбиту) заявки в односерверной системе с повторными вызовами без буфера для ожидания и с постоянной интенсивностью возвращения заявок с орбиты на сервер. Показано, что данный метод применим как в области стационарности, так и в области нестационарности системы, когда величина орбиты неограниченно растет. В выводе соотношений для вероятности блокировки существенно использован метод доказательства обобщенной формулы Литтла для классических систем обслуживания с буфером. Рассмотрены различные оценки вероятности блокировки. Для уменьшения дисперсии используется выпуклая комбинация *прямой оценки* (отношения числа отказов к общему числу приходов) и оценки доли времени занятости сервера. Приведены результаты численного моделирования для ряда систем.

## Описание системы и условие стационарности

Рассмотрим односерверную систему с повторными вызовами типа $M/G/1/1$ (без буфера) с входным пуассоновским потоком с интенсивностью $\lambda$ и произвольным (типичным) временем обслуживания $S$ со средним $\mathsf{E}S := 1/\mu$. Заявки, поступающие в систему, когда сервер занят, уходят на орбиту бесконечного объема, а затем вновь пытаются попасть на обслуживающее устройство. В отличие от классических систем, поток повторных заявок (попыток) является пуассоновским с параметром $\mu_0$, *не зависящим от величины (непустой) орбиты*. Такая модель успешно применена при моделировании телефонных линий [Fayolle, 1986], протоколов множественного доступа ALOHA [Choi et al., 1993], протокола TCP с короткими сообщениями (short TCP transfers) [Avrachenkov, Yechiali, 2008]. В частности, рассматриваемая в данной работе модель односерверной системы без буфера описывает протокол множественного доступа ALOHA, когда при наличии на орбите $n$ заявок каждая поступает (повторно) в систему с интенсивностью $\mu_0/n$. Как хорошо известно, тогда суммарный поток является пуассоновским с параметром $\mu_0$.

Таким образом, суммарный поток заявок, поступающих на сервер, складывается из двух (вообще говоря, зависимых) потоков: $\lambda$-потока первичных заявок ($\lambda$-заявок) и потока вторичных заявок с интенсивностью $\tilde{\mu}_0 \leqslant \mu_0$. (Величина $\tilde{\mu}_0$ строго определяется ниже см. (21).) Для удобства далее обозначим исходную систему через $\Sigma$.

Рассмотрим также следующую вспомогательную односерверную систему $\hat{\Sigma}$ с потерями, с нулевым буфером и без орбиты. Система $\hat{\Sigma}$ имеет тот же входной $\lambda$-поток, то же распределение времени обслуживания, что и исходная система $\Sigma$, но, в отличие от $\Sigma$, имеет еще один (независимый) пуассоновский входной поток заявок с параметром $\mu_0$. (Назовем их $\mu_0$-заявки.) Таким образом, если в момент прихода (любой) новой заявки сервер занят, то она *покидает систему и не оказывает влияния на ее будущее состояние.* Для удобства можно считать, что такие заявки уходят на виртуальную орбиту, описываемую как система вида $\cdot/M/1$ с интенсивностью обслуживания $\mu_0$, и покидают систему навсегда после обслуживания.

Можно ожидать, что сервер в системе $\hat{\Sigma}$ более загружен, чем в исходной системе, так как интенсивность потока повторных заявок $\tilde{\mu}_0 \leqslant \mu_0$. Более того, можно ожидать, что с ростом величины орбиты $\tilde{\mu}_0 \uparrow \mu_0$, и поэтому распределение состояний сервера (занят или пуст) в системе $\Sigma$ должно сближаться с соответствующим распределением в системе $\hat{\Sigma}$. Следовательно, поток потерь в системе $\hat{\Sigma}$ (или поток заявок, идущих на виртуальную орбиту) должен быть более интенсивным, чем поток блокированных заявок, идущих на орбиту в исходной системе. Отметим, что источником неустойчивости (нестационарности) исходной системы (с конечным буфером) может быть только неограниченно растущая орбита. Тогда условие устойчивости виртуальной орбиты должно гарантировать устойчивость орбиты в (менее загруженной) системе $\Sigma$.

Для пояснения рассмотрим орбиту как односерверную систему с входным потоком заявок, получивших отказ. Интуитивно понятно, что верхняя граница интенсивности входного потока на сервер равна $\lambda + \mu_0$ (что соответствует постоянно непустой орбите). Можно ожидать, что поток с бóльшей интенсивностью приводит к бóльшему числу потерь, и таким образом, более интенсивному потоку заявок на орбиту. Этот вывод опирается на свойство монотонности систем с потерями относительно интервалов входного потока, которое кратко рассматривается ниже. Суммарный поток заявок на сервер в системе $\hat{\Sigma}$ (в отличие от исходной системы $\Sigma$) представляет собой суперпозицию *независимых пуассоновских потоков* с интенсивностями $\mu_0$ и $\lambda$. В исходной системе поток повторных заявок прерывается, когда начинается *период простоя* орбиты. Этот период заканчивается при поступлении очередной повторной заявки.

Обозначим через $\mathsf{P}_{loss}$ стационарную вероятность потери в системе $\hat{\Sigma}$. (Точное определение $\mathsf{P}_{loss}$ приведено ниже.) Из вышесказанного следует ожидать, что условие

$$(\lambda + \mu_0)\mathsf{P}_{loss} < \mu_0 \qquad (1)$$

является достаточным условием устойчивости орбиты в системе $\Sigma$. Действительно, этот результат доказан в работе [Avrachenkov, Morozov, 2010].

**Замечание 1.** На самом деле данный результат доказан в [Avrachenkov, Morozov, 2010] для систем с повторными вызовами более общего вида, имеющих несколько серверов, ненулевой буфер и входной $\lambda$-поток восстановления. Более того, условие (1) является *критерием стационарности*, по крайней мере, для систем с пуассоновским $\lambda$-потоком, в том числе для рассматриваемой в данной статье системы.

Обсудим кратко свойство монотонности системы с потерями, установленное в [Sonderman, 1979, 1981]. (Этот результат верен в соответствующей форме и для многосерверной системы с ненулевым буфером.)

Обозначим через $t_n$ момент прихода $n$-й заявки, $B_n$ – $n$-й фактический момент прихода (когда заявка поступает на обслуживание), через $S_n$ – время $n$-го обслуживания, а через $D_n$ – $n$-й момент завершения обслуживания, $n \geqslant 1$. Пусть $n$-я заявка (пришедшая в момент $t_n$) попала в систему, и это есть $m$-й фактический вход в систему во временном интервале $[0, t_n]$. Обозначим этот момент через $t(m)$, таким образом, введенные величины связаны как $t_n = t(m) := B_m$, $n \geqslant m \geqslant 1$. Напомним, что $\nu_n$ – число заявок в системе перед $n$-м приходом. Имеют место следующие соотношения (в предположении, что система изначально пуста):

$$\begin{aligned} B_{n+1} &= \min(t_i : t_i \geqslant t(n), \nu_i = 0), \ \ B_0 = 0, \\ D_1 &= B_1 + S_1, \\ D_{n+1} &= (\max(D_n, B_{n+1}) + S_{n+1}), n \geqslant 1. \end{aligned} \quad (2)$$

В очевидных обозначениях рассмотрим две системы с потерями $\Sigma_1$ и $\Sigma_2$, у которых интервалы между приходами (стохастически) равны, $\tau_n^{(1)} =_{st} \tau_n^{(2)}$, а времена обслуживания упорядочены как $S_n^{(1)} \leqslant S_n^{(2)}$, $n \geqslant 1$ (с вероятностью 1 или стохастически). В [Sonderman, 1979] по индукции доказано, что $D_n^{(1)} \leqslant D_n^{(2)}$, $B_n^{(1)} \leqslant B_n^{(2)}$, $\nu_n^{(1)} \leqslant \nu_n^{(2)}$, $n \geqslant 1$. Этот результат непосредственно переносится на случай непрерывного времени (на число событий в любом интервале $[0,t]$). Также отсюда легко следует, что число потерь в обеих системах в интервале $[0,t]$ связано как $R^{(1)}(t) \leqslant R^{(2)}(t)$, $t \geqslant 0$. Именно этот результат является ключевым при доказательстве условия стационарности в работе [Avrachenkov, Morozov, 2010]. (Аналогично, в [Sonderman, 1979] доказано, что при $\tau_n^{(1)} \geqslant \tau_n^{(2)}$ и $S_n^{(1)} = S_n^{(2)}$ также $\nu_n^{(1)} \leqslant \nu_n^{(2)}$, $n \geqslant 1$.) Заметим, что для экспоненциальных распределений времен обслуживания упорядоченность $S_n^{(1)} \leqslant S_n^{(2)}$ сводится к неравенству $\mu_1 \geqslant \mu_2$ для параметров этих распределений.

Чтобы применить результаты о монотонности к анализу вероятности блокировки в исходной системе с повторными вызовами и *без потерь*, заметим, что различие между системами $\Sigma$ и $\hat{\Sigma}$ в этом анализе проявляется лишь на *периодах простоя* орбиты в исходной системе. Напомним, что $\lambda$-потоки у обеих систем совпадают. Чтобы синхронизировать поток с орбиты и пуассоновский (с параметром $\mu_0$) в системе $\hat{\Sigma}$, мы продолжаем отсчитывать *фиктивные* интервалы выходного потока с орбиты и на периодах простоя (эти интервалы завершаются приходом фиктивных заявок). В моменты завершения каждого периода простоя орбиты (в системе $\Sigma$) разыгрываются заново незавершенные интервалы в пуассоновских $\mu_0$-потоках в обеих системах $\Sigma$ и $\hat{\Sigma}$. По свойству потери памяти экспоненциального распределения эта процедура не меняет распределения потоков. Кроме того, фиктивным заявкам назначается *нулевое время обслуживания*. В результате моменты (суммарного) входного потока в обеих системах оказываются полностью синхронизированными, т. е. (в очевидных обозначениях) $\tau_n = \hat{\tau}_n$, в то время как времена обслуживания упорядочены как $S_n \leqslant \hat{S}_n$, $n \geqslant 1$. Остается применить свойство монотонности системы с потерями, приведенное выше.

Для некоторых систем с потерями известны аналитические формулы для вероятности потери, в частности, для обобщенной $c$-серверной системы Эрланга $M/G/c/c$ (без буфера)

$$\mathsf{P}_{loss} = \frac{\Big((\lambda+\mu_0)/\mu\Big)^c}{c!}\Big[\sum_{n=0}^{c}\frac{\Big((\lambda+\mu_0)/\mu\Big)^n}{n!}\Big]^{-1}. \quad (3)$$

Это позволяет записать условие (1) для системы $M/G/1/1$ в виде

$$\frac{\lambda+\mu_0}{\lambda+\mu_0+\mu} < \frac{\mu_0}{\lambda+\mu_0}, \quad (4)$$

что эквивалентно

$$\frac{\lambda^2}{\mu_0} + \lambda < \mu. \quad (5)$$

Полагая $\lambda = 1$, последнее условие можно записать в форме

$$\frac{1}{\mu_0} + 1 < \mu, \quad (6)$$

удобной для численного исследования области стационарности по двум параметрам $\mu_0$, $\mu$, характеризующим скорости ухода заявок с орбиты и обслуживания на сервере, соответственно. Напомним, что проведенный анализ верен для произвольного времени обслуживания, в частности, имеющего (стандартное) распределение Парето интервалов $P(S > x) = x^{-\alpha}$, $\alpha > 1$, $x \geqslant x_0 = 1$. Тогда $\mathsf{E}S = \frac{\alpha}{\alpha-1} := 1/\mu$ и условие (5) принимает вид

$$\lambda\Big(\frac{\lambda}{\mu_0} + 1\Big) < 1 - 1/\alpha. \tag{7}$$

(В последней формуле нельзя положить $\lambda = 1$, что связано с выбором дополнительного параметра $x_0 = 1$.)

## Регенеративная структура системы с повторными вызовами

В данном разделе описана регенеративная структура процессов в исходной системе с повторными вызовами.

Пусть $N(t)$ – число заявок на орбите (величина орбиты) в момент времени $t$, а $\nu(t)$ – состояние сервера (0 или 1) в момент $t$. Пусть $\{t_n\}$ – моменты прихода $\lambda$-заявок, $N(t_n^-) = N_n$, $\nu(t_n^-) = \nu_n$, $n \geqslant 1$. Введем общее число заявок в системе $X = \{X(t) = N(t) + \nu(t),\ t \geqslant 0\}$, и положим $X(t_n^-) = X_n := N_n + \nu_n$, $n \geqslant 1$. Заметим, что система $\hat{\Sigma}$ регенерирует в моменты, когда $\lambda$-заявки встречают пустой сервер (здесь используется свойство потери памяти входного потока $\mu_0$-заявок). Исходная же система $\Sigma$ регенерирует в моменты, когда заявка поступает в полностью пустую систему (пустой сервер и пустая орбита). Таким образом, моменты регенерации основного процесса $\{X_n\}$ в дискретном времени (полученного из исходного процесса $X$ путем наблюдения лишь в моменты прихода $\lambda$-заявок) определяются следующим образом:

$$\beta_{n+1} = \inf_k (k > \beta_n : X_k = 0),\ n \geqslant 0\ \ (\beta_0 = 0). \tag{8}$$

Процесс регенераций $\{\beta_n\}$ (и основной процесс $X$) называется положительно возвратным, если $\mathsf{E}\beta < \infty$ [Morozov, 2004]. (Эта терминология, ставшая уже общепризнанной, заимствована из теории цепей Маркова, возвратных по Харрису, у которых положительная возвратность также означает конечность средней длины цикла регенерации. По этому поводу (см. [Wolf, 1989; Morozov, 2004]).

Обозначим через $R(t)$ – общее число неудачных попыток заявок попасть на сервер в интервале $[0, t]$ (т. е. это поток, идущий на орбиту), а через $A(t)$ – общее число попыток попасть на сервер в этом интервале (порожденное первичными $\lambda$-заявками и повторными $\mu_0$-заявками). Рассмотрим вначале стационарный режим. Иными словами, пусть условие (1) выполнено, тогда процесс $X$ – положительно возвратен. Обозначим через $R$ и $A$ типичное число орбитальных заявок и (типичное) общее число попыток на цикле регенерации, соответственно. Пусть $I_n = 1$, если $n$-я попытка неудачна (иначе, $I_n = 0$). Поскольку, как легко увидеть, $\mathsf{P}(\tau > S) > 0$, то длина цикла регенерации $\beta$ – непериодическая случайная величина. Поэтому существует предел по распределению $I_n \Rightarrow I$. Более того, ввиду равномерной интегрируемости индикаторов, имеет место также сходимость в среднем

$$\mathsf{P}(I_n = 1) \to \mathsf{P}_{orb}, \qquad n \to \infty, \tag{9}$$

где $\mathsf{P}_{orb} := \mathsf{P}(I = 1)$ есть стационарная вероятность того, что вновь поступающая заявка идет на орбиту. Рассмотрим выборочные средние

$$B_N(t) := \frac{R(t)}{A(t)},\ \mathsf{P}_{orb}(n) := \frac{\sum_{k=1}^n I_k}{n}. \tag{10}$$

Из теории регенерации для этих оценок с вероятностью 1 следует сходимость

$$\lim_{t\to\infty} B_N(t) = \lim_{n\to\infty} \mathsf{P}_{orb}(n) = \frac{\mathsf{E}R}{\mathsf{E}A} = \mathsf{P}_{orb}. \tag{11}$$

Так как исходная система менее загружена, чем система $\hat{\Sigma}$, то можно ожидать, что в области стационарности исходной системы

$$\mathsf{P}_{orb}(n) \to \mathsf{P}_{orb} < \mathsf{P}_{loss}, \quad \text{при} \quad n \to \infty. \tag{12}$$

С другой стороны, в области нестационарности исходной системы $\Sigma$ можно ожидать неограниченного роста величины ее орбиты. В этом случае поток повторных заявок (идущих с орбиты на сервер) через некоторое время $t_0$ *склеивается* с пуассоновским потоком с параметром $\mu_0$. Иными словами, после момента $t_0$ поток повторных заявок стохастически эквивалентен пуассоновскому потоку. (Такое склеивание называется *каплингом* [Asmussen, 2002].) Поэтому можно ожидать, что $\mathsf{P}_{orb}(n) \to \mathsf{P}_{loss}$. Но при условии нестационарности процесс $X$ *не является положительно возвратным* (т. е. $\mathsf{E}\beta = \infty$), и применение регенеративного метода для получения предела $\mathsf{P}_{orb}$, вообще говоря, становится проблематичным. Однако в данном случае для оценки $\mathsf{P}_{orb}$ можно использовать другой тип регенераций, а именно, *квази-регенерации*. Квази-регенерации исходной (нестационарной) системы определяются как моменты, когда поступающая в систему первичная заявка встречает пустой сервер (в то время как орбита, вообще говоря, *может быть не пустой*).

Более точно, положим $\alpha_0 = 0$ и определим рекурсивно моменты квази-регенераций следующим образом:

$$\alpha_{n+1} = \inf_k (k > \alpha_n : \nu_k = 0), \; n \geqslant 0. \tag{13}$$

Возможность использования данных моментов в нестационарном режиме для оценки $\mathsf{P}_{orb}$ рассмотрена в [Avrachenkov, Morozov, 2010; Avrachenkov et al., 2011]. Важно заметить, что в области нестационарности классические моменты регенерации через некоторое время после начала работы системы перестают происходить. Точнее это можно сформулировать следующим образом. Пусть выполнено условие нестационарности

$$(\lambda + \mu_0)\mathsf{P}_{loss} > \mu_0. \tag{14}$$

(Напомним, что для рассматриваемой системы с пуассоновским входным потоком условие (1) является критерием стационарности.) Тогда, после конечного (с вероятностью 1) момента времени $t_0$, орбита никогда не опустошается, и, с точки зрения оценки вероятности $\mathsf{P}_{loss}$, исходная система ведет себя как система $\hat{\Sigma}$, на вход которой поступают два независимых потока с параметрами $\lambda$ и $\mu_0$. Другими словами, квази-регенерации становятся классическими регенерациями, но для процесса $\{\nu_n\}$, рассматриваемого изолированно, а оценка $\mathsf{P}_{orb}(n)$ сходится к $\mathsf{P}_{loss}$ с вероятностью 1. Заметим, что определенная проблема, связанная с частотой квази- и классических регенераций, может проявиться на границе области устойчивости (определяемой равенством в (1) и (14)) или вблизи нее. Можно ожидать, что чем больше разность между левой и правой частями неравенства (14) (т. е., чем *глубже* параметры $\lambda$, $\mu$, $\mu_0$ находятся *внутри области нестационарности*), тем быстрее (и устойчивее) должна быть сходимость $\mathsf{P}_{orb}(n) \to \mathsf{P}_{loss}$. Совершенно другую ситуацию следует ожидать в области стационарности системы. Именно, чем *глубже внутри области стационарности* находятся параметры $\lambda$, $\mu$, $\mu_0$ (т. е. чем больше разность правой и левой частей неравенства (1)), тем меньше должно быть значение предела $\mathsf{P}_{orb}(n) \to \mathsf{P}_{orb}$ по сравнению с $\mathsf{P}_{loss}$.

Эти соображения подтверждаются приводимыми ниже результатами численного оценивания вероятности $\mathsf{P}_{orb}$ методом регенеративного (или квази-регенеративного) моделирования. Заметим, что детальное изложение регенеративного метода моделирования можно найти в [Bratley, Fox, 1987; Law, Kelton, 1991; Glynn, Iglehart, 1993].

## Вероятность блокировки: применение формулы Литтла

Обобщенная формула Литтла является ключевым элементом анализа при выводе соотношений для вероятности блокировки $\mathsf{P}_{orb}$. Поэтому, следуя [Whitt, 1981], ниже приведена как сама обобщенная формула Литтла, так и пояснения к ее выводу. Кроме того, рассмотрены и исследованы две оценки вероятности блокировки: введенная ранее прямая (классическая) оценка $B_N(t)$ и (определенная ниже) альтернативная (непрямая) оценка $B_I(t)$, предложенная в [Spricant, Whitt, 1999].

Рассмотрим $s$-серверную систему с потерями общего вида $GI/G/s/k+s$ (здесь $k$-размер буфера для ожидания). Пусть $U(t)$ – число поступивших в систему заявок, а $D(t)$ – общее число обслуженных заявок, в интервале $[0, t]$. Пусть $W_k$ – время, проведенное $k$-й заявкой в системе (для систем без буфера эта величина равна времени обслуживания $S_k$), и $\nu(t)$ – число заявок в системе в момент $t$ (в случае одного сервера $\nu(t) = 0$ или 1). Обобщенная формула Литтла содержится в следующем утверждении из [Spricant, Whitt, 1999].

**Теорема 1.** *Пусть с вероятностью 1 существуют пределы*

$$t^{-1}U(t) \to \lambda, \;\; t^{-1}D(t) \to \lambda, \;\; t \to \infty, \tag{15}$$

*где* $0 < \lambda < \infty$. *Пусть также*

$$k^{-1} \sum_{j=1}^{k} W_j \to W, \;\; k \to \infty. \tag{16}$$

*Тогда с вероятностью 1*

$$\frac{1}{t}\int_0^t \nu(s)ds \to L \quad \text{при} \quad t \to \infty, \tag{17}$$

*и кроме того,*

$$L = \lambda W. \tag{18}$$

Для пояснения заметим, что интеграл в (17) представляет собой суммарное время, которое провели в системе $D(t)$ заявок (уже обслужившихся к моменту $t$) плюс время, проведенное в системе к моменту $t$ остальными $A(t) - D(t)$ заявками. Также в доказательстве используются следующие неравенства

$$\sum_{j=1}^{D(t)} W_j \leqslant \int_0^t \nu(s)ds \leqslant \sum_{j=1}^{U(t)} W_j, \tag{19}$$

откуда утверждение теоремы следует после деления на $t$ и перехода к пределу при $t \to \infty$.

Вернемся к системе $\Sigma$ с повторными вызовами и без буфера, описанной в разделе 1. Рассмотрим обслуживающее устройство как изолированную систему, к которой и применим Теорему 1. Входной поток можно разделить на *эффективный* поток заявок, поступающих на обслуживание, и поток отказов, идущий на орбиту. Таким образом, эффективный поток есть число поступлений $U(t)$ на сервер в интервале $[0,t]$, $t \geqslant 0$. Так как все заявки эффективного потока получают обслуживание, то с вероятностью 1

$$\lim_{t\to\infty} \frac{U(t)}{t} = \lim_{t\to\infty} \frac{D(t)}{t} := \lambda_e. \tag{20}$$

(Как доказано в Теореме 2 ниже, данный предел, т. е. интенсивность $\lambda_e$, действительно существует). Таким образом, условия Теоремы 1 выполнены, причем время, проведенное заявкой в системе (на сервере), является ее временем обслуживания $S$. Заметим также, что приращения на циклах регенерации процесса накопления $\{R(t)\}$ в стационарном режиме образуют положительно возвратный регенерирующий процесс. Поэтому с вероятностью 1 существует упоминавшийся выше предел

$$\lim_{t\to\infty} \frac{R(t)}{t} := \tilde{\mu}_0, \tag{21}$$

который является интенсивностью потока неудачных попыток (т. е. потока, идущего на орбиту) и, одновременно, потока попыток (удачных и неудачных) попасть с орбиты на сервер. (Анализ процессов накопления можно найти в [Smith, 1955].)

Теперь сформулируем и докажем следующую теорему, обозначив $\tilde{\rho} = (\lambda + \mu_0)\mathsf{E}S$.

**Теорема 2.** *В рассматриваемой системе $\Sigma$ с повторными вызовами предел*

$$\lim_{t\to\infty} \frac{D(t)}{t} = \lambda_e \tag{22}$$

*всегда существует, и выполнено следующее равенство:*

$$\lambda_e \mathsf{E}S = \mathsf{P}_{busy}, \tag{23}$$

*где $\mathsf{P}_b$ – стационарная вероятность занятости сервера. При этом*

*а) в стационарном режиме, т. е. при выполнении условия* (1)

$$\lambda_e = \lambda, \ \ \mathsf{P}_{orb} = \frac{\tilde{\mu}_0}{\lambda + \tilde{\mu}_0}; \tag{24}$$

*б) в нестационарном режиме, т. е. при выполнении условия* (14)

$$\lambda_e = (\lambda + \mu_0)(1 - \mathsf{P}_{orb})$$

*и*

$$\mathsf{P}_{orb} = 1 - \frac{\mathsf{P}_b}{\tilde{\rho}}. \tag{25}$$

*Доказательство.* Напомним, что $A(t)$ – общее число обращений к серверу (включая повторные попытки орбитальных заявок), $U(t)$ – число *успешных* попыток, а $R(t)$ – число уходов на орбиту (т. е. число неудачных попыток), все – в интервале $[0,t]$. Очевидно, что $A(t) = U(t) + R(t)$. Определим исходную последовательность (независимых, одинаково распределенных) времен обслуживания заявок $J = \{S_1, S_2, \dots\}$, из которой реальные времена обслуживания выбираются не в порядке поступления $\lambda$-заявок, а в порядке успешных попыток попасть на сервер. Таким образом, если $j$-я попытка оказывается не успешной, т. е. соответствующая ей заявка уходит на орбиту (возможно, не в первый раз), то величина $S_j$ оказывается фиктивным временем обслуживания и выбрасывается из последовательности $J$. Это правило назначения приводит к тому, что последовательность реальных времен обслуживания, вообще говоря, не совпадает с исходной последовательностью $J$, что объясняет появление стохастических неравенств ниже (вместо неравенств с вероятностью 1). Пусть $\nu(t)$ – число заявок на сервере в момент $t$ ($\nu(t) = 0$ или 1). Независимо от режима работы системы выполняются следующие неравенства:

$$\sum_{i=1}^{D(t)} S_i \leqslant_{st} \int_0^t \nu(u)du \leqslant \sum_{j=1}^{A(t)} I_j S_j, \tag{26}$$

где индикатор $I_j = 1$, если $j$-я попытка успешна, $j \geqslant 1$. (Отметим, неравенства (19) также можно трактовать как стохастические.) Заметим, что оба неравенства переходят в равенство, если $\nu(t) = 0$. Иначе значение суммы слева меньше значения интеграла на уже *прошедшую* часть времени обслуживания заявки, находящейся на сервере в момент $t$, а сумма справа больше интеграла на *оставшееся* в момент $t$ время обслуживания этой заявки. Поскольку в системе лишь один сервер, то процесс $\nu(t)$ фактически является индикатором занятости, т. е. $\nu(t) = I(\nu(t) > 0)$. Следовательно, предполагая, что существует слабый предел $\nu(t) \Rightarrow \nu$, получим

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \nu(u)du = \mathsf{E}\nu = \mathsf{P}(\nu > 0) := \mathsf{P}_{busy}. \tag{27}$$

Далее доказательство проводится отдельно для стационарного и нестационарного режимов.

*Стационарный режим.* Пусть выполнено условие (1). Перепишем (26), усилив правое неравенство,

$$\sum_{i=1}^{D(t)} S_i \leqslant_{st} \int_0^t \nu(u)du \leqslant \sum_{i=1}^{A(t)} S_i. \qquad (28)$$

Ввиду стационарности все $\lambda$-заявки, поступающие в систему, в конце концов получают обслуживание (сразу или после возвращения с орбиты). Таким образом, с вероятностью 1,

$$A(t) - D(t) \leqslant N(t) + 1 = o(t). \qquad (29)$$

(Условие $N(t) = o(t)$ следует из положительной возвратности процесса величины орбиты $\{N(t)\}$, см. [Smith, 1955; Morozov, 2004]). Отсюда следует что

$$\lim_{t\to\infty} \frac{D(t)}{t} = \lim_{t\to\infty} \frac{U(t)}{t} \to \lambda, \ \ t\to\infty, \qquad (30)$$

и $\lambda_e = \lambda$. Поскольку с вероятностью 1

$$\frac{1}{t}\sum_{i=1}^{D(t)} S_i = \frac{D(t)}{t}\frac{\sum_{i=1}^{D(t)} S_i}{D(t)} \to \lambda \mathsf{E}S,$$

$$1/t\sum_{i=1}^{U(t)} S_i \to \lambda \mathsf{E}S, \quad t\to\infty, \qquad (31)$$

то первое утверждение Теоремы 2 доказано. Далее, очевидно, что

$$\sum_{i=1}^{U(t)} S_i =_{st} \sum_{j=1}^{U(t)+R(t)} I_j S_j. \qquad (32)$$

(Напомним, что реализованные интервалы обслуживания $S_j$ при $I_j = 1$ в (32) справа, вообще говоря, не совпадают с $S_i$ слева.) Кроме того, $S_j I_j \Rightarrow SI_\infty$, при $j\to\infty$, где $S$ – типичное время обслуживания и $I_j \Rightarrow I_\infty$. Заметим, что событие $I_\infty = 1$ означает поступление заявки на сервер в стационарном режиме. Поэтому $\mathsf{P}(I_\infty = 1) = 1 - \mathsf{P}(I = 1) = 1 - \mathsf{P}_{orb}$ (см. (9)). Поскольку случайные величины $S_j$ и $I_j$ независимы для каждого $j$, то получаем $\mathsf{E}(SI_\infty) = \mathsf{E}S(1 - \mathsf{P}_{orb})$. Таким образом, учитывая также (21), можно было бы ожидать, что

$$\frac{1}{t}\sum_{j=1}^{U(t)+R(t)} I_j S_j \to (\lambda + \tilde\mu_0)\mathsf{E}(SI_\infty)$$
$$= (\lambda + \tilde\mu_0)\mathsf{E}S(1 - \mathsf{P}_{orb}), \quad t\to\infty. \qquad (33)$$

Однако случайные величины $\{S_j I_j,\, j \geqslant 1\}$ зависимы. Поэтому для корректного применения усиленного закона больших чисел будем использовать другое представление суммы $\sum_{j=1}^{U(t)+R(t)} I_j S_j$, опираясь на равенство (32). Именно, запишем (для $t > 0$)

$$\frac{\sum_{i=1}^{U(t)} S_i}{t}$$

$$= \frac{\sum_{i=1}^{U(t)} S_i}{U(t)} \cdot \frac{U(t)}{R(t) + U(t)} \cdot \frac{U(t) + R(t)}{t}. \qquad (34)$$

Очевидно, что

$$\frac{U(t) + R(t)}{t} \to \lambda + \tilde\mu_0,$$
$$\frac{U(t)}{U(t) + R(t)} \to 1 - P_{orb},\ t\to\infty, \qquad (35)$$

Отсюда, а также из (32) следует утверждение (33). Приравняв пределы (31) и (33), получаем

$$\lambda \mathsf{E}S = (\lambda + \tilde\mu_0)\mathsf{E}S(1 - \mathsf{P}_{orb}), \qquad (36)$$

откуда следует формула (24) для вероятности ухода на орбиту в стационарном режиме.

*Нестационарный режим.* Пусть выполняется условие (14). Так как тогда величина орбиты $N(t)\to\infty$ с вероятностью 1, то $R(t)/t \to \mu_0$, и легко видеть, что теперь соотношение (33) выполнено с заменой $\tilde\mu_0$ на $\mu_0$. Далее, как и выше, получим

$$\mathsf{P}_{busy} = (\lambda + \mu_0)\mathsf{E}S(1 - \mathsf{P}_{orb}). \qquad (37)$$

Отсюда и из (26) следует $\lambda_e = (\lambda + \mu_0)(1 - \mathsf{P}_{orb})$ и выражение (25). Теорема доказана. □

Заметим, что в случае пуассоновского $\lambda$-потока в нестационарном режиме суммарный входной поток в систему также будет пуассоновским с параметром $\lambda + \mu_0$. Тогда, опираясь на свойство PASTA (по которому $\mathsf{P}_{orb} = \mathsf{P}_{busy}$), для стационарной вероятности занятости из ((37)) получаем хорошо известное выражение

$$\mathsf{P}_{busy} = \frac{\tilde\rho}{\tilde\rho + 1}. \qquad (38)$$

(Свойство PASTA рассматривается, например, в [Asmussen, 2002]).

### Альтернативные оценки вероятности блокировки. Метод уменьшения дисперсии

Этот раздел в значительной мере опирается на важную работу [Spricant, Whitt, 1999]. Напомним, что *прямая оценка* вероятности блокировки имеет следующий вид:

$$B_N(t) = \frac{R(t)}{A(t)}. \qquad (39)$$

Поскольку $B_N(t) \to \mathsf{P}_{orb}$ при $t \to \infty$, то эта оценка состоятельная. Однако она представляет собой отношение двух величин, а такие оценки, вообще говоря, являются смещенными, т. е. $\mathsf{E}B_N(t) \neq \mathsf{P}_{orb}$. Но если процессы $\{R(t)\}$ и $\{A(t)\}$ имеют стационарные приращения, то, очевидно, $\mathsf{E}R(t)/\mathsf{E}A(t) = \mathsf{P}_{orb}$ для любого $t$.

Близкая к прямой *простая оценка*, предложенная в [Spricant, Whitt, 1999], опирается на асимптотику процессов восстановления и принимает для рассматриваемой здесь системы вид

$$B_S(t) = \frac{R(t)}{(\lambda + \tilde{\mu}_0)t}. \qquad (40)$$

(В нестационарном режиме $\tilde{\mu}_0$ заменяется на $\mu_0$.) Оценка $B_S(t)$ может быть предпочтительнее прямой оценки в системе с потерями, где параметр $\mu_0$ задан. Однако присутствие параметра $\tilde{\mu}_0$, требующего оценивания (наряду с $R(t)$), лишает ее данного преимущества в системе с повторными вызовами.

Гораздо более интересная *непрямая оценка* вероятности $\mathsf{P}_{orb}$, предложенная в [Spricant, Whitt, 1999], имеет вид

$$B_I(t) = 1 - \frac{\mathsf{P}_{busy}(t)}{\tilde{\rho}}, \qquad (41)$$

где

$$\mathsf{P}_{busy}(t) := \frac{1}{t}\int_0^t \nu(u)du, \qquad (42)$$

есть доля времени занятости сервера в интервале $[0, t]$.

В работе [Spricant, Whitt, 1999] показано, что оценка (41) является более (менее) эффективной нежели оценка $B_N(t)$ в случае большой (малой) нагрузки. (Бóльшая эффективность означает меньшую дисперсию.) Это приводит к рассмотрению следующей *комбинированной оценки*

$$B_C(t) = pB_N(t) + (1-p)B_I(t), \qquad (43)$$

где параметр $p$ выбирается таким образом, чтобы минимизировать дисперсию. Идея комбинирования основана на том, что с ростом $\mathsf{P}_{busy}(t)$ оценка $B_I(t)$ убывает (это видно из (41)), в то время как оценка $B_N(t)$ должна расти, так как увеличение доли времени занятости сервера в системе с потерями увеличивает потери. Таким образом, оценки $B_I(t)$ и $B_N(t)$ должны быть отрицательно коррелированы.

Для системы с повторными вызовами было бы естественно использовать аналог этой оценки,

$$\tilde{B}_I(t) = 1 - \frac{\mathsf{P}_{busy}(t)}{(\lambda + \mu_0(t))\mathsf{E}S}, \qquad (44)$$

где параметр $\mu_0$ заменен на оцениваемый параметр $\mu_0(t) = R(t)/t$ – локальную интенсивность потока попыток, идущих с орбиты (и на орбиту) в интервале $[0, t]$. Заметим, что в стационарном режиме (см. (24))

$$\tilde{B}_I(t) \to 1 - \frac{\rho}{\tilde{\rho}} = \frac{\tilde{\mu}_0}{\lambda + \tilde{\mu}_0} = \mathsf{P}_{orb}, \ \ t \to \infty. \qquad (45)$$

Преимущество комбинированной оценки (43)) в случае большой нагрузки в системе с потерями достигается за счет непрямой оценки (41). В рассматриваемой модели без потерь большая нагрузка, как правило, соответствует нестационарному режиму, в котором, как показано в разделе 5, комбинированная оценка действительно является более эффективной.

Ниже приведено краткое обоснование этого утверждения, следуя [Spricant, Whitt, 1999]. Для анализа эффективности оценок удобно представить отношение их дисперсий в виде

$$\frac{DB_C(t)}{DB_N(t)} = \frac{DB_C(t)}{DB_I(t)}\frac{DB_I(t)}{DB_N(t)}. \qquad (46)$$

В [Spricant, Whitt, 1999] показано, что отношение дисперсий $DB_C(t)/DB_I(t)$ зависит от отношения $r^2 := DB_I(t)/DB_N(t)$ и корреляции $\rho := Corr(B_I(t), B_N(t))$. (Для удобства мы иногда опускаем зависимость от $t$.) Можно показать, что

$$\begin{aligned} V(p) &:= DB_C(t) \\ &= DB_I(t)\Big(\frac{p^2}{r^2} + (1-p)^2 \\ &+ 2p(1-p)\frac{\rho}{r}\Big). \end{aligned} \qquad (47)$$

Несложно также показать, что вторая производная $V''(p) > 0$ при всех $p$, а значит минимум достигается при $p = p^\star$ таком, что $V'(p^\star) = 0$, причем

$$p^\star = \frac{r(r-\rho)}{1 + r^2 - 2r\rho}. \qquad (48)$$

При этом $0 < p^\star < 1$, так как мы предполагаем неположительную корреляцию, $\rho \leqslant 0$. Окончательно минимальная дисперсия имеет вид:

$$V(p^\star) = \frac{DB_I(t)(1-\rho^2)}{1+r^2-2r\rho}. \tag{49}$$

Поэтому *коэффициент уменьшения дисперсии*

$$R(p) := \frac{V(p)}{DB_I(t)} \tag{50}$$

принимает при $p = p^\star$ значение

$$R(p^\star) = \frac{1-\rho^2}{1+r^2-2r\rho}. \tag{51}$$

При $r \leqslant 1$ получим следующую нижнюю оценку выигрыша за счет использования комбинированной оценки

$$R(p^\star) \geqslant \frac{1-\rho^2}{4}, \tag{52}$$

а при равных дисперсиях $DB_I = DB_N$, т. е. при $r = 1$, получим $R(p^\star) = (1+\rho)/2$. Эта величина принимает большие значения лишь при $\rho \approx -1$. В этом случае нижняя граница (52) может быть аппроксимирована как

$$\frac{1-\rho^2}{4} \approx \frac{1+\rho}{2}. \tag{53}$$

Таким образом, при равных дисперсиях прямой и непрямой оценок и сильной отрицательной корреляции получаем $R(p^*) \approx 0$, т. е. очень значительное уменьшение дисперсии. При малом $r$ (т. е. при $DB_I \ll DB_N$) из (51) следует

$$\lim_{r\to 0} R(p^*) = 1-\rho^2, \tag{54}$$

независимо от знака $\rho$, что дает большой выигрыш при $|\rho| \approx 1$. Предел (54) отличается от нижней границы для $r$ в (52), достигаемой при $r = 1$ и $\rho \approx -1$, в 4 раза. Согласно [Spricant, Whitt, 1999], в случае малых значений $r$ уменьшение дисперсии в (46) можно приближенно выразить через произведение $1-\rho^2$ и $r$. Поэтому, как было сказано выше, комбинированная оценка особенно эффективна, когда $\rho \approx -1$.

Как показано в следующем разделе, приведенные выше результаты для системы с потерями хорошо согласуются с результатами численного моделирования исходной системы с повторными вызовами без потерь. Это вновь подтверждает определенную близость двух данных систем, что уже было показано при анализа стационарности.

## Численное моделирование системы типа $M/G/1$

В данном разделе представлены результаты оценивания вероятности блокировки $\mathsf{P}_{orb}$ в исходной системе $\Sigma$ типа $M/G/1/1$ с повторными вызовами с использованием регенеративного метода. Имитационное моделирование проведено для 20 000 заявок. Заметим, что оценки построены на основе соответствующего числа регенераций (или квази-регенераций), полученных в результате процедуры моделирования. Однако для удобства, зависимость оценок от числа заявок (или регенераций) опущена. Положим $\lambda = 1$ и обозначим через

$$\Gamma = \mu - 1 - \frac{1}{\mu_0}, \tag{55}$$

разность правой и левой частей в условии (6). Эта величина характеризует *меру стационарности* системы, поскольку $\Gamma > 0$ ($\Gamma < 0$) в области стационарности (нестационарности).

Зоны стационарности/нестационарности данной системы представлены на рис. 1, где граница зон определяется уравнением $\Gamma = 0$.

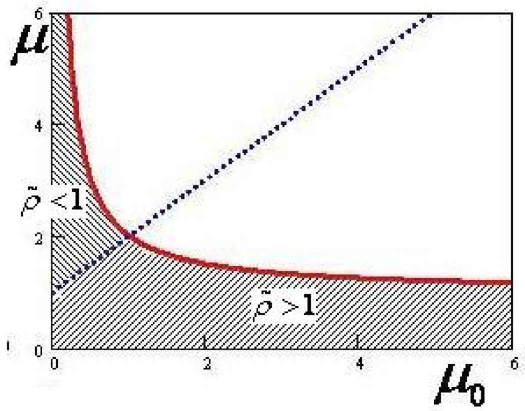



*Рис. 1.* Зоны стационарности/ нестационарности системы типа $M/M/1/1$ при $\lambda = 1$

Обсудим результаты, представленные в табл. 1. В *области стационарности* (т. е. при $\Gamma > 0$), значение прямой оценки $B_N \equiv B_N(t)$ отличается от значения вероятности потери $\mathsf{P}_{loss}$, получаемого по формуле Эрланга. При этом, чем больше значение $\Gamma$ (т. е. чем «глубже» в области стационарности выбраны параметры $\mu$, $\mu_0$), тем больше значение разности $\mathsf{P}_{loss} - B_N > 0$. Как уже отмечалось в разделе 1, это вызвано убыванием интенсивности потока заявок, идущих на орбиту, что, в свою очередь, вызывает убывание вероятности

*Таблица 1.* Оценивание вероятности $\mathsf{P}_{orb}$ в системе $M/M/1/1$ при $\lambda = 1$

| № | $\mu_0$ | $\tilde{\mu}_0$ | $\mu$ | $\tilde{\rho}$ | $\Gamma$ | $\mathsf{P}_{loss}$ | $B_N$ | $B_I$ | $B_C$ | $r^2$ | $\rho$ | $\frac{DB_C}{DB_N}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10,0 | 0,11 | 15,0 | 0,07 | 13,90 | 0,423 | 0,100 | 0,007 | -0,035 | 1,562 | 0,907 | 1,467 |
| 2 | 5,0 | 0,44 | 5,0 | 0,29 | 3,80 | 0,545 | 0,296 | 0,081 | 0,112 | 0,879 | 0,913 | 0,876 |
| 3 | 17,0 | 1,36 | 4,0 | 0,59 | 2,94 | 0,818 | 0,595 | 0,254 | 0,054 | 0,531 | 0,908 | 0,449 |
| 4 | 2,0 | 1,99 | 1,0 | 2,99 | -0,50 | 0,750 | 0,748 | 0,751 | 0,750 | 0,071 | -0,556 | 0,036 |
| 5 | 6,0 | 5,91 | 0,3 | 23,04 | -0,87 | 0,959 | 0,958 | 0,959 | 0,959 | 0,001 | -0,702 | 0,000 |
| 6 | 0,1 | 0,10 | 2,0 | 0,55 | -9,00 | 0,355 | 0,355 | 0,354 | 0,354 | 3,309 | -0,358 | 1,702 |
| 7 | 10,0 | 5,36 | 1,5 | 4,24 | 0,40 | 0,880 | 0,840 | 0,781 | 0,735 | 0,325 | 0,932 | 0,163 |

$\mathsf{P}_{orb}$ и возрастание разности $\mathsf{P}_{loss} - \mathsf{P}_{orb}$. Отметим, что прямая и непрямая оценки в стационарном режиме положительно коррелируют, $\rho > 0$. Кроме того, заметим, что в стационарном режиме загрузка может быть низкой даже в случае достаточного больших значений $\mu_0$, поскольку реальная интенсивность $\tilde{\mu}_0$ может быть достаточно мала (см. табл. 1, строка 3). Отметим, что в «глубине» области стационарности $r^2 > 1$, т. е. дисперсия непрямой оценки больше дисперсии прямой оценки (см. табл. 1, строка 1). Поэтому для оценивания $\mathsf{P}_{orb}$ следует использовать лишь прямую оценку $B_N$. (Это полностью соответствует случаю малой нагрузки в системе с потерями [Spricant, Whitt, 1999]). В *нестационарном случае* прямая и непрямая оценки отрицательно коррелируют, $\rho < 0$. При этом максимальное снижение дисперсии достигается при больших значениях $\tilde{\rho}$, т. е. при большой загрузке, что вновь соответствует результатам работы [Spricant, Whitt, 1999] для системы с потерями. Заметим, что при малом $r^2$ и $\rho \approx -1$, дисперсия комбинированной оценки снижается практически до нуля (табл. 1, строка 5), что соответствует выводу из формулы (54). Однако *сильная нестационарность* (т. е. большое отрицательное значение $\Gamma$) не гарантирует большую загрузку системы (табл. 1, строка 6). В частности, при выполнении двойного неравенства

$$1 + \mu_0 < \mu < 1 + \frac{1}{\mu_0}, \tag{56}$$

система нестационарна при малой загрузке, т. е. при $\tilde{\rho} < 1$. В этом случае дисперсия непрямой оценки превышает дисперсию прямой оценки $B_N$ (строка 6) и комбинирование не имеет смысла.

Таким образом, в системе типа $M/M/1/1$ комбинированная оценка $B_C$ эффективна (и близка к прямой) только при большом значении $\tilde{\rho}$ в нестационарном режиме. В противном случае более эффективна прямая оценка $B_N$.

Области значений параметров $\mu$ и $\mu_0$, определяющие малую/большую загрузку, представлены на рис. 2.

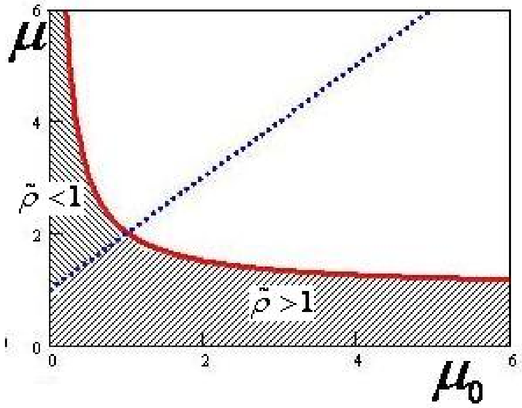



*Рис. 2.* Области малой ($\tilde{\rho} < 1$) и большой ($\tilde{\rho} > 1$) загрузки в зоне нестационарности системы типа $M/M/1/1$ при $\lambda = 1$

Обратимся к анализу результатов оценивания в системе типа $M/Pareto/1/1$, представленных в табл. 2. В данном случае мера стационарности равна

$$\Gamma = 1 - \lambda - \frac{1}{\alpha} - \frac{\lambda^2}{\mu_0}. \tag{57}$$

Эта величина (при фиксированном $\lambda$) увеличивается с ростом $\mu_0$ и $\alpha$. Коэффициент загрузки в данной системе равен

$$\tilde{\rho} = (\lambda + \tilde{\mu}_0)\frac{\alpha}{\alpha - 1} \tag{58}$$

и возрастает (при фиксированном $\lambda$) с ростом $\tilde{\mu}_0$.

Как видно из табл. 2 (см. табл. 2, строки 1–3), изменения параметра $\mu_0$ (в области больших значений) мало влияют на величину $\tilde{\mu}_0$. (Это связано с тем, что *в глубине области стационарности* орбита имеет большой запас *мощности* равный $\mu_0 - \tilde{\mu}_0$.) Таким обра-

*Таблица 2.* Оценивание вероятности $\mathsf{P}_{orb}$ в системе $M/Pareto/1/1$ при $\lambda = 0,5$

| № | $\mu_0$ | $\tilde{\mu}_0$ | $\alpha$ | $\tilde{\rho}$ | $\Gamma$ | $\mathsf{P}_{loss}$ | $B_N$ | $B_I$ | $B_C$ | $r^2$ | $\rho$ | $\frac{DB_C}{DB_N}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10,00 | 3,01 | 10 | 3,90 | 0,37 | 0,921 | 0,856 | 0,770 | 0,695 | 0,302 | 0,953 | 0,108 |
| 2 | 5,00 | 2,22 | 5 | 3,40 | 0,25 | 0,873 | 0,814 | 0,735 | 0,663 | 0,320 | 0,953 | 0,122 |
| 3 | 2,00 | 1,66 | 3 | 3,23 | 0,04 | 0,789 | 0,770 | 0,713 | 0,657 | 0,332 | 0,958 | 0,119 |
| 4 | 1,00 | 1,00 | 3 | 2,25 | -0,08 | 0,692 | 0,694 | 0,694 | 0,694 | 0,099 | -0,085 | 0,085 |
| 5 | 0,05 | 0,05 | 3 | 0,83 | -4,83 | 0,452 | 0,452 | 0,452 | 0,452 | 0,938 | -0,058 | 0,456 |
| 6 | 0,05 | 0,05 | 10 | 0,61 | -4,60 | 0,379 | 0,380 | 0,378 | 0,379 | 1,892 | -0,008 | 1,228 |

зом, значение коэффициента загрузки $\tilde{\rho}$ остается стабильным и достаточно большим. Это, в свою очередь, влечет эффективность оценки $B_C$ (см. последний столбец), что вновь согласуется с результатами [Spricant, Whitt, 1999]. По аналогии с системой $M/M/1/1$, можно получить следующее двойное неравенство, соответствующее зоне нестационарности при малой загрузке ($\tilde{\rho} < 1$),

$$1 - \frac{\lambda}{\mu_0}(\lambda + \mu_0) < \frac{1}{\alpha} < 1 - (\lambda + \mu_0). \qquad (59)$$

Следовательно, в зоне нестационарности система имеет малую загрузку только при $\lambda < \mu_0$. В этом случае эффективность оценки $B_C$ уменьшается (см. строки 4-5), что вновь согласуется с результатами для системы с потерями.

Области значения параметров $\alpha$ и $\mu_0$, определяющие малую/большую загрузку в зоне нестационарности системы $M/Pareto/1/1$, представлены на рис. 3.

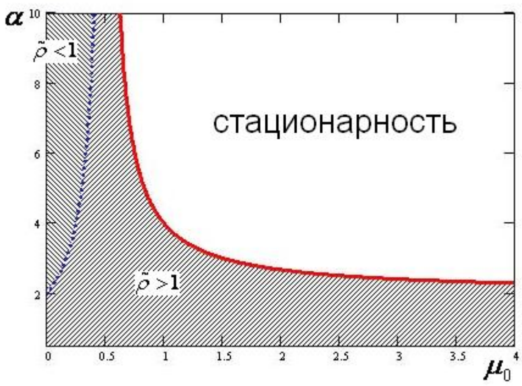



*Рис. 3.* Области малой и большой загрузки в зоне нестационарности системы типа $M/Pareto/1/1$ при $\lambda = 0,5$

Заметим, что при определенной конфигурации параметров система находится в зоне стационарности при большой загрузке ($\tilde{\rho} > 1$) (табл. 1, строка 7; табл. 2 строки 1–3). В этом случае оценка $B_C$ обладает меньшей дисперсией, однако ее значение систематически меньше значения прямой оценки $B_N$. Это говорит о том, что в зоне стационарности использовать комбинированную оценку нежелательно. Причина такой несогласованности оценок пока остается неясной и требует дальнейшего исследования. Скорее всего, это связано с различием между исходной системой с повторными вызовами и системой с потерями, для которой изначально были построены данные оценки.

Подчеркнем, что в обеих исследованных системах малая загрузка сочетается с нестационарностью, и это отражает их существенное отличие от системы с потерями.

## Заключение

В статье рассмотрена проблема оценивания вероятности блокировки в системе с повторными вызовами и постоянной скоростью возвращения заявок с орбиты на сервер. Приведено условие стационарности такой системы. Обсуждается обобщенная формула Литтла, на основе которой найдены соотношения для вероятности блокировки как в стационарном, так и в нестационарном режимах. Рассмотрены две основные альтернативные статистические оценки вероятности блокировки, имеющие отрицательную корреляцию в нестационарном режиме. Это позволяет построить комбинированную оценку, имеющую существенно меньшую дисперсию. Проведено статистическое оценивание вероятности блокировки в стационарном режиме с использованием классической регенерации, а также в нестационарном режиме с использованием newline квази-регенерации. Представленные численные расчеты хорошо согласуются с теоретическими условиями стационарности, а также с результатами работы [Spricant, Whitt, 1999] (для системы с потерями) об эффективности комбинированной оценки при большой загрузке.

Работа выполнена при поддержке РФФИ, проект 10-07-00017.

## Литература

*Asmussen S.* Applied Probability and Queues // Springer, 2002. 440 p.

*Avrachenkov K., Goricheva R. S., Morozov E. V.* Verification of stability region of a retrial queuing system by regenerative method // Proceedings of the Intenational Conference «Modern Probabilistic Methods for Analysis and optimization of Information and Telecommunication Networks». Minsk, 2011. P. 22–28.

*Avrachenkov K., Morozov E. V.* Stability analysis of GI/G/c/K Retrial Queue with Constant Retrial Rate // INRIA(Sophia Antipolis), Research Report. 2010. N. 7335.

*Avrachenkov K., Yechiali U.* Retrial networks with finite buffers and their application to Internet data traffic // Probability in the Engineering and Informational Sciences. 2008. Vol. 22. P. 519–536.

*Bratley P., Fox B. L.* A guide to simulation // New York: Springer-Verlag, 1987. 424 p.

*Choi B. D., Rhee K. H., Park K. K.* The M/G/1 retrial queue with retrial rate control policy // Probability in the Engineering and Informational Sciences, 1993. Vol. 7. P. 29–46.

*Fayolle G.* A simple telephone exchange with delayed feedback / In O. J. Boxma, J. W. Cohen and H. C. Tijms (ed.) // Teletraffic Analysis and Computer Performance Evaluation. 1986. Vol. 7. P. 245–253.

*Glynn P. W., Iglehart D. L.* Conditions for the applicability of the regenerative method // Management Science. 1993. Vol. 39. P. 1108–1111.

*Law A. M. , Kelton W. D.* Simulation modeling and analysis. New York: McGraw-Hill, 2000. 560 p.

*Morozov E. V.* Weak regeneration in modeling of queueing processes // Queueing Systems. 2004. Vol. 46. P. 295–315.

*Smith W. L.* Regenerative stochastic processes // Proc. Roy. Soc. 1955. Ser. A 232. P. 6–31.

*Sonderman D.* Comparing multi-server queues with finite waitng rooms, I: Same number of servers // Advances in Applied Probability. 1979. Vol. 11. P. 439–447.

*Sonderman D.* Comparing multi-server queues with finite waiting rooms, II: Different number of servers // Advances in Applied Probability. 1981. Vol. 11. P. 448–455.

*Spricant R., Whitt W.* Variance reduction in simulations of loss models // Operation research. 1999. Vol. 47, N. 4. P. 509–523.

*Whitt W.* Comparing counting processes and queues // Advances in Applied Probability. 1981. Vol. 13. P. 207–220.

*Wolff R. W.* Stochastic Modeling and the Theory of Queues // Prentice-Hall, 1989. 560 p.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Морозов Евсей Викторович**
ведущий научный сотрудник, д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: emorozov@karelia.ru
тел.: (8142) 763370

**Morozov, Evsey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: emorozov@karelia.ru
tel.: (8142) 763370

**Некрасова Руслана Сергеевна**
аспирантка
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: ruslana.nekrasova@mail.ru
тел.: (8142) 763370

**Nekrasova, Ruslana**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: ruslana.nekrasova@mail.ru
tel.: (8142) 763370

Труды Карельского научного центра РАН
№ 5. 2011. С. 75–85

*УДК 519.872.8: 004.382.2*

# МОДЕЛИ МНОГОСЕРВЕРНЫХ СИСТЕМ ДЛЯ АНАЛИЗА ВЫЧИСЛИТЕЛЬНОГО КЛАСТЕРА

**Е. В. Морозов, А. С. Румянцев**

*Институт прикладных математических исследований
Карельского научного центра РАН*

Работа посвящена изучению свойств процесса загрузки в многосерверных системах обслуживания, в том числе при наличии тяжелых хвостов. Проведен анализ моментных свойств процесса загрузки. Предложена модель вычислительного кластера как стохастической многосерверной системы обслуживания. Подробно обсуждается связь модели кластера с классическими моделями многосерверных систем. Приведены результаты статистических экспериментов, которые показывают хорошее согласие модели с работой кластера. Работа поддержана РФФИ, проект 07-10-00017.

Ключевые слова: вычислительный кластер, многосерверные системы, стационарность, имитационное моделирование, распределения с тяжелым хвостом.

**E. V. Morozov, A. S. Rumyantsev. MULTI-SERVER MODELS TO ANALYZE HIGH PERFORMANCE CLUSTER**

The paper is dedicated to the properties of the workload process in multi-server systems, i.a. in the presence of heavy tails. The moment properties of the workload process are analyzed. A model of a high performance cluster as a multi-server stochastic system is suggested. A connection between the model of the cluster and the classical multi-server system models is discussed in detail. The results of numerical experiments are provided, demonstrating an agreement between the model and the cluster performance.

Key words: high performance cluster, multi-server systems, stability, simulation, heavy-tailed distributions.

## Введение

Анализу стохастических моделей многосерверных систем обслуживания посвящено большое число работ. Теоретическое изучение мотивировано запросами производителей вычислительных мощностей, так как в последнее десятилетие имеет место повсеместная тенденция внедрения многоядерных и многопроцессорных архитектур. В то же время точный анализ таких моделей затруднен и аналитические результаты доступны только для узкого класса простейших моделей.

С точки зрения практики, исследователя могут интересовать характеристики системы, влияющие на качество обслуживания (QoS), например, величина средней задержки в системе, величина дисперсии задержки. Важной проблемой является определение моментных свойств сетевых процессов, в частности, так называемого моментного индекса, т. е. максимального конечного момента. Другим важным аспектом анализа является зависимость моментных свойств сетевых процессов от дисциплины обслуживания клиентов [Borst et al., 2002]. Особенно сложен анализ в случае, когда заявки в систему обслуживания поступают в виде пачек (см., напр., работы [Keilson, Seidmann, 1987; Wolff, 1991; Liu, 1993]), или в виде регенеративного процесса [Huang, Sigman, 1999]. Вычислительные кластеры и Грид представляют в этом смысле особую сложность, так как обладают вышеперечисленными свойствами, такими как многопроцессорность, сложные дисциплины обслуживания потока заявок, возможность поступления пачек заявок, необходимость в больших временах обслуживания на процессорах (тяжелые хвосты).

Теоретический интерес последних лет к исследованию моделей систем обслуживания с распределением с тяжелыми хвостами связан с обнаружением этого эффекта в траффиках компьютерных сетей (см., напр., [Crovella, Bestavos, 2002]). Тяжелый хвост означает более медленное, чем экспоненциальное, убывание хвостовой вероятности $\overline{F}(x) := P(X > x)$ случайной величины $X$, т. е.

$$\lim_{x\to\infty} e^{-ax}\overline{F}(x) = \infty, \quad \forall a > 0.$$

Это свойство радикально отличается от свойства экспоненциального распределения, попрежнему широко используемого для моделирования процессов в современных компьютерных системах. Присутствие тяжелых хвостов эмпирически обнаружено в распределениях размеров файлов на жестких дисках, времен вычисления задач на процессоре, времен передачи файлов между сервером и клиентом в высокоскоростных сетях и т. д. [Crovella, Bestavos, 2002]. Важный подкласс тяжелохвостых распределений составляют *субэкспоненциальные распределения*, которые (для неотрицательных с.в.) определяются как

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} = 2,\ x \geqslant 0. \tag{1}$$

Другой важный подкласс составляют *правильно меняющиеся распределения*, характеризуемые соотношением

$$\lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = t^{-\alpha},\ \alpha \geqslant 0.$$

(При $\alpha = 0$ функция называется *медленно меняющейся*). Наиболее часто используемым распределением такого рода является распределение Парето, имеющее (стандартный) вид

$$F(x) = 1 - x^{-\alpha},\ x \geqslant 1. \tag{2}$$

В то же время, обнаружение тяжелых хвостов в компьютерных системах обозначило ряд проблем, недооценивание которых приводит к серьезным последствиям как для компаний-производителей сетевого оборудования, так и для потребителей. (Частичный анализ такого рода ошибок приведен в работе [Morozov et al., 2008].) Одна из причин возникающих проблем состоит в том, что многие алгоритмы управления сетевыми устройствами и процессами в вычислительных системах строились с учетом предположения об экспоненциальности соответствующих распределений [Harchol-Balter, 1999]. Все это требует тщательного анализа многосерверных систем обслуживания и построения подходящих моделей.

## Модели многосерверных систем обслуживания

В этом разделе дан обзор ряда важных известных результатов для многосерверных систем обслуживания, в том числе с бесконечным числом серверов. Эти результаты в основном опираются на классические результаты для односерверных систем, которые также приводятся для сравнения.

### Система G/G/$s$ FIFO

Пусть в систему обслуживания в моменты времени $\{t_n, n \geqslant 1\}$, образующие процесс восстановления, поступают заявки с независимыми, одинаково распределенными временами обслуживания $S_n$ на одном из $s$ идентичных серверов (процессоров). Обозначим функцию распределения времени между приходами заявок $T_n := t_{n+1} - t_n$ через $A(x) = P(T \leqslant x)$, соответственно $B(x) = P(S \leqslant x)$ — распределение времени обслуживания заявки (индекс опущен, когда рассматривается типичный представитель последовательности). Заявка ожидает время $D_n \geqslant 0$ в очереди, формируемой в порядке поступления (FIFO), и поступает на обслуживание в наименее загруженный сервер. Вектор загрузки $W_n$, состоящий из оставшегося времени обслуживания на каждом процессоре в момент прихода заявки $n$, удовлетворяет рекурсии Кифера-Вольфовица [Scheller-Wolf, Sigman, 1997a]):

$$W_{n+1} = R\begin{pmatrix} W_n(1) + S_n - T_n \\ W_n(2) - T_n \\ \dots \\ W_n(s) - T_n \end{pmatrix}^+, \tag{3}$$

где оператор $R(\cdot)$ располагает компоненты в возрастающем порядке, $W_n(1) \leqslant \ldots \leqslant W_n(s)$, а $(x)^+ = \max(0, x)$. Таким образом, время ожидания $n$-й заявки $D_n := W_n(1)$. При условии $\rho := ES/ET < s$ имеет место слабая сходимость вектора загрузки к стационарному значению, $W_n \Longrightarrow W := (W(1), \ldots, W(s))$.

Обозначим через $U_n = W_n(2) - W_n(1)$ время, оставшееся на втором наименее загруженном сервере в тот момент, когда заявка $n$ поступит на обслуживание. Тогда первый из этих двух серверов освободится через время $P_n = \min(U_n, S_n)$, и время ожидания $D_n$ удовлетворяет модифицированной рекурсии Линдли [Scheller-Wolf, Sigman, 1997a]:

$$D_{n+1} = (D_n + P_n - T_n)^+, \quad n \geqslant 1. \qquad (4)$$

Заметим, что при $s = 1$ формула (4) определяет классическую рекурсию Линдли

$$D_{n+1} = (D_n + S_n - T_n)^+, \quad n \geqslant 1.$$

**Условия конечности моментов компонент вектора загрузки**

Следующее необходимое и достаточное условие конечности моментов времени ожидания в односерверной системе G/G/1 получено в работе [Kiefer, Wolfowitz, 1956]: при условии $ET < \infty$,

$$ES^{\alpha+1} < \infty \quad \Leftrightarrow \quad ED^{\alpha} < \infty. \qquad (5)$$

Обобщение результата (5) для вектора загрузки $W$ в $s$-серверной системе было получено в работе [Scheller-Wolf, Vesilo, 2011]. Так, при условии устойчивости $\rho = ES/ET < s$, для всех компонент вектора с индексами $i \leqslant \lceil \rho \rceil$ (наименьшее целое большее $\rho$) имеют место *одни и те же* достаточные условия конечности моментов порядка $\alpha \geqslant 1$:

$$E\left(S^{1+\alpha/(s-\lfloor \rho \rfloor)}\right) < \infty \quad \Rightarrow \quad E\,[W(i)]^{\alpha} < \infty.$$

В то же время для компонент с индексами $i > \lceil \rho \rceil$ моментные свойства зависят от индекса компоненты $W(i)$:

$$E\left(S^{1+\alpha/(s-i)}\right) < \infty \quad \Rightarrow \quad E\,[W(i)]^{\alpha} < \infty.$$

Частным случаем этого результата является достаточное условие конечности момента порядка $\alpha$ стационарного времени ожидания $D = W(1)$ в системе G/G/$s$ FIFO, полученное ранее в работе [Scheller-Wolf, Vesilo, 2006]:

$$E\left(S^{1+\alpha/(s-\lfloor \rho \rfloor)}\right) < \infty \;\Rightarrow\; E(D^{\alpha}) < \infty. \qquad (6)$$

При дополнительных ограничениях на распределение $B$, приведенные выше условия являются также необходимыми [Scheller-Wolf, Vesilo, 2011].

Таким образом, имеет место зависимость моментов компонент вектора $W$ как от индекса $i$, так и от коэффициента загрузки $\rho$. Фактически имеет место *взаимопомощь* среди первых $\lceil \rho \rceil$ наименее загруженных серверов, моментные свойства которых лучше. Это объясняется *запасом мощности*, состоящим из $s - \lceil \rho \rceil$ серверов, которые могут быть удалены из системы (или заняты обслуживанием очень длинных заявок) без потери устойчивости оставшейся части системы [Scheller-Wolf, Vesilo, 2011].

Частный случай результата (6) при малой загрузке $\rho < 1$ в системе G/G/$s$ был получен в работе [Scheller-Wolf, 2000]:

$$E\left(S^{1+1/s}\right) < \infty \quad \Rightarrow \quad ED < \infty.$$

В пределе при $s \to \infty$ отсюда следует классическое условие $ES < \infty$, влекущее конечность среднего стационарного времени пребывания заявки в системе G/G/$\infty$. В работе [Scheller-Wolf, Sigman, 1997b] приведена также верхняя граница для стационарной средней задержки в системе G/G/$s$.

На практике формула (6) позволяет обосновать выбор между *s медленными* серверами (работающими со скоростью $1/s$) и одним *быстрым* (работающим со скоростью 1) в случае, когда распределение времени обслуживания имеет тяжелый хвост. Действительно, при числе серверов $s > 1/(1-\rho)$, одном и том же входном потоке и временах обслуживания $\tilde{S}$ в $s$-серверной системе, пропорционально увеличенных в соответствии со скоростями серверов (т. е. $E\tilde{S} = s \cdot ES$), в $s$-серверной системе задержки будут иметь конечный момент более высокого порядка, чем в системе $G/G/1$. Для доказательства переформулируем соотношение (6), обозначив $\beta = 1 + \alpha/(s - \lfloor \rho \rfloor)$. Получим

$$E(\tilde{S}^{\beta}) < \infty \Rightarrow E\left(D^{(s-\lfloor \tilde{\rho} \rfloor)(\beta-1)}\right) < \infty. \qquad (7)$$

Для односерверной системы имеем

$$ES^{\beta} < \infty \quad \text{влечет} \quad ED^{\beta-1} < \infty.$$

Коэффициент загрузки в $s$-серверной системе $\tilde{\rho} = E\tilde{S}/ET = sES/ET < s$, т. е. система устойчива. С учетом выбора $s > 1/(1-\rho)$ получаем $s - s\rho > 1$, и поэтому $s - \lfloor \tilde{\rho} \rfloor = s - \lfloor s\rho \rfloor = \lfloor s - s\rho \rfloor \geqslant 1$.

Таким образом, если времена обслуживания имеют бесконечный момент порядка $\beta$ (например, если они распределены по закону Парето (2) с параметром $\beta > 0$), то целесообразнее использовать несколько медленных серверов, нежели один быстрый [Scheller-Wolf, Vesilo, 2006].

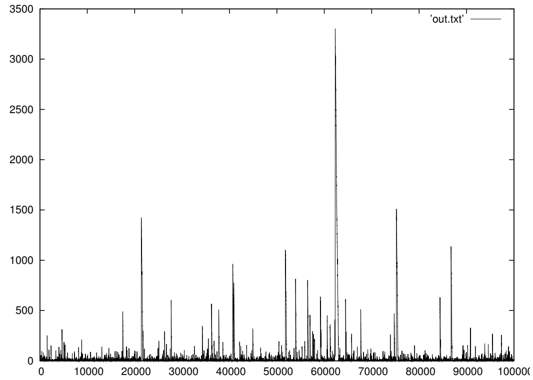



*Рис. 1.* Задержки в системе M/G/1, $\rho = 0,3$

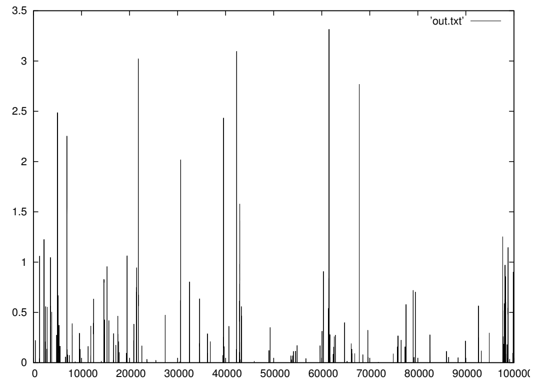



*Рис. 2.* Задержки в системе M/G/8, $\rho = 2,4$

Рассмотрим численный пример. Пусть система с процессором частотой 3 ГГц обслуживает процессы, времена выполнения которых имеют тяжелый хвост, $\overline{B}(x) = x^{-1,5}$. Пусть загрузка системы равна $\rho = 0,3$. Согласно формуле (5), стационарное время ожидания в системе имеет конечный момент только лишь порядка $0,5$, $ED^{0,5} < \infty$, т. е. бесконечное среднее. Заменим исходную систему на двухпроцессорную, где каждый процессор имеет частоту 1,5 ГГц. Общая загрузка системы возрастет до $0,6$. В то же время согласно (7), $ED^{2\cdot 0,5} = ED < \infty$ (поскольку $\lfloor\tilde{\rho}\rfloor = 0$), и средняя задержка конечна. Если же заменить систему на восьмипроцессорную, то $\lfloor\tilde{\rho}\rfloor = 2$. Поэтому $ED^3 < \infty$ и задержка имеет конечный третий момент. Иллюстрацией к численному примеру служат рисунки 1, 2, где приведены величины задержек в односерверной и восьмисерверной системах, соответственно, для выборки 100 000 заявок (задержки указаны по оси ординат).

**Двухсерверная система: случай тяжелого хвоста**

Классический результат (5) имеет интуитивное объяснение. Определяющим параметром для времени ожидания в системе является оставшееся время обслуживания $S_I$ заявки $n$ в момент прихода заявки $n+1$. Известно, что стационарное незавершенное время обслуживания имеет распределение

$$\overline{B}_I(x) = \frac{1}{ES}\int_x^\infty \overline{B}(y)\,dy,\ x \geqslant 0.$$

При этом условие $ES_I^{k-1} < \infty$ выполнено (при некотором $k > 1$), если $ES^k < \infty$, т. е. моментные свойства $S_I$ хуже, чем у $S$.

В работе [Foss, Korshunov, 2006] приведены асимптотические результаты для хвоста распределения стационарной задержки $D$ в двухсерверной системе $G/G/2$ при условии, что незавершенное время обслуживания имеет субэкспоненциальное распределение (1). Так, в случае *максимальной устойчивости*, т. е. при $\rho < 1$,

$$C_1 \leqslant \liminf_{x\to\infty} \frac{P(D > x)}{(\overline{B}_I(x))^2}$$

$$\leqslant \limsup_{x\to\infty} \frac{P(D > x)}{(\overline{B}_I(x))^2} \leqslant C_2.$$

В частности, в случае правильно меняющейся функции $\overline{B}(x)$ имеет место асимптотика

$$P(D > x) \sim C(\overline{B}_I(x))^2,\ x \to \infty,$$

где постоянная $C$ также зависит от $ES, ET$ [Foss, Korshunov, 2006]. В случае *минимальной устойчивости*, $1 < \rho < 2$, для правильно меняющейся функции распределения $B$ имеет место асимптотика

$$P(D > x) \sim C\overline{B}_I(\frac{ES}{ES - ET}x),\ x \to \infty.$$

Заметим, что постоянные $C_1, C_2, C$ выше зависит от параметров $ES, ET$.

## Модель вычислительного кластера

Вычислительный кластер – это система, состоящая из нескольких узлов, объединенных быстрой коммуникационной сетью и предоставляющая пользователю вычислительные ресурсы. Для построения модели кластера сделаем следующие предположения:

1. поток заявок представляет собой процесс восстановления (времена между приходами независимы и одинаково распределены);

2. разделяемым ресурсом является количество процессоров и время вычисления; времена вычисления заявок и требуемое каждой заявке число процессоров независимы;

3. заявки поступают в буфер неограниченной емкости и обслуживаются в порядке прихода (FIFO);

4. заявка поступает на обслуживание только в момент освобождения требуемого ей числа процессоров.

**Теоретический анализ**

В дополнение к обозначениям предыдущей секции, пусть $N_n$ — требуемое число процессоров для заявки $n$, на каждом из которых она будет выполняться в течение $S_n$ единиц времени. Предполагается, что (целочисленные) случайные величины $\{N_n\}$ независимы и одинаково распределены. Очевидно, что $1 \leqslant N_n \leqslant s$. Тогда вектор (упорядоченных по возрастанию) времен ожидания на серверах удовлетворяет следующей модифицированной рекурсии Кифера-Вольфовица

$$W_{n+1} = R \begin{pmatrix} W_n(N_n) + S_n - T_n \\ \dots \\ W_n(N_n) + S_n - T_n \\ W_n(N_n+1) - T_n \\ \dots \\ W_n(s) - T_n \end{pmatrix}^{+}. \qquad (8)$$

Действительно, $n$-я заявка дожидается в буфере (в очереди) освобождения всех $N_n$ требуемых ей процессоров. Поэтому загруженность всех серверов с номерами $[1, N_n]$ становится равной $W_n(N_n)$, поскольку, по крайней мере, часть этих серверов будет простаивать до освобождения наиболее загруженного из них, и в это время не смогут принять другие заявки. Это объясняет равенство первых $N_n$ строк в формуле (8). Отметим, что на практике такая неэффективная дисциплина используется редко. Более эффективным является алгоритм Backfill, позволяющий заполнять промежутки простоя отдельных процессоров заявками без существенного нарушения приоритетов и очередности.

Заметим, что в худшем случае при $N_n = s$ все компоненты вектора загрузки станут равными, поэтому при условии $W_n(s) + S_n < T_n$ заявка с номером $n+1$ начнет обслуживание в полностью пустой системе (все процессоры будут освобождены с уходом заявки $n$). Поскольку входной поток является процессом восстановления, то приходы таких заявок являются *моментами регенерации* системы. Это открывает возможность применения регенеративного моделирования для оценивания стационарных характеристик системы (см., напр., [Morozov, Rumyantsev, 2010]). Отметим, что суммарное время до освобождения всех необходимых $n$-й заявке процессоров (в момент ее прихода) равно

$$N_n \cdot (W_n(N_n) + S_n - T_n)^+ + \sum_{k=N_n+1}^{s} (W_n(k) - T_n)^+.$$

Эта величина является ключевой для определения условий стационарности рассматриваемой системы. Хотя такой анализ пока провести не удалось, проведенные эксперименты позволяют предположить, что если значения $N_n$ распределены равномерно в интервале $[1, s]$, то условие стационарности имеет вид $\rho = ES/ET < 1$, являясь существенно более ограничительным чем хорошо известное условие $\rho < s$ для классической системы $G/G/s$. Это предположение подкрепляется в частном случае, когда каждая заявка требует ровно $s$ процессоров. Тогда исходная система переходит в классическую систему $G/G/1$ с условием стационарности $\rho < 1$.

**Теорема 1.** *В рассматриваемой модели при $\rho = ES/ET < 1$ условие $ES^{\alpha+1} < \infty$ является достаточным для конечности момента порядка $\alpha$ стационарного времени ожидания в системе, т. е. $ED^{\alpha} < \infty$.*

*Доказательство.* Отметим, что задержка в системе $D_n := W_n(1)$ ограничена снизу значением задержки $D_n^{(low)} := W_n^{(low)}(1)$ для соответствующей системы G/G/$s$ с идентичными входным потоком с интервалами $T_n$ и временами обслуживания $S_n$. Действительно, из рекурсии (8) следует, что, в предположении индукции $D_n \geqslant D_n^{(low)}$

$$\begin{aligned} D_{n+1} &= W_{n+1}(1) = (W_n(N_n) + S_n - T_n)^+ \\ &\geqslant (W_n(1) + S_n - T_n)^+ = (D_n + S_n - T_n)^+ \\ &\geqslant (D_n^{(low)} + S_n - T_n) \geqslant (D_n^{(low)} + P_n - T_n)^+ \\ &= D_{n+1}^{(low)}. \end{aligned}$$

Одновременно задержка $D_n$ ограничена сверху задержкой $D_n^{(up)}$ для *доминирующей* *системы*, где число требуемых каждой заявке процессоров $N_n = s, n \geqslant 1$. Действительно,

$$\begin{aligned} D_{n+1} = W_{n+1}(1) = (W_n(N_n) + S_n - T_n)^+ \\ \leqslant (W_n(s) + S_n - T_n)^+ = (D_n^{(up)} + S_n - T_n)^+ \\ = D_{n+1}^{(up)}. \end{aligned}$$

Доминирующая система по сути представляет систему G/G/1, с теми же интервалами между приходами $T_n$ и временами обслуживания $S_n$, что и в исходной системе. Таким образом, для любых реализаций $\{T_n, S_n\}$

$$D_n^{(low)} \leqslant D_n \leqslant D_n^{(up)}. \quad (9)$$

Переформулируя (9) на уровне моментов порядка $\alpha$ и устремляя $n \to \infty$, получим

$$E(D^{(low)})^\alpha \leqslant ED^\alpha \leqslant E(D^{(up)})^\alpha.$$

Однако у *нижней* и *верхней* систем обслуживания, при условии $ES/ET < 1$, одно и то же достаточное условие конечности моментов времени ожидания, именно, $ES^{\alpha+1} < \infty$. Следовательно, это условие является достаточным для конечности момента $ED^\alpha$ в рассматриваемой модели. □

## Численный эксперимент

Моделирование рассматриваемых систем обслуживания проводилось на вычислительном кластере ЦКП КарНЦ РАН [ЦВОД ЦКП КарНЦ РАН, 2010], причем исходные данные описывают функционирование самого кластера. Именно данные для анализа получены из log-файлов системы управления заданиями Cleo, которая эксплуатировалась на кластере ЦКП КарНЦ РАН в период с 03.06.2009 г. по 04.02.2011 г.

Опишем подробнее структуру кластера и условия эксперимента. Кластер содержит 80 вычислительных ядер, сгруппированных по 8 ядер на каждый вычислительный узел. В процессе вычислений каждому пользователю разрешено занимать все узлы, но не менее одного (узел занимается пользователем целиком, даже если для вычислений используется лишь одно ядро этого узла). Кроме того, имеется ряд административных ограничений для пользователей кластера. В частности, введен запрет на постановку более чем 3 задач в очередь одновременно, введено ограничение на максимальное время вычисления задачи (3 суток), также есть ряд ограничений для студентов, использующих кластер в учебных целях.

За период эксплуатации системы Cleo на вычислительном кластере произведены расчеты 8292 заданий в однопроцессорном и многопроцессорном режимах. Общее время вычисления составило 28760 часов 50 минут 49 секунд. Среднее время вычисления одной задачи составило $ES \approx 3,4685$ часа или $0,1445$ суток. В то же время, астрономическое время эксплуатации системы составило 611 суток, т. е. среднее время между поступлениями заданий равно $ET \approx 1,7684$ часа или $0,0737$ суток.

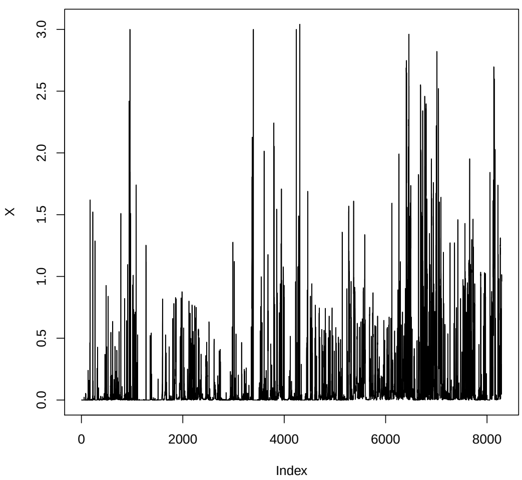



*Рис. 3.* Времена обслуживания заявок на кластере

**Времена обслуживания.** На рис. 3 изображены величины времен обслуживания заявок за рассматриваемый период. Из графика эмпирической функции распределения (рис. 4) видно, что большинство значений сконцентрировано в окрестности нуля. Это связано с тем, что пользователи кластера выполняют *отладку* задач, что часто приводит к почти мгновенному завершению задачи (в случае какой-либо ошибки в программе) и появлению (практически) нулевого времени вычисления в лог-файле. Для исключения отладочных запусков учитывались заявки, которые вычислялись более 1 часа, т. е. $0,04$ суток (2767 заявки).

Времена вычисления $\{S_n\}$ часто принимают значения, близкие к максимально допустимому (3 суток). Это, а также достаточно медленный рост эмпирической функции распределения может указывать на присутствие распределения с тяжелым хвостом. Однако ограниченность $S$ (типичное время вычисления) исключает такую возможность. В такой ситуации для моделирования времен обслуживания в компьютерных системах можно использовать усеченное распределение Парето [Harchol-Balter, 1999], имеющее функцию распределения

$$F(x) = \frac{\left(\frac{k}{x}\right)^{\alpha} - 1}{\left(\frac{k}{p}\right)^{\alpha} - 1}, \ 0 < k \leqslant x \leqslant p, \ \alpha > 0. \quad (10)$$

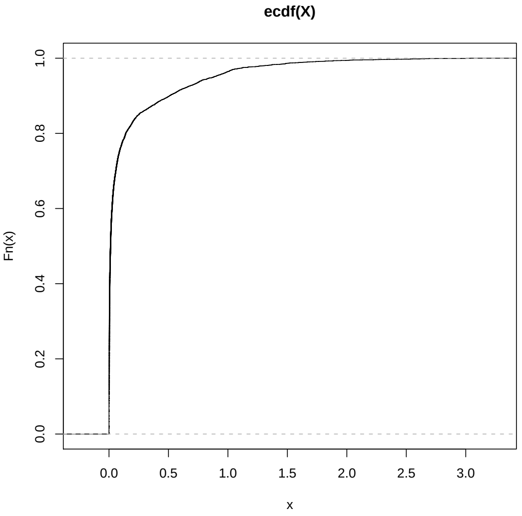



*Рис. 4.* Эмпирическая ф.р. времени вычисления

Случайная величина с таким распределением принимает значения в интервале $[k, p]$ и, кроме того, имеет конечные моменты всех порядков. Однако с убыванием параметра $\alpha$ имеет место экспоненциальный рост значений моментов [Harchol-Balter, 1999], (например, дисперсии при $\alpha < 2$). Это свойство до некоторой степени отражает свойство распределений с тяжелым хвостом. На рис. 5 показаны плотности распределения времен обслуживания (сплошная линия) и времен, имеющих усеченное распределение Парето (10) с параметрами $k = 0,04$, $p = 3,04$, $\alpha = 0,3$ (штриховая линия, выборка из 2767 значений). Для различения графиков в области малых значений $S$ выбран логарифмический масштаб по оси абсцисс. В качестве гипотезы $H_0$ рассматривалось утверждение об одинаковом распределении двух выборок. Значение критерия Колмогорова-Смирнова для этих двух выборок равно $D = 0,0517$, а соответствующее значение статистического уровня значимости, при котором гипотеза $H_0$ принимается, не больше $0,001228$. Это значение мало, так как для практических целей обычно используют уровни значимости $\geqslant 0,01$ (чтобы увеличить мощность критерия, т. е. уменьшить вероятность ошибки второго рода).

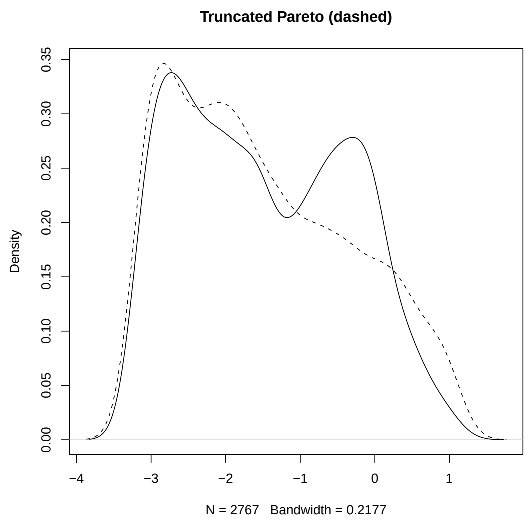



*Рис. 5.* Плотность распределения времени вычисления (усеченное Парето, $\alpha = 0,3$)

Ввиду широкого применения марковских моделей, естественно было бы также проверить применимость (усеченного) экспоненциального распределения, плотность которого имеет вид

$$f(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda k} - e^{-\lambda p}}, \ 0 < k \leqslant x \leqslant p, \qquad (11)$$

где постоянная $\lambda > 0$. Однако это распределение не подходит для моделирования времен $\{S_n\}$, как хорошо видно из рис. 6. Критерий Колмогорова-Смирнова дает в этом случае значение уровня значимости менее $10^{-16}$. На рис. 6 в логарифмическом масштабе по оси абсцисс показаны плотности исходной выборки (сплошная линия) и выборки из распределения (11) (штриховая линия, параметры $k = 0,04$, $p = 3,04, \lambda = 2,25$).

Таким образом, использование распределения Парето статистически более обосновано. Полезно отметить, что распределение с тяжелым хвостом возможно сколь угодно точно аппроксимировать с помощью свертки экспоненциальных распределений, см. [Feldmann, Whitt, 1997]. Рассмотрим более подробно основные элементы рассматриваемой системы.

**Входной поток.** Рассмотрим времена между приходами заявок. Для устранения влияния этапа ввода кластера в эксплуатацию использовалась выборка, содержащая последние 2000 заявок. Кроме того, для исключения *отладочных* задач анализ был ограничен заявками, интервалы между приходами которых составляют не менее 50 секунд, т. е., $6 \cdot 10^{-4}$ суток (1632 заявок), см. рис. 7.

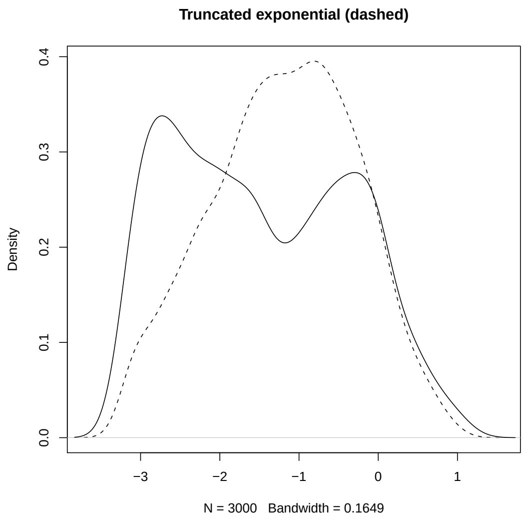



*Рис. 6.* Плотность распределения времени вычисления (усеченное экспоненциальное, $\lambda = 2,25$)

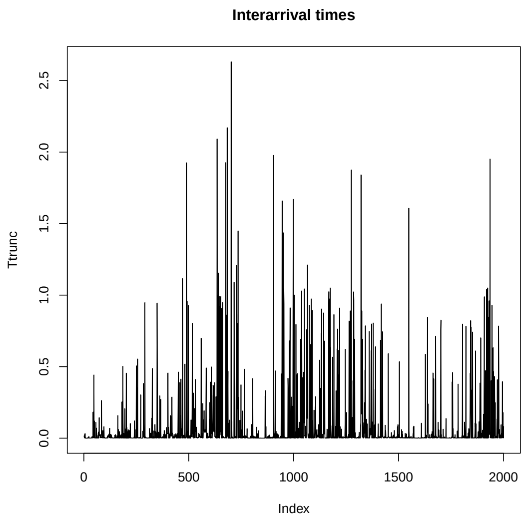



*Рис. 7.* Выборка интервалов между приходами (более 50 секунд)

Рассматривалась основная гипотеза $H_0$ о том, что интервалы $\{T_n\}$ между заявками имеют усеченное распределение Парето против альтернативной гипотезы (о том, что эти распределения различны). Рис. 8 иллюстрирует хорошее согласие между эмпирическими данными и усеченным распределением Парето с параметрами $k = 6 \cdot 10^{-4}, p = 2,63$, $\alpha = 0,22$. Графики плотностей построены по выборке из 2000 значений и представлены в логарифмическом масштабе по оси абсцисс. Критерий Колмогорова-Смирнова дает значение $D = 0,0384$ и значение уровня значимости не больше $0,1420$. Это значение достаточно для практических целей, так как позволяет принять гипотезу $H_0$, например, при уровне значимости $0,05$. (Отметим, что выбор параметров распределения можно производить с учетом методов, предложенных в работе [Aban et al., 2006].)

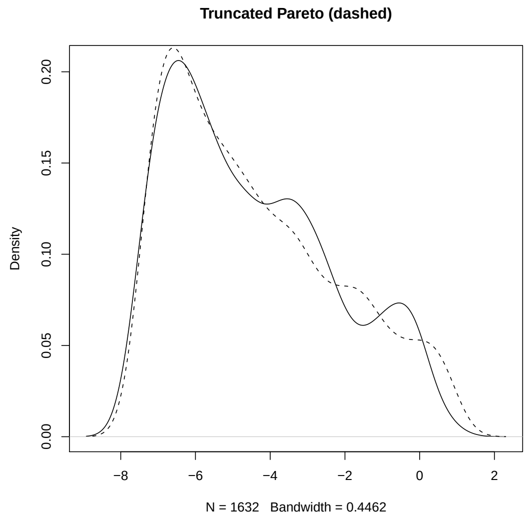



*Рис. 8.* Плотность распределения времени между приходами (усеченное Парето, $\alpha = 0,22$)

**Число процессоров.** На рис. 9 представлена гистограмма распределения (типичного) числа процессоров $N$, требуемых заявкой. Видно, что подавляющее число пользователей кластера ведет расчеты в однопроцессорном режиме, а программы, реализующие параллельные вычисления, используют число процессоров 8, 16, 32 и 64, кратное размеру узла.

С учетом наблюдений в модели принято равномерное распределение числа процессоров $N$ в диапазоне от 8 до 64, кратных 8. Однако в дальнейшем целесообразно использовать иные распределения. В частности, анализ гистограммы показывает, что распределение Зипфа (дискретный аналог распределения Парето), имеющее вид

$$P(N = i) = \frac{i^{-\alpha}}{\sum_{k=1}^{s} k^{-\alpha}},\ \alpha > 0, 1 \leqslant i \leqslant s,$$

может достаточно адекватно описать распределение числа требуемых процессоров.

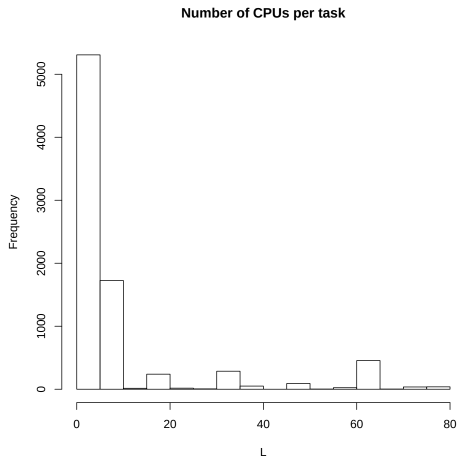



*Рис. 9.* Гистограмма числа процессоров в заявке

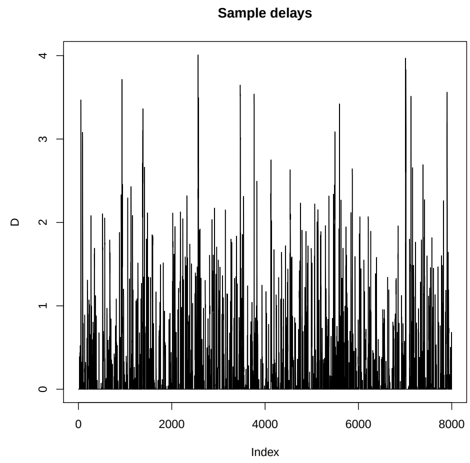



*Рис. 10.* Задержки в системе, число процессоров $s = 80$

**Результаты экспериментов.** Численное исследование модели велось на основе рекуррентного соотношения (8) для вектора загрузки системы. На рис. 10 представлен пример траектории задержек в очереди $\{D_n\}$ по выборке из 8000 значений (заявок). Средняя задержка равна $0,21$ суток. Максимальная задержка составила примерно 4 суток. Отметим, что хотя достаточное условие стационарности нарушено, $ES/ET > 1$, но неограниченное нарастание задержки не наблюдается. На гистограмме рис. 11 видно, что большинство задержек малы (меньше $0,5$ суток).

Проведенные эксперименты подтверждают, что модель адекватно отражает некоторые основные характеристики системы. Адекватность модели (несмотря на ее очевидную неполноту) позволяет, в частности, оценить эффект, вызываемый увеличением числа процессоров. Хорошо известно, что в классических системах обслуживания, в режимах, близких к полной загрузке, даже небольшое увеличение пропускной способности приводит к значительному сокращению времен ожидания. Предположим, что число узлов кластера увеличилось на два, т. е. имеется $s = 96$ процессоров. На рис. 12 видно, что число нулевых задержек в очереди увеличилось (6607 значений против 5755 в случае 80 процессоров). В этом случае средняя задержка составила приближенно $0,12$ суток, т. е., 57 % от значения для случая $s = 80$. Таким образом, незначительное увеличение числа процессоров позволяет существенно снизить среднюю задержку.

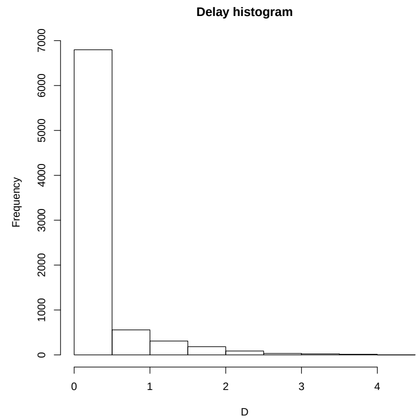



*Рис. 11.* Гистограмма задержек в системе, число процессоров $s = 80$

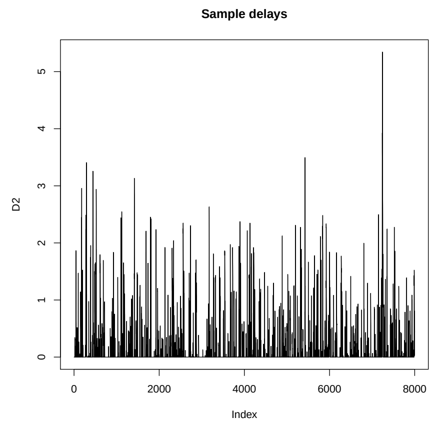



*Рис. 12.* Задержки в системе, число процессоров $s = 96$

## Заключение

В работе рассмотрены некоторые важные результаты для многосерверных систем обслуживания, в том числе, когда время обслуживания имеет тяжелый хвост. Подробно проанализированы моментные свойства процесса задержки в очереди. Предложена модель вычислительного кластера на основе модифицированной рекурсии Кифера-Вольфовица для многосерверной системы $G/G/s$, для которой получены моментные свойства стационарного времени ожидания в очереди. Полученные результаты могут применяться для оценивания качества обслуживания вычислительных кластеров и Грид. Проведенный на кластере вычислительный эксперимент показал хорошее согласие предложенной модели с реальной работой кластера.

Дальнейшая верификация модели, в том числе проверка условий ее стационарности, может быть проведена с использованием лог-файлов других вычислительных кластеров.

Кроме того, для лучшей адаптации модели к реальным кластерам целесообразно рассмотреть модели с конечным буфером для ожидания, с возможностью появления групп заявок, а также с более сложными дисциплинами выбора заявок на обслуживание (например, алгоритм Backfill и некоторые другие алгоритмы планировщиков). Поскольку аналитическое исследование таких моделей представляется трудно реализуемым, основным методом анализа может быть имитационное моделирование.

## Литература


*Центр высокопроизводительной обработки* данных ЦКП КарНЦ РАН. URL: http://cluster.krc.karelia.ru (дата обращения: 13.12.2010 г.).

*Aban I., Meerschaert M., Panorska A.* Parameter Estimation for the Truncated Pareto Distribution // Journal of the American Statistical Association. 2006. Vol. 101, N. 473. P. 270–277.

*Borst S., Boxma O., Núñez-Queija R.* Heavy Tails:The Effect of the Service Discipline // Proceedings of Performance TOOLS 2002. 2002. P. 1–30.

*Crovella M., Bestavos A.* Self-similarity in World Wide Web traffic: evidence and possible causes. // IEEE/ACM Transactions on Networking. 2002. Vol. 5, N. 6. P. 835–845.

*Feldmann A., Whitt W.* Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models // INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. 1997. P. 1096–1104.

*Foss S., Korshunov D.* Heavy Tails in Multi-Server Queue // Queueing Systems. 2006. Vol. 52. P. 31–48.

*Harchol-Balter M.* The Effect of Heavy-Tailed Job Size Distributions on Computer System Design // Proceedings of ASA-IMS Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics, Washington, DC. June 1999.

*Huang T., Sigman K.* Steady-state Asymptotics for Tandem, Split-Match and Other Feedforward Queues with Heavy-Tailed Service // Queueing Systems. 1999. Vol. 33. P. 233–259.

*Kiefer J., Wolfowitz J.* On the characteristics of the general queueing process, with applications to random walks // Ann. Math. Statist. 1956. Vol. 27. P. 147–161.

*Keilson J., Seidmann A.* $M/G/\infty$ with Batch Arrivals // Operations Research Letters October 1988. Vol. 7, N. 5. P. 219–222.

*Liu L.* Batch arrival infinite server queues with constant service times // Proceedings of the First Symposium on Queueing Theory in China September 1993. P. 47–53.

*Morozov E., Pagano M., Rumyantsev A.* Heavy-tailed Distributions with Applications to Broadband Communication Systems // Proceedings of AMICT'2007. 2008. Vol. 9. P. 157–174.

*Morozov E., Rumyantsev A.* Moment properties of queueing systems and networks // Proceedings of ICUMT'2010 (Ультрасовременные телекоммуникации и системы управления). 2010.

*Scheller-Wolf A.* Further delay moment results for FIFO multiserver queues // Queueing Systems. 2000. Vol. 34. P. 387–400.

*Scheller-Wolf A., Sigman K.* Delay Moments for FIFO GI/GI/s queues // Queueing Systems. 1997a. Vol. 25. P. 77–95.

*Scheller-Wolf A., Sigman K.* New bounds for expected delay in FIFO GI/GI/c queues // Queueing Systems. 1997b. Vol. 26. P. 169–186.

*Scheller-Wolf A., Vesilo R.* Sink or Swim Together: Necessary and Sufficient Conditions for Finite Moments of Workload Components in FIFO Multiserver Queues // Queueing Systems. January 2011. Vol. 67, N. 1. P. 47–61. March 2008.

*Scheller-Wolf A., Vesilo R.* Structural interpretation and derivation of necessary and sufficient conditions for delay moments in FIFO multiserver queues // Queueing Systems. 2006. Vol. 54. P. 221–232.

*Wolff R.* On Finite Delay-Moment Conditions in Queues // Operations Research. September-October 1991. Vol. 39, N. 5. P. 771–775.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Морозов Евсей Викторович**
ведущий научный сотрудник, д. ф.-м. н., профессор
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: ar0@krc.karelia.ru
тел.: (8142) 763370

**Morozov, Evsey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: ar0@krc.karelia.ru
tel.: (8142) 763370

**Румянцев Александр Сергеевич**
аспирант
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: ar0@krc.karelia.ru
тел.: (8142) 763370

**Rumyantsev, Alexandr**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia,
e-mail: ar0@krc.karelia.ru
tel.: (8142) 763370

Труды Карельского научного центра РАН
№ 5. 2011. С. 86–96

*УДК 519.2*

# О ТИПИЧНОЙ СТРУКТУРЕ КОНФИГУРАЦИОННОГО ИНТЕРНЕТ-ГРАФА С ИЗВЕСТНЫМ ЧИСЛОМ СВЯЗЕЙ

**Ю. Л. Павлов**

*Институт прикладных математических исследований Карельского научного центра РАН*

Рассматривается конфигурационная модель случайного графа с $N$ вершинами. Степени вершин выбираются независимо из степенного закона распределения с показателем $\tau \in (1,2)$ при условии, что сумма степеней вершин равна $n$. Показано, что если $N, n \to \infty$ так, что $n - \zeta(\tau)N = O(N^{1/\tau})$, где $\zeta(x)$ – дзета-функция Римана, то предельная структура графа не отличается от типичной структуры случайного графа без условия на сумму степеней.

Ключевые слова: случайные графы, Интернет, конфигурационная модель, случайная структура.

## Yu. L. Pavlov. ON THE TYPICAL STRUCTURE OF CONFIGURATION INTERNET GRAPH WITH KNOWN NUMBER OF LINKS

We consider the configuration model of random graph consisting of $N$ vertices. The degrees of vertices are drawn independently from power-law distribution with the exponent $\tau \in (1,2)$ under the condition that the sum of vertex degree is equal to $n$. We show that if $N, n \to \infty$ in such a way that $n - \zeta(\tau)N = O(N^{1/\tau})$, where $\zeta(x)$ is the Rimann's zeta-function, then the limit structure of the graph is the same typical structure as the random graph without the condition on the sum of degrees.

Key words: random graph, Internet, configuration model, random structure.

## Введение

За последние 10–15 лет специалисты по случайным графам уделяют значительное внимание разработке моделей сложных телекоммуникационных и социальных сетей, в частности, сети Интернет (см. например, [Durrett, 2007]). В настоящей работе мы рассмотрим одну из наиболее известных таких моделей, получившую название конфигурационной модели со случайными степенями (см. [Reittu, Norros, 2004; Durrett, 2007; Hofstad, 2011]). Пусть граф содержит $N$ вершин, занумерованных числами от 1 до $N$. Степени вершин являются независимыми одинаково распределенными случайными величинами $\eta_1, \ldots, \eta_N$ такими, что

$$\mathsf{P}\{\eta_1 \geqslant k\} = k^{-\tau}, \qquad (1)$$

где $\tau$ – положительный параметр. Если сумма $\nu_N = \eta_1 + \cdots + \eta_N$ нечетна, то в граф вводится дополнительная вершина с единичной степенью. Формирование графа происходит в два этапа. На первом этапе определяются степени вершин в соответствии с распределением (1), при этом считается, что из вершин выходят «полуребра» [Reittu, Norros, 2004], т. е. ребра, для которых смежные вершины еще не определены (все полуребра графа различны). На втором этапе полуребра соединяются равновероятно для образования ребер.

В ряде работ (см. [Faloutsos et al., 1999; Reittu, Norros, 2004; Durrett, 2007]) показано, что типичные для многих реальных сетей значения параметра $\tau$ распределения (1) принадлежат интервалу (1,2). Поскольку такие значения характерны и для сети Интернет, соответствующие графы иногда называются графами Интернет-типа или Интернет-графами.

Предельное поведение структуры и различных числовых характеристик Интернет-графов изучалось во многих публикациях (см. [Reittu, Norros, 2004; Durrett, 2007] и ссылки в них). Обнаружено, что граф имеет единственную гигантскую компоненту связности, число вершины которой пропорционально $N$, а объемы других компонент бесконечно малы по сравнению с $N$. Найдена асимптотика математического ожидания объема гигантской компоненты и получена оценка сверху ее диаметра. Предельное распределение объема гигантской компоненты получено в [Павлов, 2007].

В графах Интернет-типа сумма степеней вершин $\nu_N$ является случайной величиной. Однако представляют интерес и условные графы такого вида при условии, что величина $\nu_N$ известна и равна $n$. Такие графы полезны для моделирования сетей, в которых число связей можно оценить. Кроме того, результаты об условных графах можно использовать для нахождения предельных распределений числовых характеристик Интернет-графов без ограничений на $\nu_N$ путем усреднения условных распределений этих характеристик по распределению общего числа ребер графа. Изучение свойств условных случайных графов Интернет-типа начато в статьях [Павлов, Чеплюкова, 2008; Павлов, 2009]. Основным методом исследования в этих работах была обобщенная схема размещения частиц по ячейкам, введенная и подробно изученная В. Ф. Колчиным (см., например, [Колчин,, 2000]).

Согласно (1),

$$p_k = \mathsf{P}\{\eta_1 = k\} = \frac{1}{k^\tau} - \frac{1}{(k+1)^\tau},\ k = 1, 2, ... \tag{2}$$

Отсюда следует, как легко видеть, что

$$\mathsf{E}\,\eta_1 = \zeta(\tau), \tag{3}$$

где $\zeta(\tau)$ — значение дзета-функции Римана в точке $\tau$. Поэтому при изучении динамики структуры Интернет-графов типичным является случай $N, n \to \infty$ так, что $N/n \to \zeta(\tau)$. Естественно ожидать, что в этом случае конфигурация условного графа асимптотически эквивалентна структуре графа без ограничений на сумму $\nu_N$ и представляет интерес вопрос о том, при каких значениях $n$ такая ситуация сохраняется. В данной статье показано, что это имеет место при условии $n - \zeta(\tau)N = O(N^{1/\tau})$. По-видимому, такое утверждение сохраняет силу и в более широкой зоне изменения параметров, однако этот вопрос требует дальнейшего исследования. Идеи, на которых основаны приведенные ниже результаты, близки к изложенным в работе [Reittu, Norros, 2004], однако необходимость учета условия $\nu_N = n$ потребовала изменения техники доказательств. Это стало возможным после получения предельных распределений максимальной степени и числа вершин заданной степени в условных Интернет-графах [Павлов, Чеплюкова, 2008; Павлов, 2009].

## Основной результат

Введем следующее распределение

$$q_k = (k+1)p_{k+1}\zeta^{-1}(\tau), \quad k = 0, 1, 2, ... \tag{4}$$

Пусть $\varphi(z)$ и $F(z)$ означают производящие функции распределений (2) и (4) соответственно, следовательно,

$$\varphi(z) = \sum_{k=1}^{\infty} p_k z^k, \quad F(z) = \sum_{k=0} q_k z^k \tag{5}$$

Обозначим также:

$$\alpha = \left[\frac{\ln\ln N}{-\ln(\tau-1)}\right],\ \beta = \left[\exp\left\{\frac{2w}{2-\tau}\right\}\right], \tag{6}$$

где $[x]$ означает целую часть числа $x$, а $w = w(N)$ – медленно растущая функция [Ибрагимов, Линник, 1965], такая, что $w > 0$ и

$$w = o(\ln\ln\ln N), \quad \ln\ln\ln N = o(w). \tag{7}$$

Обозначим $A_N$ некоторое событие, зависящее от $N$. Будем говорить, что событие $A_N$ *асимптотически достоверно* (а. д.), если $\mathsf{P}\{A_N\} \to 1$ при $N \to \infty$. Автор выражает благодарность А. М. Зубкову за обсуждение этого определения.

Справедливо следующее утверждение.

**Теорема.** *Пусть $N, n \to \infty$ так, что $n - \zeta(\tau)N = O(N^{1/\tau})$. Тогда граф содержит компоненту связности объема $V$, для которой*

$\mathsf{E}\, V/N \to 1 - \varphi(q)$, *где* $q$ *– вероятность вырождения ветвящегося процесса Гальтона-Ватсона с распределением (4) числа прямых потомков каждой частицы. Диаметр этой компоненты а.д. не превосходит* $2(\alpha+\beta)$.

Ниже приводятся леммы 1–6, с помощью которых в конце статьи будет доказана сформулированная теорема.

Далее будем использовать символ $j$ (с индексом или без) для обозначения вершин графа. В тех случаях, когда это не может вызвать недоразумений, также будем обозначать номера вершин, отождествляя их с соответствующими вершинами.

## Связь между подмножествами вершин

Пусть $j$ – произвольная вершина графа, имеющая степень $d = d(N,n)$. Рассмотрим некоторое множество вершин $W$, сумма степеней которых равна $D = D(N,n)$. Пусть $j \notin W$, $d \leqslant D$. Обозначим $A$ событие, состоящее в том, что вершина $j$ соединена ребром с какой-нибудь вершиной множества $W$.

**Лемма 1.** *Пусть* $N, n \to \infty$ $n - \zeta(\tau)N = O(N^{1/\tau})$. *Тогда справедливы следующие утверждения.*

1. *Если* $dD/n \to \infty$, *то* $\mathsf{P}\{A\} \to 1$.
2. *Если* $dD/n \to 0$, *то* $\mathsf{P}\{A\} \to 0$.

*Доказательство.* Мы используем идею доказательства леммы 3.8 статьи [Reittu, Norros, 2004]. Пусть выполнено условие 1. Нетрудно видеть, что из множества $d$ ребер, выходящих из вершины $j$, можно выбрать подмножество из $d'$ ребер, удовлетворяющую в условиях леммы следующим соотношениям:

$$d' \leqslant d, \quad d'D/n \to \infty, \quad d'^2/n \to 0. \qquad (8)$$

Обозначим $A_1$ событие, состоящее в том, что все $d'$ ребер не идут в $W$ и не образуют петель, а через $A_2$ событие, состоящее в наличии петель, образованных какими-то из $d'$ ребер. Нетрудно видеть, что

$$\mathsf{P}\{\bar{A}\} \leqslant \mathsf{P}\{A_1\} + P\{A_2\} \leqslant$$

$$\frac{(n-d'-D)(n-d'-D-1)...(n-2d'-D+1)}{(n-1)(n-3)...(n-2d'+1)}$$

$$+1 - \frac{(n-d')(n-d'-1)...(n-2d'+1)}{(n-1)(n-3)\dots(n-2d'+1)} <$$

$$\left(\frac{n-d'-D}{n-2d'+1}\right)^{d'} + 1 - \left(\frac{n-2d'+1}{n-1}\right)^{d'} = O\left(\frac{d'^2}{n}\right).$$

Отсюда и из (8) следует первое утверждение. Второе доказывается аналогично:

$$\mathsf{P}\{\bar{A}\} =$$

$$\frac{(n-d-D)(n-d-D-1)...(n-2d-D+1)}{(n-1)(n-3)...(n-2d+1)}$$

$$> \left(\frac{n-2d-D+1}{n-2d+1}\right)^{d} \to 1.$$

□

## Алгоритм построения графа

Рассмотрим один из возможных алгоритмов построения графа на втором этапе его формирования. Заметим, что для безусловных графов аналогичный алгоритм описан в [Reittu, Norros, 2004] (см. также [Павлов, 2007]). Идея алгоритма следующая. Граф строится с начальной вершины, например, с вершины 1. Далее для всех полуребер начальной вершины по очереди и равновероятно ищутся пары и, таким образом, образуются ребра, инцидентные первой вершине. Вершины, соединенные ребром с начальной (и не совпадающие с ней), называются вершинами первого поколения. Далее для всех свободных полуребер вершин первого поколения опять последовательно и равновероятно ищутся пары и образуются новые ребра. В результате вершины, соединенные путями длины 2 с начальной вершиной, образуют второе поколение. Этот процесс продолжается до тех пор, пока у вершин очередного поколения не окажется свободных полуребер. Обозначим $Z_t$ число вершин $t$-го поколения, $t = 0, 1, 2, \ldots$. Очевидно, что $Z_0 = 1$. Если $Z_t = 0$, а построенный граф еще не содержит все $N$ вершин, то завершено построение только одной компоненты связности и для продолжения формирования графа среди всех не рассмотренных вершин выбирается вершина с наименьшим номером в качестве начальной для построения следующей компоненты связности, и далее процедура осуществляется аналогично. Повторив, если нужно, эти действия несколько раз до исчерпания всех $N$ вершин, мы построим одну из возможных реализаций графа. Нетрудно видеть, что любая такая реализация возникает с вероятностью, равной соответствующей величине, индуцируемой вероятностным пространством, заданным на множестве рассматриваемых графов. Ниже приведем алгоритм построения одной компоненты связности. По ходу работы этого алгоритма составляется последовательность номеров вершин, которые одна за другой присоединяются к графу по мере образования ребер: $j_1, j_2, ..., j_s$, где $j_1$ – номер начальной частицы, $j_2$ – номер первой вершины, соединенной ребром с $j_1$, а $j_s$ – номер последней вершины последнего поколения данной компоненты связности. Процесс образования такой последовательности номеров подобен естественной нумерации частиц в ветвящихся процессах Гальтона – Ватсона (так называемая «история семейств», см. [Ватутин, 1993]). Отличие состоит в том, что номера частиц процесса не повторяются, а в последовательности $j_1, j_2, ..., j_s$ номера повторяться могут в силу возможности образования петель, циклов и кратных ребер.

В тексте алгоритма мы будем рассматривать вершины как $N$ различных ящиков, содержащих, соответственно, $k_1, ..., k_N$ различных шаров ($k_1 + ... + k_N = n$). Понятно, что эти шары обозначают полуребра вершин. Удаление двух шаров подряд подразумевает образование очередного ребра графа путем соединения соответствующих полуребер. В этом алгоритме $U$ означает последовательность номеров присоединяемых вершин.

Алгоритм

1. $U := \{j_1\}, \quad i = 1, \ k = 2.$
2. Удаляется первый шар из ящика $j_i$.
3. Выбирается равновероятно шар из оставшихся и удаляется, номер его ящика обозначается $j_k$.
4. $U := U \cup \{j_k\}$.
5. $k = k + 1$.
6. Если в $j_i$ есть еще шары, то переход на шаг 2.
7. $i = i + 1$.
8. Если в ящиках из $U$ есть еще шары, то переход на шаг 6.
9. Стоп.

## Древовидность компоненты связности

Далее мы покажем, что если $N, n \to \infty$, то при больших $t$ (и даже при $t \to \infty$, если это стремление происходит достаточно медленно) построенный по приведенному алгоритму подграф до поколения $t$ включительно является деревом, степени вершин которого сравнительно невелики. Следующие леммы 2 и 3 являются развитием соответствующих результатов работ [Reittu, Norros, 2004] и [Павлов, 2007]. Обозначим $A_t = A_t(h)$, $t = 0, 1, ...,$ событие, состоящее в том, что такой подграф является деревом и в нем нет вершин, степень которых больше или равна $h$, где $h = h(N)$ – медленно растущая функция такая, что $h = O(\ln N)$.

**Лемма 2.** *Пусть* $N, n \to \infty$ *так, что* $n - \zeta(\tau)N = O(N^{1/\tau})$, $t = o(\ln\ln h)$. *Тогда* $\mathsf{P}\{A_t\} = 1 + O(h^{-(\tau-1)^t})$.

*Доказательство.* Нетрудно видеть, что

$$\mathsf{P}\{A_0\} = \mathsf{P}\{\eta_1 < h | \nu_N = n\}$$

$$= 1 - (\mathsf{P}\{\nu_N = n\})^{-1} \sum_{k \geqslant h} p_k \, \mathsf{P}\{\nu_{N-1} = n - k\}. \tag{9}$$

Согласно лемме 3 из [Павлов, 2007] при $N \to \infty$

$$\sup_k |N^{1/\tau} \, \mathsf{P}\{\nu_N = k\} - f((k - N\zeta(\tau))/N^{1/\tau})| \to 0, \tag{10}$$

где $f(x)$ – плотность устойчивого распределения с показателем $\tau$ и характеристической функцией

$$\exp\left\{\frac{\Gamma(2-\tau)}{\tau - 1}|t|^\tau \left(1 + i\frac{t}{|t|} tg\frac{\pi\tau}{2}\right) \cos\frac{\pi\tau}{2}\right\}.$$

Учитывая одновершинность устойчивых распределений [Ибрагимов, Линник, 1965] и (1), из (9) находим, что

$$\mathsf{P}\{A_0\} = 1 + O(h^{-\tau}),$$

следовательно, для $t = 0$ утверждение леммы 2 доказано.

Обозначим $\Psi(t)$ характеристическую функцию случайной величины $\eta_1$. Из (2) нетрудно получить, что

$$\Psi(t) = 1 + (e^{it} - 1)\Phi(e^{it}, \tau, 1), \tag{11}$$

где $\Phi(z, s, a)$ – транцендентная функция Лерча (см., например, [Flajolet, Sedgewick, 2009]), задаваемая соотношениями

$$\Phi(z, s, a) = \sum_{i=0}^{\infty} \frac{z^i}{(i + a)^s}$$

$$= \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1} e^{-ax}}{1 - ze^{-x}} dx, \tag{12}$$

$\Gamma(s)$ – значение гамма-функции в точке $s$. Используя (11) и формулу обращения для решетчатых распределений, можно уточнить (10) следующим образом:

$$\mathsf{P}\{\nu_N = k\}$$

$$= N^{-1/\tau}(f((k - N\zeta(\tau))/N^{1/\tau}) + O(N^{-1/\tau})), \tag{13}$$

причем это равенство выполняется равномерно относительно $k$ таких, что величина $(k - N\zeta(\tau))N^{-1/\tau}$ лежит в любом фиксированном конечном интервале.

Рассмотрим $\mathsf{P}\{A_1\}$. Ясно, что

$$\mathsf{P}\{A_1\} = \sum_{k<h} \mathsf{P}\{A_1, \eta_1 = k\}. \qquad (14)$$

Обозначим $\xi_1^{(1)}$ степень первой вершины первого поколения. Тогда

$$\mathsf{P}\{A_1, \eta_1 = 1\} = \sum_{s<h} \mathsf{P}\{A_1, \eta_1 = 1, \xi^{(1)} = s\}. \qquad (15)$$

Пусть $\mu_r$ означает число вершин степени $r$ в графе. В теореме 6 статьи [Павлов, 2009] показано, что при $N, n \to \infty$ так, что $n - \zeta(\tau)N = o(N^{1/\tau})$ для целых неотрицательных $k$

$$\mathsf{P}\{\mu_r = k\}$$

$$= \frac{1 + o(1)}{\sqrt{2\pi N p_r(1 - p_r)}} \exp\left\{-\frac{(k - Np_r)^2}{2Np_r(1 - p_r)}\right\} \qquad (16)$$

равномерно относительно $\frac{(k-Np_r)}{\sqrt{2\pi Np_r(1-p_r)}}$ в любом фиксированном конечном интервале. Повторив доказательство этой теоремы при условии
$0 < C_1 \leqslant (n - \zeta(\tau)N)/N^{1/\tau} \leqslant C_2 < \infty$, здесь и далее символы $C_1, C_2, ...$ означают некоторые положительные постоянные, легко получить, что для таких же $k$

$$\mathsf{P}\{\mu_r = k\}$$

$$= \frac{(1 + o(1))f((n - \zeta(\tau)N)/(N(1 - p_r))^{1/\tau})}{(1 - p_r)^{1/\tau}\sqrt{2\pi N(1 - p_r)}f((n - \zeta(\tau)N^{1/\tau})}$$

$$\times \exp\left\{-\frac{(k - Np_r)^2}{2Np_r(1 - p_r)}\right\}. \qquad (17)$$

Учитывая соотношения $N^{-1/2} = o(N^{1/\tau-1})$, $n - \zeta(\tau)N = O(N^{1/\tau})$, процедуру построения графа и свойства функции $h$, из (4), (16) и (17) получаем, что

$$\mathsf{P}\{A_1, \eta_1 = 1, \xi_1^{(1)} = 1\}$$

$$= p_1 N p_1 \frac{(n-3)(n-5)...3 \cdot 1}{(n-1)(n-3)...3 \cdot 1}\left(1 + O(N^{-1/2})\right)$$

$$= p_1^2\zeta^{-1}(\tau)(1+O(N^{1/\tau-1})) = p_1 q_0(1+O(N^{1/\tau-1})).$$

Рассуждая аналогично для $2 \leqslant s \leqslant h$, находим, что

$$\mathsf{P}\{A_1, \eta_1 = 1, \xi_1^{(1)} = s\} = p_1 q_{s-1}(1+O(N^{1/\tau-1})). \qquad (18)$$

Отсюда и из (15) следует:

$$\mathsf{P}\{A_1, \eta_1 = 1\} = p_1(1 - \sum_{s\geqslant h} q_{s-1})(1+O(N^{1/\tau-1})). \qquad (19)$$

Из (2) при $k \to \infty$ получаем :

$$p_k \sim \tau/k^{\tau+1}, \qquad (20)$$

поэтому с учетом (4),

$$\sum_{k\geqslant h} q_{s-1} \leqslant C_3 \sum_{s\geqslant h} s^{-\tau} \leqslant C_4 h^{1-\tau}. \qquad (21)$$

Отсюда и из (19) вытекает :

$$\mathsf{P}\{A_1, \eta_1 = 1\} = p_1(1 + O(h^{1-\tau})). \qquad (22)$$

Подобным же образом можно рассмотреть поведение вероятности $\mathsf{P}\{A_1, \eta_1 = 2\}$. Тогда

$$\mathsf{P}\{A_1, \eta_1 = 2\} = \sum_{s<h} \mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = s\}. \qquad (23)$$

Заметим, что

$$\mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = s\}$$

$$= \sum_{i<h} \mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = 1, \xi_2^{(1)} = i\}, \qquad (24)$$

где $\xi_2^{(1)}$ – степень второй вершины первого поколения (такая вершина, очевидно, существует, поскольку событие $A_1$ подразумевает, что у начальной частицы нет петель). Легко видеть, опять учитывая (4), (16), (17) и условие $\nu_N = n$, что

$$\mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = 1, \xi_2^{(1)} = 1\}$$

$$= p_2 \left(\begin{array}{l} Np_1 \\ 2 \end{array}\right) 2\frac{(n-5)(n-7)...3 \cdot 1}{(n-1)(n-3)...3 \cdot 1}$$

$$\times \left(1 + O(\frac{1}{\sqrt{N}})\right) = p_2 q_0^2(1 + o(N^{1/\tau-1}))^2$$

и, аналогично, для $2 \leqslant i < h$

$$\mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = 1, \xi_2^{(1)} = i\}$$

$$= p_2 q_0 q_{i-1}\left(1 + o\big(N^{1/\tau-1}\big)\right)^2.$$

Отсюда и из (21), (24) находим, что

$$\mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = 1\}$$

$$= p_2 q_0\left(1 + O\big(h^{1-\tau}\big)\right)\left(1 + o\big(N^{1/\tau-1}\big)\right)^2.$$

Продолжая этот процесс, получаем, что при $2 \leqslant s < h$

$$\mathsf{P}\{A_1, \eta_1 = 2, \xi_1^{(1)} = s\}$$

$$= p_2 q_{s-1}\left(1+O(h^{1-\tau})\right)\left(1+o(N^{1/\tau-1})\right)^2.$$

Поэтому из (21) и (23) следует, как легко видеть, что

$$\mathsf{P}\{A_1, \eta_1 = 2\} = p_2\left(1+O(h^{1-\tau})\right)^2.$$

Теперь становится понятным, что для $k = 1, 2, ...$

$$\mathsf{P}\{A_1, \eta_1 = k\} = p_k\left(1+O(h^{1-\tau})\right)^k.$$

Отсюда и из (1), (3), (14) и (32) следует:

$$\mathsf{P}\{A_1\} = \sum_{k<h} p_k\left(1+O(kh^{1-\tau})\right) = 1+O(h^{1-\tau}) \tag{25}$$

Далее рассмотрим $\mathsf{P}\{A_1\}$. Проводя выкладки, подобные приведенным выше, находим, что

$$\mathsf{P}\{A_2, \eta_1 = 1\} = p_1 \sum_{i<h} q_{i-1}\big(1-\sum_{s\geqslant h} q_s\big)^{i-1}$$

$$= p_1\zeta^{-1}(\tau)\sum_{i<h} ip_i\big(1-\sum_{s\geqslant h} q_s\big)^{i-1}. \tag{26}$$

Рассмотрим выражение

$$\zeta^{-1}(\tau)\sum_{i<h} ip_i\big(1-\sum_{s\geqslant h} q_s\big)^{i-1}$$

$$= \zeta^{-1}(\tau)\Big(\sum_{i=1}^{\infty} ip_i\big(1-\sum_{s\geqslant h} q_s\big)^{i-1} - \sum_{i\geqslant h} ip_i\big(1-\sum_{s\geqslant h} q_s\big)^{i-1}\Big). \tag{27}$$

Обозначим $\lambda = 1 - \sum\limits_{s\geqslant h} q_s$ и рассмотрим вспомогательную случайную величину $\xi$, имеющую распределение

$$\mathsf{P}\{\xi = k\} = \lambda^k p_k(1-(1-\lambda)\Phi(\lambda,\tau,1))^{-1}, \tag{28}$$

где $k = 1, 2, ...,$ а $\Phi(\lambda,\tau,1)$ определено в (12). Нетрудно найти, что

$$\mathsf{E}\,\xi = \frac{\Phi(\lambda,\tau,1)-(1-\lambda)\Phi(\lambda,\tau,1)}{1-(1-\lambda)\Phi(\lambda,\tau,1)}$$

и в наших условиях $\lambda \to 1, \mathsf{E}\,\xi \to \zeta(\tau)$. Для того, чтобы получить более точную оценку $\mathsf{E}\,\xi$, воспользуемся асимптотическим разложением $\Phi(\lambda,\tau,1)$ при $\lambda \to 1$ (см. [Flajolet, Sedgewick, 2009]):

$$\Phi(\lambda,\tau,1) = \lambda^{-1}(\zeta(\tau)+O((1-\lambda)^{\tau-1})).$$

Отсюда и из явного вида $\lambda$ находим:

$$\mathsf{E}\,\xi = \zeta(\tau) + O(h^{-(\tau-1)^2}). \tag{29}$$

Заметим, что в силу (20)

$$\sum_{i\geqslant h} ip_i\big(1-\sum_{s\geqslant h} q_s\big)^i < C_5 \sum_{i\geqslant h} i^{-\tau} = O\big(h^{1-\tau}\big),$$

поэтому из (26) – (29) следует:

$$\mathsf{P}\{A_2, \eta_1 = 1\} = p_1\big(1+O(h^{(\tau-1)^2})\big).$$

Таким же образом находим, что для $i = 1, 2, ...$

$$\mathsf{P}\{A_2, \eta_1 = i\} = p_i\big(1+O(h^{-(\tau-1)^2})\big),$$

следовательно

$$\mathsf{P}\{A_2\} = 1+O(h^{-(\tau-1)^2}).$$

Учитывая (25) и продолжая этот процесс по индукции, получаем утверждение леммы. □

## Связь с ветвящимися процессами

Рассмотрим начинающийся с одной частицы ветвящийся процесс Гальтона – Ватсона $G$, в котором число прямых потомков начальной частицы имеет распределение (2), а число прямых потомков остальных частиц имеет распределение (4). Обозначим $Z_t(G)$ число частиц $t$-го поколения. Связь между рассмотренными случайными графами и процессом $G$ можно описать следующим образом.

**Лемма 3.** *Пусть $N, n \to \infty$ так, что $n - \zeta(\tau)N = O(N^{1/\tau})$, $t = o(\ln N/\ln\ln N)$. Тогда для $m = 0, 1, 2, ...$*

$$\mathsf{P}\{Z_t = m\} = \mathsf{P}\{Z_t(G) = m\}(1+O(h^{-(\tau-1)^{t-1}})).$$

*Доказательство.* Ясно, что

$$\mathsf{P}\{Z_0 = 1\} = \mathsf{P}\{Z_0(G) = 1\} = 0.$$

Из леммы 2 и (13) следует :

$$\mathsf{P}\{Z_1 = m\} = \mathsf{P}\{\eta_1 = m | \nu_N = n\}$$

$$= p_m(1+O(N^{-1/\tau})), \tag{30}$$

поэтому, в силу равенства $\mathsf{P}\{Z_1(G) = m\} = p_m$, получаем утверждение леммы для $t = 1$. Пусть $t = 2$. Согласно лемме 2,

$$\mathsf{P}\{Z_2 = m\} = \mathsf{P}\{Z_2 = m, A_1\} + O(h^{1-\tau}). \tag{31}$$

Рассмотрим

$$\mathsf{P}\{Z_2 = m, A_1\} = \sum_{k<h} P\{\eta_1 = k, Z_2 = m, A_1\}. \tag{32}$$

Из (18) и (21) находим, что

$$P\{\eta_1 = 1, Z_2 = m, A_1\}$$

$$= p_1 q_{m-1}(1 + o(N^{1/\tau - 1}))(1 + O(h^{1-\tau})).$$

Учитывая (16), (17) и условие $n - \zeta(\tau)N = o(N^{1/\tau})$, получаем :

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 0, A_1\} = p_2 2 \binom{Np_1}{2}$$

$$\times(1 + O(N^{-1/2}))\frac{(n-5)(n-7)...3\cdot 1}{(n-1)(n-3)...3\cdot 1}$$

$$= p_2 q_0^2\big(1 + o\big(N^{1/\tau-1}\big)\big). \qquad (33)$$

Нетрудно видеть далее, что

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 1, A_1\} =$$

$$2\sum_{1<k\leqslant h} \mathsf{P}\{\eta_1 = 2, Z_2 = 1, \xi_1^{(1)} = k, \xi_2^{(1)} = 1, A_1\}, \qquad (34)$$

где обозначения $\xi_1^{(1)}$ и $\xi_2^{(1)}$ имеют тот же смысл, что и в доказательстве леммы 2. Используя (16), (17), получаем:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 1, \xi_1^{(1)} = 2, \xi_2^{(1)} = 1, A_1\}$$

$$= p_2 \binom{Np_1}{2} Np_2 4\frac{(n-7)(n-9)...3\cdot 1}{(n-1)(n-3)...3\cdot 1}$$

$$\times\left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right)^2 = p_2 q_1 q_0^2(1 + o(N^{1/\tau-1}))^3.$$

Подобным же образом для $2 \leqslant k \leqslant h$ приходим к равенствам:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 1, \xi_1^{(1)} = k, \xi_2^{(1)} = 1, A_1\}$$

$$= p_2 q_1 q_0 q_{k-1}(1 + o(N^{1/\tau-1}))^3,$$

поэтому из леммы 2 и (34) находим:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 1, A_1\}$$

$$= 2p_2 q_0 q_1(1 + o(N^{1/\tau-1}))^3(1 + O(h^{1-\tau})). \qquad (35)$$

Дальнейшие рассуждения проводятся по той же схеме. Ясно, что

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, A_1\}$$

$$= 2\,\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, A_1\}$$

$$+\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 2, \xi_2^{(1)} = 2, A_1\}. \qquad (36)$$

Обозначим $\xi_1^{(2)}$ и $\xi_2^{(2)}$ первую и вторую вершины второго поколения. Нетрудно видеть, что

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, A_1\}$$

$$= \sum_{k=1}^{\infty} \mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3,$$

$$\xi_2^{(1)} = 1, \xi_1^{(2)} = k, A_1\}. \qquad (37)$$

Очевидно также, что

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 1, A_1\}$$

$$= \sum_{s=1}^{\infty} \mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)}$$

$$= 1, \quad \xi_2^{(2)} = s, A_1\}. \qquad (38)$$

Далее получаем:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 1,$$

$$\xi_2^{(2)} = s, A_1\} = p_2 q_0^2 q_2 q_{s-1}(1 + o(N^{1/\tau-1}))^4,$$

поэтому из (38) следует:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 1, A_1\}$$

$$= p_2 q_0^2 q_2(1 + o(N^{1/\tau-1}))^4. \qquad (39)$$

Теперь рассмотрим

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 2, A_1\}$$

$$= \sum_{s=1}^{\infty} \mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)}$$

$$= 1, \quad \xi_2^{(2)} = s\}. \qquad (40)$$

Как обычно, находим:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 2,$$

$$\xi_2^{(2)} = s, A_1\} = p_2 q_0 q_1 q_2 q_{s-1}(1 + o(N^{1/\tau-1}))^4$$

и из (40) получаем:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = 2,$$

$$A_1\} = p_2 q_0 q_1 q_2(1 + o(N^{1/\tau-1}))^4. \qquad (41)$$

Продолжая этот процесс и учитывая (39), (41), видим, что

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, \xi_1^{(2)} = k, A_1\}$$

$$= p_2 q_0 q_2 q_{k-1}(1 + o(N^{1/\tau-1}))^4,$$

поэтому из (37) следует равенство:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 3, \xi_2^{(1)} = 1, A_1\}$$

$$= p_2 q_0 q_2(1 + o(N^{1/\tau-1}))^4. \qquad (42)$$

Рассматривая следующую вероятность по той же схеме, находим:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, \xi_1^{(1)} = 2, \xi_2^{(1)} = 2, A_1\}$$
$$= p_2 q_1^2 (1 + o(N^{1/\tau - 1}))^4.$$

Отсюда и из (36), (42) вытекает:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = 2, A_1\}$$
$$= (2p_2 q_0 q_2 + p_2 q_1^2)(1 + o(N^{1/\tau - 1}))^4. \qquad (43)$$

Подобно тому, как были получены равенства (33), (35) и (42), из леммы (2) находим:

$$\mathsf{P}\{\eta_1 = 2, Z_2 = m, A_1\}$$
$$= p_2 \sum_{\substack{k_1, k_2 \geqslant 0, \\ k_1 + k_2 = m}} q_{k_1} q_{k_2} (1 + o(h^{\tau - 1})).$$

Аналогичным образом можно вывести, что

$$\mathsf{P}\{\eta_1 = k, Z_2 = m, A_1\}$$
$$= p_k \sum_{\substack{k_1, ..., k_k \geqslant 0, \\ k_1 + ... + k_k = m}} q_{k_1} q_{k_2} ... q_{k_k} (1 + O(h^{1 - \tau})).$$

Отсюда и из (1), (31), (32) следует:

$$\mathsf{P}\{Z_2 = m\}$$
$$= \sum_{k=1}^{\infty} p_k \sum_{\substack{k_1, ..., k_k \geqslant 0, \\ k_1 + ... + k_k = m}} q_{k_1} q_{k_2} ... q_{k_k} (1 + O(h^{1 - \tau})). \qquad (44)$$

Из определения ветвящегося процесса $G$ и (30), (44) ясно, что утверждение леммы верно при $t \leqslant 2$. Как и при доказательстве леммы 2, продолжим рассмотрение вероятностей $\mathsf{P}\{Z_t = m\}$ по индукциии и придем к утверждению леммы 3. $\square$

## Ядро графа

В работе [Reittu, Norros, 2004] показано, что в Интернет-графе (без ограничений на число связей) вершины, имеющие достаточно большие степени, концентрируются вокруг вершины с максимальной степенью $j^*$ (если таких вершин несколько, то в качестве $j^*$ можно взять любую из них) и соединены с ней путями небольшой (порядка $\ln \ln N$) длины. Такой подграф получил название ядра. Не удивительно, что подобная структура возникает и в рассматриваемых условных графах.

Обозначим $g_1 = (\tau - 1)/\tau$, $\eta_{(N)}$ – степень вершины $j^*$, $\varepsilon = \varepsilon(N) = w/\ln N$, где $w$ определено в (7). Пусть $U_1$ означает множество вершин следующего вида:

$$U_1 = \{j : \eta_j \geqslant N^{g_1 + \varepsilon}\}. \qquad (45)$$

Согласно теореме 4 из [Павлов, 2009], в наших условиях при $r = (N/\gamma)^{1/\tau}(1 + o(1))$, где $\gamma$ – некоторая положительная постоянная, справедливо соотношение

$$\mathsf{P}\{\eta_{(N)} \leqslant r\} \to e^{-\gamma}, \qquad (46)$$

откуда ясно, что $j^* \in U_1$. Нетрудно видеть, что

$$N^{\varepsilon} \to \infty. \qquad (47)$$

Учитывая (45) – (47), с помощью леммы 1 получаем, что любая вершина из $U_1$ а.д. соединена ребром с $j^*$. Обозначим

$$g_i = (\tau - 1)^i/\tau + \varepsilon \sum_{k=0}^{i-1} (\tau - 1)^k, \quad i = 2, 3, ..., \qquad (48)$$

и пусть

$$U_i = \{j : N^{g_i + \varepsilon} \leqslant \eta_i < N^{g_{i-1} + \varepsilon}\}, \quad i = 2, 3, ..., . \qquad (49)$$

Множество вершин графа

$$C = \bigcup_{i=1}^{\alpha} U_i \qquad (50)$$

назовем *ядром.*

**Лемма 4.** *Справедливы следующие утверждения.*

1. *Любая вершина ядра, принадлежащая множеству $U_{i+1}$, $i = 1, 2, ...$, а.д. соединена ребром с какой-нибудь вершиной множества $U_i$.*
2. *Любая вершина ядра, не принадлежащая множеству $\bigcup_{i=1}^{s+2} U_i$, $s = 1, 2, ...$, а.д. не имеет ребер, соединяющих ее с вершиной множества $U_s$.*

*Доказательство.* Из (1), (45), (49) нетрудно показать, что при $N \to \infty$ справедливо соотношение $\mathsf{P}\{j \in U_i\} \sim N^{-(g_i + \varepsilon)\tau}$. Используя известное приближение биномиального распределения нормальным, находим, что число вершин в $U_i$ асимптотически эквивалентно $N^{1 - (g_i + \varepsilon)\tau}$. Отсюда и из (49) получаем, что сумма степеней вершин из $U_i$ а. д. пропорциональна $N^{1 - (\tau - 1)(g_i + \varepsilon)}$. Из (48) следует, что

$$g_{i+1} = (\tau - 1) g_i + \varepsilon, \qquad (51)$$

поэтому из (47), (49) с помощью леммы 1 и соотношения $n/N \to \zeta(\tau)$ приходим к первому утверждению леммы 4.

Как мы видим, число степеней вершин в $U_s$ пропорционально $N^{1-(\tau-1)(g_s+\varepsilon)}$. Из (49) следует, что степень любой вершины, не принадлежащей $\bigcup\limits_{i=1}^{s+2} U_i$, меньше, чем $N^{g_{s+2}+\varepsilon}$. Отсюда с помощью (51) и леммы 1 получаем второе утверждение леммы 4. □

Из леммы 4 очевидным образом вытекают такие следствия.

**Следствие 1.** *Ядро а.д. состоит из непересекающихся слоев $U_1, U_2, ...,$ концентрически расположенных вокруг вершины $j^*$. Кратчайший путь из вершины $j^*$ в любую вершину ядра, принадлежащую слою $U_i$, $i = 2, 3, ...,$ проходит последовательно через вершины, принадлежащие слоям $U_1, U_2, ..., U_{i-1}$, и он не содержит ребер, соединяющих непосредственно вершины из слоев, номера которых отличаются больше, чем на 2.*

**Следствие 2.** *Длина кратчайшего пути, соединяющего любые две вершины ядра, а.д. не превосходит $2\alpha$.*

Хорошо известно [Севастьянов, 1971], что определенная в теореме вероятность вырождения $q$ удовлетворяет уравнению

$$q = F(q), \quad 0 < q < 1. \tag{52}$$

Выберем произвольную вершину $j^*$ и обозначим $D_j$ событие, состоящее в том, что эта вершина принадлежит компоненте связности, содержащей ядро.

**Лемма 5.** *Пусть $N \to \infty$. Тогда $\mathsf{P}\{D_j\} \to 1 - \varphi(q)$.*

*Доказательство.* Найдем сначала вероятность вырождения процесса $G$. Следуя [Севастьянов, 1971], введем обозначение

$$F_t(z, G) = \sum_{k=0}^{\infty} \mathsf{P}\{Z_t(G) = k\} z^k, \quad t = 0, 1, 2, ...$$

Известно [Севастьянов, 1971], что $F_t(z, G) = \varphi(F_t(z, G))$. Полагая $z = 0$, отсюда получаем:

$$\mathsf{P}\{Z_{t+1}(G) = 0\} = \varphi(\mathsf{P}\{Z_t(G) = 0\}).$$

Переходя в этом равенстве к пределу при $t \to \infty$ и учитывая (5), (52), находим вероятность вырождения $G$:

$$\lim_{t\to\infty} \mathsf{P}\{Z_t(\alpha) = 0\} = \varphi(q). \tag{53}$$

Обозначим $D_{j,t}$ событие, состоящее в том, что к моменту $t$ закончилось построение первой связной компоненты графа, не содержащей вершин из ядра, с помощью описанного выше алгоритма, работа которого началась с вершины $j$. Ясно, что в этом случае

$$\mathsf{P}\{D_{j,t}\} \leqslant \mathsf{P}\{Z_t = 0\}. \tag{54}$$

Положим $h = \ln N$, $t \leqslant \ln\ln\ln\ln N$. Из леммы 2 следует :

$$\mathsf{P}\{D_{j,t}\} \geqslant \mathsf{P}\{D_{j,t}, A_t\}$$
$$= \mathsf{P}\{Z_t = 0\} + O((\ln N)^{-(\tau-1)^2}). \tag{55}$$

Утверждение леммы 5 следует из леммы 3 и соотношений (53) – (55). □

В лемме 3 установлено соответствие между рассматриваемыми случайными графами и ветвящимся процессом $G$. С помощью этого соответствия ниже, в лемме 6, будет показано, что асимптотически эквивалентны события, состоящие в том, что процесс $G$ выродился и что соответствующий подграф, построенный с помощью алгоритма, не содержит вершин ядра.

**Лемма 6.** *Пусть $N, n \to \infty$ так, что $n - \zeta(\tau)N = O(N^{1/\tau})$. Если вершина графа принадлежит компоненте связности, содержащей ядро, то расстояние от нее до ядра, а. д. не превосходит $\beta$.*

*Доказательство.* Мы будем следовать схеме рассуждений, приведенных в [Reittu, Norros, 2004] при доказательстве аналогичного утверждения о безусловных случайных графах.

Для последовательности номеров вершин $j_1, j_2, ...,$ полученной с помощью алгоритма, введем обозначения:

$$\sigma = \sigma(j_1, j_2, ...,) = \inf\{k : \ j_k = j_l, \ l < k\},$$

$$\Delta = \inf\{k : \ \sum_{i=1}^{k}(\eta_{j_i} - 1) \geqslant \beta\}, \quad S_k = \sum_{i=1}^{k} \eta_{j_i},$$
$$m = [4\beta / \mathsf{P}\{\eta_1 > 1\}].$$

Докажем, что

$$\mathsf{P}\{\Delta \leqslant m | \nu_N = n\} \to 1. \tag{56}$$

Действительно, нетрудно видеть, что в силу леммы 3 [Павлов, 2007] при $n \to \infty$

$$\mathsf{P}\{\eta_{j_1} > 1 | \nu_N = n\} \to \mathsf{P}\{\eta_1 > 1\}. \tag{57}$$

Обозначим $\mathbf{I}(A)$ индикатор события $A$. Для случайной величины $X$, имеющей биномиальное распределение с параметрами $n, p$, справедлива оценка [Феллер, 1984]:

$$\mathsf{P}\{X \leqslant np/2\} \leqslant 4(1 - p/2)/np, \tag{58}$$

поэтому с помощью (57), (58) находим, что

$$\mathsf{P}\{\Delta \leqslant m|\nu_N = n\}$$

$$\geqslant \mathsf{P}\left\{\sum_{i=1}^{m} \mathbf{I}(\eta_{j_i} > 1) > \beta|\nu_N = n\right\}$$

$$\geqslant \mathsf{P}\left\{\sum_{i=1}^{m} \mathbf{I}(\eta_{j_i} > 1) > m\,\mathsf{P}\{\eta_1 > 1\}/4|\nu_N = n\right\}$$

$$\geqslant \mathsf{P}\left\{\sum_{i=1}^{m} \mathbf{I}(\eta_{j_i} > 1) > m\,\mathsf{P}\{\eta_1 > 1|\nu_N = n\}/2|\nu_N\right.$$

$$\left. = n\right\} - \mathsf{P}\{\mathsf{P}\{\eta_1 > 1|\nu_N = n\} > \mathsf{P}\{\eta_1 > 1\}/2|\nu_N$$

$$= n\} \geqslant 1 - 8/m\,\mathsf{P}\{\eta_1 > 1|\nu_N = n\} + o(1) \to 1,$$

откуда и следует (56).

По определению $\Delta$ для любого $u$ эквивалентны неравенства $u \geqslant \Delta$ и $S_u \geqslant u + \beta$. Отсюда нетрудно видеть, что неравенства

$$u \geqslant \Delta \quad \text{и} \quad S_u \geqslant \Delta + \beta \tag{59}$$

также эквивалентны. Для всех $l = o(N)$

$$\mathsf{P}\{S_{\sigma-1} < l|\nu_N = n\}$$

$$= \sum_{i=1}^{l} \mathsf{P}\{\sigma = i, S_{i-1} < l|\nu_N = m\}. \tag{60}$$

Используя лемму 3 [Павлов, 2007], находим, что

$$\mathsf{P}\{\sigma = i, S_{i-1} < l|\nu_N = n\} \leqslant C_1 l/n,$$

поэтому из (60) следует:

$$\mathsf{P}\{S_{\sigma-1} < l|\nu_N = n\} \leqslant C_1 l^2/n.$$

Отсюда и из (56), (59) получаем:

$$\mathsf{P}\{\sigma > \Delta|\nu_N = n\} = \mathsf{P}\{S_{\sigma-1} \geqslant \Delta + \beta|\nu_N = n\}$$

$$\geqslant \mathsf{P}\{S_{\sigma-1} \geqslant m+\beta|\nu_N = n\} - \mathsf{P}\{\Delta > m|\nu_N = n\}$$

$$\geqslant 1-(m+\beta)^2/n - \mathsf{P}\{\Delta > m|\nu_N = n\} \to 1. \tag{61}$$

Обозначим $L = \inf\{i : j_i \in G\}$. Оценим эту величину сверху. Для того, чтобы вершина принадлежала ядру, необходимо (см. (6), (48), (50), (51)), чтобы ее степень была не меньше, чем $N^{g_\alpha+\varepsilon}$. Как легко видеть из (51), это значит, что при достаточно больших $N$ степень такой вершины должна превосходить $N^{(3-\tau)\varepsilon/(2-\tau)}$. Отсюда и из (1) следует, как нетрудно проверить, что для любой вершины $j$ справедливо соотношение $\mathsf{P}\{j \in G\} \sim N^{-(3-\tau)\tau\varepsilon/(2-\tau)}$. Учитывая нормальное приближение биномиального распределения, для случайной величины $v$, равной объему ядра, получаем неравенство $v \geqslant C_2 N^{1-(3-\tau)\tau\varepsilon/(2-\tau)}$, откуда следует, что число степеней вершин ядра больше, чем $C_3 N^{1-(\tau-1)(3-\tau)\varepsilon/(2-\tau)}$. Тогда из (6), (49) и условия $n/N \to \varsigma(\tau)$ находим:

$$\mathsf{P}\{j_1, ..., j_\beta\} \bigcap C = \varnothing|\nu_N = n\}$$

$$\leqslant (1 - C_4 N^{(1-\tau)(3-\tau)\varepsilon/(2-\tau)})^\beta \to 0,$$

откуда

$$\mathsf{P}\{L \leqslant \beta\} \to 1. \tag{62}$$

□

## Доказательство теоремы

Предположим, что построенный с помощью алгоритма и начинающийся с рассматриваемой вершины подграф содержит $\delta < \infty$ поколений вершин. Положим $h = o(N^{(3-\tau)\varepsilon/(2-\tau)})$. В силу лемм 2 и 3 это значит, что ветвящийся процесс $G$ к моменту $\delta$ выродился, а построенный подграф не содержит вершин ядра. Если же, наоборот, процесс $G$ не выродился до рождения $\beta$ частиц, то из (61) и определения $\sigma$ следует, что мы можем считать идентичными траекторию процесса до этого момента и соответствующую реализацию подграфа и, в силу (62), это значит, что ядро достигается.

Теперь нетрудно доказать теорему. Первое утверждение очевидным образом следует из леммы 5 и свойств биномиального распределения, а второе – из следствия 2 и леммы 6.

## Литература

*Ватутин В. А.* Распределение расстояния до корня минимального поддерева, содержащего все вершины данной высоты // Теория вероятностей и ее применения. 1993. Т. 38, № 2. С. 273–287.

*Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные величины. М.: Наука, 1965. 424 с.

*Колчин В. Ф.* Случайные графы. М.: Физматлит, 2000. 256 с.

*Павлов Ю. Л.* Предельное распределение объема гигантской компоненты в случайном графе Интернет-типа // Дискретная математика. 2007. Т. 19, № 3. С. 22–34.

*Павлов Ю. Л.* О предельных распределениях степеней вершин в условных Интернет-графах // Дискретная математика. 2009. Т. 21, № 3. С. 14–23.

*Павлов Ю. Л., Чеплюкова И. А.* Случайные графы Интернет-типа и обобщенная схема размещения // Дискретная математика. 2008. Т. 20, № 3. С. 3–18.

*Севастьянов Б. А.* Ветвящиеся процессы. М.: Наука, 1971. 436 с.

*Феллер В.* Введение в теорию вероятностей и ее применения. Т. 1. М.: Мир, 1984. 527 с.

*Durrett R.* Random Graph Dynamics. N. Y.: Cambridge University Press, 2007. 221 p.

*Faloutsos C., Faloutsos P., Faloutsos M.* On power-law relationship of the Internet topology // Computer Communications Rev. 1999. Vol. 29. P. 251–262.

*Flajolet P., Sedgewick R.* Analytic Combinatorics. Cambridge: Cambridge University Press, 2009. 824 p.

*Hofstad R.* Random Graph and Complex Networks. Department of Mathematics and Computer Sciences. Eindhoven University of Thechnology. 2011. URL:http://www.uin.tue.nl/ rhofstad/ Notes RGGN.pdf (дата обращения: 11.04.2011)

*Reittu H., Norros I.* On the power-low random graph model of massive data networks // Performance Evaluation. 2004. Vol. 55, N. 1. P. 3–23.

## СВЕДЕНИЯ ОБ АВТОРЕ:

**Павлов Юрий Леонидович**
зав. лаб. теории вероятностей и компьютерной статистики, д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: pavlov@krc.karelia.ru
тел.: (8142) 761218

**Pavlov, Yury**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: pavlov@krc.karelia.ru
tel.: (8142) 761218

Труды Карельского научного центра РАН
№ 5. 2011. С. 97–102

*УДК 519.6:539.2*

# РАЗНОСТНАЯ СХЕМА ДЛЯ КРАЕВОЙ ЗАДАЧИ ВОДОРОДОПРОНИЦАЕМОСТИ ПРИ НАЛИЧИИ ДЕФЕКТА ЗАЩИТНОГО ПОКРЫТИЯ

**Н. И. Родченкова, Е. К. Костикова**

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассматривается водородопроницаемость цилиндрической перегородки при наличии на входной поверхности дефекта защитного покрытия. Для различных конструкционных материалов, когда лимитирующими являются процессы диффузии и десорбции, представлены соответствующие математические модели в форме краевых задач с условиями I рода и нелинейными динамическими граничными условиями. Разработана разностная схема для численного моделирования проникающего потока.

Ключевые слова: водородопроницаемость, нелинейные краевые задачи, разностные схемы.

**N. I. Rodchenkova, E. K. Kostikova. DIFFERENCE SCHEME FOR THE BOUNDARY-VALUE PROBLEM OF HYDROGEN PERMEABILITY IN THE PRESENCE OF PROTECTIVE COATING DEFECT**

We consider the permeability of cylindrical barrier to hydrogen in the presence of protective coating defect on the inlet surface. For different structural materials, when diffusion and desorption are the limiting processes, corresponding mathematical models in the form of boundary-value problems with first-type conditions and nonlinear dynamic boundary conditions are presented. The difference scheme for numerical modelling of the permeation flux has been developed.

Key words: hydrogen permeability, nonlinear boundary-value problems, difference schemes.

## Постановка задачи

Снижение проникновения водорода и его изотопов сквозь стенки из конструкционных материалов является важнейшей задачей при решении комплексных проблем хранения и транспортировки водорода, защиты от водородного охрупчивания, контроля содержания трития в защитных системах будущих термоядерных реакторов (проект ITER). Конструкция из металла или сплава обеспечивает необходимую механическую прочность перегородки, а нанесенное защитное покрытие должно препятствовать миграции изотопов водорода. Дефекты защитной пленки могут подвергать соответствующую область конструкционного материала прямому воздействию водорода [Писарев и др., 2008]. В статье [Zajec, 2011] поставлена задача математического моделирования водородопроницаемости цилиндрического образца радиуса $L$ и высоты

$H$ в случае, когда диффузия является единственным лимитирующим процессом. На входной поверхности $z = 0$, покрытой тонкой защитной пленкой, присутствует дефект малого радиуса $r_0$ (булавочное отверстие), через который проникает водород. Остальная часть входной поверхности водородонепроницаема, как и боковая поверхность. На выходной стороне $z = H$ поддерживается вакуум. В начальный момент времени $t = 0$ образец обезводорожен. Затем на входной стороне скачкообразно повышается давление молекулярного водорода до уровня $p$. Если пренебречь относительно быстрым (это зависит от $p$, материала и размеров образца) переходным процессом, то можно считать, что концентрация растворенного водорода под дефектом поддерживается на постоянном уровне $c_0$ (находится в равновесии с газообразной фазой по закону Сивертса, $c_0 \propto \sqrt{p}$). Растворенный (атомарный) водород диффундирует к выходной поверхности. С помощью масс-спектрометра регистрируется проникающий поток.

Аналитический анализ краевой задачи без учета поверхностных процессов проведен лишь для случая полупространства ($r \to +\infty$) [Warrick et al., 1992; Rajendran, Sangaranarayanan, 1995]. Основным недостатком такой постановки задачи является то, что поверхностные процессы, которым в последнее время уделяется повышенное внимание, в модели не учитываются.

Цель работы — построить разностную схему для моделей водородопроницаемости цилиндрического образца заданных размеров при наличии дефекта как без учета, так и с учетом влияния поверхности.

## Диффузионная модель

Рассмотрим краевую задачу водородопроницаемости цилиндрической перегородки только с учетом диффузии в объеме:

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial r^2} + \frac{1}{r}\cdot\frac{\partial c}{\partial r} + \frac{\partial^2 c}{\partial z^2}\right), \qquad (1)$$

$$r \in (0, L),\ z \in (0, H),\ t \in (0, t_*),$$

$$c(t, r, 0) = c_0, \quad r \in [0, r_0],\ r_0 < L, \qquad (2)$$

$$\frac{\partial c}{\partial z}(t, r, 0) = 0, \quad r \in (r_0, L],\ t \geqslant 0, \qquad (3)$$

$$c(t, r, H) = 0, \quad r \in [0, L],\ t \geqslant 0, \qquad (4)$$

$$\frac{\partial c}{\partial r}(t, L, z) = 0, \quad \frac{\partial c}{\partial r}(t, +0, z) = 0, \qquad (5)$$

$$c(0, r, z) = 0, \quad r \in [0, L],\ z \in [0, H].$$

Здесь $c(t, r, z)$ – концентрация атомарного водорода в конструкционном материале (металле или сплаве); $D$ – коэффициент диффузии. Момент времени $t_*$ определяется выходом проникающего потока на стационарное значение. Следует отметить, что установление носит асимптотический характер. Но $t_*$ не следует выбирать слишком большим, чтобы переходные процессы «не потерялись» на фоне стационара. Условие $c_r(t, +0, z)$ следует из симметрии распределения $c(t, r, z)$.

*Уточнение поставки задачи.* Цель состоит в разработке разностной схемы для моделирования потока водорода с выходной поверхности цилиндрического образца:

$$J(t) = -D\int_0^L \frac{\partial c}{\partial z}\Big|_{z=H} 2\pi r\, dr.$$

Изменения выражения $J(t)$ с учетом рекомбинации атомов водорода в молекулы на поверхности (в приповерхностном объеме) приведены в последующих разделах статьи.

Численное моделирование позволяет выявить особенности кинетики проникновения водорода и оценить влияние параметров модели, включая размеры образца и дефекта.

## Разностная аппроксимация краевой задачи

Будем следовать методике, разработанной для задачи термодесорбции [Родченкова, Заика, 2010]. Следуя стандартной технике [Самарский, 1971], введем пространственную сетку

$$\Omega_h = \left\{(r_i, z_j)\,\middle|\, \begin{array}{l} r_i = i\,h_r,\ i = 0, 1, ..., N_1 = [L/h_r] \\ z_j = j\,h_z,\ j = 0, 1, ..., N_2 = [H/h_z] \end{array}\right\}$$

и сетку по времени

$$\omega_\tau = \{\, t_k = k\tau,\ k = 0, 1, ..., K = [t_*/\tau]\}.$$

Обозначим через $c_{i,j}^k$ приближенные значения объемной концентрации $c(t_k, r_i, z_j)$. При переходе с $k$ на $k+1$ слой по времени будем использовать следующие обозначения, опуская номер слоя по времени: $c_{i,j} = c_{i,j}^k$, $\bar{c}_{i,j} = c_{i,j}^{k+1/2}$, $\hat{c}_{i,j} = c_{i,j}^{k+1}$, где $(r_i, z_j) \in \Omega_h$, $t_k \in \omega_\tau$. Для уравнения (1) рассмотрим неявную разностную схему метода переменных направлений, называемую продольно-поперечной (схемой Писмена-Рэкфорда) [Косарев, 2000]. Переход от слоя $k$ к слою $k+1$ осуществляется в два этапа. На первом этапе определяются промежуточные значения $\bar{c}_{i,j}$ из соотношений

$$\frac{\bar{c}_{i,j} - c_{i,j}}{0{,}5\,\tau D} = \frac{c_{i,j-1} - 2\,c_{i,j} + c_{i,j+1}}{h_z^2} + \frac{\bar{c}_{i-1,j} - 2\,\bar{c}_{i,j} + \bar{c}_{i+1,j}}{h_r^2} + \frac{\bar{c}_{i+1,j} - \bar{c}_{i-1,j}}{2\,r_i h_r}.$$

На втором этапе, пользуясь найденными $\bar{c}_{i,j}$, находим $\hat{c}_{i,j}$ из соотношений

$$\frac{\hat{c}_{i,j}-\bar{c}_{i,j}}{0,5\,\tau D}=\frac{\hat{c}_{i,j-1}-2\,\hat{c}_{i,j}+\hat{c}_{i,j+1}}{h_z^2}+\frac{\bar{c}_{i-1,j}-2\,\bar{c}_{i,j}+\bar{c}_{i+1,j}}{h_r^2}+\frac{\bar{c}_{i+1,j}-\bar{c}_{i-1,j}}{2\,r_i h_r}. \quad (6)$$

Данные соотношения рассматриваются во внутренних узлах сетки ($i=2,...,N_1-1$, $j=1,...,N_2-1$). Суммарная погрешность аппроксимации $O(\tau^2+h_r^2+h_z^2)$.

**Прогонка по радиусу $r$**

Рассмотрим переход с $k$-го на $(k+1/2)$-й слой. В обозначениях

$$A_i=\frac{1-[2i]^{-1}}{\varkappa},\ B_i=\frac{1+[2i]^{-1}}{\varkappa},\ \varkappa=2\Big[1+\frac{h_r^2}{D\tau}\Big],$$
$$F_{i,j}=\Big(\frac{h_r}{h_z}\Big)^2\cdot\frac{c_{i,j-1}-2c_{i,j}+c_{i,j+1}}{\varkappa}+\Big[1-\frac{2}{\varkappa}\Big]c_{i,j}$$

при каждом фиксированном $j=1,2,\ldots,N_2-1$ получаем при $k\geqslant 0$

$$A_i\bar{c}_{i-1,j}-\bar{c}_{i,j}+B_i\bar{c}_{i+1,j}+F_{i,j}=0.$$

Значения в начальный момент времени (на нулевом слое) известны: $c_{i,j}^0=0$. Следуя методу прогонки, ищем приближенные значения концентрации в узлах сетки на $(k+1/2)$-м слое ($k\geqslant 0$) по времени в виде

$$\bar{c}_{i,j}=\alpha_{i+1,j}\,\bar{c}_{i+1,j}+\beta_{i+1,j},\ \ i=0,...,N_1-1. \quad (7)$$

Прогоночные коэффициенты: $i=2,3,\ldots,N_1$,

$$\alpha_{i,j}=\frac{B_{i-1}}{1-A_{i-1}\,\alpha_{i-1,j}},\ \ \beta_{i,j}=\frac{F_{i-1,j}+A_{i-1}\,\beta_{i-1,j}}{1-A_{i-1}\alpha_{i-1,j}}.$$

При $r\to+0$ имеем

$$c_r/r=(c_r(t,r,z)-c_r(t,0,z))/r\approx c_{rr}.$$

Начальные коэффициенты $\alpha_{1,j}$, $\beta_{1,j}$ находим из аппроксимации уравнения $c_t=D(2c_{rr}+c_{zz})$ на $(k+1/2)$-м слое, $i=1$, и условия $c_r|_{+0}=0$: $\alpha_{1,j}=(6-\varkappa)/4$, $\beta_{1,j}=F_{1,j}\,\varkappa/4$.

*Ближайшая цель* — найти значение $\bar{c}_{N_1,j}$, необходимое для реализации обратного хода прогонки. Запишем аппроксимацию первого из граничных условий (5) ($r=L$). В граничном узле с точностью до $O(h_r^2)$ имеем:

$$2h_r c_r(t_{k+1/2},L,z_j)\approx\bar{c}_{N_1-2,j}-4\bar{c}_{N_1-1,j}+3\bar{c}_{N_1,j}.$$

Подставляя значения $\bar{c}_{N_1-2,j}$, $\bar{c}_{N_1-1,j}$ из соотношения (7), имеем

$$\bar{c}_{N_1,j}=\frac{\beta_{N_1,j}(4-\alpha_{N_1-1,j})-\beta_{N_1-1,j}}{3+\alpha_{N_1,j}(\alpha_{N_1-1,j}-4)}.$$

По формуле (7) определяем искомые значения $\bar{c}_{i,j}$, $i=0,\ldots,N_1-1$, $j=1,\ldots,N_2-1$.

Теперь найдем оставшиеся приближения $\bar{c}_{i,j}$ при $j=0$ и $j=N_2$, $i=0,1,\ldots,N_1$. Значения $\bar{c}_{i,N_2}=0$ согласно граничному условию (4) ($z=H$). Обозначим $i_0=\max\{i: r_i\leqslant r_0\}$. Тогда из граничных условий (2), (3) ($z=0$) получаем соотношения $\bar{c}_{i,0}=c_0$, $i=0,1,\ldots,i_0$ и $\bar{c}_{i,0}=(4\bar{c}_{i,1}-\bar{c}_{i,2})/3$, $i=i_0+1,\ldots,N_1$.

**Прогонка по переменной $z$**

Поскольку в цилиндрических координатах возникает особенность при $r\to+0$, и на границе $z=0$, задано смешанное краевое условие, то переход с $(k+1/2)$-го слоя на $(k+1)$-й совершается в два шага.

Первый шаг: $i=1$, $r\to+0$ (аппроксимируем $c_r/r\approx c_{rr}$), реализуется алгоритм прогонки для уравнения $c_t=D(2c_{rr}+c_{zz})$.

Второй шаг: $i=2,\ldots,N_1-1$, $r>0$, прогонка для уравнения $c_t=D(c_{rr}+c_r/r+c_{zz})$.

Аппроксимация уравнения $c_t=D(2c_{rr}+c_{zz})$:

$$\frac{\hat{c}_{1,j}-\bar{c}_{1,j}}{0{,}5\,\tau D}=\frac{\hat{c}_{1,j-1}-2\,\hat{c}_{1,j}+\hat{c}_{1,j+1}}{h_z^2}+2\,\frac{\bar{c}_{0,j}-2\,\bar{c}_{1,j}+\bar{c}_{2,j}}{h_r^2}.$$

В обозначениях $G=2(1+h_z^2/[D\tau])$,

$$\bar{F}_{1,j}=2\Big(\frac{h_z}{h_r}\Big)^2\big(\bar{c}_{0,j}-2\bar{c}_{1,j}+\bar{c}_{2,j}\big)+(G-2)\bar{c}_{1,j}$$

получаем

$$\hat{c}_{1,j-1}-G\,\hat{c}_{1,j}+\hat{c}_{1,j+1}+\bar{F}_{1,j}=0,\ \ k\geqslant 0. \quad (8)$$

Ищем приближенные значения концентрации на $(k+1)$-м слое по времени в виде: $k\geqslant 0$,

$$\hat{c}_{1,j}=\alpha_{1,j+1}\,\hat{c}_{1,j+1}+\beta_{1,j+1},\ j=0,1,...,N_2-1. \quad (9)$$

Прогоночные коэффициенты: $j=2,3,\ldots,N_2$,

$$\alpha_{1,j}=\frac{1}{G-\alpha_{1,j-1}},\ \ \beta_{1,j}=\frac{\bar{F}_{1,j-1}+\beta_{1,j-1}}{G-\alpha_{1,j-1}}.$$

Начальные коэффициенты $\alpha_{1,1}$, $\beta_{1,1}$ находим из формулы (9) при $j=0$ и условия (2) ($\hat{c}_{1,0}=c_0$): $\alpha_{1,1}=0$, $\beta_{1,1}=c_0$.

Для разностного соотношения (6) в обозначениях

$$\bar{F}_{i,j}=\Big(\frac{h_z}{h_r}\Big)^2\Big[\Big(1-\frac{1}{2i}\Big)\bar{c}_{i-1,j}-2\Big(1-\frac{h_r^2}{D\tau}\Big)\bar{c}_{i,j}+\Big(1+\frac{1}{2i}\Big)\bar{c}_{i+1,j}\Big]$$

при каждом $i = 2, 3, \ldots, N_1 - 1$ получаем

$$\hat{c}_{i,j-1} - G\hat{c}_{i,j} + \hat{c}_{i,j+1} + \bar{F}_{i,j} = 0, \;\; k \geqslant 0. \quad (10)$$

Ищем значения концентрации в узлах сетки на $(k+1)$-м слое по времени в виде: $k \geqslant 0$,

$$\hat{c}_{i,j} = \alpha_{i,j+1}\, \hat{c}_{i,j+1} + \beta_{i,j+1},\; j = 0, 1, ..., N_2 - 1. \quad (11)$$

Прогоночные коэффициенты: $j = 2, 3, \ldots, N_2$,

$$\alpha_{i,j} = \frac{1}{G - \alpha_{i,j-1}}, \;\; \beta_{i,j} = \frac{\bar{F}_{i,j-1} + \beta_{i,j-1}}{G - \alpha_{i,j-1}}.$$

Начальные коэффициенты при $r \leqslant r_0$ $(i \leqslant i_0)$ находим из (11) при $j = 0$ и условия (2): $\alpha_{i,1} = 0$, $\beta_{i,1} = c_0$. Начальные коэффициенты при $r > r_0$ $(i > i_0)$ определяем из (10) при $j = 1$ и условия (3): $\alpha_{i,1} = 2 - G/2$, $\beta_{i,1} = \bar{F}_{i,1}/2$.

Граничные значения, необходимые для реализации обратного хода метода прогонки, находим из условия (4): $\hat{c}_{i,\,{}_{N_2}} = 0$. По формулам (9), (11) вычисляем $\hat{c}_{i,j}$, $i = 1, \ldots, N_1 - 1$, $j = 0, \ldots, N_2 - 1$.

Найдем оставшиеся приближения $\hat{c}_{i,j}$ при $i = 0$ и $i = N_1$, $j = 0, 1, \ldots, N_2$. Используя второе из граничных условий (5) $(c_r|_{+0} = 0)$, получаем $\hat{c}_{0,j} = (4\hat{c}_{1,j} - \hat{c}_{2,j})/3$. Согласно первому условию (5) $\hat{c}_{{}_{N_1},j} = (4\hat{c}_{{}_{N_1-1},j} - \hat{c}_{{}_{N_1-2},j})/3$.

## Модификация модели с учетом объемной десорбции

Вместо краевых условий (2), (4) используем

$$\mu sp - bc^2(t, r, 0) = -D\, \frac{\partial c}{\partial z}\Big|_{z=0}, \;\; r \in [0, r_0], \quad (12)$$

$$bc^2(t, r, H) = -D \frac{\partial c}{\partial z}\Big|_{z=H}, \;\; r \in [0, L], \quad (13)$$

$$J(t) = \int_0^L b\, c^2(t, r, H)\; 2\pi r\, dr.$$

Здесь $b$ — коэффициент объемной десорбции (эффективной рекомбинации [Писарев и др., 2008]), $\mu$ — кинетическая константа, $p$ — давление молекулярного водорода, $s$ — коэффициент прилипания водорода к поверхности. При необходимости можно учесть различие входной и выходной поверхностей: $b = b_1$ при $z = 0$ и $b = b_2$ при $z = H$.

*1. Метод встречных прогонок.* Рассмотрим переход с $(k+1/2)$-го слоя на $(k+1)$-й. Для соотношений (8), (10), используя метод встречных прогонок [Самарский, Николаев, 1978], будем искать приближенные значения концентрации в следующем виде $(i = 1, \ldots, i_0,\, k \geqslant 0)$:

$$\hat{c}_{i,j} = \alpha_{i,j+1}\, \hat{c}_{i,j+1} + \beta_{i,j+1} + \gamma_{i,j+1}\, \hat{c}_{i,0}, \quad j = 0, 1, \ldots, N_2 - 1;$$

$$\hat{c}_{i,j} = \hat{\alpha}_{i,j-1}\, \hat{c}_{i,j-1} + \hat{\beta}_{i,j-1} + \hat{\gamma}_{i,j-1}\, \hat{c}_{i,N_2}, \quad j = 1, 2, \ldots, N_2.$$

Прогоночные коэффициенты:

$$\alpha_{i,j+1} = (G - \alpha_{i,j})^{-1}, \qquad \hat{\alpha}_{i,j-1} = \alpha_{i,\,{}_{N_2-(j-1)}},$$
$$\beta_{i,j+1} = (\beta_{i,j} + \bar{F}_{i,j})\alpha_{i,j+1}, \quad \hat{\beta}_{i,j-1} = (\hat{\beta}_{i,j} + \bar{F}_{i,j})\hat{\alpha}_{i,j-1},$$
$$\gamma_{i,j+1} = \gamma_{i,j}\alpha_{i,j+1}, \qquad \hat{\gamma}_{i,j-1} = \gamma_{i,\,{}_{N_2-(j-1)}}.$$

Отметим, что выражения для величин $\bar{F}_{1,j}$ и $\bar{F}_{i,j}$, $i > 1$, различны. Начальные прогоночные коэффициенты: $\alpha_{i,1} = \beta_{i,1} = \hat{\beta}_{i,\,{}_{N_2-1}} = 0$, $\gamma_{i,1} = 1$.

В обозначениях $\Delta = \alpha_{i,{}_{N_2-1}} - 4$,

$$\mathbb{A} = 3\; + \alpha_{i,{}_{N_2}}\Delta, \qquad \mathbb{B} = \beta_{i,{}_{N_2-1}} + \beta_{i,{}_{N_2}}\Delta,$$
$$\hat{\mathbb{B}} = \hat{\beta}_{i,1} + \hat{\beta}_{i,0}\Delta, \qquad \mathbb{G} = \gamma_{i,{}_{N_2-1}} + \gamma_{i,{}_{N_2}}\Delta$$

получаем аппроксимацию граничных условий (12), (13) с точностью до $O(h_z^2)$ в виде

$$\begin{cases} \dfrac{2h_z}{D}\, b\, \hat{c}_{i,0}^2 + \mathbb{A}\hat{c}_{i,0} \; + \Big[\mathbb{G}\hat{c}_{i,{}_{N_2}} + \hat{\mathbb{B}} - \dfrac{2h_z}{D}\mu sp\Big] = 0, \\ \dfrac{2h_z}{D}\, b\, \hat{c}_{i,{}_{N_2}}^2 + \mathbb{A}\hat{c}_{i,{}_{N_2}} + \big[\mathbb{G}\hat{c}_{i,0} + \mathbb{B}\big] = 0. \end{cases}$$

Система уравнений имеет единственное решение $\hat{c}_{i,0} > 0$, $\hat{c}_{i,{}_{N_2}} > 0$, по крайней мере, при малых $h_r \sim h_z$ (сравнимых по порядку).

Осталось получить значения концентрации $\hat{c}_{i,j}$ при $i > i_0$. Изменения коснутся лишь границы $z = H$. В объеме выполняется соотношение (10), прогоночные коэффициенты остаются без изменений. Модификация появляется после выполнения прямого хода прогонки. Для нахождения $\hat{c}_{i,{}_{N_2}}$ аппроксимируем (13) с точностью до $O(h_z^2)$ и используем (11):

$$2h_z b D^{-1} \hat{c}_{i,{}_{N_2}}^2 + \mathbb{A}\hat{c}_{i,{}_{N_2}} + \mathbb{B} = 0,$$

где $i > i_0$, $\mathbb{A} = 3 + \alpha_{i,{}_{N_2}}\Delta$, $\mathbb{B} = \beta_{i,{}_{N_2-1}} + \beta_{i,{}_{N_2}}\Delta$, $\Delta = \alpha_{i,{}_{N_2-1}} - 4$. При малых $h_r \sim h_z$ корни квадратного уравнения имеют разные знаки, по физическому смыслу задачи выбираем положительный.

При переходе с $k$-го на $(k+1/2)$-й слой оставшиеся граничные значения $\bar{c}_{i,0}$, $\bar{c}_{i,{}_{N_2}}$ для $i = 0, ..., N_1$ определяются как положительные корни квадратных уравнений, полученных после подстановки в (12), (13) выражений

$$\bar{c}_z(t, r_i, 0) \approx \frac{-\bar{c}_{i,2} + 4\bar{c}_{i,1} - 3\bar{c}_{i,0}}{2h_z}, \quad (14)$$

$$\bar{c}_z(t, r_i, H) \approx \frac{\bar{c}_{i,{}_{N_2-2}} - 4\bar{c}_{i,{}_{N_2-1}} + 3\bar{c}_{i,{}_{N_2}}}{2h_z},$$

где $\bar{c}_{i,1}$, $\bar{c}_{i,2}$, $\bar{c}_{i,{}_{N_2-1}}$, $\bar{c}_{i,{}_{N_2-1}}$ уже известны по результатам прогонки по $r$.

*2. Итерационный метод.* На $(k+1)$-м слое по времени аппроксимируем $c_z(t_{k+1}, r, 0) \approx [-3\hat{c}_{i,0} + 4\hat{c}_{i,1} - \hat{c}_{i,2}]/2h_z$. Подставляя в граничное условие (12) при $z = 0$, находим $\hat{c}_{i,0} = f_0(\hat{c}_{i,1}, \hat{c}_{i,2})$ как положительный корень квадратного уравнения. Аналогичным образом определяем $\hat{c}_{i,N_2} = f_{N_2}(\hat{c}_{i,N_2-1}, \hat{c}_{i,N_2-2})$ как положительный корень из (13) при $z = H$. Значения $\hat{c}_{i,1}$, $\hat{c}_{i,2}$, $\hat{c}_{i,N_2-1}$, $\hat{c}_{i,N_2-2}$ предварительно подсчитываются по явной разностной схеме, аппроксимирующей уравнение диффузии (1). С текущими $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ решаем методом прогонки по переменной $z$ трехдиагональную систему линейных алгебраических уравнений и находим новые приближения концентраций $\hat{c}_{i,1}$, $\hat{c}_{i,2}$, $\hat{c}_{i,N_2-1}$, $\hat{c}_{i,N_2-2}$ (и остальные значения $\hat{c}_{i,j}$ для $j = 3, \ldots, N_2 - 3$, $i = 0, ..., i_0$). Снова решаем квадратные уравнения относительно $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ и повторяем вычисления до установления граничных значений (обычно 2–3 итерации). Нахождение приближений $\hat{c}_{i,j}$ при $i > i_0$ описано в предыдущем пункте. Затем переходим к следующему слою по времени.

## Модификация модели с учетом поверхностной десорбции

Вместо условий (2), (4) используем

$$\frac{\partial q_0}{\partial t} = \mu sp - bq_0^2(t,r) + D\frac{\partial c}{\partial z}\Big|_0, \ r \in [0, r_0], \quad (15)$$

$$\frac{\partial q_H}{\partial t} = -bq_H^2(t,r) - D\frac{\partial c}{\partial z}\Big|_{z=H}, \ r \in [0, L], \quad (16)$$

$$c(t,r,0) = gq_0(t,r), \ c(t,r,H) = gq_H(t,r),$$

$$J(t) = \int_0^L b\, q_H^2(t,r)\ 2\pi r\, dr.$$

Здесь $q_0$, $q_H$ — поверхностные концентрации на входной и выходной поверхности, $g = \mathrm{const}$ — коэффициент соответствия концентраций атомов водорода в объеме и на поверхности (коэффициент быстрого растворения).

*1. Метод встречных прогонок.* Остановимся на вычислении граничных концентраций $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$, $i \leqslant i_0$. Чтобы сохранить порядок аппроксимации $O(\tau^2 + h_z^2)$, для краевых условий (15), (16), будем использовать схему с весами:

$$\frac{\hat{c}_{i,0} - c_{i,0}}{g\tau} = \sigma\Big[\mu sp - b\Big(\frac{\hat{c}_{i,0}}{g}\Big)^2 + D\partial_z\hat{c}_{i,0}\Big] + (1-\sigma)\Big[\mu sp - b\Big(\frac{c_{i,0}}{g}\Big)^2 + D\partial_z c_{i,0}\Big],$$

$$\frac{\hat{c}_{i,N_2} - c_{i,N_2}}{g\tau} = \sigma\Big[-b\Big(\frac{\hat{c}_{i,N_2}}{g}\Big)^2 - D\partial_z\hat{c}_{i,N_2}\Big] + (1-\sigma)\Big[-b\Big(\frac{c_{i,N_2}}{g}\Big)^2 - D\partial_z c_{i,N_2}\Big].$$

Здесь $\partial_z c_{i,0}$, $\partial_z c_{i,N_2}$ определяются выражениями, аналогичными (14). В дальнейшем полагаем $\sigma = 1/2$. Как и в модификации модели для объемной десорбции (сохраняя обозначения), нахождение $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ сводится к решению системы квадратных уравнений:

$$\begin{cases} \dfrac{b}{g^2}\hat{c}_{i,0}^2 + \Big(\dfrac{D\mathbb{A}}{2h_z} + \dfrac{2}{g\tau}\Big)\hat{c}_{i,0} + \dfrac{D\mathbb{G}}{2h_z}\hat{c}_{i,N_2} + C_1 = 0, \\ \dfrac{b}{g^2}\hat{c}_{i,N_2}^2 + \Big(\dfrac{D\mathbb{A}}{2h_z} + \dfrac{2}{g\tau}\Big)\hat{c}_{i,N_2} + \dfrac{D\mathbb{G}}{2h_z}\hat{c}_{i,0} + C_2 = 0, \end{cases}$$

$$C_1 \equiv \frac{D\hat{\mathbb{B}}}{2h_z} + \frac{b}{g^2}c_{i,0}^2 - \frac{2}{g\tau}c_{i,0} - D\partial_z c_{i,0} - 2\mu sp,$$

$$C_2 \equiv \frac{D\mathbb{B}}{2h_z} + \frac{b}{g^2}c_{i,N_2}^2 - \frac{2}{g\tau}c_{i,N_2} + D\partial_z c_{i,N_2}.$$

Все значения $c_{i,j}$ на $k$-м слое известны.

*2. Итерационный метод.* Для краевых условий (15), (16) используется схема с весами (при $t = t_{k+1/2} + \tau/4$ порядок аппроксимации $O(\tau^2 + h_z^2)$):

$$\frac{\hat{c}_{i,0} - \bar{c}_{i,0}}{0{,}5\,\tau g} = 0{,}5\Big[\mu sp - b\Big(\frac{\hat{c}_{i,0}}{g}\Big)^2 + D\partial_z\hat{c}_{i,0}\Big] + 0{,}5\Big[\mu sp - b\Big(\frac{\bar{c}_{i,0}}{g}\Big)^2 + D\partial_z\bar{c}_{i,0}\Big],$$

$$\frac{\hat{c}_{i,N_2} - \bar{c}_{i,N_2}}{0{,}5\,\tau g} = 0{,}5\Big[-b\Big(\frac{\hat{c}_{i,N_2}}{g}\Big)^2 - D\partial_z\hat{c}_{i,N_2}\Big] + 0{,}5\Big[-b\Big(\frac{\bar{c}_{i,N_2}}{g}\Big)^2 - D\partial_z\bar{c}_{i,N_2}\Big].$$

Граничные значения $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ $(i \leqslant i_0)$ определяются как положительные корни квадратных уравнений:

$$\frac{b}{g^2}\hat{c}_{i,0}^2 + \Big(\frac{3D}{2h_z} + \frac{4}{g\tau}\Big)\hat{c}_{i,0} + C_1 = 0,$$

$$\frac{b}{g^2}\hat{c}_{i,N_2}^2 + \Big(\frac{3D}{2h_z} + \frac{4}{g\tau}\Big)\hat{c}_{i,N_2} + C_2 = 0,$$

$$C_1 \equiv \frac{D}{2h_z}\big(\hat{c}_{i,2} - 4\hat{c}_{i,1}\big) + \frac{b}{g^2}\bar{c}_{i,0}^2 - \frac{4}{g\tau}\bar{c}_{i,0} - D\partial_z\bar{c}_{i,0} - 2\mu sp,$$

$$C_2 \equiv \frac{D}{2h_z}\big(\hat{c}_{i,N_2-2} - 4\hat{c}_{i,N_2-1}\big) + \frac{b}{g^2}\bar{c}_{i,N_2}^2 + D\partial_z\bar{c}_{i,N_2} - \frac{4}{g\tau}\bar{c}_{i,N_2}.$$

В выражениях, аналогичных (14), значения $\hat{c}_{i,\{1,2\}}$, $\hat{c}_{i,\{N_2-1,N_2-2\}}$ предварительно подсчитываются по явной разностной схеме, аппроксимирующей уравнение диффузии (1). С текущими $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ решаем методом прогонки трехдиагональную систему линейных алгебраических уравнений и находим новые приближения $\hat{c}_{i,\{1,2\}}$, $\hat{c}_{i,\{N_2-1,N_2-2\}}$ (и остальные значения $\hat{c}_{i,j}$, $j = 3,\ldots,N_2-3$, $i = 0,...,i_0$). Снова решаем квадратные уравнения относительно $\hat{c}_{i,0}$, $\hat{c}_{i,N_2}$ и повторяем вычисления до установления граничных значений (обычно 2–3 итерации). Для нахождения граничной концентрации $\hat{c}_{i,N_2}$ при $i > i_0$ аппроксимируем (16) с точностью до $O(\tau^2 + h_z^2)$ и используем прогоночные коэффициенты, найденные прямой прогонкой при использовании условия (3):

$$bg^{-2}\hat{c}^2_{i,N_2} + [D(2h_z)^{-1}\mathbb{A} + 2(\tau g)^{-1}]\hat{c}_{i,N_2} + \mathbb{B} = 0,$$

где $\mathbb{A} = 3 + \alpha_{i,N_2}\Delta$, $\Delta = \alpha_{i,N_2-1} - 4$, $\mathbb{B} = -2\bar{c}_{i,N_2}/(\tau g) + D(\beta_{i,N_2-1} + \beta_{i,N_2}\Delta)/(2h_z)$. При малых $h_z$, $\tau$ корни квадратного уравнения разных знаков, по физическому смыслу задачи выбираем положительный. Вычисленные аппроксимации $c_z(t,r,H)$ в модели (1)–(5); концентрация $c(t,r,H)$ в случае граничных условий (12), (13); поверхностная концентрация $q_{_H}(t,r)$ для модификации модели (15), (16) дают возможность приближенно вычислить $J(t)$.

## Заключение

Таким образом, представленный вычислительный алгоритм позволяет по входным данным краевых задач водородопроницаемости (с учетом различных лимитирующих факторов) численно моделировать проникающий поток водорода сквозь перегородку из конструкционного материала при наличии дефекта защитного покрытия.

Работа выполнена при поддержке РФФИ (грант № 09-01-00439).

## Литература

*Косарев В. И.* 12 лекций по вычислительной математике. М.: МФТИ, 2000. 224 с.

*Писарев А. А., Цветков И. В., Маренков Е. Д., Ярко С. С.* Проницаемость водорода через металлы. М.: МИФИ, 2008. 144 с.

*Родченкова Н. И., Заика Ю. В.* Численное моделирование десорбции водорода с цилиндрической поверхности // Труды Карельского научного центра РАН. 2010. № 3. С. 72–82.

*Самарский А. А.* Введение в теорию разностных схем. М.: Наука, 1971. 553 с.

*Самарский А. А., Николаев Е. С.* Методы решения сеточных уравнений. М.: Наука, 1978. 592 с.

*Rajendran L., Sangaranarayanan M. V.* A two-point Pade approximation for the non-steady-state chronoamperometric current at ultramicrodisc electrodes // Journal of Electroanalytical Chemistry. Elsevier, 1995. Vol. 392. P. 75–78.

*Warrick A. W., Broadbridge P., Lomen D. O.* Approximations for diffusion from a disc source // Applied Mathematical Modelling. Elsevier, 1992. Vol. 16. P. 155–161.

*Zajec B.* Hydrogen permeation barrier – recognition of defective barrier film from transient permeation rate // International Journal of Hydrogen Energy. Elsevier, 2011. Vol. 36. P. 7353–7361.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Родченкова Наталья Ивановна**
научный сотрудник, к. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: nirodchenkova@yandex.ru
тел.: (8142) 766312

**Rodchenkova, Natalia**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: nirodchenkova@yandex.ru
tel.: (8142) 766312

**Костикова Екатерина Константиновна**
младший научный сотрудник
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: fedorova@krc.karelia.ru
тел.: (8142) 766312

**Kostikova, Ekaterina**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: fedorova@krc.karelia.ru
tel.: (8142) 766312

Труды Карельского научного центра РАН
№ 5. 2011. С. 103–110

*УДК 004.01:006.72 (470.22)*

# ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ПРИОРИТЕТНОЙ ОЧЕРЕДЬЮ В ПАМЯТИ ОДНОГО УРОВНЯ

**А. В. Соколов, А. В. Драц**

*Институт прикладных математических исследований Карельского научного центра РАН*

В работе предложены новые математические модели для представления приоритетной очереди в виде $n$ последовательных FIFO-очередей. На основе этих моделей предложены алгоритмы и комплекс программ, которые позволяют для заданных вероятностей выполнения основных операций находить оптимальный способ представления приоритетной очереди в смысле максимизации среднего времени до переполнения памяти и в смысле минимизации доли потерянных при переполнении элементов.

Ключевые слова: приоритетная очередь, FIFO-очереди, случайное блуждание, цепи Маркова, динамические структуры данных.

**A. V. Sokolov, A. V. Drats. OPTIMAL CONTROL OF PRIORITY QUEUE IN SINGLE LEVEL MEMORY**

The paper contains new mathematical models concerning representation of a priority queue in single level memory as $n$ serial FIFO queues. Proceeding from these models we propose the algorithms and programs which allow to find the optimal way of representation of a priority queue in the sense of maximizing the average time until memory overflow and in the sense of minimizing the proportion of lost elements when the probabilities are known.

Key words: Priority queue, FIFO-queues, random walks, Markov chains, dynamic data structures.

## Введение

Во многих приложениях используется структура данных, в которой основными операциями являются вставка элемента и удаление элемента с наибольшим приоритетом (ключом). Такую структуру данных называют приоритетной очередью. Основными методами реализации такой структуры данных являются упорядоченные и неупорядоченные списки, массивы, бинарные деревья, пирамиды ([Кормени др., 2000; Кнут, 2001; Боллапаргада и др., 2002]).

В этих и других работах различные методы представления приоритетной очереди сравниваются с точки зрения стоимости (времени) выполнения основных операций. Целью данной работы является изучение динамики изменения длины очереди в разных методах представления. При этом допускается возможность ее переполнения, что, например, в маршрутизаторах является обычной ситуацией. В работе будут рассмотрены представления в виде массива и в виде $n$ последовательных FIFO-очередей. Предлагаются новые математические модели в виде случайных блужданий по целочисленной $n$-мерной решетке, описывающие работу приоритетной очереди, представленной в виде $n$ последовательных FIFO-очередей. Для представления в виде массива использована предложенная ранее модель в виде одномерного случайного блуждания. На основе этих моделей разработаны программы, которые позволяют для заданных значений параметров проектируемой или настраиваемой программной или аппаратной системы выбирать оптимальный метод представления приоритетной очереди.

В случае представления в виде массива вместе с каждым элементом необходимо хранить его приоритет. В этом случае переполнение наступит только тогда, когда вся свободная память закончится, но часть памяти тратится на хранение приоритетов.

В случае представления приоритетной очереди в виде $n$ последовательных FIFO-очередей все элементы с одинаковыми приоритетами помещаются в одну FIFO-очередь. (Число FIFO-очередей равно количеству приоритетов в приоритетной очереди). Каждой FIFO-очереди выделен свой участок памяти. В этом случае память не тратится на хранение приоритетов, но могут возникнуть потери памяти тогда, когда одна из FIFO-очередей полностью исчерпает свою часть памяти, а остальные заполнят ее не полностью. Здесь возникает дополнительная задача оптимального разбиения памяти между FIFO-очередями.

Если использовать связанное представление очередей, т. е. хранить каждую FIFO-очередь не в виде массива, а в виде связного списка, то этот метод представления приоритетной очереди сводится к представлению в виде массива. Вместе с каждым элементом хранить нужно будет не его приоритет, а указатель на следующий элемент FIFO-очереди. Поэтому этот случай представления приоритетной очереди мы отдельно рассматривать не будем.

В качестве первого критерия оптимальности рассмотрим среднее время работы до переполнения $T$. Этот критерий применяется в том случае, если переполнение является аварийной ситуацией.

Переполнение памяти не всегда является аварийной ситуацией. Если допускаются потери элементов, то в случае попытки включения элемента, он просто отбрасывается, если нет свободной памяти для него. Так продолжается до тех пор, пока не появится свободная память, т. е. до исключения элемента. Такое поведение очереди называется «сбросом хвоста» [Боллапаргада и др., 2002]. Такая схема применяется, например, в работе сетевых маршрутизаторов. В этом случае разумно будет минимизировать долю времени, которую приоритетная очередь проводит в состоянии «сброса хвоста» $P^*$. Тем самым минимизируется доля потерянных элементов. Эта величина рассмотрена в качестве второго критерия. Будем считать, что такая схема работы применяется в течение длительного промежутка времени, поэтому время можно условно считать бесконечным.

## Постановка задачи

Рассмотрим очередь с $n$ приоритетами, расположенную памяти размера $m$ $(m > n)$. Будем считать, что время дискретно, и в каждый момент времени происходит одна из следующих операций:

- Вставить элемент с приоритетом $i$ с вероятностью $p_i$ $(1 \leqslant i \leqslant n)$;
- Удалить элемент с наибольшим приоритетом с вероятностью $q$;
- Найти элемент с наибольшим приоритетом с вероятностью $r$.

$p_1 + \cdots + p_n + q + r = 1$.

Вероятности не зависят от времени и количества элементов в очереди. Наивысший приоритет равен $n$, наименьший – 1. Работа начинается с пустой очереди. Предположим также, что не происходит завершения работы в случае исключения элемента из пустой структуры данных.

Требуется построить математические модели рассматриваемых методов представления приоритетной очереди, на их основе найти среднее время работы $T$ и долю времени $P^*$ и разработать программы выбора лучшего метода представления приоритетной очереди для данных критериев оптимальности.

## Критерий максимизации времени работы до переполнения. Представление в виде массива

Введем обозначения: $p = p_1 + \cdots + p_n$ – суммарная вероятность включения элемента в приоритетную очередь, $x = x_1 + \cdots + x_n$ – общее количество элементов, $l$ – отношение размера памяти, которую занимает ключ к размеру информационной части.

Будем считать, что информационная часть занимает одну единицу памяти. $M = \left\lfloor \dfrac{m}{1+l} \right\rfloor$ – размер памяти, который тратится на размещение информационных частей, он же равен общему количеству элементов, которые можно разместить в памяти. Например, если элемент массива объявлен как struct node{DataT info; DataKey key;}; sizeof(info)= 4*sizeof(key); sizeof(info) = 4 байт и размер памяти равен 40 байт, то $m = 10$ единиц памяти, $l = 1/4$ и $M = \frac{m}{1+l} = 8$ единиц памяти (единица памяти равна 4 байта). В работах [Феллер, 1964], [Тарасюк, 2008], [Аксенова, Соколов, 2008] в качестве математической модели было предложено случайное блуждание в интервале длины $M$. $x = -1$ – отражающий экран. $x = M + 1$ – поглощающий экран. Переходы между состояниями:

$$
\begin{aligned}
&(x) \xrightarrow{\ p\ } (x') = \begin{cases} (x+1) & 0 \leqslant x \leqslant M-1 \\ \text{Поглощающее состояние} & x = M \end{cases} \\
&(x) \xrightarrow{\ q\ } (x') = \begin{cases} (x) & x = 0 \\ (x-1) & 1 \leqslant x \leqslant M \end{cases} \\
&(x) \xrightarrow{\ r\ } (x') = (x) \quad 0 \leqslant x \leqslant M
\end{aligned}
\tag{1}
$$

Было показано, что среднее время работы до переполнения равно

$$
T = \begin{cases} \dfrac{-M-1}{q-p} + \dfrac{q}{(q-p)^2}\left(\left(\dfrac{q}{p}\right)^{M+1} - 1\right), & p \neq q \\ \dfrac{1}{2q}(M+1)(M+2), & p = q \end{cases}
\tag{2}
$$

## Критерий максимизации времени работы до переполнения. Представление в виде $n$ последовательных FIFO-очередей

Пусть $k_1, \ldots, k_n$ – фиксированное разбиение памяти между очередями. В качестве математической модели рассмотрим блуждание по целочисленному $n$-мерному параллелепипеду, ограниченному гиперплоскостями $x_i = 0$, $x_i = k_i$, $(1 \leqslant i \leqslant n)$. $x_i = -1$ $(1 \leqslant i \leqslant n)$ — отражающие барьеры. $x_i = k_i + 1$ — поглощающие барьеры. Количество состояний равно $\prod_{i=1}^{n}(k_i + 1)$.

Перенумеруем состояния в лексикографическом порядке следующим образом:

$$
\begin{array}{l}
(0,0,0,\ldots,0,0),(0,0,0,\ldots,0,1),\ldots,(0,0,0,\ldots,0,k_n-1),(0,0,0,\ldots,0,k_n) \\
(0,0,0,\ldots,1,0),(0,0,0,\ldots,1,1),\ldots,(0,0,0,\ldots,1,k_n-1),(0,0,0,\ldots,1,k_n) \\
\ldots \\
(0,0,0,\ldots,k_{n-1}-1,0),(0,0,0,\ldots,k_{n-1}-1,1),\ldots,(0,0,0,\ldots,k_{n-1}-1,k_n-1), \\
(0,0,0,\ldots,k_{n-1}-1,k_n) \\
(0,0,0,\ldots,k_{n-1},0),(0,0,0,\ldots,k_{n-1},1),\ldots,(0,0,0,\ldots,k_{n-1},k_n-1), \\
(0,0,0,\ldots,k_{n-1},k_n) \\
\ldots \\
(1,0,0,\ldots,0,0),(1,0,0,\ldots,0,1),\ldots,(1,0,0,\ldots,0,k_n-1),(1,0,0,\ldots,0,k_n) \\
\ldots \\
(k_1,0,0,\ldots,0,0),(k_1,0,0,\ldots,0,1),\ldots,(k_1,0,0,\ldots,0,k_n-1),(k_1,0,0,\ldots,0,k_n) \\
\ldots \\
(k_1,k_2,k_3,\ldots,k_{n-1},0),(k_1,k_2,k_3,\ldots,k_{n-1},1),\ldots,(k_1,k_2,k_3,\ldots,k_{n-1},k_n-1), \\
(k_1,k_2,k_3,\ldots,k_{n-1},k_n)
\end{array}
\tag{3}
$$

Пример нумерации состояний показан на рис. 1. Переход системы из состояния $(x_1, ..., x_n)$ в состояние $(x'_1, \ldots, x'_n)$ происходит в соответствии с правилами (4).

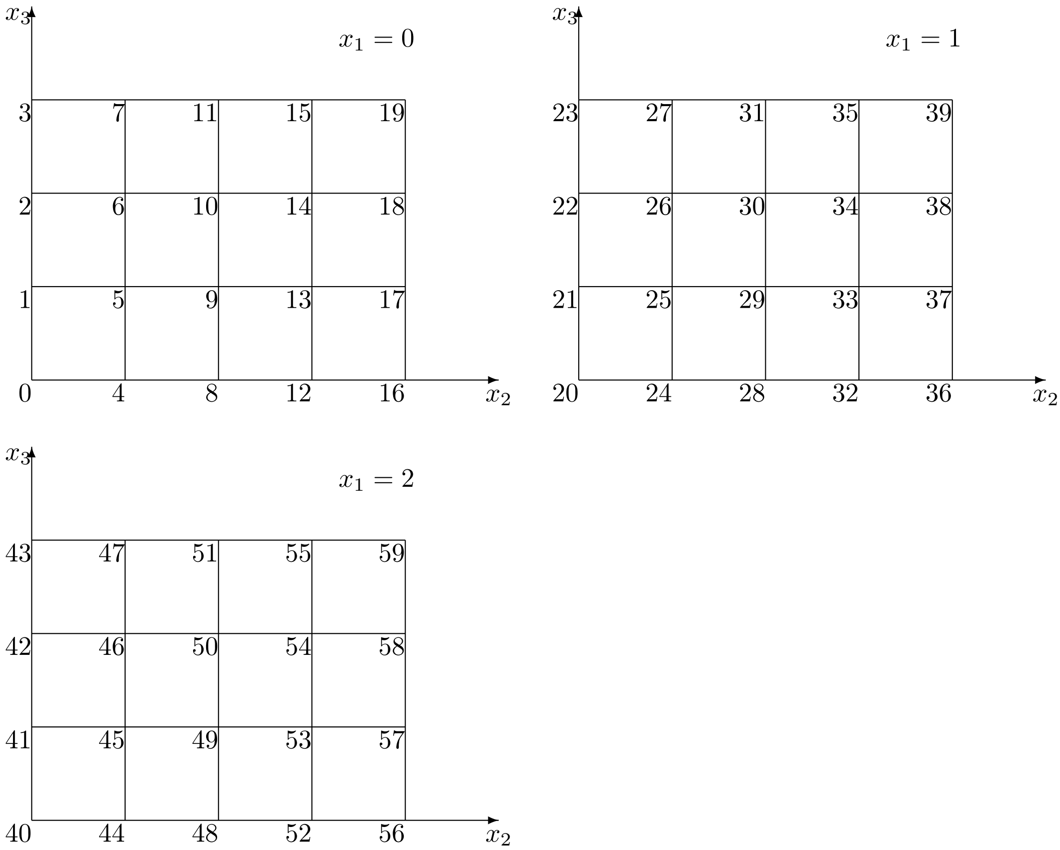



*Рис. 1.* Нумерация состояний при $m = 9$, $n = 3$, $k_1 = 2$, $k_2 = 4$, $k_3 = 3$

$$(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{p_i} \begin{cases} (\ldots, x_i + 1, \ldots, x_j, \ldots) & 0 \leqslant x_i < k_i \\ \text{Поглощающее состояние} & x_i = k_i \end{cases}$$

$$(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{q} \begin{cases} (\ldots, x_i, \ldots, x_j, \ldots) & x_1 = \cdots = x_n = 0 \\ (\ldots, x_i - 1, \ldots, x_j, \ldots) & 1 \leqslant x_i \leqslant k_i,\ x_{i+1} = \cdots = x_n = 0 \end{cases} \quad (4)$$

$$(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{r} (\ldots, x_i, \ldots, x_j, \ldots)$$

$$1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant n, \quad i \neq j.$$

Для описания процесса построим соответствующую поглощающую цепь Маркова. Необходимо построить матрицу вероятностей переходов $Q$. Для этого нужно знать номер каждого состояния. Для вычисления номера состояния построим функцию $F(x_1, x_2, \ldots, x_n) = I$, где $I$ – номер состояния.

Если есть только одна очередь, то функция имеет вид: $F(x_1) = x_1$. Если появляется вторая очередь, то состояния $(x_i)$ становятся эквивалентными $(x_i, 0)$. Между состояниями $(x_i, 0)$ и $(x_{i+1}, 0)$ нужно вставить $k_2 - 1$ состояний: $(x_i, 1), (x_i, 2), \ldots, (x_i, k_{n-1} - 1)$, поэтому функция будет иметь вид $F(x_1, x_2) = k_2 x_1 + x_2$.

При появлении третьей очереди получаем функцию
$F(x_1, x_2, x_3) = k_3(k_2 x_1 + x_2) + x_3$.

Рассуждая таким образом, получим для $n$ очередей функцию
$F(x_1, x_2, \ldots, x_n) = k_n(k_{n-1}(\ldots(k_3(k_2 x_1 + x_2) + x_3) + \cdots + x_{n-1}) + x_n$.

Для построения матрицы переходных вероятностей нужно в цикле перебрать все состояния (от $(0, 0, \ldots, 0, 0)$ до $(0, 0, \ldots, 0, M)$) и вычислить вероятности перехода из $I$-того состояния в остальные ($I = F(x_1, x_2 \ldots, x_n)$):

$$
\begin{array}{l}
Q[F(x_1,x_2,\ldots,x_n)][F(x_1+1,x_2,\ldots,x_n)]=p_1\\
\ldots\\
Q[F(x_1,x_2,\ldots,x_n)][F(x_1,x_2,\ldots,x_n+1)]=p_n\\
Q[F(x_1,\ldots,x_i,\ldots,x_n)][F(x_1,\ldots,x_i-1,\ldots,x_n)]=q,\ \text{где } i=\max\limits_{1\leqslant j\leqslant n}\{j:x_j>0\}\\
Q[F(x_1,x_2,\ldots,x_n)][F(x_1,x_2,\ldots,x_n)]=r
\end{array}
\tag{5}
$$

Для вычисления среднего времени блуждания нужно вычислить фундаментальную матрицу $N=(E-Q)^{-1}$ [Кемени, Снелл, 1970]. $N_{ij}$ – среднее время, которое процесс проведет в состоянии $j$, если он начался в состоянии $i$. Поскольку работа начинается с пустых структур данных, то $i=0$ и $T=\sum\limits_j N_{0j}$.

Чтобы найти оптимальное разбиение памяти, необходимо перебрать все возможные разбиения памяти, для каждого из них найти среднее время работы до переполнения и выбрать наибольшее время. Всего для размера памяти $m$ и $n$ очередей существует $C_{m-1}^{n-1}=\frac{(m-1)!}{(n-1)!(m-n)!}$ способов разбиения [Риордан, 1963].

## Управление на бесконечном времени. Представление в виде массива

Как и в предыдущем критерии часть памяти тратится на хранение приоритетов. Всего можно хранить $M=\left\lfloor\dfrac{m}{l+1}\right\rfloor$ элементов в памяти. В качестве математической модели рассмотрим блуждание в одномерном интервале длины $M$. Переход из состояния $(x)$ в состояние $(x')$ происходит в соответствии со следующими правилами:

$$
(x)\ \xrightarrow{\quad p\quad}\ (x')=\begin{cases}(x+1) & 0\leqslant x\leqslant M\\ (x) & x=M+1\end{cases}
$$

$$
(x)\ \xrightarrow{\quad q\quad}\ (x')=\begin{cases}(x) & x=0\\ (x-1) & 1\leqslant x\leqslant M\\ (x-2) & x=M+1\end{cases}
$$

$$
(x)\ \xrightarrow{\quad r\quad}\ (x')=\begin{cases}(x) & 0\leqslant x\leqslant M\\ (x-1) & x=M+1\end{cases}
$$

Задача сводится к вычислению доли времени, которую проводит одна FIFO-очередь в состоянии «сброса хвоста». Эта задача была решена в работе [Аксенова и др., 2009]:

$$
P^*=\begin{cases}\dfrac{q-p}{\left(\dfrac{q}{p}\right)^{M+1}-1} & q\neq p\\[2ex] \dfrac{p}{M+1} & q=p\end{cases}
\tag{6}
$$

## Управление на бесконечном времени. Представление в виде $n$ последовательных FIFO-очередей

Пусть $k_1,\ldots,k_n$ – фиксированное разбиение памяти между очередями. В качестве математической модели рассмотрим блуждание по целочисленному $n$-мерному параллелепипеду, ограниченному гиперплоскостями $x_i=0$, $x_i=k_i$, $(1\leqslant i\leqslant n)$. Дополнительно к каждому состоянию на гиперплоскостях $x_j=k_j$ введем состояние «сброса хвоста». Количество обычных состояний равно $\prod_{i=1}^{n}(k_i+1)$.

Количество состояний «сброса хвоста», соответствующих $j$-й очереди равно

$$
\prod_{\substack{i=1\\ i\neq j}}^{n}(k_i+1).
$$

Общее количество состояний «сброса хвоста» равно:

$$
\sum_{j=1}^{n}\left(\prod_{\substack{i=1\\ i\neq j}}^{n}(k_i+1)\right).
$$

Перенумеруем сначала обычные состояния в лексикографическом порядке, как в предыдущем параграфе, затем перенумеруем состояния «сброса хвоста», соответствующие гиперплоскости $x_1=k_1$, затем соответствующие гиперплоскости $x_2=k_2$ и т. д. На каждой гиперплоскости $x_i=k_i$ перенумеруем состояния в лексикографическом порядке, по аналогии с обычными состояниями. Пример нумерации состояний показан на рис. 2. Состояния с номерами больше 35 являются состояниями «сброса хвоста».

Перенумеруем состояния в лексикографическом порядке следующим образом:

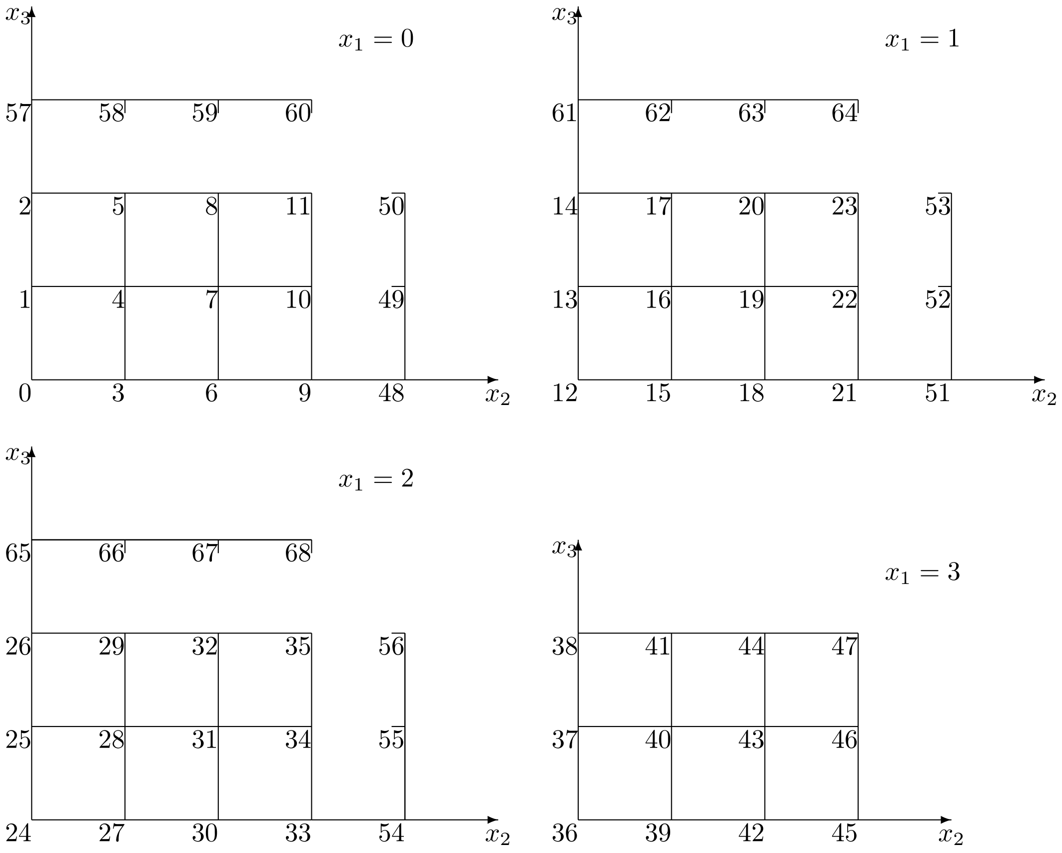



*Рис. 2.* Нумерация состояний при $m = 7$, $n = 3$, $k_1 = 2$, $k_2 = 3$, $k_3 = 2$

Переход системы из состояния $(x_1, \ldots, x_n)$ в состояние $(x'_1, \ldots, x'_n)$ происходит в соответствии с правилами (7).

Для описания процесса построим соответствующую регулярную цепь Маркова. Построим матрицу вероятностей переходов $Q$. Для обычных состояний ее вид будет такой же, как и в предыдущем параграфе. Для состояния «сброса хвоста» на плоскости $x_i = k_i$ матрица вероятностей переходов принимает вид (8).

$$
\begin{gathered}
(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{p_i} \begin{cases}
(\ldots, x_i + 1, \ldots, x_j, \ldots) & 0 \leqslant x_i \leqslant k_i, x_j \leqslant k_j \\
(\ldots, x_i + 1, \ldots, x_j - 1, \ldots) & 0 \leqslant x_i \leqslant k_i, x_j = k_j + 1 \\
(\ldots, x_i, \ldots, x_j, \ldots) & x_i = k_i + 1, x_j \leqslant k_j \\
(\ldots, x_i, \ldots, x_j - 1, \ldots) & x_i = k_i + 1, x_j = k_j + 1
\end{cases} \\
(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{q} \begin{cases}
(\ldots, x_i, \ldots, x_j, \ldots) & x_1 = \cdots = x_n = 0 \\
(\ldots, x_i, \ldots, x_j - 1, \ldots) & 0 \leqslant x_i < k_i,\ x_j > 0,\ x_{j+1} = \cdots = x_n = 0 \\
(\ldots, x_i - 1, \ldots, x_j - 1, \ldots) & x_i = k_i + 1,\ x_j > 0,\ x_{j+1} = \cdots = x_n = 0 \\
(\ldots, x_i, \ldots, x_j - 2, \ldots) & 0 \leqslant x_i < k_i,\ x_j = k_j + 1,\ x_{j+1} = \cdots = x_n = 0 \\
(\ldots, x_i - 1, \ldots, x_j - 2, \ldots) & x_i = k_i + 1,\ x_j = k_j + 1,\ x_{j+1} = \cdots = x_n = 0
\end{cases} \\
(\ldots, x_i, \ldots, x_j, \ldots) \xrightarrow{r} \begin{cases}
(\ldots, x_i, \ldots, x_j, \ldots) & 0 \leqslant x_i \leqslant k_i, x_j \leqslant k_j \\
(\ldots, x_i, \ldots, x_j - 1, \ldots) & 0 \leqslant x_i \leqslant k_i, x_j = k_j + 1
\end{cases}
\end{gathered}
\tag{7}
$$

$$
\begin{gathered}
F(x_1, \ldots, x_{i-1}, k_i + 1, x_{i-1}, \ldots, x_n) = \prod_{\substack{i=1 \\ i \neq j}}^{n} (k_i + 1) + \sum_{j=1}^{i-1} \left( \prod_{\substack{l=1 \\ l \neq j}}^{n} (k_l + 1) \right) \\
+ k_n(k_{n-1} \ldots k_{i+1}(k_{i-1}(\ldots(k_3(k_2 x_1 + x_2) + x_3) + \cdots + x_{i-1}) + x_{i+1} + \cdots + x_{n-1}) + x_n
\end{gathered}
\tag{8}
$$

*Таблица 1.* Среднее время работы до переполнения $T$, $n = 4$, $m = 24$

| $r$ | 0 | 0 | 0 | 0 | 0.5 | 0 | 0 | 0.1 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| $p_1$ | 0,15 | 0,2 | 0,05 | 0,2 | 0,1 | 0,4 | 0,05 | 0,18 | 0,1 |
| $p_2$ | 0,15 | 0,15 | 0,1 | 0,2 | 0,1 | 0,1 | 0,05 | 0,1 | 0,1 |
| $p_3$ | 0,15 | 0,1 | 0,15 | 0,2 | 0,1 | 0,05 | 0,1 | 0,1 | 0,1 |
| $p_4$ | 0,15 | 0,05 | 0,2 | 0,2 | 0,1 | 0,05 | 0,4 | 0,02 | 0,1 |
| $q$ | 0,4 | 0,5 | 0,5 | 0,2 | 0,1 | 0,4 | 0,4 | 0,5 | 0,6 |
| $k_1$ | 9 | 13 | 6 | 7 | 7 | 17 | 5 | 13 | 9 |
| $k_2$ | 6 | 6 | 7 | 7 | 7 | 3 | 4 | 5 | 6 |
| $k_3$ | 5 | 3 | 6 | 6 | 6 | 2 | 6 | 4 | 5 |
| $k_4$ | 4 | 2 | 5 | 4 | 4 | 2 | 9 | 2 | 4 |
| $T$ | 64,3 | 234,98 | 200,46 | 27,32 | 54,65 | 78,32 | 68,37 | 1165,8 | 2868,62 |
| $l_1$ | 55,05 | 182 | 182 | 21,11 | 42,22 | 55,05 | 55,05 | 729,5 | 112102,41 |
| $l_2$ | 75,01 | 306 | 306 | 27,78 | 55,56 | 75,01 | 75,01 | 2000,45 | 49763,85 |
| $l_4$ | 90 | 420 | 420 | 32,78 | 65,56 | 90 | 90 | 4086,81 | 14678,92 |
| $l_8$ | 100 | 506 | 506 | 36,11 | 72,22 | 100 | 100 | 6506,26 | 2839,29 |

*Таблица 2.* Доля времени, которую приоритетная очередь проводит в состоянии «сброса хвоста» $T$, $n = 4$, $m = 16$

| $r$ | 0 | 0 | 0 | 0 | $0,5$ | 0 | 0 | $0,1$ | 0 |
|---|---|---|---|---|---|---|---|---|---|
| $p_1$ | 0,15 | 0,2 | 0,05 | 0,2 | 0,1 | 0,4 | 0,05 | 0,18 | 0,1 |
| $p_2$ | 0,15 | 0,15 | 0,1 | 0,2 | 0,1 | 0,1 | 0,05 | 0,1 | 0,1 |
| $p_3$ | 0,15 | 0,1 | 0,15 | 0,2 | 0,1 | 0,05 | 0,1 | 0,1 | 0,1 |
| $p_4$ | 0,15 | 0,05 | 0,2 | 0,2 | 0,1 | 0,05 | 0,4 | 0,02 | 0,1 |
| $q$ | 0,4 | 0,5 | 0,5 | 0,2 | 0,1 | 0,4 | 0,4 | 0,5 | 0,6 |
| $k_1$ | 9 | 10 | 4 | 12 | 12 | 13 | 7 | 9 | 6 |
| $k_2$ | 3 | 3 | 5 | 2 | 1 | 1 | 2 | 3 | 4 |
| $k_3$ | 3 | 2 | 4 | 1 | 1 | 1 | 3 | 3 | 3 |
| $k_4$ | 2 | 1 | 3 | 1 | 2 | 1 | 4 | 1 | 3 |
| $P^*$ | 0,2011 | 0,0433 | 0,0456 | 0,6 | 0,3 | 0,2 | 0,2009 | 0,01 | 0,0041 |
| $l_1$ | 0,20534 | 0,0556 | 0,0556 | 0,6 | 0,3 | 0,1155 | 0,1155 | 0,0155 | 0,00543 |
| $l_2$ | 0,20234 | 0,0455 | 0,0455 | 0,6 | 0,3 | 0,1094 | 0,1094 | 0,0094 | 0,00234 |
| $l_4$ | 0,201 | 0,0385 | 0,0385 | 0,6 | 0,3 | 0,1058 | 0,1058 | 0,00582 | 0,00103 |
| $l_8$ | 0,2005 | 0,0333 | 0,0333 | 0,6 | 0,3 | 0,1036 | 0,1036 | 0,00365 | 0,00046 |

У построенной цепи Маркова существует предельный вектор $\alpha$, который удовлетворяет уравнению $\alpha * Q = \alpha$ [Кемени, Снелл, 1970]. По закону больших чисел значение $\alpha_i$ является долей времени, которое процесс проводит в состоянии $i$. Чтобы найти долю времени, которую система проводит в состоянии «сброса хвоста», необходимо вычислить сумму тех компонент вектора $\alpha$, которые соответствуют состояниям «сброса хвоста». В примере на рис. 2 необходимо будет просуммировать компоненты с номерами от 36 до 68.

## Численные результаты

Для решения задачи были написаны программы, которые вычисляют матрицу $Q$, среднее время работы до переполнения и долю времени, которую система проводит в состоянии «сброса хвоста» и находят оптимальное разбиение памяти между очередями в случае представления приоритетной очереди в виде последовательных FIFO-очередей. Некоторые результаты работы программ представлены в таблицах 1 и 2. В строке $T$ находится среднее время работы до переполнения в случае представления в виде $n$ FIFO-очередей, в строках $k_i$ $(1 \leqslant i \leqslant 4)$ – оптимальное разбиение памяти. В строке $P^*$ находится доля времени, которую система проводит в состоянии «сброса хвоста» в случае представления в виде $n$ FIFO-очередей. В строке $l_1$ указано время работы в случае представления в виде массива, когда размер ключа равен размеру информационной части. (На ключи тратится половина памяти.)
В строке $l_2$ – когда размер ключа равен 1/2 размеру информационной части. (На ключи тратится 1/3 часть памяти.)
В строке $l_4$ – когда размер ключа равен 1/4 размеру информационной части. (На ключи тратится 1/5 часть памяти.)
В строке $l_8$ – когда размер ключа равен 1/8 размеру информационной части. (На ключи тратится 1/9 часть памяти.)

Предложенные в статье модели и разработанные на их основе программы могут дать для известных значений вероятностей выполнения операций оптимальные размеры очередей разных приоритетов, которые системный администратор может учитывать при настройке маршрутизатора.

Работа выполнена при финансовой поддержке РФФИ, грант 09-01-00330.

## Литература

*Аксенова Е. А., Драц А. В., Соколов А. В.* Оптимальное управление $n$ FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2009. № 6. С. 46–54.

*Аксенова Е. А., Соколов А. В.* Анализ некоторых методов реализации приоритетной очереди // Межвуз. сб. «Стохастическая оптимизация в информатике» Изд-во С.-Петербургского университета, 2008. Вып. 4. С. 61–71.

*Боллапаргада В., Мэрфи К., Расс У.* Структура операционной системы Cisco IOS. М.: Вильямс, 2002. 208 с.

*Кемени Дж., Снелл Дж.* Конечные цепи Маркова. М.: Наука, 1970. 272 с.

*Кнут Д.* Искусство программирования для ЭВМ. М.: Вильямс, 2001. Т. 3. 434 с.

*Кормен Е., Лейзерсон Ч., Ривест Р.* Алгоритмы: построение и анализ. М.: МЦНМО, 2000. 960 с.

*Риордан Дж.* Введение в комбинаторный анализ. М.: Издательство иностранной литературы, 1963. 375 с.

*Седжвик Р.* Фундаментальные алгоритмы на C++. Анализ. Структуры данных. Сортировка. Поиск пер. с англ. Роберт Седжвик. К.: Издательство «ДиаСофт», 2001. 688 с.

*Тарасюк А. В.* Оптимальная реализация N FIFO-очередей в памяти одного уровня в случае одновременного выполнения операций // Системы управления и информационные технологии. 2008. № 1 (31). С. 75–78.

*Феллер В.* Введение в теорию вероятностей и ее приложения. М.: Мир, 1964. 499 с.

## СВЕДЕНИЯ ОБ АВТОРАХ:

**Соколов Андрей Владимирович**
д. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: avs@krc.karelia.ru
тел.: (8142) 766313

**Sokolov, Andrey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: avs@krc.karelia.ru

**Драц Андрей Владимирович**
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: adeon88@mail.ru
тел.: (8142) 763313

**Drats, Andrey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: adeon88@mail.ru

Труды Карельского научного центра РАН
№ 5. 2011. С. 111–114

*УДК 519.23*

# ОБ УСЛОВИЯХ ГЛОБАЛЬНОЙ ИДЕНТИФИЦИРУЕМОСТИ ДЛЯ МОДЕЛЕЙ ФАКТОРНОГО АНАЛИЗА

**С. В. Стафеев**

*Институт прикладных математических исследований Карельского научного центра РАН*

В статье рассматривается модель факторного анализа с зависимыми остатками. Получены условия глобальной идентифицируемости и разработан метод проверки этих условий.

К л ю ч е в ы е  с л о в а: факторный анализ, идентифицируемость, инварианты.

## S. V. Stafeev. ON GLOBAL IDENTIFIABILITY CONDITIONS OF FACTOR ANALYSIS MODELS

A factor analysis model with dependent residuals is considered. Global identifiability conditions were obtained and a method for testing these conditions was developed.

K e y  w o r d s: factor analysis, identifiability, invariants.

---

## Введение

В статье рассматривается модель факторного анализа с зависимыми остатками [Стафеев, 2007]. В отличие от стандартной модели факторного анализа [Harman, 1976], в данной модели допускаются зависимости между остатками. Одной из основных проблем, возникающих при работе с моделями, содержащими латентные (скрытые) переменные, является проблема параметрической идентифицируемости, состоящей в ответе на вопрос о возможности однозначного восстановления параметров модели по матрице ковариаций наблюдаемых случайных величин. Отсутствие идентифицируемости приводит к невозможности состоятельного оценивания параметров модели. Рассматриваемая модель является частным случаем систем структурных уравнений с латентными переменными. Заметим, что в работе [Drton et al., 2011] были получены условия глобальной идентифицируемости для систем структурных уравнений, не содержащих латентные переменные. Ранее для различных моделей факторного анализа с зависимыми остатками были получены условия почти всюду (п.в.) локальной и глобальной идентифицируемости [Стафеев, 2005, 2007; Stafeev, 2007]. В данной работе получены достаточные условия глобальной идентифицируемости рассматриваемой модели, и разработан метод проверки этих условий по имеющимся данным.

## Модель факторного анализа

Рассмотрим модель факторного анализа:

$$\mathbf{X} = A\mathbf{H} + \mathbf{Y}, \tag{1}$$

где $\mathbf{X} = \{X_1, ..., X_n\}^t$ является вектором с матрицей ковариаций $\Sigma = (\sigma_{ij})$; $\mathbf{H} = \{H_1, ..., H_k\}^t$ есть множество независимых нормально распределенных латентных (скрытых) случайных величин с $\mathbf{M}(H_j) = 0$ и $\mathbf{Var}(H_j) = 1$,

$j = 1, ..., k$; $A = (a_{ij})$ – матрица факторных нагрузок. Вектор остатков $\mathbf{Y} = \{Y_1, ..., Y_n\}^t$ имеет нормальное распределение с нулевым вектором математических ожиданий и матрицей ковариаций $\Theta_G = (\theta_{ij})$. Векторы $\mathbf{Y}$ и $\mathbf{H}$ являются независимыми. Взаимосвязь между компонентами вектора $\mathbf{Y}$ представляется в виде ковариационной графовой модели со структурой $G = (V, E)$, где $V = \{1, .., n\}$ и $(i, j) \notin E$, если $\theta_{ij} = 0$.

Параметры модели (1) состоят из матрицы факторных нагрузок $A$ и ненулевых элементов матрицы $\Theta_G$.

Модель (1) называется глобально идентифицируемой, если по матрице $\Sigma$ матрица $A$ определяется с точностью до ортогонального вращения, а элементы матрицы $\Theta_G$ определяются однозначно.

Матрица ковариаций наблюдаемых случайных $\Sigma$ величин допускает следующее представление:

$$\Sigma = AA^t + \Theta_G. \qquad (2)$$

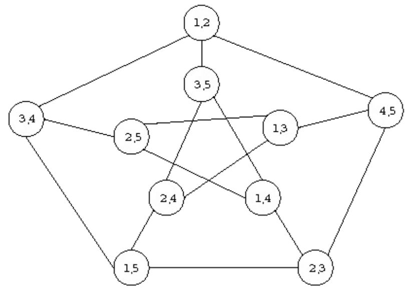
Пусть $P_G$ – множество всех неотрицательно определенных $n \times n$ матриц, допускающих разложение (2). Наша задача состоит в нахождении множества $P'_G \subset P_G$ ковариационных матриц, которые соответсвуют глобально идентифицируемым моделям. Задача будет решаться путем наложения ограничений на элементы матрицы $\Sigma$ и граф $G$.

## Условия глобальной идентифицируемости

Пусть $\overline{G} = (V, \overline{E})$ – дополнительный для $G$ граф. Определим граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, состоящий из множества вершин $\mathbf{V} = \{\mathbf{i} = (i_1, ..., i_k), 1 \leqslant i_1 < i_2 < ... < i_k \leqslant k\}$ и множества ребер $\mathbf{E} = \{(\mathbf{i}, \mathbf{j}) | \mathbf{i} = (i_1, ..., i_k), \mathbf{j} = (j_1, ..., j_k), (i_s, j_l) \in \overline{E}, s, l = 1, ...k\}$. (Впервые граф $\mathbf{G}$ был определен в работе [Стафеев, 2007] в контексте п.в. локальной идентифицируемости. Граф $\mathbf{G}$ также играет ключевую роль при нахождении полиномиальных инвариантов модели 1 [Стафеев, 2010; Stafeev, 2010].) На рисунке представлен пример графа $\mathbf{G}$ для $k = 2$, $n = 5$ и некоррелированых остатков.

Каждому ребру графа $\mathbf{G}$ сопоставим в соответсвие детерминант $|\Sigma_{\mathbf{i},\mathbf{j}}|$ матрицы $\Sigma_{\mathbf{i},\mathbf{j}} = (\sigma_{i_m, j_l})_{m,l=1}^k$. Данный $k \times k$ минор матрицы $\Sigma$ будем называть весом ребра $(\mathbf{i}, \mathbf{j})$.

В следующей теореме получены условия глобальной идентифицируемости модели (1).




Граф $\mathbf{G}$ для $k = 2$ и $n = 5$

**Теорема 1.** *Модель (1) является глобально идентифицируемой, если граф, полученный из графа $\mathbf{G}$ с помощью удаления ребер, соответствующих нулевым весам, содержит компоненту связности $\mathbf{G}' = (\mathbf{V}', \mathbf{E}')$ с нечетным простым циклом и $\cup_{\mathbf{i} \in \mathbf{V}'} \mathbf{i} = V$.*

*Доказательство.* Введем обозначения: $\mathbf{a_i} = \{a_{i1}, ..., a_{ik}\}^t$, $A_{\mathbf{i}} = \{\mathbf{a}_{\mathbf{i}_1}, ..., \mathbf{a}_{\mathbf{i}_\mathbf{k}}\}$ и $\Psi = AA^t$.

Очевидно, что $\Psi_{\mathbf{i},\mathbf{j}} = \Sigma_{\mathbf{i},\mathbf{j}}$ для $(\mathbf{i}, \mathbf{j}) \in \mathbf{E}$. В работе [Стафеев, 2010] показано, что

$$\Sigma_{\mathbf{i},\mathbf{j}} = \Psi_{\mathbf{i},\mathbf{j}} = A_{\mathbf{i}}(A_{\mathbf{j}})^t, (\mathbf{i}, \mathbf{j}) \in \mathbf{E}. \qquad (3)$$

Так как граф $\mathbf{G}'$ не содержит ребер с нулевым весом, то все матрицы из набора

$$\Psi_{\mathbf{i},\mathbf{j}}, (\mathbf{i}, \mathbf{j}) \in \mathbf{E}' \qquad (4)$$

являются обратимыми. Докажем, что, используя множество (4), можно однозначно определить множество матриц

$$\Psi_{\mathbf{i},\mathbf{j}}, \mathbf{i}, \mathbf{j} \in \mathbf{V}'. \qquad (5)$$

Предположим, что граф $\mathbf{G}'$ содержит простую цепь с множеством вершин $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3, \mathbf{i}_4\}$ и множеством ребер $\{(\mathbf{i}_1, \mathbf{i}_2), (\mathbf{i}_2, \mathbf{i}_3), (\mathbf{i}_3, \mathbf{i}_4)\}$. Используя (3), получаем

$$\Psi_{\mathbf{i}_1,\mathbf{i}_2}(\Psi_{\mathbf{i}_3,\mathbf{i}_2})^{-1}\Psi_{\mathbf{i}_3,\mathbf{i}_4}$$
$$= A_{\mathbf{i}_1}(A_{\mathbf{i}_2})^t[A_{\mathbf{i}_3}(A_{\mathbf{i}_2})^t]^{-1}A_{\mathbf{i}_3}(A_{\mathbf{i}_4})^t = A_{\mathbf{i}_1}(A_{\mathbf{i}_4})^t.$$

Таким образом, получаем, что

$$\Psi_{\mathbf{i}_1,\mathbf{i}_2}(\Psi_{\mathbf{i}_3,\mathbf{i}_2})^{-1}\Psi_{\mathbf{i}_3,\mathbf{i}_4} = \Psi_{\mathbf{i}_1,\mathbf{i}_4}. \qquad (6)$$

Теперь предположим, что граф $\mathbf{G}'$ содержит простой нечетный цикл с множеством вершин $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$ и множеством ребер $\{(\mathbf{i}_1, \mathbf{i}_2), (\mathbf{i}_2, \mathbf{i}_3), (\mathbf{i}_1, \mathbf{i}_3)\}$. Снова используя (3), получаем

$$\Psi_{\mathbf{i}_1,\mathbf{j}_2}(\Psi_{\mathbf{i}_3,\mathbf{j}_2})^{-1}\Psi_{\mathbf{i}_3,\mathbf{j}_1}$$
$$= A_{\mathbf{i}_1}(A_{\mathbf{i}_2})^t[A_{\mathbf{i}_3}(A_{\mathbf{i}_2})^t]^{-1}A_{\mathbf{i}_3}(A_{\mathbf{i}_1})^t = A_{\mathbf{i}_1}(A_{\mathbf{i}_1})^t.$$

Отсюда следует, что

$$\Psi_{\mathbf{i}_1,\mathbf{i}_2}(\Psi_{\mathbf{i}_3,\mathbf{i}_2})^{-1}\Psi_{\mathbf{i}_3,\mathbf{i}_1} = \Psi_{\mathbf{i}_1,\mathbf{i}_1}. \qquad (7)$$

Соотношения (6) и (7) допускают следующие обобщения. Пусть граф $\mathbf{G}'$ содержит простую цепь с четным числом вершин $\{\mathbf{i}_1, ..., \mathbf{i}_l\}$ и множеством ребер $\{(\mathbf{i}_1, \mathbf{i}_2), ..., (\mathbf{i}_{l-1}, \mathbf{i}_l)\}$. Получаем:

$$\Psi_{\mathbf{i}_1,\mathbf{i}_l} = [\prod_{s=1}^{s=l-3} \Psi_{\mathbf{i}_s,\mathbf{i}_{s+1}}(\Psi_{\mathbf{i}_{s+2},\mathbf{i}_{s+1}})^{-1}]\Psi_{\mathbf{i}_{l-1},\mathbf{i}_l}. \quad (8)$$

Теперь пусть граф $\mathbf{G}'$ содержит простой нечетный цикл с множеством вершин $\{\mathbf{i}_1, ..., \mathbf{i}_l\}$ и множеством ребер $\{(\mathbf{i}_1, \mathbf{i}_2), ..., (\mathbf{i}_l, \mathbf{i}_l)\}$. В этом случае получаем:

$$\Psi_{\mathbf{i}_1,\mathbf{i}_1} = [\prod_{s=1}^{s=l-2} \Psi_{\mathbf{i}_s,\mathbf{i}_{s+1}}(\Psi_{\mathbf{i}_{s+2},\mathbf{i}_{s+1}})^{-1}]\Psi_{\mathbf{i}_l,\mathbf{i}_1}. \quad (9)$$

По условиям теоремы граф $\mathbf{G}'$ содержит нечетный цикл. Пусть $\mathbf{V}_1$ – множество вершин этого цикла и пусть $\mathbf{V}_2 = \mathbf{V}' \setminus \mathbf{V}_1$. Неизвестные матрицы множества (5) будем определять в три этапа. Очевидно, что в графе $\mathbf{G}'$ любые две различные вершины $\mathbf{i}$ и $\mathbf{j}$, $\mathbf{i}, \mathbf{j} \in \mathbf{V}_1$, $(\mathbf{i}, \mathbf{j}) \notin \mathbf{E}'$, соединены простой цепью с четным числом вершин. Таким образом, используя (8), по набору (4) однозначно определяем множество матриц $\{\Psi_{\mathbf{ij}}, \mathbf{i} \neq \mathbf{j}, \mathbf{i}, \mathbf{j} \in \mathbf{V}_1\}$.

Образуем граф $\mathbf{G}_1 = (\mathbf{V}', \mathbf{E}_1)$, где $\mathbf{E}_1 = \mathbf{E}' \cup \{(\mathbf{i}, \mathbf{j}), \mathbf{i} \neq \mathbf{j}, \mathbf{i}, \mathbf{j} \in \mathbf{V}_1\}$. В графе $\mathbf{G}_1$ любые две вершины $\mathbf{i} \in \mathbf{V}_1$ и $\mathbf{j} \in \mathbf{V}_2$, $(\mathbf{i}, \mathbf{j}) \notin \mathbf{E}_1$, соединены простой цепью с четным числом вершин. Таким образом, мы однозначно определяем матрицы $\{\Psi_{\mathbf{ij}}, \mathbf{i} \neq \mathbf{j}, \mathbf{i} \in \mathbf{V}_1, \mathbf{j} \in \mathbf{V}_2\}$.

Теперь образуем граф $\mathbf{G}_2 = (\mathbf{V}', \mathbf{E}_2)$, где $\mathbf{E}_2 = \mathbf{E}_1 \cup \{(\mathbf{i}, \mathbf{j}), \mathbf{i} \neq \mathbf{j}, \mathbf{i}\mathbf{V}_1, \mathbf{j} \in \mathbf{V}_2\}$. По аналогии с предыдущими случаями, в графе $\mathbf{G}_2$ для любых двух различных вершин $\mathbf{i}$ и $\mathbf{j}$, $\mathbf{i}, \mathbf{j} \in \mathbf{V}_2$, $(\mathbf{i}, \mathbf{j}) \notin \mathbf{E}_2$ мы определяем $\Psi_{\mathbf{ij}}$.

Образуем граф $\mathbf{G}_3 = (\mathbf{V}', \mathbf{E}_3)$, где $\mathbf{E}_3 = \mathbf{E}_2 \cup \{(\mathbf{i}, \mathbf{j}), \mathbf{i} \neq \mathbf{j}, \mathbf{i}, \mathbf{j} \in \mathbf{V}_2\}$. Легко видеть, что граф $\mathbf{G}_3$ полный. Теперь, используя (10), однозначно определяем множество матриц $\{\Psi_{\mathbf{ii}}, \mathbf{i} \in \mathbf{V}'\}$.

Используя набор матриц (4), мы нашли все элементы матриц из набора (5). Ввиду того, что по условию $\cup_{\mathbf{i} \in \mathbf{V}'} \mathbf{i} = V$, получаем, что мы однозначно определи все элементы матрицы $\Psi$. Матрица $\Psi$ имеет ранг равный $k$, поэтому матрица $A$ определяется с точностью до ортогонального вращения. Зная матрицу $\Psi$, мы однозначно определяем матрицу $\Theta = \Sigma - \Psi$. Теорема доказана.

Верно следующее утверждение.

**Утверждение 1.** *Пусть* $\mathbf{C} = (\mathbf{V_C}, \mathbf{E_C})$ *простой цикл графа* $G$. *Тогда*

$$\prod_{(\mathbf{i},\mathbf{j}) \in \mathbf{E_C}} |\Sigma_{\mathbf{i},\mathbf{j}}| \geqslant 0. \quad (10)$$

*Доказательство.* Пусть $\mathbf{V_C} = \{\mathbf{i}_1, ..., \mathbf{i}_l\}$ и $\mathbf{E_C} = \{(\mathbf{i}_1, \mathbf{i}_2), ..., (\mathbf{i}_l, \mathbf{i}_l)\}$. Используя (4), получаем

$$\prod_{(\mathbf{i},\mathbf{j}) \in \mathbf{E_C}} |\Sigma_{\mathbf{i},\mathbf{j}}| = |\Sigma_{\mathbf{i}_1,\mathbf{i}_s}| \prod_{s=1}^{l-1} |\Sigma_{\mathbf{i}_s,\mathbf{i}_{s+1}}|$$

$$= |A_{\mathbf{i}_1}(A_{\mathbf{i}_s})^t| \prod_{s=1}^{l-1} |A_{\mathbf{i}_s}(A_{\mathbf{i}_{s+1}})^t|$$

$$= \prod_{s=1}^{l} |A_{\mathbf{i}_s}(A_{\mathbf{i}_s})^t| \geqslant 0.$$

При доказательстве мы воспользовались тем, что матрица $A_{\mathbf{i}_s}(A_{\mathbf{i}_s})^t$ является неотрицательно определенной.

Утверждение 1 допускает следующее обобщение. Пусть $\mathbf{V}' = \{\mathbf{i}_1, ..., \mathbf{i}_L\}$. Определим элементы $L \times L$ матрицы $D$ следующим образом:

$$d_{fm} = \begin{cases} sign|\Sigma_{\mathbf{i}_f,\mathbf{i}_m}|, & \text{если } (\mathbf{i}_f, \mathbf{i}_m) \in \mathbf{E}'; \\ 0, & \text{в противном случае.} \end{cases}$$

**Утверждение 2.** *Для матрицы* $D$ *найдется такая диагональная (с 1 и -1 на диагонали) матрица* $U = diag\{u_1, ..., u_L\}$ *, что матрица*

$$UDU \quad (11)$$

*имеет только неотрицательные элементы.*

Утверждение 2 является простым критерием для проверки возможности произвольной матрице быть матрицей ковариаций наблюдаемых случайных величин модели (1).

## Проверка условий идентифицируемости

Пусть $S = (s_{ij})$ – выборочная (подправленная на несмещенность) матрица ковариаций, построенная по $N$ независимым реализациям наблюдаемых случайных величин модели (1). Для проверки условий идентифицируемости нам необходимо проверять гипотезы, имеющие вид: $H_{\Sigma_{\mathbf{i},\mathbf{j}}} : |\Sigma_{\mathbf{i},\mathbf{j}}| \neq 0$. Для проверки таких гипотез можно использовать статистику $|S_{\mathbf{i},\mathbf{j}}|$, которая является асимптотически несмещенной оценкой для $|\Sigma_{\mathbf{i},\mathbf{j}}|$. Подправленная на несмещенность оценка для $|\Sigma_{\mathbf{i},\mathbf{j}}|$ имеет следующий вид [Drton et al., 2008]:

$$\widetilde{|S_{\mathbf{i},\mathbf{j}}|} = \frac{(N-1)^{k-1}}{(N-k)...(N-2)} |S_{\mathbf{i},\mathbf{j}}|. \quad (12)$$

Статистика $\widetilde{|S_{\mathbf{i},\mathbf{j}}|}$ ассимптотически (при $N \to \infty$) имеет нормальное распределение с математическим ожиданием $|\Sigma_{\mathbf{i},\mathbf{j}}|$ и дисперсией $\mathbf{Var}_{\Sigma_{\mathbf{i},\mathbf{j}}} \widetilde{|S_{\mathbf{i},\mathbf{j}}|}$, которая является полиномом, зависящем от элементов матрицы $\Sigma$ [Drton et al., 2007]. В работе [Drton et al., 2008] получен точный вид этой дисперсии:

$$\mathbf{Var}_{\Sigma_{\mathbf{i},\mathbf{j}}}|\widetilde{S_{\mathbf{i},\mathbf{j}}}|$$

$$= C|\Sigma_{\mathbf{i},\mathbf{j}}|^2\{\frac{(N+1)!}{(N+1-k)!} - \frac{(N-1)!}{(N-k-1)!}\}$$

$$+C|\Sigma_{\mathbf{i}\cup\mathbf{j},\mathbf{i}\cup\mathbf{j}}|(\sum_{m=0}^{k-1}(k-m)!\frac{(N+1)!}{(N+1-m)!}$$

$$\times(-1)^m tr\{(\Sigma_{\mathbf{i},\mathbf{j}}\Sigma^{\mathbf{i},\mathbf{j}})^{(m)}\}),$$

где $C = \frac{1}{(N-k)\ldots(N-1)}$, $\Sigma^{\mathbf{i},\mathbf{j}}$ – соответствующая подматрица матрицы $(\Sigma_{\mathbf{i}\cup\mathbf{j},\mathbf{i}\cup\mathbf{j}})^{-1}$, а матрица $(\Sigma_{\mathbf{i},\mathbf{j}}\Sigma^{\mathbf{i},\mathbf{j}})^{(m)}$ состоит из всех $m \times m$ миноров матрицы $\Sigma_{\mathbf{i},\mathbf{j}}\Sigma^{\mathbf{i},\mathbf{j}}$.

Заменяя в дисперсии $\mathbf{Var}_{\Sigma_{\mathbf{i},\mathbf{j}}}|\widetilde{S_{\mathbf{i},\mathbf{j}}}|$ элементы матрицы $\Sigma_{\mathbf{i},\mathbf{j}}$ соответствующими состоятельными оценками $S_{\mathbf{i},\mathbf{j}}$, мы получим состоятельную оценку $\mathbf{Var}_{S_{\mathbf{i},\mathbf{j}}}|\widetilde{S_{\mathbf{i},\mathbf{j}}}|$ для дисперсии статистики $|\widetilde{S_{\mathbf{i},\mathbf{j}}}|$. Легко видеть, что

$$\frac{|\widetilde{S_{\mathbf{i},\mathbf{j}}}| - |\Sigma_{\mathbf{i},\mathbf{j}}|}{\sqrt{\mathbf{Var}_{S_{\mathbf{i},\mathbf{j}}}|\widetilde{S_{\mathbf{i},\mathbf{j}}}|}} \quad (13)$$

ассимптотически (при $N \to \infty$) имеет стандартное нормальное распределение.

Используя (13), построим приближенный $\alpha$-доверительный интервал $[q^1_\alpha, q^2_\alpha]$ для $|\Sigma_{\mathbf{i},\mathbf{j}}|$. Гипотеза $H_{\Sigma_{\mathbf{i},\mathbf{j}}}$ будет (на уровне значимости $\alpha$) отвергаться, если интервал $[q^1_\alpha, q^2_\alpha]$ содержит 0. Проверка условий идентифицируемости заключается в следующем. Пусть $[\mathbf{E}]$ – число ребер графа $\mathbf{G}$. Задаемся уровнем значимости $\alpha$. Затем строим $\alpha/[\mathbf{E}]$ – доверительные интервалы $[q^1_\alpha, q^2_\alpha]_{\mathbf{i},\mathbf{j}}$ для $|\Sigma_{\mathbf{i},\mathbf{j}}|$, $(\mathbf{i},\mathbf{j}) \in \mathbf{E}$. Если доверительный интервал $[q^1_\alpha, q^2_\alpha]_{\mathbf{i},\mathbf{j}}$ содержит 0, то удаляем соответствующее ребро из графа $\mathbf{G}$. Если после удаления ребер получившийся граф $\mathbf{G}'' = (\mathbf{V}'', \mathbf{E}'')$ удовлетворяет условиям теоремы 1, то считаем, что гипотеза о том, что модель является глобально идентифицируемой, не отвергается на уровне значимости $\alpha$.

Выборочный вариант критерия (11) выглядит следующим образом. Вместо графа $\mathbf{G}'$, для построения матрицы $D$ используем граф $\mathbf{G}'' = (\mathbf{V}'', \mathbf{E}'')$, полученный при проверке условий идентифицируемости по имеющимся данным:

$$d_{fm} = \begin{cases} sign|S_{\mathbf{i}_f,\mathbf{i}_m}|, & \text{если } (\mathbf{i}_f, \mathbf{i}_m) \in \mathbf{E}''; \\ 0, & \text{в противном случае.} \end{cases}$$

Если получившаяся матрица не удовлетворяет условиям утверждения 2, то гипотеза о том, что выборка была сгенерирована моделью 1, на уровне значимости $\alpha$ отвергается.

## Литература


*Стафеев С. В.* Факторный анализ с зависимыми остатками: проблема идентифицируемости и оценка параметров // Труды ИПМИ КарНЦ РАН. 2005. Вып. 6. С. 119–130.

*Стафеев С. В.* О модели факторного анализа с зависимыми остатками // Обозрение прикладной и промышленной математики. 2007. Т. 14, вып. 6. С. 1058–1064.

*Стафеев С. В.* Полиномиальные инварианты для моделей с латентными переменными // Труды Карельского научного центра Российской академии наук. 2010. № 3. С. 83–86.

*Drton M., Sturmfels B., Sullivant S.* Algebraic factor analysis: tetrads, pentads and beyond // Probability Theory and Related Fields. 2007. Vol. 138. P. 463–493.

*Drton M., Massam H., Olkin I.* Moments of minors of Wishart matrices // Annals of Statistics. 2008. Vol. 36(5). P. 2261–2283.

*Drton M., Foygel R., Sullivant S.* Global identifiability of linear structural equation models // Annals of Statistics. 2011. Vol. 39(2). P. 865–886.

*Harman H.* Modern Factor Analysis. University of Chicago Press, third edition, 1976. 450 p.

*Stafeev S.* On the parameter estimation of recursive "bow-free" models with latent variables // Proceedings of the Eighth International Minsk Conference Computer Data Analysis and Modeling: Complex Stochastic Data and Systems. Minsk, 2007. Vol. 1. P. 178–181.

*Stafeev S.* On the method for finding invariants of factor analysis models // Computer Data Analysis and Modeling. Proceedings of the Ninth International Conference. Minsk, 2010. Vol. 1. P. 211–214.


## СВЕДЕНИЯ ОБ АВТОРЕ:


**Стафеев Сергей Вячеславович**
научный сотрудник, к. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: stafeev@krc.karelia.ru
тел.: (8142) 763370

**Stafeev, Sergey**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: stafeev@krc.karelia.ru
tel.: (8142) 763370

Труды Карельского научного центра РАН
№ 5. 2011. С. 115–119

*УДК 51-72*

# МАТЕМАТИЧЕСКОЕ ОБОСНОВАНИЕ МЕТОДА МОРФОЛОГИЧЕСКИХ ТРАЕКТОРИЙ

**И. А. Чернов**

*Институт прикладных математических исследований
Карельского научного центра РАН*

В статье приведено математическое обоснование метода морфологических траекторий, применяемого в физической химии. С математической точки зрения метод сводится к задаче оптимального управления системой, описываемой обыкновенными дифференциальными уравнениями. Показана разрешимость задачи и устойчивость модельной кривой к вариациям управлений.

Ключевые слова: метод морфологических траекторий, существование решения, корректность задачи Коши.

## I. A. Chernov. MATHEMATICAL SUBSTANTIATION OF THE METHOD OF MORPHOLOGICAL TRAJECTORIES

In the paper we present the mathematical basis for the method of morphological trajectories used in physical chemistry. From the mathematical point of view the method is an optimal control problem for the system described by ordinary differential equations. We prove the solvability of the problem and the stability of the solution with respect to perturbations of the controls.

Key words: morphological trajectory method, existence of solution, well-posedness of the Cauchy problem.

## Введение

Метод «морфологических траекторий» [Evard, Voyt, 2011], предложенный в Санкт-Петербургском государственном университете, призван объяснять экспериментальные результаты по формированию и разложению гидридов металлов в условиях быстрой диффузии, но может применяться и в других системах с фазовым переходом. Суть метода в описании динамики комплекса зародышей новой фазы тремя величинами, значимыми для кинетики: доля объема $V$, занимаемая новой фазой, площадь $S_o$, занимаемая новой фазой на поверхности частицы, и площадь $S_i$ поверхности раздела фаз. При этом обе площади выражены как функции объема: $S_o = S_o(V)$, $S_i = S_i(V)$; эти зависимости и носят название «морфологических траекторий». Консервативная математическая модель описывает процесс фазового перехода; если заданы морфологические траектории и кинетические параметры, то модельная кривая может быть рассчитана и сравнена с экспериментальной.

Целью настоящей работы является математическое обоснование метода. Математически имеем задачу оптимального управления системой, описываемой обыкновенными дифференциальными уравнениями и с квадратичным минимизируемым функционалом. Здесь возникают следующие вопросы. Управляема ли система? Иными словами, существуют ли положительные морфологические траектории, обеспечивающие совпадение модельной кривой экспериментальной? Устойчиво ли такое решение, т. е. не приводят ли малые вариации морфологических траекторий к значительным отклонениям модельной кривой? Как правило, морфологические траектории выбираются из некоторого параметрического семейства; это сводит задачу к конечномерной оптимизации. Разрешима ли эта задача в такой постановке?

Ответ на эти вопросы дает даже больше. Слишком хорошая управляемость является и сильной, и слабой стороной метода. Ситуация отсутствия решения может быть неотличима от неединственности решения. Например, траектории $S_o(V)$ достаточно для полной управляемости системы при любой $S_i(V)$ и любых кинетических параметрах, что делает бессмысленной задачу определения оптимального управления как пары $(S_o, S_i)$. При наличии ответа на поставленные вопросы можно понять возможности и слабые стороны метода, что, несомненно, будет способствовать его успешному применению на практике.

## Метод морфологических траекторий

Рассматриваем эксперимент дегидрирования порошка; для простоты предполагаем, что температура постоянна, а десорбция идет в вакуум. При этом в вакуумной камере находится изучаемый порошкообразный образец (гидрид); температура такова, что идет фазовый переход и десорбция водорода с поверхности. Рассматриваем одну частицу порошка единичного объема. Часть $V$ ее объема занимает новая фаза (металл с растворенным водородом), остальное — гидрид. Объем, занимаемый новой фазой, ограничен поверхностью, которую можно разделить на две части: это часть $S_o$ внешней поверхности частицы и поверхность раздела фаз, имеющая площадь $S_i$. Концентрацию водорода в фазе гидрида (она постоянна) примем за единицу. Концентрация $c(t)$ в фазе металла пространственно-постоянна, но может в общем случае меняться во времени.

Плотность потока десорбции квадратично зависит от концентрации: $bc^2$, коэффициент пропорциональности $b$ является одним из кинетических параметров. Считаем, что десорбция идет только с поверхности металла; таким образом, полный поток имеет вид $bc^2 S_o$. Сказанное позволяет записать закон сохранения:

$$\frac{d}{dt}\big(cV + (1 - V)\big) = -bc^2 S_o\big(V(t)\big). \qquad (1)$$

Динамика объема также описывается законом сохранения в предположении о нормальном росте зародышей, скорость которого линейно зависит от концентрации:

$$\frac{dV}{dt} = \Big(1 - \frac{c}{\bar{c}}\Big)\, kS_i\big(V(t)\big). \qquad (2)$$

Здесь $\bar{c}$ — равновесная концентрация, т. е. такая, при которой распад гидрида приостанавливается. Коэффициент $k$ — второй кинетический параметр; пусть $K = k/\bar{c}$.

Элементарные преобразования приводят к системе двух обыкновенных уравнений (точкой здесь и далее обозначаем производную по времени $t$):

$$V\dot{c} = (1 - c)\,(\bar{c} - c)\,KS_i - bc^2 S_o, \qquad (3)$$

$$\dot{V} = (\bar{c} - c)\,KS_i. \qquad (4)$$

Возможна запись уравнения для фазовой траектории $C(V) = c\big(V(t)\big)$:

$$V\frac{dC}{dV} = (1 - C) - \frac{bC^2}{(\bar{c} - C)K}U. \qquad (5)$$

Здесь $U = S_o/S_i$. Из этого уравнения следует, что фазовая кривая зависит только от отношения площадей, а не от них самих.

Уравнения сингулярны при $V = 0$. Физически это означает невозможность реакции, если образец исходно был однофазным. Это связано с тем, что в модели десорбция с фазы гидрида отсутствует. Поэтому мы примем $V(0) = V_0 > 0$ за начальное состояние; вторым начальным условием является $c(0) = c_0$ с очевидным ограничением $0 < c_0 < \bar{c}$.

Суть метода морфологических траекторий — в выборе функций $S_o(V)$ и $S_i(V)$ и кинетических параметров $b$ и $k$ так, чтобы доставить минимум функционалу квадратичной невязки

$$F = \int\limits_0^1 \Big(bc^2(t) S_o\big(V(t)\big) - \hat{J}(t)\Big) dt, \qquad (6)$$

где $\hat{J}(t)$ — измеряемый в эксперименте поток десорбции, а интервал времени единичен.

Морфологические траектории могут выбираться из различных семейств функций. Наиболее общая задача — накладывающая только требования положительности и гладкости. Будем считать, что обе функции $S_o$ и $S_i$ липшицевы на $[0, 1]$, причем $S_o > 0$ на $[V_0, 1]$, $S_i > 0$ на $[V_0, 1)$ и $S_i(1) = 0$.

Измерения потока $\hat{J}$ считаем также положительной липшицевой функцией на $[0, 1]$.

### Разрешимость задачи

Изучим вопрос об управляемости такой системы: минимизации функционала (6) при ограничениях (3) и (4) выбором липшицевых функций $S_\mathrm{o}(V)$ и $S_\mathrm{i}(V)$ при заданных параметрах $b$ и $k$. Покажем, что такая система полностью управляема:

**Утверждение 1.** *Для любых положительных $b$ и $k$ и любой функции $S_\mathrm{i}(V)$, липшицевой на $[V_0, 1]$, положительной на $[V_0, 1)$ и $S_\mathrm{i}(1) = 0$, существует липшицева положительная функция $S_\mathrm{o}(V)$ на $[V_0, 1]$ такая, что функционал $F = 0$.*

Докажем сначала очевидное

**Утверждение 2.** *При любых начальных данных $V_0 > 0$, $c_0 \in (0, \bar{c})$ решение системы* (3), (4) *обладает свойствами $V \in [V_0, 1]$, $\dot{V} > 0$, $c \in (0, \bar{c})$.*

*Доказательство.* Множество точек отрезка $[0, 1]$, в которых $c = \bar{c}$ замкнуто и потому компактно, и имеет минимальный элемент $t^*$. Утверждение следует из того, что если $c(t^*) = \bar{c}$, то $\dot{c}(t^*) < 0$. Это означает, что $c(t) > \bar{c}$ при каких-то $t < t^*$, что противоречит минимальности $t^*$. Возможность $t^* = 0$ отпадает в силу $c_0 < \bar{c}$. Рассуждение для $c = 0$ аналогично. Свойство монотонности $V$ следует из доказанного и вида уравнения (4). Тогда $V \geqslant V_0$ очевидно; а $V \leqslant 1$ следует из свойства $S_\mathrm{i}(1) = 0$ и, стало быть, $\dot{V}(t) = 0$ если $V(t) = 1$. □

Утверждение может быть усилено.

**Утверждение 3.** *Существует такая константа $r \in (0, \bar{c}/2)$, что если $V_0 > 0$, $c_0 \in [r, \bar{c} - r)$, то $V \in [V_0, 1]$, $\dot{V} > 0$, $c \in (r, \bar{c} - r)$.*

*Доказательство.* В самом деле, в правой части (3) — квадратный трехчлен относительно $c$. Если множитель при $c^2$ положителен, то либо вещественные корни отсутствуют, либо оба положительны (теорема Виета). В первом случае $\dot{c} > 0$ и в качестве $r$ можно выбрать $c_0$. Во втором $\dot{c} > 0$ при $c \in [0, c_-)$, где $c_-$ – меньший корень; выбираем $r$ произвольно из $(0, c_-)$. Если же множитель при $c^2$ меньше нуля, то корни вещественны и разных знаков, причем $\dot{c} > 0$ при $c \in [0, c_+)$, где $c_+$ теперь — больший корень. Оценка сверху получается проще: поскольку правая часть (3) отрицательна при $c = \bar{c}$, то отрицательна она и в некоторой окрестности, т. е. при $c < \bar{c}$. □

Теперь докажем утверждение 1.

*Доказательство.* Требование $F = 0$ сразу приводит к системе

$$
\begin{aligned}
V\dot{c} &= (1 - c)\,(\bar{c} - c)\,KS_\mathrm{i} - \hat{J}(t), \\
\dot{V} &= (\bar{c} - c)\,KS_\mathrm{i}.
\end{aligned}
$$

Правые части этой системы липшицевы по переменным $c$ и $V$ в силу доказанного утверждения. Следовательно, существует единственное решение системы. Более того, производные $\dot{V}$ и $\dot{c}$ ограничены, поэтому решение может быть продолжено на $[0, 1]$. Это позволяет определить $S_\mathrm{o}$ как функцию времени формулой

$$S_\mathrm{o} = \hat{J}(t)\big(bc^2(t)\big)^{-1}.$$

Однако функция $V(t)$ монотонна и потому инъективно отображает отрезок $[0, 1]$ в $[V_0, 1]$. Поэтому на $[V_0, V(1)]$ определена монотонная обратная функция $t = t(V)$, позволяющая определить $S_\mathrm{o}$ как функцию $V$. Если $V(1) = 1$, то $S_\mathrm{o}(V)$ определена на $[V_0, 1]$ однозначно; если же $V(1) < 1$, то на отрезке $[V(1), 1]$ функцию $S_\mathrm{o}(V)$ можно доопределить произвольным образом (с сохранением свойств — например, константой по непрерывности). Аналогично при необходимости доопределяется $S_\mathrm{o}(V)$ на $[0, V_0]$.

Более того, производная $\dot{V}$ непрерывна и отделена от нуля в силу доказанного утверждения; поэтому $t(V)$ непрерывно дифференцируема и производная ее ограничена. Поэтому полученная $S_\mathrm{o}(V)$ липшицева и положительна на $[V_0, 1]$. □

Отметим, что из физического смысла следуют и другие требования к функции $S_\mathrm{o}(V)$, имеющей смысл площади, занимаемой зародышами новой фазы на поверхности. Например, монотонное возрастание (впрочем, в ряде случаев возможно перераспределение водорода, приводящее к формированию гидрида в отдельных областях за счет интенсивного разложения в других, и в этом случае $S_\mathrm{o}$, в принципе, может убывать) или асимптотика при малых и больших $V$. Расхождения с физическим смыслом свидетельствуют об ошибочном выборе $S_\mathrm{i}$ и кинетических параметров.

### Непрерывная зависимость решения от морфологических траекторий

Рассмотрим вопрос о непрерывности зависимости решения от морфологических траекторий. Она важна, по крайней мере, по двум причинам. Во-первых, устойчивость решения по отношению к управляющим воздействиям дает основания искать решение в том или ином параметрическом семействе или решать задачу приближенными методами, поскольку ошибка мала при небольшом отклонении приближенного решения от точного. Во-вторых, при поиске морфологических траекторий в параметрическом семействе, $F$ оказывается непрерывной функцией параметров и имеет минимум на любом компактном подмножестве этих параметров.

Отметим, что правая часть уравнения (5) непрерывно дифференцируема по переменным $C$ и $U$, причем производные отрицательны, ограничены и отделены от нуля.

Пусть функциям $U(V)$ и $U + \Delta U$ отвечают соответственно решения $C(V)$ и $C + \Delta C$.

**Утверждение 4.** *Существует такое $L > 0$, что $|\Delta C| \leqslant L|\Delta U|$ на $[V_0, 1]$.*

*Доказательство.* Для $\Delta C(V)$ имеем уравнение

$$V\frac{d(\Delta C)}{dV} = -\big(U^*(V)A(V) + 1\big)\Delta C - B(V)\Delta U,$$

причем $A > 0$, $B > 0$ ограничены, отделены от нуля и не зависят от $U$, $\Delta U$ и $\Delta C$, а $U^*$ лежит между $U$ и $U + \Delta U$. Если экстремальное значение $\Delta C$ достигнуто в точке $V \in (V_0, 1)$, то производная обращается в ноль, откуда следует оценка $|\Delta C| \leqslant L|\Delta U|$ для некоторой константы $L$, зависящей только от $\bar{c}$, $b$, $k$, и $r$. Если положительное максимальное значение $\Delta C$ достигнуто при $V = 1$, то производная в $V = 1$ неотрицательна, что приводит к неравенству

$$\Delta C \leqslant -\frac{B(V)}{A(V)}\Delta U \leqslant L|\Delta U|,$$

аналогичное рассуждение для отрицательного минимума доказывает справедливость оценки при $V = 1$. Начальное условие $\Delta C(V_0) = 0$ распространяет ее на весь отрезок $[V_0, 1]$. □

Уравнение (4) определяет функцию $t(V)$; оценим ее изменение при вариации морфологических траекторий. Введем обозначения: черта сверху означает максимум модуля по $[V_0, 1]$; $S_\mathrm{i}^- = \min(S_\mathrm{i}, S_\mathrm{i} + \Delta S_\mathrm{i})$; $\delta S_\mathrm{i} = \Delta S_\mathrm{i}/S_\mathrm{i}^-$ (относительная вариация); $R = Kr$.

**Утверждение 5.** *Пусть $S_\mathrm{i}(V)$ такова, что интеграл от $S_\mathrm{i}^{-1}$ по $[V_0, 1]$ сходится. Существуют такие $M_\mathrm{U} > 0$ и $M_\mathrm{i} > 0$, что*

$$\overline{\Delta t} \leqslant M_\mathrm{U}\overline{\Delta U} + M_\mathrm{i}\overline{\delta S_\mathrm{i}}.$$

*Доказательство.* Из уравнения (4) следует, что

$$t(V) = \int\limits_{V_0}^{V} \frac{dV}{(\bar{c} - C(V))\,KS_\mathrm{i}(V)}.$$

Если паре функций $C + \Delta C$, $S_\mathrm{i} + \Delta S_\mathrm{i}$ отвечает решение $t + \Delta t$, то

$$|\Delta t| \leqslant \overline{|\Delta C|}\int\limits_{V_0}^{V}\frac{dV}{rRS_\mathrm{i}^-(V)} + \overline{|\delta S_\mathrm{i}|}\int\limits_{V_0}^{V}\frac{dV}{RS_\mathrm{i}^-(V)}.$$

Сходимость интегралов завершает доказательство. □

Перейдем в функционале (6), к переменной $V$, приняв соглашение $dt = 0$ при $V > V(1)$:

$$F = \int\limits_{V_0}^{1} \frac{bC^2(V)S_\mathrm{o}(V) - \hat{J}\big(t(V)\big)}{(\bar{c} - C)\,KS_\mathrm{i}(V)}dV$$

Пусть $S_\mathrm{o}(V)$ липшицева и положительна на $[V_0, 1]$.

**Утверждение 6.** *В предположениях предыдущего утверждения существуют такие $A_\mathrm{o} > 0$, $A_\mathrm{i} > 0$ и $A_\mathrm{U} > 0$, что*

$$|\Delta F| = A_\mathrm{o}\overline{|\Delta S_\mathrm{o}|} + A_\mathrm{i}\overline{|\delta S_\mathrm{i}|} + A_\mathrm{U}\overline{|\Delta U|}.$$

*Доказательство.* Аналогичное рассуждение для $\Delta F$ приводит с учетом ограниченности $S_\mathrm{o}$ к оценке вида

$$|\Delta F| = A_\mathrm{o}\overline{|\Delta S_\mathrm{o}|} + A_\mathrm{i}\overline{|\delta S_\mathrm{i}|} + \frac{A_\mathrm{U}}{L}\overline{|\Delta C|},$$

где $A_\mathrm{o}$, $A_\mathrm{i}$ и $A_\mathrm{U}$ все содержат интеграл от $S_\mathrm{i}^{-1}$ по $[V_0, 1]$, который по предположению сходится. Учитывая оценку для $\Delta C$, получаем доказываемое неравенство. □

В общем случае малость $\Delta S_\mathrm{o}$ и $\Delta S_\mathrm{i}$ (и даже $\delta S_\mathrm{i}$) не влечет малости $\Delta U$, поскольку $S_\mathrm{i}(1) = 0$ и, следовательно, вблизи точки $V = 1$ малые отклонения $S_\mathrm{o}$ могут влечь большие отклонения $U = S_\mathrm{o}/S_\mathrm{i}$. Поэтому оценка содержит $\Delta U$ в правой части. Однако дополнительные физические предположения позволяют оценить $\Delta U$ через $\Delta S_\mathrm{o}$ и $\delta S_\mathrm{i}$. Точки $V \approx 1$ отвечают ситуации «почти полного дегидрирования», когда объем старой фазы $1 - V$ мал. Физические соображения (кристаллографические или принцип минимизации энергии поверхности) позволяют предположить, что форма этой области приблизительно постоянна; а соображения размерности приводят к асимптотической зависимости $S_\mathrm{i} = O\big((1 - V)^{2/3}\big)$ при $V \to 1$, аналогично для $S_\mathrm{o}$. Это предположение влечет сходимость интеграла от $S_\mathrm{i}^{-1}$, которая предполагалась в утверждениях, доказанных выше.

Уточним предположение: будем считать, что отношение $S_\mathrm{i}$ к $(1-V)^{2/3}$ имеет при $V \to 1$

один и тот же предел, аналогично для $S_{\rm o}$. Иными словами, вариация морфологической траектории не нарушает асимптотику. Тогда $\Delta S_{\rm o} = o\big((1-V)^{2/3}\big)$, $\Delta S_{\rm i} = o\big((1-V)^{2/3}\big)$, что приводит к малости $\Delta U$ вблизи $V = 1$. На полуинтервале $[V_0, 1)$ функция $S_{\rm i}$ положительна и проблемы малого знаменателя не возникает.

## Заключение

В работе рассмотрены вопросы, связанные с математическим обоснованием метода морфологических траекторий, применяемого в физической химии: доказана разрешимость соотвествующих математических задач и устойчивость решения.

Работа выполнена при поддержке РФФИ (грант 09-03-00947-а).

## Литература

*Evard E. A., Voyt A. P.* Hydride decomposition characterization by means of «morphological trajectory» method — applied to $AlH_3$ // Journal of Alloys and Compounds. 2011. В печати.

## СВЕДЕНИЯ ОБ АВТОРЕ:

**Чернов Илья Александрович**
старший научный сотрудник, к. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: IAChernov@yandex.ru
тел.: (8142) 766312

**Chernov, Ilya**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: IAChernov@yandex.ru
tel.: (8142) 766312

Труды Карельского научного центра РАН
№ 5. 2011. С. 120–124

*УДК 519.7*

# МУЛЬТИНОМИАЛЬНЫЙ ЛОГИТ-АНАЛИЗ И КОНКУРЕНТНОЕ ПОВЕДЕНИЕ НА РЫНКЕ

**А. В. Щипцова**

*Институт прикладных математических исследований Карельского научного центра РАН*

В статье рассматривается применение мультиномиального логит-анализа для моделирования спроса на рынке одного товара, зависящего от цены и расстояния от потребителя до участника рынка. Исследуется конкурентное поведение игроков на рынке с помощью понятия равновесия по Нэшу. Представлены результаты численного моделирования.

Ключевые слова: мультиномиальный логит-анализ, дуополия Хотеллинга на плоскости, равновесие по Нэшу.

## A. V. Shchiptsova. MULTINOMIAL LOGIT ANALYSIS AND COMPETITIVE BEHAVIOUR IN THE MARKET

The paper considers application of the multinomial logit model for estimation of the customer demand for a product in the market. Customer demand depends on the price and distance from a customer to a market actor. We examine the players' competitive behaviour in the market using Nash equilibrium. Computational modeling with multinomial logit analysis is presented.

Key words: multinomial logit analysis, Hotelling's duopoly on the plane, Nash equilibrium.

## Введение

Дуополия Хотеллинга [Hotelling, 1929] описывает поведение участников рынка – продавцов одного товара. В модели предполагается, что на величину потребительского спроса влияют установленная игроком цена и транспортные расходы потребителя, равные расстоянию от него до игрока. Хотеллинг нашел равновесные цены на линейном рынке и поставил задачу о размещении игроков на рынке. В дальнейшем было доказано [d'Aspremont et al., 1979], что в такой постановке равновесие в задаче о размещении не существует.

Салоп [Salop, 1979] распространил дуополию Хотеллинга на модель «кругового» города, где участники рынка располагаются вдоль окружности на одинаковом расстоянии друг от друга. В работе [Mazalov, Sakaguchi, 2003] показано, что решение задачи о размещении существует на плоском рынке с квадратичными транспортными расходами.

В работах [Щипцова, 2009; Мазалов и др., 2010] исследовалась модель дуополии Хотеллинга на плоскости. Ценовое равновесие и решение задачи о размещении были построены для случая рыночной конкуренции между двумя участниками рынка. Увеличение числа игроков в модели Хотеллинга ведет к существенному усложнению задачи.

Данная статья посвящена задаче поиска ценового равновесия для $n$ игроков в рамках мультиномиальной логит-модели, предложенной МакФадденом [McFadden, 1973]. Товар, предлагаемый участником рынка, является одной из потребительских альтернатив. Как и в дуополии Хотеллинга, будем считать, что выбор потребителя зависит от цены и расстояния между ним и участником рынка.

## Мультиномиальная логит-модель

Пусть на рынке есть $m$ потребителей из множества $I = \{1 \dots m\}$. Каждый потребитель делает выбор из конечного множества альтернатив $J = \{1 \dots n\}$. Альтернатива для потребителя состоит в приобретении товара у $j$-го участника рынка. Выбор $i$-го потребителя опишем с помощью решающей функции $d(i) : I \to J$.

Будем считать, что $i$-й потребитель стремится максимизировать полезность $u_{ij}$, которую он получает от приобретения товара у $j$-го участника рынка. Функция полезности $u_{ij}$ имеет вид

$$u_{ij} = V_{ij} + \varepsilon_{ij},$$

где $V_{ij}$ – детерминированная составляющая, зависящая от свойств самой альтернативы и предпочтений потребителя, $\varepsilon_{ij}$ – стохастическая.

Вероятность выбора $i$-м потребителем альтернативы $j$ равна вероятности того, что полезность $u_{ij}$ наибольшая из возможных. Таким образом,

$$\begin{aligned} &P\left(d(i) = j\right) \\ &= P\left(u_{ij} \geqslant u_{ir}, \forall r \in \{1 \dots n\} : r \neq j\right) \\ &= P\left(u_{ij} = \max_{r \in \{1 \dots n\}} u_{ir}\right). \end{aligned}$$

МакФадден [McFadden, 1973] предложил мультиномиальную логит-модель, в которой стохастические составляющие функции полезности $\varepsilon_{ij}$ есть независимые случайные величины, распределенные по закону Гумбеля (экстремальное распределение I-го типа) с соответствующими функциями распределения и плотности

$$F(\varepsilon) = e^{-e^{-\beta(\varepsilon-\alpha)}}, \beta > 0,$$
$$f(\varepsilon) = \beta e^{-\beta(\varepsilon-\alpha) e^{-e^{-\beta(\varepsilon-\alpha)}}}.$$

Тогда в предположении о независимости $\varepsilon_{ij}$ вероятность выбора альтернативы $j$ представима в явном виде

$$P\left(d(i) = j\right) = \frac{e^{V_{ij}}}{\sum\limits_{r=1}^{n} e^{V_{ir}}}. \qquad (1)$$

Формула вероятности выбора делает мультиномиальную логит-модель привлекательной для практического применения. Увеличение числа альтернатив не ведет к усложнению модели.

## Ценовое равновесие в мультиномиальной логит-модели

Пусть рынок потребительских услуг представлен кругом радиуса единица с равномерным распределением населения. Плотность населения равна единице. Каждый из участников рынка в точке $(x_j, y_j)$ $(j = 1 \dots n)$ предлагает потребителям один и тот же товар по цене $p_j$ и стремится получить наибольшую прибыль от продажи. Спрос является абсолютно неэластичным. Без потери общности будем считать, что себестоимость товара для участников рынка равна нулю.

Как и в дуополии Хотеллинга, будем предполагать, что кроме цены за приобретение товара потребитель также уплачивает транспортные расходы за его доставку. Таким образом, полезность приобретенного товара у $j$-го участника рынка представима в виде

$$u_j(p_j, x, y) = -\beta_1 p_j - \beta_2 \rho_j(x, y) + \varepsilon_j,\ j = 1 \dots n,$$

где $\beta_1, \beta_2 \geqslant 0$ – некоторые константы, $\rho_j(x, y) = \sqrt{(x - x_j)^2 + (y - y_j)^2}$ – расстояние от потребителя до продавца и $\varepsilon_j$ – стохастическая составляющая полезности, одинаковая для всех потребителей. Будем предполагать, что $\varepsilon_j$ – независимые случайные величины, распределенные по закону Гумбеля.

Каждый потребитель стремится получить максимальную полезность при выборе $j$-го участника рынка.

Таким образом, из (1) вероятность приобретения товара у $j$-го участника при установленных ценах $p_1, \dots p_n$ для потребителя в точке $(x, y)$ составит

$$\begin{aligned} &P_j(x, y) = \frac{e^{-\beta_1 p_j - \beta_2 \sqrt{(x-x_j)^2 + (y-y_j)^2}}}{\sum\limits_{i=1}^{n} e^{-\beta_1 p_i - \beta_2 \sqrt{(x-x_i)^2 + (y-y_i)^2}}}, \\ &\qquad j = 1 \dots n-1, \\ &P_n(x, y) = 1 - \sum_{i=1}^{n-1} P_i(x, y). \end{aligned} \qquad (2)$$

Доля потребителей, выбирающих товар $j$-го игрока, будет равна

$$S_j = \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} P_j(x,y)dxdy, \quad j = 1 \dots n-1, \tag{3}$$
$$S_n = \pi - \sum_{i=1}^{n-1} S_i.$$

Мы определили бескоалиционную игру

$$\Gamma = \Big\langle \{1 \dots n\}, \{p_j \in [0, +\infty),\ j = 1 \dots n\}, \{H_j = p_j S_j,\ j = 1 \dots n-1, \quad H_n = \pi - \sum_{i=1}^{n-1} S_i\} \Big\rangle. \tag{4}$$

Из вида (2) получаем, что

$$\frac{\partial P_j(x,y)}{\partial p_j} = -\beta_1 P_j(x,y)(1 - P_j(x,y)).$$

Точка равновесия по Нэшу $(p_1^*, \dots, p_n^*)$ в игре (4) удовлетворяет системе уравнений:

$$\begin{cases} \frac{\partial H_j}{\partial p_j} = S_j + p_j \frac{\partial S_j}{\partial p_j} = 0,\ j = 1 \dots n-1, \\ \frac{\partial H_n}{\partial p_n} = \pi - \sum\limits_{i=1}^{n-1} S_i + p_n \sum\limits_{i=1}^{n-1} \frac{\partial S_i}{\partial p_i} = 0. \end{cases} \tag{5}$$

Из (2), (3) и (5) получаем, что равновесие на рынке потребительских услуг, территория которого представлена единичным кругом, отвечает условиям

$$\begin{cases} \frac{\partial H_j}{\partial p_j} = \int\limits_{-1}^{1} \int\limits_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} P_j(x,y)\Big(1 - \beta_1 p_j \big(1 - P_j(x,y)\big)\Big)dxdy = 0,\ j = 1 \dots n-1, \\ \frac{\partial H_n}{\partial p_n} = \pi - \int\limits_{-1}^{1} \int\limits_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum\limits_{i=1}^{n-1} P_i(x,y)\Big(1 + \beta_1 p_n \big(1 - \sum\limits_{i=1}^{n-1} P_i(x,y)\big)\Big)dxdy = 0. \end{cases}$$

## Моделирование

Мультиномиальная логит-модель использовалась при проведении численного моделирования конкурентного поведения игроков на двух типах рынков: рынке по предоставлению парикмахерских услуг (цена на мужскую модельную стрижку) и рынке услуг АЗС (цена на дизельное топливо). Расстояние между потребителем и поставщиком услуги рассчитывалось в евклидовой метрике.

Рынок парикмахерских услуг рассматривался в рамках одного городского микрорайона (мкр. Древлянка г. Петрозаводска) с количеством игроков $n = 4$. Территория микрорайона была смоделирована кругом радиуса единица (рис. 1).

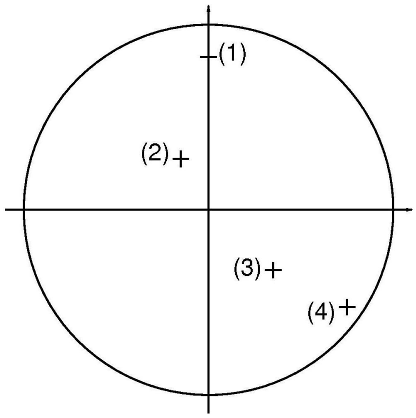



*Рис. 1.* Расположение игроков на рынке парикмахерских услуг ($n = 4$)

Конкурентное поведение игроков на рынке услуг АЗС ($n = 16$) исследовалось в пределах г. Петрозаводска (исключая мкр.-ны Соломенное, Сулажгора и Птицефабрика), модель города была представлена полукругом радиуса единица (рис. 2).

Ценовое равновесие было найдено как решение системы (5) для соответствующей области. Оценка параметров мультиномиальной логит-модели получена с помощью метода максимального правдоподобия. В качестве единицы цены было взято среднее отклонение от минимальной цены из существующих реальных цен, предлагаемых участниками рынка. Себестоимость услуги была принята как минимальная реальная цена.

Результаты расчетов приведены в таблицах 1 и 2. Полученные данные показывают, что поведение игроков на рынке парикмахерских услуг близко к оптимальному. Участники рынка услуг АЗС отклоняются от равновесия по Нэшу в рамках мультиномиальной логит-модели.

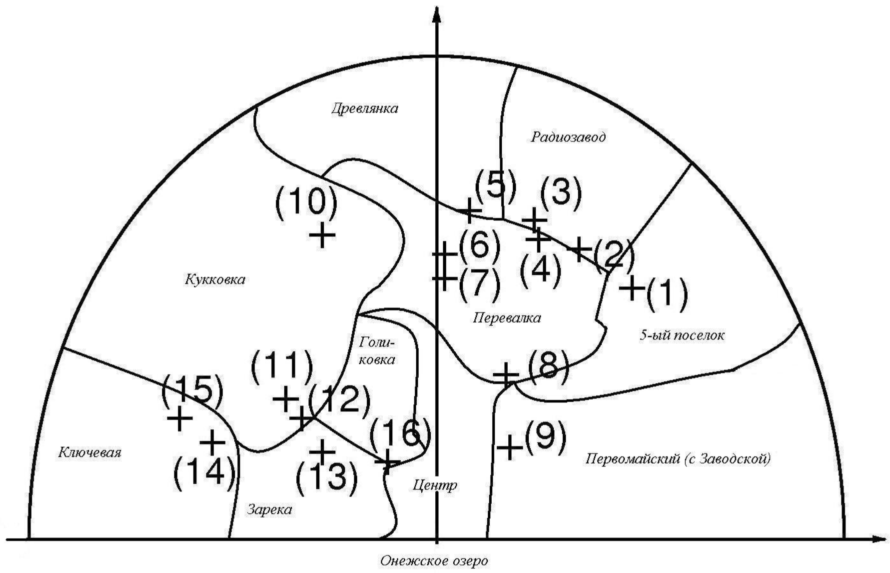



*Рис. 2.* Расположение игроков на рынке услуг АЗС ($n = 16$)

*Таблица 1.* Ценовое равновесие на рынке парикмахерских услуг($n = 4$, $\beta_1 = 292,755$, $\beta_2 = 292,609$)

| Игрок | Координаты | Цена в равновесии | Реальная цена (руб.) | Модельная цена (руб.) |
|---|---|---|---|---|
| 1 | (0, 0,825) | 0,449006 | 230 | 212,348 |
| 2 | (–0,15, 0,275) | 0,433218 | 250 | 211,913 |
| 3 | (0,35, –0,325) | 0,447982 | 230 | 212,32 |
| 4 | (0,75, –0,525) | 0,214147 | 200 | 205,89 |

*Таблица 2.* Ценовое равновесие на рынке услуг АЗС ($n = 16$, $\beta_1 = 702,496$, $\beta_2 = 702,405$)

| Игрок | Координаты | Цена в равновесии | Реальная цена (дизель) (руб.) | Модельная цена (руб.) |
|---|---|---|---|---|
| 1 | (0,48, 0,52) | 0,158352 | 26,10 | 25,35 |
| 2 | (0,35, 0,6) | 0,0580351 | 26,95 | 25,25 |
| 3 | (0,24, 0,66) | 0,0452477 | 26,30 | 25,25 |
| 4 | (0,25, 0,62) | 0,0237764 | 26,00 | 25,20 |
| 5 | (0,08, 0,68) | 0,111941 | 26,10 | 25,30 |
| 6 | (0,02, 0,59) | 0,289 | 26,10 | 25,50 |
| 7 | (0,02, 0,54) | 0,125509 | 26,90 | 25,35 |
| 8 | (0,17, 0,34) | 0,0540679 | 26,00 | 25,25 |
| 9 | (0,18, 0,19) | 0,123205 | 25,60 | 25,35 |
| 10 | (–0,28, 0,63) | 0,20681 | 27,30 | 25,40 |
| 11 | (–0,37, 0,29) | 0,0804116 | 25,40 | 25,30 |
| 12 | (–0,33, 0,25) | 0,111806 | 26,30 | 25,30 |
| 13 | (–0,28, 0,18) | 0,0952577 | 26,25 | 25,30 |
| 14 | (–0,55, 0,2) | 0,0459539 | 26,00 | 25,25 |
| 15 | (–0,63, 0,25) | 0,0883109 | 25,20 | 25,30 |
| 16 | (–0,12, 0,16) | 0,142341 | 26,30 | 25,35 |

## Заключение

В статье найдены условия, которым удовлетворяет оптимальное по Нэшу конкурентное поведение игроков на рынке одного товара. Применение мультиномиального логит-анализа для моделирования потребительского спроса позволило получить условия ценового равновесия для произвольной размерности задачи по количеству участников рынка ($n \geqslant 2$). Проведено численное моделирование для рынка потребительских услуг при $n = 4$ и $n = 16$.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (10-01-00089-а) и Отделения Математических наук РАН «Алгебраические и комбинаторные методы математической кибернетики и информационные системы нового поколения».

## Литература

*Мазалов В. В., Щипцова А. В., Токарева Ю. С.* Дуополия Хотеллинга и задача о размещении на плоскости // Экономика и математические методы. 2010. Т. 46, вып. 4. С. 91–100.

*Шандор З.* Мультиномиальные модели дискретного выбора // Квантиль. 2009. № 7. С. 9–19.

*Щипцова А. В.* Задача о размещении // Методы математического моделирования и информационные технологии. Труды ИПМИ КарНЦ РАН. 2009. Вып. 9. С. 53–69.

*d'Aspremont C., Gabszewicz J., Thisse J. F.* On Hotelling's "Stability in competition" // Econometrica. 1979. Vol. 47, N. 5. P. 1145–1150.

*Heiss F.* Structural choice analysis with nested logit models // The Stata Journal. 2002. Vol. 2, N. 3. P. 227–252.

*Hotelling H.* Stability In Competition // The Economic Journal. 1929. Vol. 39. Issue 153. P. 41–57.

*Mazalov V. V., Sakaguchi M.* Location Game On The Plane // International Game Theory Review. 2003. Vol. 5, N. 1. P. 1–13.

*McFadden D.* Conditional logit analysis of qualitative choice behavior / Ed. P. Zarembka // Frontiers in econometrics. New York: Academic Press, 1973. P. 105–142.

*Salop S.* Monopolistic competition with outside goods // Bell journal of Economics. 1979. Vol. 10. P. 141–156.

## СВЕДЕНИЯ ОБ АВТОРЕ:

**Щипцова Анна Владимировна**
аспирантка
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: ann_sh@inbox.ru
тел.: (8142) 766312

**Shchiptsova, Anna**
Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: ann_sh@inbox.ru
tel.: (8142) 766312

Труды Карельского научного центра РАН
№ 5. 2011. С. 125–131

# ХРОНИКА

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДЕЯТЕЛЬНОСТИ ЦКП КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН «ЦЕНТР ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ»

В феврале 2009 г. в Карельском научном центре РАН компанией Т-Платформы были проведены работы по установке и вводу в эксплуатацию высокопроизводительной системы, состоящей из вычислительного кластера и системы хранения данных. В июне того же года был создан Центр коллективного пользования (ЦКП) КарНЦ РАН «Центр высокопроизводительной обработки данных». На основании постановления Президиума КарНЦ РАН от 30.06.2009 г. № 38 базовой организацией ЦКП установлено Учреждение Российской академии наук Институт прикладных математических исследований (ИПМИ) КарНЦ РАН.

Согласно Положению [Положение о ЦКП КарНЦ РАН, 2011], основными направлениями деятельности ЦКП являются:

- обеспечение доступа пользователей ЦКП к высокопроизводительному вычислительному оборудованию и системе хранения данных;
- техническая поддержка оборудования, системных и прикладных программных средств;
- осуществление методического обеспечения, техническая поддержка пользователей и руководство организацией работ в области высокопроизводительных вычислений;
- участие в обучении студентов и специалистов по вопросам применения и эффективного использования современных методов высокопроизводительных и параллельных вычислений для решения сложных задач математического и имитационного моделирования;
- разработка новых и совершенствование существующих методов высокопроизводительных и параллельных вычислений для решения задач математического моделирования;
- текущее содержание и развитие материально-технической базы ЦКП путем дооснащения имеющихся комплексов современным техническим оборудованием и программными средствами для обеспечения и развития исследований в научных учреждениях КарНЦ РАН.

Высокопроизводительная вычислительная система КарНЦ РАН располагается в специальном помещении, оборудованном мощным источником бесперебойного питания APC Smart-UPS 15KVA/12kW 400V и системой кондиционирования воздуха, обеспечивающей необходимые климатические условия.

**Аппаратное обеспечение.** Вычислительный кластер — это группа компьютеров, объединенных высокоскоростными каналами связи и представляющая с точки зрения пользователя единый аппаратный ресурс. Кластер ЦКП состоит из одного управляющего и десяти вычислительных узлов, характеристики которых приведены в таблице. Узлы объединены управляющей сетью на базе Gigabit Ethernet и высокопроизводительной сетью передачи данных на базе интерфейса InfiniBand 4X DDR, обеспечивающего пропускную способность до 20 Гбит/с.

Пользователям ЦКП также доступна система хранения данных (СХД) ReadyStorage SAN 3994, состоящая из 16 жестких дисков емкостью по 146 Гбайт каждый, объединенных в RAID-массив 5 уровня. Доступ к СХД осуществляется через два высокоскоростных оптических интерфейса Fibre Channel.

Пиковая производительность кластера согласно проектной документации составляет 851 Гфлопс, производительность, показанная на тесте Linpack, — 637.2 Гфлопс.

Характеристики узлов кластера

| Тип узла | Управляющий | Вычислительный |
|---|---|---|
| CPU | 2xQuad-Core Intel Xeon 5430 2,66 ГГц | 2xQuad-Core Intel Xeon 5430 2,66 ГГц |
| RAM | 4 Гб FB-DIMM DDR2-667 REG ECC | 4 Гб FB-DIMM DDR2-667 REG ECC |
| HDD | 6x146 Гб SAS 15000 rpm | 2x250 Гб SCSI |

**Программное обеспечение и доступ к ресурсам.** Вычислительный кластер работает под управлением операционной системы SUSE Linux Enterprise Server 10, которая установлена как на управляющем, так и на вычислительных узлах. Для осуществления круглосуточного мониторинга доступности кластера и предотвращения нештатных ситуаций весной 2010 г. была установлена специализированная система мониторинга Nagios, которая в автоматическом режиме оповещает администратора кластера о возникающих неполадках. Сбор статистики по интенсивности использования аппаратных ресурсов (загрузка процессоров, использование оперативной памяти, пропускной способности сети и др.) осуществляется системой Ganglia. Для управления очередью задач на кластере установлена система SLURM.

Для пользователей ЦКП на кластере доступен набор инструментов разработки параллельных программ Intel Cluster Toolkit Compiler Edition, включающий в себя:

- производительные оптимизирующие компиляторы языков C, C++, Fortran;
- библиотеку математических функций Intel Math Kernel Library;
- средство отладки Intel Debugger;
- средство оптимизации Intel Trace Analyzer and Collector.

Для отладки также можно использовать систему Total View Debugger. Программистам доступны несколько версий библиотеки MPI: Intel MPI, MVAPICH2, MVAPICH, OpenMPI, MPICH; установлена библиотека для разработки программ с общей памятью Intel OpenMP.

Доступ к кластеру для пользователей ЦКП организован с использованием протокола ssh из локальной сети КарНЦ и глобальной сети Интернет. Для удобства пользователей создан и поддерживается сайт ЦКП [ЦКП КарНЦ РАН, 2011], на котором содержится информация о правилах доступа, имеются краткие руководства пользования кластером и другая полезная информация, публикуются новости Центра высокопроизводительной обработки данных. На сайте также поддерживается раздел, позволяющий в реальном времени отслеживать занятость узлов кластера, размер очереди задач и примерное время до завершения выполнения задач.

**Образовательная деятельность.** Работа, направленная на подготовку специалистов по параллельным вычислениям, велась сотрудниками ИПМИ КарНЦ РАН в рамках спецкурсов «Методы и алгоритмы параллельных вычислений», «Введение в параллельные вычисления» (д. ф.-м. н., проф. А. В. Соколов), «Современные технологии высокопроизводительных вычислений» (к. ф.-м. н. Е. Е. Ивашко). Всего за 2009–2010 учебный год обучение прошли 55 человек.

В целях популяризации высокопроизводительных и параллельных вычислений, а также для повышения заинтересованности в использовании ресурсов вычислительного кластера, силами сотрудников ЦКП с привлечением пользователей ЦКП были организованы и проведены семинары: «Ресурсы вычислительного кластера ЦКП КарНЦ РАН» (к. ф.-м. н. Е. Е. Ивашко, А. С. Румянцев, ИПМИ КарНЦ РАН), «GRID-сегмент КарНЦ РАН и основы разработки программ для GRID-платформы BOINC» (Н. Н. Никитина, ИПМИ КарНЦ РАН), «Опыт использования пакета Firefly для моделирования расплавов» (О. В. Кременецкая, ПетрГУ), «Опыт разработки параллельных программ» (А. М. Караваев, ПетрГУ).

Все семинары проходили в «открытом» режиме, объявления о проводимых мероприятиях размещались на сайте ЦКП, на информационных досках институтов КарНЦ РАН и направлялись по электронной почте согласно списка рассылки пользователей ЦКП.

**Исследования, проводимые с использованием высокопроизводительной системы КарНЦ РАН.** В этом разделе представлены краткие описания некоторых задач, решаемых пользователями ЦКП с использованием высокопроизводительной системы КарНЦ РАН. Описания приводятся согласно отчетам пользователей Центра высокопроизводительной обработки данных.

- Численное моделирование крупномасштабной гидродинамики Белого моря

(к. ф.-м. н. И. А. Чернов, ИПМИ КарНЦ РАН)

В основу исследования положена модель крупномасштабной циркуляции Арктики, разработанная д. ф.-м. н. Н. Г. Яковлевым (ИВМ РАН, г. Москва) и адаптированная для Белого моря. Модель представляет собой систему дифференциальных уравнений в частных производных для полей скоростей течений, температуры и солености в области сложной формы, в ней учтены метеорологические данные, солнечная радиация, приливы, сток рек. Численное решение такой системы требует значительных вычислительных ресурсов.

В настоящее время модель воспроизводит качественное состояние вод и льда. Ведется разработка параллельной версии программной реализации модели.

- Свойства природных липидных мембран (д. ф.-м. н. А. Л. Рабинович, ИБ КарНЦ РАН, Д. В. Журкин, ПетрГУ)

  Проект направлен на исследование структурной организации и физических свойств природных мембран методами имитационного компьютерного моделирования. Основными задачами проекта являются:

  1. исследование компонентов природных мембран — липидных бислоев методом молекулярной динамики;
  2. исследование отдельных молекул липидных бислоев различными методиками метода Монте-Карло.

  При исследовании отдельных молекул липидных бислоев основной задачей является установление взаимосвязей и закономерностей между химической структурой и физическими свойствами молекул в разных условиях, а также их функциями в мембране.

  С использованием кластера были исследованы температурные зависимости некоторых свойств основных молекул липидных бислоев методом Монте-Карло в рамках простой и существенной выборок. Исследовались средние размеры молекул (средний квадрат расстояния между концевыми атомами цепи, средний квадрат радиуса инерции и т. п.), термодинамические свойства молекул (средние значения отдельных компонентов энергии конформации, теплоемкость и т. п.).

  Работа поддержана РФФИ, грант № 10-03-00201а, а также грантом Президента РФ для ведущих научных школ НШ-3731.2010.4 и Visby programme 00961/2008.

- Модели многосерверных систем обслуживания (д. ф.-м. н., проф. Е. В. Морозов, А. С. Румянцев, ИПМИ КарНЦ РАН)

  На основе классической модели Кифера-Вольфовица для вектора загруженности системы построена упрощенная модель вычислительного кластера. Исследуются моментные и корреляционные свойства векторов состояния системы, в т. ч. времени ожидания заявок в очереди, что в свою очередь позволяет характеризовать качество обслуживания в системе (QoS).

  Работа поддержана РФФИ, грант № 10-07-00017а.

- Прогнозирование состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантовохимических расчетов модельных систем (О. В. Кременецкая, ПетрГУ)

  Объектом исследования являются расплавы фторидов и хлоридов щелочных металлов Na, K, Cs, содержащих небольшие добавки фторидных и хлоридных комплексов переходных металлов. Интерес к данным объектам обусловлен тем, что они являются средой, широко используемой для технологических процессов, таких как получение чистых и высокочистых металлов; получение защитных и каталитически активных покрытий; синтез соединений, которые невозможно получить из водных и неводных сред при низких температурах.

  Знание состава комплексов в расплаве позволяет гораздо точнее моделировать химические реакции и перенос заряда в расплаве. А знание механизма и влияния на него разных факторов (состава, температуры, поверхности) даст возможность управления реакцией. Работы ведутся совместно с лабораторией высокотемпературной электрохимии Института химии Кольского НЦ РАН.

  Расчеты проводятся квантовохимическими методами HF, MP2, DFT с помощью программы Firefly. Экспериментально установлено, что процесс переноса заряда в исследуемых объектах протекает по-разному в зависимости не только от катионного, но и от анионного состава электролита расплава. Это обстоятельство делает необходимым при расчете различных характеристик включение в модельную систему не только второй, но и третьей координационной сферы комплексов переходных металлов.

Работа выполняется при финансовой поддержке РФФИ, грант № 08-03-00397а).

Всего со ссылкой на вычислительный кластер КарНЦ РАН пользователями ЦКП было опубликовано 13 статей в научных журналах и трудах конференций.

**Грид-сегмент КарНЦ РАН.** Другим перспективным способом организации ресурсов для проведения высокопроизводительных вычислений является технология Грид. В августе 2010 г. на базе ЦКП КарНЦ РАН был создан Грид-сегмент для использования свободных вычислительных ресурсов кластера, серверов и настольных компьютеров сотрудников ИПМИ КарНЦ РАН. В качестве промежуточного программного обеспечения Грид используется широко распространенная платформа BOINC [BOINC, 2011] с открытым исходным кодом. Платформа имеет архитектуру «клиент-сервер», при этом клиентская часть может работать на компьютерах с различным аппаратным и программным обеспечением. К маю 2011 г. в состав Грид были включены управляющий и вычислительные узлы кластера, четыре сервера и один персональный компьютер ИПМИ КарНЦ РАН (всего 16 вычислительных узлов). На текущий момент вычислительные ресурсы Грид доступны всем пользователям Центра высокопроизводительной обработки данных.

Для обеспечения доступа пользователей к ресурсам Грид-сегмента разработан и запущен в тестовую эксплуатацию веб-интерфейс для запуска вычислений в Грид-сегменте КарНЦ РАН. Помимо запуска вычислений, пользователи также имеют возможность получить через веб-интерфейс информацию о статусе выполнения рабочих заданий а также при необходимости добавить новые задания в Грид-проект.

На базе Грид нами разрабатывается «облачный» сервис организации и проведения математических вычислений. В основе «облачных» вычислений лежит подход, согласно которому пользователь по требованию получает удаленный доступ к вычислительным ресурсам в виде Интернет-сервиса, при этом структура организации ресурсов от него скрыта. Специальный веб-интерфейс позволяет «прозрачно» для пользователя инициировать Грид-проект, выполняющий на одном из вычислительных узлов заданную пользователем последовательность команд на математическом языке высокого уровня. В настоящее время поддерживается язык математического пакета Octave. Результаты вычислений также предоставляются пользователю через веб-интерфейс.

В конце 2010 — начале 2011 гг. организована тестовая эксплуатация Грид-сегмента и выполнен ряд расчетов в интересах пользователей ЦКП, проводящих научные исследования:

- для проекта по математическому моделированию кинетики гидридного фазового перехода (к. ф.-м. н. И. А. Чернов, ИПМИ КарНЦ РАН) на Грид было выполнено более 250 тыс. рабочих заданий, в рамках которых независимо друг от друга осуществлялась проверка наборов параметров, обеспечивающих совпадение модельной кривой с экспериментальной в пределах заданной погрешности. Обработка результатов вычислений на стороне Грид-сервера включала в себя автоматический отбор результатов, удовлетворяющих критерию, заданному пользователем;
- для проекта по прогнозированию состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантовохимических расчетов модельных систем (О. В. Кременецкая, ПетрГУ) вычисления проводились с использованием программного пакета для квантовохимических расчетов Firefly. При помощи Грид был проведен ряд экспериментов по расчету энергий систем частиц с различными параметрами;
- проведены первые расчеты в рамках проекта по разработке эволюционной модели и алгоритмов для решения потоковой задачи Штейнера в приложении к распределительным электрическим сетям (В. Д. Кукин, ИПМИ КарНЦ РАН).

**Заключение.** Следуя целям своего создания и общемировым тенденциям, ЦКП КарНЦ РАН стремится к развитию и популяризации высокопроизводительных вычислений, предоставлению своим пользователям удобного доступа, достаточных вычислительных ресурсов и современного высокопроизводительного программного обеспечения.

Таким образом, за период с момента создания ЦКП была проделана определенная научная, организационная и образовательная работа. Приобретение и освоение новых программных продуктов, отвечающих потребностям пользователей научных учреждений КарНЦ РАН, позволит повысить эффективность и расширить сферу применения кластера и системы хранения данных при проведении фундаментальных и прикладных научных исследований.

## Литература

*Положение о центре* коллективного пользования КарНЦ РАН «Центр высокопроизводительной обработки данных» // URL: http://cluster.krc.karelia.ru/doc/polozhenie.pdf (дата обращения 13.05.2011).

*Правила доступа* к кластеру и системе хранения данных ЦКП «Центр высокопроизводительных вычислений» КарНЦ РАН // URL: http://cluster.krc.karelia.ru/doc/pravila.pdf (дата обращения 13.05.2011).

*ЦКП КарНЦ РАН* Центр высокопроизводительной обработки данных // URL: http://cluster.krc.karelia.ru/ (дата обращения 13.05.2011).

BOINC: свободно распространяемое ПО для проведения Грид-вычислений // URL: http://boinc.berkeley.edu (дата обращения 13.05.2011).

*В. Т. Вдовицын, А. Д. Сорокин,*
*Е. Е. Ивашко, А. С. Румянцев,*
*Н. Н. Никитина*

# МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ «СТОХАСТИЧЕСКАЯ ТЕОРИЯ ОПТИМАЛЬНОЙ ОСТАНОВКИ»

(ПЕТРОЗАВОДСК, 12–16 СЕНТЯБРЯ 2010 Г.)

С 12 по 16 снтября 2010 г. в 35 километрах от г. Петрозаводска на скалистом берегу озера Укшезеро в отеле «Калевала» проводилась очередная международная конференция по теории оптимальной остановки. Конференция «Стохастическая теория оптимальной остановки» («Stochastic optimal stopping») была организована Институтом прикладных математических исследований Карельского научного центра РАН и Математическим институтом им. В. А. Стеклова РАН при поддержке Российского фонда фундаментальных исследований РАН и Президиума Российской академии наук.

Данная конференция продолжает серию совещаний по оптимальной остановке. Предыдущие проходили в Германии (1987), США (1990), Японии (1994) и Польше (2002). В России исследования по стохастическому анализу проводятся во многих университетах и исследовательских институтах. Первая в России конференция в этом направлении «Оптимальная остановка и стохастическое управление» была организована в г. Петрозаводске в 2005 г. Она вызвала большой интерес у специалистов в этой области, что привело к необходимости проведения еще одной встречи ученых. Наконец, настоящая конференция в Петрозаводске была шестой в данном направлении. В петрозаводской конференции приняли участие ученые из России, Германии, Японии, США, Польши, Великобритании и Китая. Общее число участников составило 51, из которых 11 зарубежных. На совещании присутствовало много молодых специалистов.

Программный комитет состоял из ведущих российских и зарубежных ученых. В него вошли: чл.-корр. РАН А. Н. Ширяев (председатель программного комитета, Математический институт им. В. А. Стеклова, Москва), В. В. Мазалов (ИПМИ КарНЦ РАН), А. Ирле (Университет Киель, Германия), Х. Д. Энгельберт (Университет им. Фридриха Шиллера, Германия), Х. Р. Лерхе (Университет Фрейбурга, Германия), Л. Шепп (Рутгерский университет, США), М. Тамаки (Университет Аичи, Япония), М. Ясуда (Университет Чиба, Япония), К. Шайовски (Технологический университет г. Вроцлав, Польша), Г. Пешкир (Университет Манчестера, Великобритания), Б. Оксендаль (Университет Осло, Норвегия), Е. Альтман (Национальный институт исследований в области вычислительной техники и автоматики, Франция), П. Салминен (Академический Университет Або, Финляндия). Тематика данной конференции была посвящена обсуждению новых результатов в области теории оптимальной остановки и стохастического управления, а также их применений к моделированию различных процессов экономики, социологии, компьтерных сетей и др. При этом работа велась по секциям: общие вопросы теории оптимальной остановки; правила многократной остановки и оптимальная остановка пространственных процессов; задачи наилучшего выбора и их обобщения; задачи оптимальной остановки с ограничениями; задача о разладке; стохастические игры; сетевые игры; оптимальная остановка и американские опционы; методы математического программирования в задачах оптимальной остановки. Значительное внимание на конференции было уделено управляемым случайным процессам (А. Н. Ширяев, А. В. Колногоров, А. С. Тихомиров). Особо стоит отметить пленарный доклад чл.-корр. РАН А. Н. Ширяева, который был посвящен применению управления для наискорейшего обнаружения изменения свойств процессов. Исследованию оптимальных правил остановки случайных процессов посвящены доклады Э. Л. Пресмана, К. Шайовского, В. В. Мазалова, Р. В. Иванова.

Большой интерес вызвали доклады, посвященные новым приложениям стохастического анализа в финансовой математике (А. Ирле, Л. Шепп, В. К. Доманский, В. Л. Крепс). Теория оптимальной остановки широко применяется при исследовании задач наилучшего выбора и задач о «разладке». Был представлен ряд работ, связанных с нахождением решений в новых задачах данной тематики (К. Ано, Т. Н. Шелонина, А. А. Ивашко, Е. Е. Ивашко, В. Г. Бурмистрова). Теоретико-игровым стохастическим моделям посвящены доклады Х. Лерхе, Е. В. Шевкопляс, И. А. Чернова. В последнее время большой интерес в науке вызвало новое направление, связанное с применением теоретико-игровых методов в информационных сетях. Оптимальному распределению потоков и маршрутизации был посвящен доклад Ю. В. Чуйко, стохастическим моделям Интернет-графов — доклады Ю. Л. Павлова и М. М. Лери. Наряду с работами известных ученых на конференции обсуждались доклады молодых участников (С. Ш. Кумачева, Е. С. Берникович, Н. В. Плаксина).

К началу конференции были изданы расширенные тезисы докладов. Презентации докладов участников выставлены на сайте http://mathem.krc.karelia.ru/event.php?id=132& plang=r. По результатам работы симпозиума принято решение подготовить к публикации «Труды конференции» и издать их в журнале «Математическая теория игр и ее приложения» (http://mgta.krc.karelia.ru).

*В. В. Мазалов, А. А. Ивашко*

Труды Карельского научного центра РАН
№ 5. 2011. С. 132–135

# ЮБИЛЕИ И ДАТЫ

## АНАТОЛИЙ ДМИТРИЕВИЧ СОРОКИН

**(к 75-летию со дня рождения)**

Сорокин Анатолий Дмитриевич родился 12 октября 1936 г. в г. Петрозаводске в семье слесаря Кировской железной дороги. После окончания железнодорожной средней школы № 9 в 1954 г. выбор будущей специальности решился однозначно – Ленинградский политехнический институт, электромеханический факультет. Закончив успешно вуз в 1960 г. и проработав 2 года в конструкторском бюро п/я 20 в г. Выборге, поступил в очную аспирантуру Карельского филиала АН ССР. После окончания аспирантуры в 1965 г. был распределен на работу в Карельский НИИ лесной промышленности, где проработал до 1975 г. сначала в должности старшего научного сотрудника, затем заведующего лабораторией автоматизации учета вычислительных работ.

В 1971 г. защитил кандидатскую диссертацию по специальности «Техническая кибернетика». В январе 1975 г. прошел по конкурсу на должность заведующего лабораторией автоматизации научных исследований Отдела математических методов и вычислительной техники КФ АН СССР, и с этого времени его научная и научно-организационная деятельность свыше 35 лет неразрывно связана с Карельским филиалом АН СССР.

Обладая хорошими организаторскими способностями как в научной, так и общественной жизни, удивительной коммуникабельностью в отношениях с сотрудниками, являясь исключительно добросовестным и инициативным работником, А. Д. Сорокин привлек внимание руководства Карельского филиала АН СССР и был выдвинут на должность заместителя председателя Президиума КФ АН СССР по научной работе. В этой должности А. Д. Сорокин проработал с 1977 по 1992 г., одновременно совмещая административную должность с заведыванием лабораторией.

Под его научным руководством и при непосредственном участии в Филиале было создано новое научное направление – автоматизация научных исследований, объединившее усилия многих и разнообразных специалистов Филиала в создании систем автоматизации экспериментов и обработки данных. Под руководством А. Д. Сорокина в филиале шло формирование центров коллективного пользования по хроматографии и радиоизотопным методам исследований с автоматизацией сбора и обработки данных. Он проводил активную работу по координации и кооперации научно-исследовательских работ в области информатизации и автоматизации научных исследований на всесоюзном уровне, участвовал в работе международных конгрессов и совещаний.

В 1989 г. решением Президиума АН СССР А. Д. Сорокину присвоено ученое звание старшего научного сотрудника по специальности 05.13.16 «применение вычислительной техники и математических методов в научных исследованиях».

С 1 февраля 1992 г., в связи с избранием в 1990 г. по конкурсу на должность заведующего Отделом ММАНИП Карельского научного центра РАН, А. Д. Сорокин полностью переходит на работу в Отдел, уделяя большое внимание подготовке и росту научных кадров, оснащенности Отдела техническими средствами, новому направлению исследований – информационно-телекоммуникационным системам, имеет грантовскую поддержку РФФИ, РГНФ и международного фонда ИНТАС.

При поддержке руководства КарНЦ РАН и Отделения математики РАН А. Д. Сорокин ведет большую подготовительную работу по преобразованию Отдела в институт. Отдел получил разрешение на подготовку аспирантов и соискателей ученой степени кандидата наук, с октября 1993 г. переведен на самостоятельный баланс с правами юридического лица, имеет свой Устав, гербовую печать, прошел аккредитацию.

Отдел осуществляет научно-методическое руководство работами по организации компьютерной сети КарНЦ РАН и созданию инфраструктуры доступа к сети Internet, поддержке WWW-сервера центра, имеет связь с вузовской наукой, проводит международные конференции и семинары, имеет финансовую поддержку фондов РФФИ, РГНФ, Правительства РФ, Института «Открытое общество».

В июне 1999 г. Постановлением Президиума РАН Отдел преобразован в Институт прикладных математических исследований КарНЦ РАН. А. Д. Сорокин назначен заместителем директора по научной работе, одновременно руководит научной темой по созданию и развитию электронных информационных ресурсов в КарНЦ РАН. Разработанный подход к созданию и использованию электронных научных информационных ресурсов в виде комплексной распределенной информационно-аналитической системы поддержки и сопровождения научных исследований представляется перспективным и соответствует современным отечественному и зарубежному уровням исследований и разработок в этой области. Было подготовлено ТЗ по приобретению кластера с системой хранения данных, в результате чего закуплен и запущен в эксплуатацию 10-узловой вычислительный кластер, создан центр коллективного пользования КарНЦ РАН «Центр высокопроизводительной обработки данных». Реализованный ряд организационно-технических мероприятий позволил добиться более надежного функционирования компьютерной сети КарНЦ РАН и обеспечить более высокую пропускную способность каналов связи.

А. Д. Сорокин имеет более 100 научных работ, из них 80 печатных, является соавтором 4-х электронных коллекций научных информационных ресурсов, зарегистрированных в депозитарии электронных изданий НТЦ «Информрегистр».

А. Д. Сорокин ведет большую общественную работу в республике. Он является членом Совета по информатизации при Главе РК, заместителем председателя научно-технического Совета по телекоммуникациям при Президиуме КарНЦ РАН.

За успехи в научной и научно-организационной деятельности А. Д. Сорокин награжден Почетной грамотой Президиума РАН и профсоюза работников РАН, Почетной грамотой РК, юбилейной медалью ФНПР «100 лет профсоюзам России», Медалью Ордена «За заслуги перед Отечеством 2 степени», ему присвоено звание «Заслуженный деятель науки РК».

А. Д. Сорокина всегда отличает ответственное отношение к работе, корректность и доброжелательность.

Анатолий Дмитриевич до сих пор «дружит» со спортом. С молодости он увлекся парусным спортом. Был «кэп»ом команды. Да и теперь обязательно раз в год вместе с такими же любителями уходит в недельный поход на яхте, «ловит ветер» на Онего, рыбачит. Коллектив института поздравляет юбиляра, желает ему здоровья и дальнейших творческих успехов.

### СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ А. Д. СОРОКИНА

**1967.** Структурно-топологический метод расчета усилителей низкой частоты на ЦВМ // Сборник научных работ. Петрозаводск: Карельское книжное издательство. (Совместно с А. М. Сучилиным).

**1969.** Применение ЦВМ для автоматизации топологического метода расчета электронных схем // Труды ЛПИ им. Калинина «Энергия». (Совместно с А. М. Сучилиным).

**1970.** Некоторые вопросы вывода отчетных документов по заработной плате с помощью ЭВМ «Минск-22» // Оптимальное планирование и управление лесопромышленными комплексами М. (Совместно с Р. А. Патёмой).

Механизация учета труда и заработной платы на базе ЭВМ // Лесная промышленность. № 12. М. (Совместно с В. А. Лебедевым).

**1973.** Эксплуатация системы механизации труда и заработной платы с помощью ЭВМ (на примере Суккозерского ЛПХ) // Автоматизированные системы управления и вопросы

экономики в лесной и лесоперерабатывающей промышленности. М. (Совместно с Л. Н. Курчиной, Л. Я. Храбуновой).

**1974.** Разработка первой очереди подсистемы бухгалтерского учета // Сборник научно-технических статей. Петрозаводск: Изд-во «Карелия».

**1975.** Решение задачи учета труда и заработной платы на ЭВМ. ВНИПИЭИлеспром // Экономика и управление. № 8. Москва. (Совместно с Р. Г. Безручко, Л. Я. Храбуновой).

**1978.** Автоматизация обработки графической информации с использованием ЭВМ «Минск-32» и АЦП типа Ф-001 / Оперативно-информационные материалы КФ АН СССР. Петрозаводск. (Совместно с В. П. Похлёбкиным).

**1979.** Об одном подходе к изучению свойств объектов автоматизации научных исследований / Оперативно-информационные материалы. Автоматизация обработки данных при исследовании сложных объектов. Петрозаводск: КФ АН СССР. (Совместно с М. И. Фёдоровым).

**1980.** Основные направления автоматизации научных исследований в Карельском филиале АН СССР // Автоматизация эксперимента и обработки данных. Петрозаводск: КФ АН СССР.

**1982.** Комплексная программа работ по автоматизации научных исследований в КФ АН СССР в 1981–1985 гг. и на период до 1990 г. Препринт. Петрозаводск: КФ АН СССР. (Совместно с Г. А. Борисовым, В. А. Лебедевым).

Автоматизация многопрофильных исследований // Труды I Междунар. школы по АНИ. (Совместно с Ю. Л. Павловым, М. И. Фёдоровым, В. В. Яковлевым).

Предпроектные исследования АСНИ регионального научного центра // Материалы XVI Всесоюз. школы по АСНИ. Институт прикладной физики АН СССР, Горький. (Совместно с В. А. Лебедевым, Ю. Л. Павловым).

**1983.** Опыт управления социалистическим соревнованием научных коллективов в Карельском филиале АН СССР. Региональные проблемы социалистического соревнования. ИАСЭП АН СССР, Ленинград. (Совместно с С. М. Яскуновым).

**1984.** Анализ использования технических средств и математических методов при автоматизации экспериментов. АСНИ регионального научного центра. Петрозаводск: Карельский филиал АН СССР. (Совместно с М. И. Фёдоровым, В. Н. Хариным).

АСНИ Карельского филиала АН СССР. Петрозаводск: Карельский филиал АН СССР. (Совместно с Г. А. Борисовым, В. А. Лебедевым).

Совершенствование управления качеством работы и социалистическим соревнованием в региональном научном центре. Петрозаводск: Карельский филиал АН СССР. (Совместно с С. М. Яскуновым).

**1986.** Подсистема автоматизации экспериментов на газожидкостных хроматографах. Информационный листок, Карельский ЦНТИ. № 24. Петрозаводск. (Совместно с В. П. Похлёбкиным, А. А. Степановым, М. И. Фёдоровым).

Подсистема автоматизации исследований эмпирических зависимостей. Разработка и внедрение комплексной автоматизированной системы научных исследований КФ АН СССР. Петрозаводск. (Совместно с Т. П. Лайкачёвой, Ю. Л. Павловым и др.).

Внедрение результатов научных исследований в производство // Годы свершений. Петрозаводск: Изд-во «Карелия». (Совместно с Г. С. Бискэ).

**1993.** Прибор для оперативного сбора информации в полевых условиях // Приборы и техника эксперимента. № 5. (Совместно с В. С. Клыпуто, Л. В. Солововой).

**1994.** Возможности и перспективы использования ЛВС в КНЦ РАН // Труды ОМАД КНЦ РАН. Вып. 1. Петрозаводск. (Совместно с В. С. Клыпуто, Н. П. Анисимовой, В. В. Яковлевым).

**1995.** Система баз по экологии в научном центре // Материалы междунар. конф. НТИ-95, ВИНИТИ. М. (Совместно с В. А. Лебедевым).

**1997.** Состояние и перспективы формирования и использования информационных ресурсов в Карельском научном центре РАН // Труды 3 Междунар. конф. «Информационные ресурсы. Интеграция. Технология». ВИНИТИ. М. (Совместно с В. Т. Вдовицыным).

**1998.** Состояние и перспективы развития информационно-телекоммуникационной среды КарНЦ РАН // Труды VI Санкт-Петербургской Междунар. конф. «Региональная информатика-98». СПб. (Совместно с В. Т. Вдовицыным).

**1999.** Вопросы организации функционирования информационно-телекоммуникационной сети KRCNet // Труды института прикладных математических исследований, вып. 1. Методы математического моделирования и информационные технологии. Петрозаводск. (Совместно с В. Т. Вдовицыным, В. С. Клыпуто).

**2000.** Создание и развитие электронных информационных ресурсов в КарНЦ РАН // Доклады второй Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Протвино. (Совместно с В. Т. Вдовицыным, Н. Б. Луговой).

**2001.** Электронная коллекция информационных ресурсов по топонимии Европейского Севера // Труды третьей Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». RCDL-2001. Петрозаводск. (Совместно с В. Т. Вдовицыным, Г. М. Кертом, Н. А. Беляевой и др.).

**2002.** Вопросы построения электронной библиотеки Карельского научного центра РАН // Труды четвертой Всерос. науч. конф. RCDL-2002. Дубна. (Совместно с В. Т. Вдовицыным).

Концепция развития системы телекоммуникаций КарНЦ РАН // Труды Института прикладных математических исследований КарНЦ РАН, вып. 3. Петрозаводск. (Совместно с В. Т. Вдовицыным, В. С. Клыпуто, А. А. Бедоревым и др.).

**2003.** Технология публикации и сопровождения документов в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН // Труды пятой Всерос. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». СПб. (Совместно с В. Т. Вдовицыным).

Разработка многоагентной системы для организации поиска данных в распределенной информационной системе // Труды Всерос. научн. конф. «Научный сервис в сети Интернет». Новороссийск. (Совместно с В. Т. Вдовицыным, Н. Б. Луговой, И. В. Чудаковой).

Вопросы построения единой информационной системы Карельского научного центра РАН // Труды «Современные технологии в информационном обеспечении науки». М.: Научный мир. (Совместно с В. Т. Вдовицыным).

Концепция создания единой информационной системы Карельского научного центра РАН // Труды ИПМИ КарНЦ РАН, вып. 4. Петрозаводск. (Совместно с В. Т. Вдовицыным).

**2004.** Вопросы формирования и использования электронных научных информационных ресурсов // Информационные ресурсы России. № 4. М. (Совместно с В. Т. Вдовицыным).

Электронная библиотека научных информационных ресурсов КарНЦ РАН: состояние и перспективы // Труды шестой Всерос. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». RCDL-2004. Пущино. (Совместно с В. Т. Вдовицыным, Н. Б. Луговой).

Развитие программных сервисов и контента ЭБ КарНЦ РАН // Труды седьмой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Ярославль. (Совместно с В. Т. Вдовицыным, Н. Б. Луговой).

**2006.** Электронные научные информационные ресурсы для поддержки инвестиционной деятельности в регионе // Информационные ресурсы России. № 4. М. (Совместно с А. В. Бархатовым, В. Т. Вдовицыным, Н. Б. Луговой).

Институт прикладных математических исследований // Академическая наука в Карелии. 1946–2006. Т. 2, Москва.: Наука. (Совместно с В. В. Мазаловым).

Электронные научные информационные ресурсы Карельского научного центра РАН: состояние, проблемы и перспективы // Труды Карельского научного центра РАН, вып. 9. Петрозаводск. (Совместно с В. Т. Вдовицыным).

**2007.** Институт прикладных математических исследований // Карелия. Энциклопедия. Т. 1. Петрозаводск.: Петропресс. (Совместно с Г. В. Воиновой).

**2010.** Юрий Васильевич Заика (к 50-летию со дня рождения) // Труды КарНЦ РАН. № 3. Сер. Математическое моделирование и информационные технологии, вып. 1. Петрозаводск.

*В. В. Мазалов, Г. В. Воинова*

# ПРАВИЛА ДЛЯ АВТОРОВ

(требования к работам, представляемым к публикации в «Трудах Карельского научного центра Российской академии наук»)

«Труды Карельского научного центра Российской академии наук» (далее – Труды КарНЦ РАН) публикуют результаты завершенных оригинальных исследований в различных областях современной науки: теоретические и обзорные статьи, сообщения, материалы о научных мероприятиях (симпозиумах, конференциях и др.), персоналии (юбилеи и даты, потери науки), статьи по истории науки. Представляемые работы должны содержать новые, ранее не публиковавшиеся данные.

Статьи проходят обязательное рецензирование. Решение о публикации принимается редакционной коллегией серии или тематического выпуска Трудов КарНЦ РАН после рецензирования, с учетом научной значимости и актуальности представленных материалов. Редколлегии серий и отдельных выпусков Трудов КарНЦ РАН оставляют за собой право возвращать без регистрации рукописи, не отвечающие настоящим правилам.

При получении редакцией рукопись регистрируется (в случае выполнения авторами основных правил ее оформления) и направляется на отзыв рецензентам. Отзыв состоит из ответов на типовые вопросы «Анкеты» и может содержать дополнительные расширенные комментарии. Кроме того, рецензент может вносить замечания и правки в текст рукописи. Авторам высылается электронная версия «Анкеты» и комментарии рецензентов. Доработанный экземпляр автор должен вернуть в редакцию вместе с первоначальным экземпляром и ответом на все вопросы рецензента не позднее, чем через месяц после получения рецензии. Перед сдачей в печать авторам высылается распечатанная версия статьи, которая вычитывается, подписывается авторами и возвращается в редакцию.

Почтовый адрес редакции: 185910, г. Петрозаводск, ул. Пушкинская, 11, КарНЦ РАН, редакция Трудов КарНЦ РАН. Телефон: (8142) 780109.

Содержание номеров Трудов КарНЦ РАН и другая полезная информация, включая настоящие Правила, доступна на сайте http://transactions.krc.karelia.ru.

## ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСИ

Статьи публикуются на русском или английском языках. Рукописи должны быть тщательно выверены и отредактированы авторами.

Статьи должны быть подписаны всеми авторами.

Объем рукописи (включая таблицы, список литературы, подписи к рисункам, рисунки) не должен превышать: для обзорных статей – 30 страниц, для оригинальных – 25, для сообщений – 15, для хроники и рецензий – 5–6. Объем рисунков не должен превышать 1/4 объема статьи. Рукописи большего объема (в исключительных случаях) принимаются при достаточном обосновании по согласованию с ответственным редактором.

Рукописи присылаются в электронном виде, а также в двух экземплярах, напечатанных на одной стороне листа формата А4 в одну колонку через 1,5 интервала (12 пунктов шрифта типа Times New Roman). Размер полей: сверху, снизу – 2,5 см, справа, слева – 2,5 см. Все страницы, включая список литературы и подписи к рисункам, должны иметь сплошную нумерацию в нижнем правом углу. Страницы с рисунками не нумеруются.

## ОБЩИЙ ПОРЯДОК РАСПОЛОЖЕНИЯ ЧАСТЕЙ СТАТЬИ

Элементы статьи должны располагаться в следующем порядке: *УДК* курсивом на первой странице, в левом верхнем углу; заглавие статьи на русском языке заглавными буквами полужирным шрифтом; инициалы, фамилии всех авторов на русском языке полужирным шрифтом; полное название организации – место работы каждого автора в именительном падеже на русском языке курсивом (если авторов несколько и работают они в разных учреждениях, то следует отметить арабскими цифрами соответствие фамилий авторов учреждениям, в которых они работают; если все авторы статьи работают в одном учреждении, можно не указывать место работы каждого автора отдельно); аннотация на русском языке; ключевые слова на русском языке; инициалы, фамилии всех авторов на английском языке полужирным шрифтом; название статьи на английском языке заглавными буквами полужирным шрифтом; аннотация на английском языке; ключевые слова на английском языке; текст статьи (статьи экспериментального характера, как правило, должны иметь разделы: ВВЕДЕНИЕ. МАТЕРИАЛЫ И МЕТОДЫ. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. ВЫВОДЫ. ЛИТЕРАТУРА); благодарности; литература (с новой страницы); таблицы (на отдельном листе); рисунки (на отдельном листе); подписи к рисункам (на отдельном листе).

На отдельном листе дополнительные сведения об авторах: фамилия, имя, отчество всех авторов полностью на русском и английском языках; полный почтовый адрес каждой организации (страна, город) на русском и английском языках; должности авторов; адрес электронной почты для каждого автора; телефон для контактов с авторами статьи (можно один на всех авторов).

ЗАГЛАВИЕ СТАТЬИ должно точно отражать содержание статьи*[1] и содержать не более 8–10 значащих слов.

АННОТАЦИЯ должна быть лишена вводных фраз, содержать только главную информацию статьи, не превышать объем – 15 строк.

Отдельной строкой приводится перечень КЛЮЧЕВЫХ СЛОВ. Ключевые слова или словосочетания отделяются друг от друга запятой, в конце фразы ставится точка.

МАТЕРИАЛЫ И МЕТОДЫ должны содержать сведения об объекте исследования с обязательным указанием латинских названий и сводок, по которым они приводятся, авторов классификаций и пр. Транскрипция географических названий должна соответствовать атласу последнего года издания. Единицы физических величин приводятся по Международной системе СИ. Желательна статистическая обработка всех количественных данных. Необходимо возможно точнее обозначать местонахождения (в идеале – с точным указанием географических координат).

ИЗЛОЖЕНИЕ РЕЗУЛЬТАТОВ должно заключаться не в пересказе содержания таблиц и графиков, а в выявлении следующих из них закономерностей. Автор должен сравнить полученную им информацию с имеющейся в литературе и показать, в чем заключается ее новизна. Для фаунистических и флористических работ следует указывать место хранения коллекционных образцов. Если в статье приводятся сведения о новых для исследованной территории таксонах, то желательно и процитировать этикетку. Следует ссылаться на табличный и иллюстративный материал так: на рисунки, фотографии и таблицы в тексте (рис. 1, рис. 2, табл. 1, табл. 2 и т. д.), фотографии, помещаемые на вклейках (рис. I, рис. II). Обсуждение завершается формулировкой основного вывода, которая должна содержать конкретный ответ на вопрос, поставленный во Введении. Ссылки на литературу в тексте даются фамилиями, например: Карху, 1990 (один автор); Раменская, Андреева, 1982 (два автора); Крутов и др., 2008 (три автора или более), и заключаются в квадратные скобки. При перечислении нескольких источников работы располагаются в хронологическом порядке, например: [Иванов, Топоров, 1965; Успенский, 1982; Erwin et al., 1989; Рыбаков, 1994; Longman, 2001].

ТАБЛИЦЫ нумеруются в порядке упоминания их в тексте, каждая таблица имеет свой заголовок. На полях рукописи (слева) карандашом указываются места расположения таблиц при первом упоминании их в тексте. Диаграммы и графики не должны дублировать таблицы. Материал таблиц должен быть понятен без дополнительного обращения к тексту. Все сокращения, использованные в таблице, должны быть пояснены в Примечании, расположенном под ней. При повторении цифр в столбцах нужно их повторять, при повторении слов – в столбцах ставить кавычки. Таблицы могут быть книжной или альбомной ориентации (при соблюдении вышеуказанных параметров страницы).

РИСУНКИ представляются отдельными файлами с расширением TIFF (*.TIF), или JPG (не встраивать в Word). Графические материалы должны быть снабжены распечатками с указанием желательного размера рисунка в книге, пожеланий и требований к конкретным иллюстрациям. На каждый рисунок должна быть как минимум одна ссылка в тексте. Иллюстрации объектов, исследованных с помощью фотосъемки, микроскопа (оптического, электронного трансмиссионного и сканирующего), должны сопровождаться масштабными линейками, причем в подрисуночных подписях надо указать длину линейки. Приводить данные о кратности увеличения необязательно, поскольку при публикации рисунков размеры изменятся. Крупномасштабные карты желательно приводить с координатной сеткой, обозначениями населенных пунктов и/или названиями физикогеографических объектов и разной фактурой для воды и суши. В углу карты желательна врезка с мелкомасштабной картой, где был бы указан участок, увеличенный в крупном масштабе в виде основной карты.

ПОДПИСИ К РИСУНКАМ должны содержать достаточно полную информацию, для того чтобы приводимые данные могли быть понятны без обращения к тексту (если эта информация уже не дана в другой иллюстрации). Аббревиации расшифровываются в подрисуночных подписях.

ЛАТИНСКИЕ НАЗВАНИЯ. В расширенных латинских названиях таксонов не ставится запятая между фамилией авторов и годом, чтобы была понятна разница между полным названием таксона и ссылкой на публикацию в списке литературы. Названия таксонов рода и вида печатаются курсивом. Вписывать латинские названия в текст от руки недопустимо. Для флористических, фаунистических и таксономических работ при первом упоминании в тексте и таблицах приводится русское название вида (если такое название имеется) и полностью – латинское, с автором и, желательно, с годом, например: водяной ослик (*Asellus aquaticus* (L. 1758). В дальнейшем можно употреблять только русское название или сокращенное латинское

---

* **Названия видов приводятся на латинском языке КУРСИВОМ**, в скобках указываются высшие таксоны (семейства), к которым относятся объекты исследования.

без фамилии автора и года опубликования, например, для брюхоногого моллюска *Margarites groenlandicus* (Gmelin 1790) – *M. groenlandicus* или для подвида *M. g. umbilicalis.*

СОКРАЩЕНИЯ. Разрешаются лишь общепринятые сокращения — названия мер, физических, химических и математических величин и терминов и т. п. Все сокращения должны быть расшифрованы, за исключением небольшого числа общеупотребительных.

БЛАГОДАРНОСТИ. В этой рубрике выражается признательность частным лицам, сотрудникам учреждений и фондам, оказавшим содействие в проведении исследований и подготовке статьи, а также указываются источники финансирования работы.

ЛИТЕРАТУРА. Пристатейные ссылки и/или списки пристатейной литературы следует оформлять по ГОСТ Р 7.0.52008. Библиографическая ссылка. Общие требования и правила составления (http://www.bookchamber.ru/GOST_P_7.0.5.2008). Список работ представляется в алфавитном порядке. Все ссылки даются на языке оригинала (названия на японском, китайском и других языках, использующих нелатинский шрифт, пишутся в русской транскрипции). Сначала приводится список работ на русском языке и на языках с близким алфавитом (украинский, болгарский и др.), а затем – работы на языках с латинским алфавитом. В списке литературы между инициалами ставится пробел.

ОБРАЗЕЦ ОФОРМЛЕНИЯ 1-Й СТРАНИЦЫ

*УДК 631.53.027.32: 635.63*

**ВЛИЯНИЕ РАЗЛИЧНЫХ РЕЖИМОВ ПРЕДПОСЕВНОГО ЗАКАЛИВАНИЯ СЕМЯН НА ХОЛОДОУСТОЙЧИВОСТЬ РАСТЕНИЙ ОГУРЦА**

**Е. Г. Шерудило[1], М. И. Сысоева[1], Г. Н. Алексейчук[2], Е. Ф. Марковская[1]**

[1] *Институт биологии Карельского научного центра РАН*
[2] *Институт экспериментальной ботаники НАН Республики Беларусь им. В. Ф. Купревича*

*Аннотация на русском языке*
Ключевые слова: *Cucumis sativus* L., кратковременное снижение температуры, устойчивость.

**E. G. Sherudilo, M. I. Sysoeva, G. N. Alekseichuk, E. F. Markovskaya. EFFECTS OF DIFFERENT REGIMES OF SEED HARDENING ON COLD RESISTANCE IN CUCUMBER PLANTS**

Аннотация на английском языке
Key words: *Cucumis sativus* L., temperature drop, resistance.

ОБРАЗЕЦ ОФОРМЛЕНИЯ ТАБЛИЦЫ

*Таблица 2*
**Частота встречаемости видов нематод в исследованных биотопах**

| Биотоп (площадка) | Кол-во видов | Встречаемость видов нематод | | | | |
|---|---|---|---|---|---|---|
| | | 100 % | 80 % | 60 % | 40 % | 20 % |
| 1Н | 26 | 8 | 4 | 1 | 5 | 8 |
| 2Н | 13 | 2 | 1 | 1 | 0 | 9 |
| 3Н | 34 | 13 | 6 | 3 | 6 | 6 |
| 4Н | 28 | 10 | 5 | 2 | 2 | 9 |
| 5Н | 37 | 4 | 10 | 4 | 7 | 12 |

*Примечание.* Здесь и в табл. 3–4: Биотоп 1Н – территория, заливаемая в сильные приливы; 2Н – постоянно заливаемый луг; 3Н – редко заливаемый луг; 4Н – незаливаемая территория; 5Н – периодически заливаемый луг.

ОБРАЗЕЦ ОФОРМЛЕНИЯ ПОДПИСИ К РИСУНКУ

*Рис. 1.* Северный точильщик (*Hadrobregmus confuses* Kraaz.)

ОБРАЗЕЦ ОФОРМЛЕНИЯ СПИСКА ЛИТЕРАТУРЫ

Ссылки на книги

*Вольф Г. Н.* Дисперсия оптического вращения и круговой дихроизм в органической химии / ред. Г. Снатцке. М.: Мир, 1970. С. 348–350.

*Илиел Э.* Стереохимия соединений углерода / пер. с англ. М.: Мир, 1965. 210 с.

*Несис К. Н.* Океанические головоногие моллюски: распространение, жизненные формы, эволюция. М.: Наука, 1985. 285 с.

*Knorre D. G., Laric O. L.* Theory and practice in affinity techniques / Eds. P. V. Sundaram, F. L. Eckstein. N. Y., San Francisco: Acad. Press, 1978. P. 169–188.

Ссылки на статьи

*Викторов Г. А.* Межвидовая конкуренция и сосуществование экологических гомологов у паразитических перепончатокрылых // Журн. общ. биол. 1970. Т. 31, № 2. С. 247–255.

*Grove D. J., Loisides L., Nott J.* Satiation amount, frequency of feeding and emptying rate in *Salmo gairdneri* // J. Fish. Biol. 1978. Vol. 12, N 4. P. 507–516.

Ссылки на материалы конференций

*Марьинских Д. М.* Разработка ландшафтного плана как необходимое условие устойчивого развития города (на примере Тюмени) // Экология ландшафта и планирование землепользования: тезисы докл. Всерос. конф. (Иркутск, 11–12 сент. 2000 г.). Новосибирск, 2000. С. 125–128**.**

Ссылки на авторефераты диссертаций

*Шефтель Б. И.* Экологические аспекты пространственно-временных межвидовых взаимоотношений землероек Средней Сибири: автореф. дис. …канд. биол. наук. М., 1985. 23 с.

Ссылки на диссертации

*Шефтель Б. И.* Экологические аспекты пространственно-временных межвидовых взаимоотношений землероек Средней Сибири: дис. …канд. биол. наук. М., 1985. С. 21–46.

Ссылки на патенты

Патент РФ № 2000130511/28, 04.12.2000.

*Еськов Д. Н., Серегин А. Г.* Оптико-электронный аппарат // Патент России № 2122745. 1998. Бюл. № 33.

Ссылки на архивные материалы

*Гребенщиков Я. П.* К небольшому курсу по библиографии: материалы и заметки, 26 февр. – 10 марта 1924 г. // ОР РНБ. Ф. 41. Ед. хр. 45. Л. 1–10.

Ссылки на Интернет-ресурсы

*Паринов С. И., Ляпунов В. М., Пузырев Р. Л.* Система Соционет как платформа для разработки научных информационных ресурсов и онлайновых сервисов // Электрон. б-ки. 2003. Т. 6, вып. 1. URL: http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP/ (дата обращения: 25.11.2006).

Ссылки на электронные ресурсы на CDROM

Государственная Дума, 19992003 [Электронный ресурс]: электронная энциклопедия/Аппарат Гос. Думы Федер. Собрания Рос. Федерации. М., 2004. 1 CDROM.

## CONTENTS

Научное издание

**Труды Карельского научного центра**
**Российской академии наук**

№ 5, 2011
Серия МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Вып. 2

*Печатается по решению*
*Президиума Карельского научного центра РАН*

Редактор М. А. Радостина
Оригинал-макет Е. Н. Спектор
Стилевой файл А. С. Румянцев

Подписано в печать 29.09.2011. Формат 60x84$^{1}/_{8}$.
Гарнитура CMR. Печать офсетная. Уч.-изд. л. 16,8. Усл. печ. л. 16,27.
Тираж 500 экз. Изд. № 198.  Заказ 985

Карельский научный центр РАН
Редакционно-издательский отдел
185003, г. Петрозаводск, пр. А. Невского, 50