МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА

Факультет
вычислительной математики
и кибернетики

Сборник статей молодых ученых факультета ВМК МГУ

Выпуск 7

МОСКВА — 2010

УДК 517.6+519.8
ББК 22
С23

Составители:
А. В. СТОЛЯРОВ, И. Г. ШЕВЦОВА

С23 **Сборник статей молодых ученых факультета ВМК МГУ** / Ред. совет: Ложкин С. А. и др. — М.: Издательский отдел факультета ВМК МГУ (лицензия ИД №05899 от 24.09.2001), выпуск №7, 2010 — 110 стр.

В настоящий сборник вошли статьи, выполненные молодыми учеными факультета Вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова в 2009–2010 г.г.

УДК 517.6+519.8
ББК 22

ISBN 978-5-89407-420-7

Данный выпуск посвящается

СОРОКАЛЕТИЮ ФАКУЛЬТЕТА ВМК МГУ

СОДЕРЖАНИЕ

УДК 519.1

О ЗАМКНУТЫХ ЛЕКСИКОГРАФИЧЕСКИХ ОТРЕЗКАХ В k–ЗНАЧНОЙ n–МЕРНОЙ РЕШЕТКЕ

© 2010 г. Т. В. Андреева, С. С. Лебедев

mathcyb@cs.msu.su

Кафедра математической кибернетики

1 Введение

В работах [5], [6] А. А. Сапоженко получил асимптотику числа антицепей в ранжированных частично упорядоченных множествах. Существенную роль в решении этой задачи играет оценка числа замкнутых 2–связных множеств в двудольных графах ([4]).

В работе [1] Т. В. Андреева исследовала вопрос о числе антицепей в n–мерной трехзначной решетке E_3^n. Была получена нижняя оценка, предположительно являющаяся асимптотически точной. Для верхней оценки достаточно получить оценку числа замкнутых 2–связных множеств с заданной мощностью границы в слоях E_3^n. В данной работе рассматриваются лексикографические отрезки в слоях n–мерной k–значной решетки E_k^n для $k \geq 2$.

Пусть на множестве $E_k = \{0, 1, \ldots, k-1\}$ задано отношение порядка "$<$", при котором $0 < 1 < \ldots < k-1$.

Множество $E_k^n = \{\tilde{a} = (a_1, \ldots, a_n) : a_r \in E_k, r = 1, \ldots, n\}$ с отношением частичного порядка "$\leq$" называется *n–мерной k–значной решеткой*. Для сокращения записи в дальнейшем индекс k будем опускать.

Множество $E_i^n = \left\{\tilde{a} \in E^n : \sum_{r=1}^n a_r = i\right\}$ называется *i–ым слоем E^n*.

Отношение "$\prec$" на множестве E^n, при котором $\tilde{a} = (a_1, \ldots, a_n) \prec \tilde{b} = (b_1, \ldots, b_n)$, если либо $a_1 < b_1$, либо $a_1 = b_1, \ldots, a_{t-1} = b_{t-1}$ и $a_t < b_t$ для некоторого $t \geq 2$, называется *отношением лексикографического порядка*. Будем использовать обозначение $\tilde{a} \preceq \tilde{b}$, если либо $\tilde{a} = \tilde{b}$, либо $\tilde{a} \prec \tilde{b}$.

Множество A, состоящее из идущих подряд в лексикографическом порядке $|A|$ вершин слоя E_i^n, называется *лексикографическим отрезком* в E_i^n, сокращенно ЛО. ЛО A называется *левым* (*правым*) в E_i^n, если он состоит из первых (соответственно, последних) вершин слоя E_i^n. Сокращенно будем называть его ЛЛО (ПЛО).

Множество $\partial A = \{\tilde{u} \in E_{i-1}^n : \exists \tilde{a} \in A, \tilde{u} \leq \tilde{a}\}$ называется *границей множества* $A \subseteq E_i^n$. Положим $\partial\tilde{a} = \partial\{\tilde{a}\}$.

Обозначим ЛЛО длины m в слое E_i^n через $L_i(m)$. Из работы Клемента и Линдстрома [8] следует

Теорема 1. *Для произвольного $A \subseteq E_i^n$ $(1 \leq i \leq 2n)$ справедливо*

(1) $|\partial A| \geq \big|\partial L_i\big(|A|\big)\big|$;

(2) *если* $A = L_i\big(|A|\big)$, *то* $\partial A = L_{i-1}\big(|\partial A|\big)$.

Замыканием $A \subseteq E_i^n$ называется множество $[A] = \{\tilde{a} \in E_i^n : \partial\tilde{a} \subseteq \partial A\}$. Множество A называется *замкнутым*, если $A = [A]$.

Пусть $A \subseteq E_i^n$. Вершину $\tilde{a} \in A$ такую, что для любого $\tilde{b} \in A$ выполнено $\tilde{b} \preceq \tilde{a}$, назовем *правой вершиной множества* A и обозначим $r(A)$. Вершину $\tilde{a} \in A$ такую, что для любого $\tilde{b} \in A$ выполнено $\tilde{a} \preceq \tilde{b}$, назовем *левой вершиной множества* A и обозначим $l(A)$.

Пусть $\tilde{\sigma} = (\sigma_1, \ldots, \sigma_p) \in E^p$. *Гранью* E^n назовем множество $E_{\tilde{\sigma}}^n = \{\tilde{a} \in E^n : a_r = \sigma_r,\ r = 1, \ldots, p\}$. Из определения грани следует, что

$$E^n = E_{(0)}^n \cup \ldots \cup E_{(k-1)}^n.$$

2 О свойствах правых лексикографических отрезков

Положим $RS_i = \{\tilde{a} = (a_1, \ldots, a_n) \in E_i^n : a_1 \geq 1\}$.

Лемма 1. *Для любых вершин $\tilde{a}, \tilde{b} \in RS_i$ $(1 \leq i \leq 2n-1)$ таких, что $\tilde{a} \prec \tilde{b}$, выполнено $l(\partial\tilde{a}) \prec l(\partial\tilde{b})$.*

Доказательство. Пусть $\tilde{a} = (a_1 \ldots, a_n)$ и $\tilde{b} = (b_1, \ldots, b_n)$, при этом $a_1 \geq 1$, $b_1 \geq 1$. Тогда

$$l(\partial\tilde{a}) = (a_1 - 1, a_2, \ldots, a_n), \qquad l(\partial\tilde{b}) = (b_1 - 1, b_2, \ldots, b_n).$$

Отсюда и из определения лексикографического порядка следует утверждение леммы.

Теорема 2. *ПЛО $A \subseteq E_i^n$ $(1 \leq i \leq 2n-1)$ замкнут тогда и только тогда, когда $A \subset RS_i$.*

Доказательство. Множество RS_i является ПЛО, это следует из определения лексикографического порядка.
Необходимость. Из определения следует, что

$$RS_i = (E_{(1)}^n \cup \ldots \cup E_{(k-1)}^n) \cap E_i^n.$$

При $\sigma \geq 1$ имеем

$$\partial(E_{(\sigma)}^n \cap E_i^n) = (E_{(\sigma-1)}^n \cup E_{(\sigma)}^n) \cap E_{i-1}^n.$$

Таким образом,

$$\partial RS_i = \bigcup_{\sigma=1}^{k-1} \partial(E_{(\sigma)}^n \cap E_i^n) = \left(\bigcup_{\sigma=0}^{k-1} E_{(\sigma)}^n\right) \cap E_{i-1}^n = E_{i-1}^n = \partial E_i^n.$$

Значит, $[RS_i] = E_i^n$. Следовательно, если $RS_i \subseteq A \subset E_i^n$, то $[A] = E_i^n$, т.е. множество A не является замкнутым.

Достаточность. Из леммы 1 следует, что для ПЛО A и B таких, что $B \subset A \subset RS_i$ выполнено $l(\partial A) \prec l(\partial B)$, т.е. $\partial A \neq \partial B$. Значит, каждый такой ПЛО является замкнутым.

Теорема доказана.

Теорема 3. *Количество замкнутых ПЛО в слое E_i^n $(1 \leq i \leq 2n-1)$ равно $|E_i^n| - |E_i^{n-1}| - 1$.*

Доказательство. В силу теоремы 2 количество замкнутых ПЛО в слое E_i^n равно $|RS_i| - 1$. Из определения RS_i следует, что

$$|RS_i| = |E_i^n \setminus E_{(0)}^n| = |E_i^n| - |E_i^n \cap E_{(0)}^n| = |E_i^n| - |E_i^{n-1}|.$$

Положим $RSS_i = \{\tilde{a} = (a_1, \ldots, a_n) \in E_i^n : a_1 = k-1, a_2 \geq 1\}$.

Теорема 4. *Пусть множество A является ПЛО в слое E_i^n $(1 \leq i \leq 2n-1)$. Множество ∂A является ПЛО в слое E_{i-1}^n тогда и только тогда, когда $RSS_i \subseteq A$.*

Доказательство. Из определения лексикографического порядка следует, что RSS_i является ПЛО.
Необходимость. Пусть $A \subset RRS_i$. Положим $\tilde{b} = l(RSS_i)$ $\tilde{w} = l(\partial A)$. Из определения следует, что $\tilde{b} = (k-1, 1, b_3, \ldots, b_n)$, $\tilde{w} = (k-2, w_2, \ldots, w_n)$. Если ∂A является ПЛО, то для каждой вершины $\tilde{v} \in E_{i-1}^n$ такой, что $\tilde{w} \prec \tilde{v}$, найдется вершина $\tilde{a} \in A$, для которой $\tilde{v} \in \partial\tilde{a}$.

Покажем что это не так. Рассмотрим вершину $\tilde{u} = (k-1, 0, b_3, \ldots, b_n) \in E_{i-1}^n$. Очевидно, что $\tilde{w} \prec \tilde{u}$. Положим $\nabla\tilde{u} = \{\tilde{c} \in E_i^n : \tilde{u} \in \partial\tilde{c}\}$. Если $\tilde{c} \in \nabla\tilde{u}$, то либо $\tilde{c} = \tilde{b}$, либо $\tilde{c} = (k-1, 0, c_3, \ldots, c_n)$. В обоих случаях $\tilde{c} \notin A$. Значит, $\tilde{A}$ не является ПЛО.

Достаточность. Если $RS_i \subseteq A$, то из теоремы 2 следует, что $\partial A = E_{i-1}^n$, т.е. является ПЛО. Предположим теперь, что $A \subset RS_i$.

1. Покажем сначала, что утверждение справедливо для $A = RSS_i$. Из определения следует, что

$$RSS_i = \left(\bigcup_{\sigma=1}^{k-1} E^n_{(k-1,\sigma)} \right) \cap E^n_i.$$

При $\sigma \geq 1$ имеем

$$\partial E^n_{(k-1,\sigma)} = (E^n_{(k-2,\sigma)} \cup E^n_{(k-1,\sigma-1)} \cup E^n_{(k-1,\sigma)}) \cap E^n_{i-1}.$$

Следовательно,

$$\partial RSS_i = \left(\bigcup_{\sigma=1}^{k-1} E^n_{(k-2,\sigma)} \cup \bigcup_{\sigma=0}^{k-1} E^n_{(k-1,\sigma)} \right) \cap E^n_{i-1} = \left(E^n_{(k-1)} \cup E^n_{(k-2)} \setminus E^n_{(k-2,0)}\right) \cap E^n_{i-1}.$$

Из определения лексикографического порядка следует, что ∂RSS_i является ПЛО.

2. Пусть теперь $RSS_i \subset A$, $l(A) = \tilde{a} = (a_1, \ldots, a_n)$ и D — ПЛО в слое E^n_{i-1} такой, что $l(D) = l(\partial\tilde{a})$. Поскольку $A \subset RS_i$, в силу теоремы 2 достаточно показать, что $D = \partial A$.

Из леммы 1 следует, что $\partial RSS_i \subset \partial A \cap D$. Значит, достаточно показать, что для всякой вершины $\tilde{u} = (u_1, \ldots, u_n) \in D \setminus \partial RSS_i$ найдется вершина $\tilde{b} \in A$ такая, что $\tilde{u} \in \partial\tilde{b}$.

Поскольку $\tilde{a} \in RS_i$, выполнено $a_1 \geq 1$, поскольку $\tilde{u} \notin \partial RSS_i$, выполнено $u_1 \leq k-2$. Из определения лексикографического порядка следует, что если

$$l(\partial\tilde{a}) = (a_1 - 1, a_2, \ldots, a_n) \prec (u_1, u_2, \ldots, u_n) = \tilde{u},$$

то

$$\tilde{a} = (a_1, a_2, \ldots, a_n) \prec (u_1 + 1, u_2, \ldots, u_n).$$

Положим $\tilde{b} = (u_1 + 1, u_2, \ldots, u_n)$. Очевидно, что $\tilde{u} \in \partial\tilde{b}$.

Теорема доказана.

Из теоремы 4 при $k = 3$ следует теорема 1 из работы [2].

3 О свойствах левых лексикографических отрезков

Лемма 2. *Для любых вершин $\tilde{a}, \tilde{b} \in E^n_i$ $(1 \leq i \leq 2n-1)$ таких, что $\tilde{a} \prec \tilde{b}$, выполнено $r(\partial\tilde{a}) \preceq r(\partial\tilde{b})$.*

Доказательство. Если вершина $\tilde{a} \in E^n_i$ имеет вид $\tilde{a} = (a_1, \ldots, a_m, 0, \ldots, 0)$, где $a_m > 0$ $(1 \leq m \leq n)$, то $r(\partial\tilde{a}) = (a_1, \ldots, a_{m-1}, a_m - 1, 0, \ldots, 0)$.

Из определения лексикографического порядка следует, что если $\tilde{a} \prec \tilde{b}$, то $\tilde{b}$ имеет вид $\tilde{b} = (a_1, \ldots, a_{t-1}, b_t, \ldots, b_l, 0, \ldots, 0)$, где $b_l > 0$, $b_t > a_t$ $(1 \leq t \leq l \leq n)$.

1. $l > t$. В этом случае

$$\begin{aligned} r(\partial\tilde{a}) &= (a_1, \ldots, a_{t-1}, a_t, \ldots, a_{m-1}, a_m - 1, 0, \ldots, 0), \\ r(\partial\tilde{b}) &= (a_1, \ldots, a_{t-1}, b_t, \ldots, b_{l-1}, b_l - 1, 0, \ldots, 0). \end{aligned}$$

Утверждение следует из того, что $a_t < b_t$.

2. $l = t$. В этом случае

$$\begin{aligned} r(\partial\tilde{a}) &= (a_1, \ldots, a_{t-1}, a_t, \ldots, a_{m-1}, a_m - 1, 0, \ldots, 0), \\ r(\partial\tilde{b}) &= (a_1, \ldots, a_{t-1}, b_t - 1, 0, \ldots, 0). \end{aligned}$$

Если $a_t < b_t - 1$, то из условия и определения лексикографического порядка следует, что $r(\partial\tilde{a}) \prec r(\partial\tilde{b})$.

Если $a_t = b_t - 1$, то $\tilde{a} = (a_1, \ldots, a_{t-1}, b_t - 1, 0, \ldots, 0, 1, 0, \ldots, 0)$, поскольку $\tilde{a}$ и $\tilde{b}$ лежат в одном слое. Следовательно, $r(\partial\tilde{a}) = r(\partial\tilde{b}) = (a_1, \ldots, a_{t-1}, b_t - 1, 0, \ldots, 0)$.

Неравенство $a_t > b_t - 1$ не имеет места, т.к. в этом случае для целых a_t, b_t должно выполняться неравенство $b_t - 1 < a_t < b_t$.

Таким образом, лемма доказана.

Следствие 1. *Пусть вершины $\tilde{a}, \tilde{b} \in E_i^n$ $(1 \leq i \leq 2n-1)$ таковы, что $\tilde{a} \prec \tilde{b}$. Равенство $r(\partial\tilde{a}) = r(\partial\tilde{b})$ имеет место тогда и только тогда, когда вершины $\tilde{a}, \tilde{b}$ имеют следующий вид:*

$$\tilde{a} = (a_1, \ldots, a_t, \sigma, 1, \underbrace{0, \ldots, 0}_{n-t-2}), \qquad \tilde{b} = (a_1, \ldots, a_t, \sigma + 1, \underbrace{0, 0, \ldots, 0}_{n-t-1}),$$

где $\sigma \in \{0, 1, \ldots, k-2\}$, $0 \leq t \leq n-2$.

Теорема 5. *ЛЛО A в слое E_i^n $(1 \leq i \leq 2n-1)$ замкнут тогда и только тогда, когда не найдется такого t $(0 \leq t \leq n-2)$, что для некоторого $\sigma \in \{0, 1, \ldots, k-2\}$ выполнено*

$$r(A) = (a_1, \ldots, a_t, \sigma, 1, \underbrace{0, \ldots, 0}_{n-t-2}). \tag{1}$$

Доказательство. Положим $r(A) = \tilde{a}$.
Необходимость. Пусть ЛЛО A замкнут. Предположим, найдется номер t $(0 \leq t \leq n-2)$ такой, что для некоторого $\sigma \in \{0, 1, \ldots, k-2\}$ имеет место $\tilde{a} = (a_1, \ldots, a_t, \sigma, 1, 0, \ldots, 0)$.

В силу теоремы 1 множество ∂A является ЛЛО. Рассмотрим $\tilde{b} = (a_1, \ldots, a_t, \sigma + 1, 0, \ldots, 0)$. Из следствия 1 вытекает, что $\partial\tilde{b} \subseteq \partial A$. Кроме того, очевидно, что $\tilde{a} \prec \tilde{b}$, следовательно, $\tilde{b} \notin A$. Это противоречит тому, что множество A замкнуто.

Достаточность. Пусть для $r(A)$ не найдется номера t такого, что выполнено (1). Предположим, $A \neq [A]$, т.е. найдется вершина $\tilde{b} \in E_i^n \setminus A$ такая, что $\partial\tilde{b} \subseteq \partial A$.

Поскольку $\tilde{a} \prec \tilde{b}$, из леммы 2 следует, что $r(\partial\tilde{a}) \preceq r(\partial\tilde{b})$. В силу условия и следствия 1 равенство невозможно. Значит, $\partial\tilde{b} \setminus \partial A \neq \emptyset$, что противоречит предположению.

Теорема доказана.

Теорема 6. *Количество замкнутых ЛЛО в слое E_i^n $(1 \leq i \leq 2n-1)$ равно $|E_i^n| - |E_i^{n-1}|$.*

Доказательство. Количество ЛЛО в слое E_i^n равно $|E_i^n|$.

Подсчитаем количество ЛЛО, не являющихся замкнутыми. По теореме 5 такими являются те и только те ЛЛО A, для которых выполнено (1) при некоторых $0 \leq t \leq n-2$ и $\sigma \in \{0, 1, \ldots, k-2\}$.

Каждой вершине

$$\tilde{a} = (a_1, \ldots, a_t, \sigma, 1, \underbrace{0, \ldots, 0}_{n-t-2}) \in E_i^n$$

однозначно сопоставляется вершина

$$\tilde{b} = (a_1, \ldots, a_t, \sigma + 1, \underbrace{0, \ldots, 0}_{n-t-2}) \in E_i^{n-1}.$$

Наоборот, каждая вершина $\tilde{b} \in E_i^{n-1}$ имеет вид $\tilde{b} = (b_1, \ldots, b_t, b_{t+1}, \underbrace{0, \ldots, 0}_{n-t-2})$, где $b_{t+1} > 0$, $0 \leq t \leq n-2$. Ей однозначно сопоставляется вершина $\tilde{a} = (b_1, \ldots, b_t, b_{t+1} - 1, 1, \underbrace{0, \ldots, 0}_{n-t-2}) \in E_i^n$.

Количество вершин $\tilde{b}$ равно мощности слоя E_i^{n-1}. Следовательно, количество ЛЛО, не являющихся замкнутыми, в слое E_i^n равно $|E_i^{n-1}|$.

Отсюда следует утверждение теоремы.

4 О свойствах лексикографических отрезков

Из теорем 2 и 5 вытекает

Теорема 7. *ЛО A в слое E_i^n $(1 \le i \le 2n-1)$ замкнут тогда и только тогда, когда $A \subset RS_i$ и не найдется такого t $(0 \le t \le n-2)$, что для некоторого $\sigma \in \{0, 1, \ldots, k-2\}$ выполнено* (3).

Из теорем 3, 6, и 7 вытекает

Теорема 8. *Количество замкнутых ЛО в слое E_i^n $(1 \le i \le 2n-1)$ равно*

$$|E_i^n| - 2|E_i^{n-1}| + |E_i^{n-2}| - 1.$$

Пусть $G_A = (A, E_G)$ — граф, в котором $\{\tilde{a}, \tilde{b}\} \in E_G$ тогда и только тогда, когда $\partial\tilde{a} \cap \partial\tilde{b} \neq \emptyset$. Множество A называется 2–*связным*, если граф G_A является связным.

Расстоянием между вершинами $\tilde{a} = (a_1, \ldots, a_n)$ и $\tilde{b} = (b_1, \ldots, b_n)$ называется величина $\rho(\tilde{a}, \tilde{b}) = \sum_{r=1}^{n} |a_r - b_r|$.

Заметим, что $\partial\tilde{a} \cap \partial\tilde{b} \neq \emptyset$ для $\tilde{a}, \tilde{b} \in E_i^n$ тогда и только тогда, когда $\rho(\tilde{a}, \tilde{b}) = 2$.

Для $\tilde{a}, \tilde{\tau} \in E^n$ таких, что $\tilde{\tau} \le \tilde{a}$ положим

$$\tilde{a} - \tilde{\tau} = (a_1 - \tau_1, a_2 - \tau_2, \ldots, a_n - \tau_n).$$

Пусть $A \subset E_i^n$ и $\tau_r^*(A) = \min\limits_{(a_1,\ldots,a_n)\in A} a_r$, $r = 1, \ldots, n$. Положим $\tilde{\tau}^*(A) = \big(\tau_1^*(A), \ldots, \tau_n^*(A)\big)$. Для $\tilde{\tau} \le \tilde{\tau}^*(A)$ пусть $\widehat{A}_{\tilde{\tau}} = \{\tilde{a} - \tilde{\tau} :\ \tilde{a} \in A\}$. Скажем, что множество $\widehat{A}_{\tilde{\tau}}$ *получено из множества A вычитанием набора $\tilde{\tau}$.*

Замечание 1. Множество A является 2–связным тогда и только тогда, когда множество $\widehat{A}$, полученное из него вычитанием некоторого набора $\tilde{\tau}$, является 2–связным. Это следует из того, что $\rho(\tilde{a}, \tilde{b}) = \rho(\tilde{a} - \tilde{\tau}, \tilde{b} - \tilde{\tau})$.

Замечание 2. Если множество A является замкнутым, то множество $\widehat{A}$ может не являться замкнутым. Например, множество $A = \{(1,1,1,2), (1,1,2,1), (1,2,1,1)\}$ является 2–связным и замкнутым в $E_{3;5}^4$. Множество $C = \{(0,0,0,1), (0,0,1,0), (0,1,0,0)\}$, полученное из A вычитанием набора $(1,1,1,1)$, является ЛЛО в $E_{3;1}^4$, 2–связно, но не замкнуто.

Пусть $(\pi_1, \pi_2, \ldots, \pi_n)$ — некоторая подстановка на множестве $\{1, 2, \ldots, n\}$, и пусть $A \subset E_i^n$. Скажем, что множество

$$\widetilde{A} = \{(a_{\pi_1}, a_{\pi_2}, \ldots, a_{\pi_n}) :\ (a_1, a_2, \ldots, a_n) \in A\}$$

изоморфно множеству A в смысле перестановки координат.

Замечание 3. Множество A является 2–связным тогда и только тогда, когда множество $\widetilde{A}$, изоморфное ему в смысле перестановки координат, является 2–связным.

Замечание 4. Множество A является замкнутым тогда и только тогда, когда множество $\widetilde{A}$ является замкнутым.

Рассмотрим множество $B = \{(1,1,1,2),\ (1,2,0,2),\ (2,0,1,2),\ (2,1,0,2),\ (2,1,1,1),\ (2,1,2,0)\} \subseteq E_{3;5}^4$. Оно является 2–связным, замкнутым, но не изоморфно в смысле перестановки координат ни одному из ЛО. Кроме того, никакой ЛО в слоях E_3^4 не может быть получен из B вычитанием некоторого набора $\tilde{\tau}$.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 10-01-00768.

Список литературы

[1] Т. В. Андреева *Развитие метода граничных функционалов и его приложение к комбинаторным задачам*, Математические вопросы кибернетики. Вып.13: Сборник статей / Под ред. О.Б. Лупанова. – М.: Физматлит, 2004. С. 147–222.

[2] К. М. Иванкин *Оценка числа антицепей в трехзначной n–мерной решетке*, Москва, 2008 (Дипломная работа).

[3] Н. Н. Катериночкина *Некоторые соотношения для подмножеств слоев n–мерной k–значной решетки*, Ж. вычисл. матем. и матем. физ., 1984 — Т. 24, N 5. — С. 782 – 786.

[4] А. А. Сапоженко *О числе связных подмножеств с заданной мощностью границы в двудольных графах*, Методы дискретного анализа в решении комбинаторных задач – Новосибирск – 1987 – вып. 45, – С. 42 – 70.

[5] А. А. Сапоженко, *О числе антицепей в ранжированных частично упорядоченных множествах*, Дискретная математика – М.: Наука – 1989 – т. 1, вып. 1, – С. 74 – 93.

[6] А. А. Сапоженко, *О числе антицепей в многослойных ранжированных частично упорядоченных множествах*, Дискретная математика – М.: Наука – 1989 – т. 1, вып. 2, – С. 110 – 128.

[7] А. А. Сапоженко *Проблема Дедекинда и метод граничных функционалов*, Математические вопросы кибернетики, Вып. 9 – М.: Наука, 2000, с. 161 – 220.

[8] G. Clements, B. Lindstrom *A generalization of a combinatorial theorem of Macaulay*, J. Comb. Theory, 1968, v.7, N2, p. 230–238.

УДК 517.955.8

ИССЛЕДОВАНИЕ КАЧЕСТВЕННЫХ СВОЙСТВ РЕШЕНИЙ ОДНОГО НЕЛИНЕЙНОГО СОБОЛЕВСКОГО УРАВНЕНИЯ

© 2010 г. А. И. Аристов

Кафедра Общей математики

1 Введение

Работа посвящена исследованию качественных свойств решений следующего уравнения:

$$\frac{\partial}{\partial t}(\Delta u - u) + a(\Delta u - u) + F(u) = 0, \tag{1}$$

где $F(u)$ – нелинейности специального вида.

Уравнения такого вида используются для описания давления жидкости в пористой среде или потенциала в полупроводнике.

Здесь исследованы задача Коши и начально-краевая задача для уравнения (1). Для задачи Коши найдены условия глобального по времени существования и единственности решения, исследовано асимптотическое поведение решений при больших временах. Для начально-краевой задачи найдены достаточные условия локальной по времени однозначной разрешимости, найдены в виде явных формул двусторонние оценки времени опрокидывания решения (т. е. обращения в бесконечность нормы решения в пространстве H_0^1), а также указаны достаточные условия глобального по времени существования и ограниченности решений.

В исследованиях обеих задач речь идет о стремлении к бесконечности какой-то физической величины (времени или функционала энергии), что существенно затрудняет численный анализ. Поэтому целесообразно выводить асимптотические формулы для ограниченных глобально по времени решений и оценки времени разрушения для опрокидывающихся.

2 Задача Коши

Рассмотрим следующую задачу Коши:

$$\begin{cases} \frac{\partial}{\partial t}(\Delta u - u) + a(\Delta u - u) + \mu(x)|u|^{\sigma} u = 0 \\ u(x,0) = u_0(x). \end{cases} \tag{2}$$

Здесь u– действительнозначная функция, зависящая от пространственной переменной $x \in R^N$ (N – любое натуральное число) и времени $t > 0$. Будем предполагать, что $\mu(\cdot) \in L_q(R^N)$, где множество возможных q зависит от размерности пространственной переменной следующим образом:

1. Если $N \leqslant 3$, то $2 \leqslant q \leqslant \infty$.
2. Если $N \geqslant 4$, то $N/2 < q \leqslant \infty$.

Пусть, кроме того, $a > 0$ и $\sigma \geqslant 1$.

В данном разделе получают развитие идеи, ранее использовавшиеся автором в [1] и в [2] для исследования задачи Коши для уравнения с нелинейностью в виде степенного ряда по u.

Будет дано определение обобщенного решения, затем будет показано, что если начальные данные достаточно малы в смысле некоторой нормы, то обобщенное решение существует и единственно, причем нелинейность не влияет на качественное поведение решения, определяя только порядок стремления к нулю остаточного члена.

Замена $u = ve^{-at}$ сводит задачу к следующей:

$$\left\{ \begin{array}{r} \frac{\partial}{\partial t}\left(\Delta v - v\right) + e^{-\sigma a t}\mu\left(x\right)\left|v\right|^{\sigma} v = 0 \\ v\left(x, 0\right) = u_0\left(x\right). \end{array} \right. \tag{3}$$

Введем обозначения, необходимые для изложения рассуждений:

$$\begin{array}{l} L_q = L_q\left(R^N\right),\ 1 \leqslant q \leqslant \infty, H^2 = H^2\left(R^N\right), \\ \overline{\mathrm{B}}\left[\varphi\right] = F^{-1}\left[\dfrac{1}{1+\left|p\right|^2}F\left[\varphi\right]\right], \\ B\left(x\right) = \dfrac{1}{\left(2\pi\right)^N}\displaystyle\int_{R^N}\dfrac{e^{i(p,x)}}{1+\left|p\right|^2}dp, \\ U = L_\infty \bigcap H^2, \\ X = C\left[0;\ \infty;\ U\right), \\ \left\|\varphi\right\| = \sup\limits_{t\geqslant 0}\left(\left\|\varphi\right\|_{L_\infty} + \left\|\varphi\right\|_{H^2}\right) \\ (\textit{в частности, если } \varphi \textit{ не зависит от } t, \\ \textit{то } \left\|\varphi\right\| = \left\|\varphi\right\|_{L_\infty} + \left\|\varphi\right\|_{H^2}), \\ X_\rho = \left\{\varphi \in X \mid \left\|\varphi\right\| < \rho\right\} \end{array}$$

Преобразование Фурье вводится по правилу

$$\hat{f}\left(p\right) = F\left[f\left(x\right)\right] = \frac{1}{\left(2\pi\right)^{N/2}}\int\limits_{R^N} e^{-i(p,x)}f\left(x\right)dx.$$

Через C будем обозначать положительные постоянные, возможно, различные.
Будем предполагать, что $u_0\left(x\right) \in U$.
Замечание. Несложно убедиться, что $\overline{\mathrm{B}}\left[\varphi\right] = \int_{R^N} B\left(x-y\right)\varphi\left(y\right)dy$

Формально применим к уравнению в задаче (3) оператор $\overline{\mathrm{B}}\left[\cdot\right]$ и проинтегрируем от 0 до t с учетом начальных условий. Получим следующее интегральное уравнение:

$$v\left(x,t\right) = u_0\left(x\right) + \int\limits_0^t e^{-\sigma a \tau}\,\overline{\mathrm{B}}\left[\mu\left|v\right|^{\sigma} v\right]d\tau \tag{4}$$

Определение 1. *Обобщенным решением задачи (3) называется решение уравнения (4) из пространства X.*

Определение 2. *Обобщенным решением задачи (2) называется элемент u пространства X, определяемый по правилу $u = ve^{-at}$, где v – обобщенное решение задачи (3). (Заметим, что если $v \in X$, то и $u \in X$.)*

Приведем несколько вспомогательных утверждений.

Лемма 1. *Существуют такие постоянные $c_{1,2} > 0$, что*

$$\begin{array}{l} 1.\ \left|B\left(x\right)\right| \leqslant c_1\left|x\right|^{(1-N)/2}e^{-\left|x\right|},\ \left|x\right| \geqslant 1, \\ 2.\ \left|B\left(x\right)\right| \leqslant c_2\int_{\left|x\right|}^1 y^{1-N}dy,\ \left|x\right| < 1. \end{array}$$

Лемма 2. *Существует такая постоянная $c_3 > 0$, что*

$$\left\|\overline{\mathrm{B}}\left[\varphi\right]\right\|_{H^2} \leqslant c_3\left\|\varphi\right\|_{L_2}$$

Леммы 1 и 2 доказаны в [3].

Лемма 3. *1). Если $N = 1$, то $B\left(\cdot\right) \in L_p$, где $p \in \left[1;\ \infty\right]$.
2). Если $N = 2$, то $B\left(\cdot\right) \in L_p$, где $p \in \left[1;\ \infty\right)$.*

3). Если $N \geqslant 3$, то $B(\cdot) \in L_p$, где $p \in [1;\ N/(N-2))$.

Оценивая $\|B(\cdot)\|_{L_p}$ с помощью леммы 1 и переходя к полярным координатам, найдем условия на N и p, при которых $\|B(\cdot)\|_{L_p} < \infty$. **Лемма доказана.**

Лемма 4. $\int_{R^N} |B(x-y)\mu(y)|\, dy$ *ограничен постоянной, не зависящей от x.*

Действительно, из неравенства Юнга следует, что

$$\int_{R^N} |B(x-y)\mu(y)|\, dy \leqslant \left\| \int_{R^N} |B(x-y)\mu(y)|\, dy \right\|_{L_\infty} \leqslant \|B\|_{L_p} \|\mu\|_{L_q},$$

где $1/p + 1/q - 1 = 0$. По условию, $\mu \in L_q$, причем если $N \leqslant 3$, то $2 \leqslant q \leqslant \infty$, если $N \geqslant 4$, то $N/2 < q \leqslant \infty$. В этих случаях получим соответственно $1 \leqslant p \leqslant 2$ и $1 \leqslant p < N/(N-2)$. Из леммы 3 следует, что в обоих случаях $\|B\|_{L_p} < \infty$. **Лемма доказана.**

Лемма 5. *Пусть $\varphi \in X$, тогда* $\sup\limits_t \left\| \int_0^t \mu |\varphi|^\sigma \varphi\, e^{-\sigma a\tau} d\tau \right\|_{L_2} \leqslant (\sigma a)^{-1} \|\mu\|_{L_q} \|\varphi\|^{\sigma+1}$.

Действительно,

$$\sup_t \left\| \int\limits_0^t \mu(x) |\varphi(x,\tau)|^\sigma \varphi(x,\tau) e^{-\sigma a\tau} d\tau \right\|_{L_2} =$$

$$= \sup_t \sqrt{\int\limits_{R^N} \left(\int\limits_0^t \mu(x) |\varphi(x,\tau)|^\sigma \varphi(x,\tau) e^{-\sigma a\tau} d\tau \right)^2 dx} \leqslant$$

$$\leqslant \sup_t \sqrt{\int\limits_{R^N} \int\limits_0^\infty |\mu| |\varphi|^{2-4/q} |\varphi|^{\sigma+1-2+4/q} e^{-\sigma a\tau} d\tau \int\limits_0^\infty |\mu| |\varphi|^{\sigma+1} e^{-\sigma a\tau} d\tau dx} \leqslant$$

$$\leqslant \sup_t \sqrt{\int\limits_{R^N} \mu^2 \sup_t \|\varphi\|_{L_\infty}^{2\sigma+4/q} \int\limits_0^\infty |\varphi|^{2-4/q} e^{-\sigma a\tau} d\tau \int\limits_0^\infty e^{-\sigma a\tau} d\tau dx} =$$

$$= \sup_t \|\varphi\|_{L_\infty}^{\sigma+2/q} \sup_t \sqrt{\int\limits_{R^N} \mu^2 \int\limits_0^\infty |\varphi|^{2-4/q} e^{-\sigma a\tau} d\tau \frac{1}{\sigma a} dx} =$$

$$= \sup_t \|\varphi\|_{L_\infty}^{\sigma+2/q} \frac{1}{\sqrt{\sigma a}} \sup_t \sqrt{\int\limits_0^\infty e^{-\sigma a\tau} \int\limits_{R^N} \mu^2 |\varphi|^{2-4/q} dx d\tau}$$

Положим $\delta = \int_{R^N} \mu^2 |\varphi|^{2-4/q} dx$. Если $q = 2$, то $\delta = \|\mu\|_{L_2}^2$. Пусть $q > 2$. Применим неравенство Гельдера:

$$\delta \leqslant \left\|\mu^2\right\|_{L_{q/2}} \left\| |\varphi|^{2-4/q} \right\|_{L_{q/(q-2)}} = \left(\int\limits_{R^N} |\mu|^q dx \right)^{2/q} \left(\int\limits_{R^N} \varphi^2 dx \right)^{(q-2)/q} = \|\mu\|_{L_q}^2 \|\varphi\|_{L_2}^{2(q-2)/q}$$

Учитывая, что $\|\varphi\|_{L_2} \leqslant \|\varphi\|_{H^2}$, приходим к требуемой оценке. **Лемма доказана.**

Теорема 1. *Существует такая окрестность нуля, что если $\|u_0\|$ находится в этой окрестности, то существует единственное обобщенное решение $v \in X$ задачи (3-4) (а значит, и решение и задачи (2) из того же класса).*

Таким образом, если начальные данные достаточно малы в смысле нормы пространства X, то исследуемая задача однозначно разрешима.

Докажем это утверждение. Пусть $\rho = \|u_0\| c_4$; ограничения на параметр $c_4 > 0$ уточним позже. Обозначим

$$Mv = u_0(x) + \int\limits_0^t e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right] d\tau$$

Докажем, что оператор M переводит шар X_ρ в себя и является сжимающим отображением. Тогда по принципу сжимающих отображений можно будет утверждать, что $\forall u_0 \in X_\rho$ существует единственное решение уравнения $Mv = v$ из пространства X.

Сначала докажем, что если $v \in X_\rho$, то и $Mv \in X_\rho$.

Заметим, что

$$\sup_t \|Mv\|_{L_\infty} \leqslant \|u_0\|_{L_\infty} + \sup_t \left\| \int\limits_0^t e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right] d\tau \right\|_{L_\infty}$$

Оценим выражение $\overline{\mathrm{B}}[\ldots]$ с помощью леммы 4:

$$\begin{aligned} \left\|\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right]\right\|_{L_\infty} &= \left\| \int\limits_{R^N} B(x-y)\,\mu(y)\,|v(y)|^\sigma v(y)\,dy \right\|_{L_\infty} \leqslant \\ &\leqslant \left\|v^{\sigma+1}\right\|_{L_\infty} \left\| \int\limits_{R^N} B(x-y)\,\mu(y)\,dy \right\|_{L_\infty} \leqslant C\rho^{\sigma+1} \end{aligned}$$

Следовательно,

$$\sup_t \left\| \int\limits_0^t e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right] d\tau \right\|_{L_\infty} \leqslant \sup_t \int\limits_0^t e^{-\sigma a\tau} \left\|\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right]\right\|_{L_\infty} d\tau \leqslant C\rho^{\sigma+1}$$

Следовательно,

$$\sup_t \|Mv\|_{L_\infty} \leqslant \|u_0\|_{L_\infty} + C\rho^{\sigma+1}$$

Докажем теперь аналогичное соотношение в пространстве H^2. Представим $\overline{\mathrm{B}}[\cdot]$ в виде свертки, поменяем порядок интегрирования и воспользуемся леммами 2 и 5:

$$\sup_t \left\| \int\limits_0^t e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\,|v|^\sigma v\right] d\tau \right\|_{H^2} \leqslant c_3 \sup_t \left\| \int\limits_0^t e^{-\sigma a\tau} \mu\,|v|^\sigma v d\tau \right\|_{L_2} \leqslant c_3 \frac{\|\mu\|_{L_q}}{\sigma a} \|v\|_X^{\sigma+1}$$

Значит,

$$\sup_t \|Mv\|_{H^2} \leqslant \|u_0\|_{H^2} + C\rho^{\sigma+1}$$

Таким образом,

$$\begin{aligned} &\|Mv\| \leqslant \sup_t \|Mv\|_{L_\infty} + \sup_t \|Mv\|_{H^2} \leqslant \|u_0\|_{L_\infty} + \|u_0\|_{H^2} + C\rho^{\sigma+1} = \\ &= \|u_0\| + C\rho^{\sigma+1} = \rho/c_4 + C\rho^{\sigma+1} \end{aligned}$$

Из этой оценки видно, что можно подобрать такое $c_4 > 0$, что для всех достаточно малых ρ будет выполняться соотношение $\|Mv\| \leqslant \rho/c_4 + C\rho^{\sigma+1} < \rho$. Значит, оператор M переводит шар X_ρ в себя.

Доказательство оценки $\|Mw - Mv\| \leqslant 1/2 \cdot \|w - v\|\ \forall w,\ v \in X_\rho$, означающей, что оператор M является сжимающим отображением, аналогично доказательству утверждения, что этот оператор переводит шар X_ρ в себя. Здесь следует принять во внимание неравенство $||w|^\sigma w - |v|^\sigma v| \leqslant (\sigma+1)\rho^\sigma |w-v|$.

Итак, оператор M переводит шар X_ρ в себя и является сжимающим отображением. Значит, существует единственное решение уравнения $Mw = w$ из пространства X. **Теорема доказана.**

Теорема 2. *Существует такая окрестность нуля, что если* $\|u_0\|$ *находится в этой окрестности, то при* $t \to \infty$ *имеет место следующая асимптотика:*

$$u\left(x,t\right) = A\left(x\right)e^{-at} + O\left(e^{-(\sigma+1)at}\right),$$

где

$$A\left(x\right) = u_0\left(x\right) + \int\limits_0^\infty e^{a\tau}\,\overline{\mathrm{B}}\left[\mu\left|u\right|^\sigma u\right]d\tau$$

Таким образом, если начальные данные достаточно малы в смысле нормы пространства X, то главная составляющая асимптотики полностью определяется линейными членами уравнения, тогда как нелинейность влияет только на порядок стремления к нулю остаточного члена. В частности, если нелинейность тождественно равна нулю, то главный член асимптотики является точным решением: $u\left(x,t\right) = A\left(x\right)e^{-at} = u_0\left(x\right)e^{-at}$.

Докажем названную асимптотическую формулу.

Исходя из уравнения (4), представим v в следующем виде:

$$v\left(x,t\right) = u_0\left(x\right) + \int\limits_0^\infty e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\left|v\right|^\sigma v\right]d\tau - \int\limits_t^\infty e^{-\sigma a\tau}\,\overline{\mathrm{B}}\left[\mu\left|v\right|^\sigma v\right]d\tau$$

Обозначим сумму первых двух слагаемых через $A(x)$. Функция $A(\cdot)$ не зависит от времени. Обозначим через R третье слагаемое. Докажем, что $R = O\left(e^{-\sigma at}\right)$. Заметим, что

$$R = \int\limits_t^\infty e^{-\sigma a\tau}\int\limits_{R^N} B\left(x-y\right)\mu\left(y\right)\left|v\left(y,\tau\right)\right|^\sigma v\left(y,\tau\right)dyd\tau$$

Следовательно,

$$|R| \leqslant \int\limits_t^\infty e^{-\sigma a\tau}d\tau\int\limits_{R^N}\left|B\left(x-y\right)\mu\left(y\right)\right|\|v\|_{L_\infty}^{\sigma+1}dy \leqslant \frac{e^{-\sigma at}}{\sigma a}\cdot I\rho^{\sigma+1},$$

где $I = \sup\limits_x \int_{R^N}\left|B\left(x-y\right)\mu\left(y\right)\right|dy$ (такое число существует в силу леммы 4). Следовательно,

$$v\left(x,t\right) = A\left(x\right) + O\left(e^{-\sigma at}\right),$$

откуда

$$u\left(x,t\right) = v\left(x,t\right)e^{-at} = A\left(x\right)e^{-at} + O\left(e^{-(\sigma+1)at}\right)$$

Теорема доказана.

3 Начально-краевая задача

Рассмотрим следующую задачу:

$$\left\{\begin{array}{r} \frac{\partial}{\partial t}(\Delta u-u)+a(\Delta u-u)+F(u)=0 \\ u(x,0)=u_0(x) \\ u(x,t)|_{\partial\Omega}=0 \end{array}\right. \tag{5}$$

Здесь $F(u)=\mu|u|^{\sigma}u+(\lambda,\nabla)u^2$, u – действительнозначная функция, зависящая от пространственной переменной $x\in\Omega$, где Ω – ограниченное подмножество R^3 с границей $\partial\Omega\in C^{(2,\delta)}$, $\delta\in(0;1]$, $u_0(x)\in H_0^1(\Omega)$, $a>0$, $\mu\in R$, $\lambda\in R^3$, $\sigma\in(0;\ 4]$.

Будем рассматривать решения этой задачи, соответствующие нетривиальным начальным данным.

В данном разделе получают развитие идеи, использовавшиеся в [5] для исследования начально-краевой задачи для уравнения Осколкова-Бенджамена-Бона-Махони-Бюргерса:

$$\frac{\partial}{\partial t}(\Delta u-u)+\Delta u+u\frac{\partial u}{\partial x_1}+u^3=0$$

Определение 3. *Обобщенным решением задачи (5) будем называть такое* $u\in C^1\left[0,\ T;\ H_0^1(\Omega)\right)$, *что*

$$\left\langle\frac{\partial}{\partial t}(\Delta u-u)+a(\Delta u-u)+F(u),\ w\right\rangle=0\ \forall w\in H_0^1(\Omega) \tag{6}$$

(смысл параметра $T>0$ *уточним позже).*

Определение 4. *Говорят, что обобщенное решение задачи (5) разрушается за конечное время, если эта задача имеет решение* $u\in C^1\left[0,\ T;\ H_0^1(\Omega)\right)$ *при некотором конечном* $T>0$, *но не имеет решения из класса* $u\in C^1\left[0,\ \infty;\ H_0^1(\Omega)\right)$. *В частности, говорят, что решение разрушается путем опрокидывания, если* $\lim\limits_{t\to T-}\|u\|_{H_0^1(\Omega)}=\infty$.

Введем обозначения, необходимые для изложения рассуждений:

$$\begin{array}{l} L_p=L_p(\Omega),\ 1\leqslant p\leqslant\infty, \\ H_0^1=H_0^1(\Omega), \\ v=ue^{at}, \\ \|\cdot\|=\|\cdot\|_{L_2}, \\ \Phi=\|v\|^2+\|\nabla v\|^2, \\ v'=\dfrac{\partial v}{\partial t}, \\ W=\|v'\|^2+\|\nabla v'\|^2, \\ \Psi=\|u\|^2+\|\nabla u\|^2, \\ \Lambda=\|\lambda\|_{l_1} \end{array}$$

Если не оговорено противное, нижний индекс "0" означает, что выражение рассматривается при $t=0$. C_p – оптимальная константа вложения H_0^1 в L_p, т. е.

$$C_p=\inf\{C|\ \|w\|_{L_p}\leqslant C\|w\|_{H_0^1}\ \forall w\}$$

Замечание. Очевидно, $v_0=u_0$ и $\Phi_0=\Psi_0$.

В обобщенном смысле задача сводится к следующему интегральному уравнению:

$$v=u_0+\int\limits_0^t H^{-1}(F(u))\,d\tau, \tag{7}$$

где $Hu = u - \Delta u$.

Теорема 3. *$\forall u_0 \in H_0^1$ $\exists T > 0$ (возможно, $T = \infty$), для которого существует единственное обобщенное решение. При этом если $T < \infty$, то $\lim\limits_{t \to T-} \|u\|_{H_0^1(\Omega)} = \infty$, т. е. имеет место опрокидывание решения.*

Доказательство существования и единственности основано на принципе сжимающих отображений и аналогично доказательству соответствующего утверждения для уравнения Осколкова-Бенджамена-Бона-Махони-Бюргерса [5].

Для нахождения оценок времени разрушения решения в виде явных формул ограничимся рассмотрением следующих двух подклассов нелинейностей, входящих в уравнение:

$$F(u) = F_1(u) \equiv \mu u^3 + (\lambda, \nabla) u^2$$

и

$$F(u) = F_2(u) \equiv \mu |u|^\sigma u$$

Теорема 4. *Пусть $F(u) = F_1(u)$, причем $\Lambda > 0$.*

Если коэффициенты и начальные данные удовлетворяют условию

$$\mu\rho^2 - 2\Lambda\rho - a > 0, \text{ где } \rho = \frac{\|u_0\|^2_{L_4}}{\sqrt{\|u_0\|^2 + \|\nabla u_0\|^2}}, \tag{8}$$

то решение разрушается за конечное время $T \in [T_1;\ T_2]$, где

$$T_1 = -\frac{1}{2a} \ln\left(1 - \frac{a}{\mu C_4^4 \left(\|u_0\|^2 + \|\nabla u_0\|^2\right)}\right),$$
$$T_2 = \frac{1}{\beta} \ln \frac{2\alpha\mu\left(\|u_0\|^2 + \|\nabla u_0\|^2\right)}{2\alpha\mu\left(\|u_0\|^2 + \|\nabla u_0\|^2\right) - \beta},$$

где $\alpha = 1 - \Lambda/(\mu\rho)$, $\beta = 2(a + \Lambda\rho)$.

Если $\mu \leqslant 0$ или $\|u_0\|^2 + \|\nabla u_0\|^2 \leqslant a\mu^{-1}C_4^{-4}$, то решение существует глобально по времени и ограничено $\forall\lambda$.

<u>Замечание.</u> Несложно убедиться, что из условия (8) следует неравенство $\mu > 0$. Кроме того, из доказательства будет видно, что из (8) следует, что $\|u_0\|^2 + \|\nabla u_0\|^2 > a\mu^{-1}C_4^{-4}$. Значит, названные оценки времени разрушения не противоречат второй части теоремы, утверждающей, что решение глобально ограничено, если $\mu \leqslant 0$ или $\|u_0\|^2 + \|\nabla u_0\|^2 \leqslant a\mu^{-1}C_4^{-4}$.

Полагая в (6) $w = v$ и $w = v'$, получим следующие энергетические равенства:

$$I.\ \frac{1}{2}\frac{d\Phi}{dt} = \mu e^{-2at}\|v\|^4_{L_4},$$
$$II.\ W = \frac{\mu}{4} e^{-2at} \frac{d}{dt}\|v\|^4_{L_4} - e^{-at}\int\limits_\Omega v^2 (\lambda, \nabla) v' dx$$

Из I энергетического равенства видно, что при $\mu \leqslant 0$ решение существует глобально по времени и ограничено в смысле нормы H_0^1 для всех начальных данных и для всех значений $\lambda \in R^3$. В дальнейшем будем считать, что $\mu > 0$.

Найдем оценку сверху для времени разрушения.

Выразим $\mu\|u\|^4_{L_4}$ из I равенства и подставим во II:

$$W = \frac{e^{-2at}}{8}\frac{d}{dt}\left(e^{2at}\frac{d\Phi}{dt}\right) - e^{-at}\int\limits_\Omega v^2 (\lambda, \nabla) v' dx \tag{9}$$

Оценим интеграл с учетом I энергетического равенства:

$$\left| e^{-at}\int\limits_\Omega v^2 (\lambda, \nabla) v' dx \right| \leqslant$$
$$\leqslant \Lambda e^{-at}\left(\frac{\varepsilon(t)}{2}\|\nabla v'\|^2 + \frac{1}{2\varepsilon(t)}\|v\|^4_{L_4}\right) \leqslant \Lambda e^{-at}\left(\frac{\varepsilon(t)}{2}W + \frac{e^{2at}}{4\mu\varepsilon(t)}\Phi'\right), \tag{10}$$

где $\varepsilon(t)$ – некоторая непрерывная функция с положительными значениями. Положим $\varepsilon(t) = \varepsilon_0 e^{at}$, $\varepsilon_0 > 0$ и подставим это выражение в (9)-(10). Следовательно,

$$\left(1 - \frac{\Lambda\varepsilon_0}{2}\right) W \leqslant \frac{\Phi''}{8} + \left(\frac{a}{4} + \frac{\Lambda}{4\mu\varepsilon_0}\right)\Phi' \tag{11}$$

Из [5] известно, что

$$\frac{1}{4}\Phi'^2 \leqslant \Phi W \tag{12}$$

Из (11) и (12) следует, что

$$\Phi\Phi'' - 2\left(1 - \frac{\Lambda\varepsilon_0}{2}\right)\Phi'^2 + 2\left(a + \frac{\Lambda}{\mu\varepsilon_0}\right)\Phi\Phi' \geqslant 0$$

Пусть коэффициент при Φ'^2 больше единицы, т. е. $\varepsilon_0 < 1/\Lambda$. Положим $\Phi = z^{-1/\alpha}$, где $\alpha = 1 - \Lambda\varepsilon_0 > 0$. После упрощений получим:

$$z'' + \beta z' \leqslant 0,$$

где $\beta = 2\left(a + \Lambda/\left(\mu\varepsilon_0\right)\right) > 0$. Следовательно, $z' + \beta z \leqslant z_0' + \beta z_0$, т. е.

$$\frac{d}{dt}\left(ze^{\beta t}\right) \leqslant \left(z_0' + \beta z_0\right)e^{\beta t}$$

Интегрируя это неравенство, получим:

$$z \leqslant \frac{z_0' + \beta z_0}{\beta} - \frac{z_0'}{\beta}e^{-\beta t},$$

откуда

$$\Phi \geqslant \left(\frac{z_0' + \beta z_0}{\beta} - \frac{z_0'}{\beta}e^{-\beta t}\right)^{-1/\alpha}$$

Заметим, что выражение в скобках положительно в некоторой окрестности $t = 0$ и обращается в 0 при

$$t = T_2 = \frac{1}{\beta}\ln\frac{z_0'}{z_0' + \beta z_0}$$

Условие $T_2 > 0$ равносильно неравенству $z_0' + \beta z_0 < 0$. Вернемся к первоначальным обозначениям, учитывая, что $\Phi_0' = 2\mu\|u_0\|_{L_4}^4$:

$$\beta < 2\alpha\mu\|u_0\|_{L_4}^4\Phi_0^{-1},$$

т. е.

$$a + \frac{\Lambda}{\mu\varepsilon_0} < \mu\left(1 - \Lambda\varepsilon_0\right)\rho^2$$

Это неравенство равносильно следующему:

$$\frac{1}{\mu\rho\varepsilon_0} + \mu\rho\varepsilon_0 < \frac{\mu\rho^2 - a}{\Lambda\rho}$$

Соотношение заведомо будет выполняться, если взять $\mu\rho\varepsilon_0 = 1$ и потребовать, чтобы правая часть была строго больше двух. Это требование равносильно условию (8). Значит, $\varepsilon_0 = (\mu\rho)^{-1}$. Возвращаясь к обозначениям, которые были введены в формулировке теоремы, получим:

$$T_2 = \frac{1}{\beta}\ln\frac{2\alpha\mu\left(\|u_0\|^2 + \|\nabla u_0\|^2\right)}{2\alpha\mu\left(\|u_0\|^2 + \|\nabla u_0\|^2\right) - \beta}$$

Найдем оценку снизу для времени разрушения.

Рассмотрим интегральное тождество:

$$\int_\Omega \left(\frac{\partial}{\partial t}\left(\Delta u - u\right) + a\left(\Delta u - u\right) + \mu u^3 + \left(\lambda, \nabla\right) u^2\right) w dx = 0 \ \forall w \in H_0^1$$

Положим в нем $w = u$ и после интегрирования по частям получим III энергетическое равенство:

$$\frac{1}{2}\frac{d\Psi}{dt} + a\Psi = \mu \|u\|_{L_4}^4$$

Воспользуемся теоремой вложения H_0^1 в L_4:

$$\frac{1}{2}\frac{d\Psi}{dt} + a\Psi \leqslant \mu C_4^4 \Psi^2,$$

откуда

$$\Psi' + A\Psi \leqslant B\Psi^2,$$

где $A = 2a$ и $B = 2\mu C_4^4$. Поделим это неравенство на Ψ^2 и положим $\frac{1}{\Psi} = y$. Тогда $-y' + Ay \leqslant B$, т. е.

$$\frac{d}{dt}\left(e^{-At} y\right) \geqslant -Be^{-At}$$

После интегрирования левой и правой частей от 0 до t получим:

$$y \geqslant y_0 e^{At} - \frac{B}{A}\left(e^{At} - 1\right) = \left(y_0 - \frac{B}{A}\right) e^{At} + \frac{B}{A}$$

Заметим, что правая часть положительна в некоторой окрестности $t = 0$. Правая граница этой окрестности T_1 и будет нижней оценкой времени разрушения T. Таким образом,

$$0 < \Psi \leqslant \frac{1}{\left(y_0 - \frac{B}{A}\right) e^{At} + \frac{B}{A}} = \frac{\Psi_0}{\frac{B}{A}\Psi_0 + \left(1 - \frac{B}{A}\Psi_0\right) e^{At}} \tag{13}$$

Если выполняется (8), то выражение в скобках в последней дроби отрицательно. Действительно, многочлен $P(\rho) = \mu\rho^2 - 2\Lambda\rho - a$ имеет корни разных знаков, поэтому условие $P(\rho) > 0$ равносильно неравенству

$$\frac{\Lambda}{\mu} + \sqrt{\frac{\Lambda^2}{\mu^2} + \frac{a}{\mu}} < \rho \equiv \frac{\|u_0\|_{L_4}^2}{\sqrt{\Psi_0}}$$

Заметим, что $\Lambda/\mu + \sqrt{\Lambda^2/\mu^2 + a/\mu} > \sqrt{a/\mu}$ и $\|u_0\|_{L_4}^2/\sqrt{\Psi_0} \leqslant C_4^2\sqrt{\Psi_0}$. Значит, (8) $\Rightarrow \sqrt{a/\mu} < C_4^2\sqrt{\Psi_0} \Leftrightarrow a/\mu < C_4^4\Psi_0$. Следовательно, функция, оценивающая Ψ в (13), возрастает и обращается в бесконечность при $t = T_1$, где

$$T_1 = -\frac{1}{A}\ln\left(1 - \frac{A}{B\Psi_0}\right)$$

Итак,

$$T_1 = -\frac{1}{2a}\ln\left(1 - \frac{a}{\mu C_4^4\left(\|u_0\|^2 + \|\nabla u_0\|^2\right)}\right)$$

Если же $\Psi_0 \leqslant a\mu^{-1}C_4^{-4}$, то из (13) следует, что величина Ψ ограничена глобально по времени, т. е. решение не разрушается. **Теорема доказана.**

Лемма 6. *Если* $0 < \omega < 1$, *то* $0 < (1-\omega)^2/2 < 1 - \omega + \omega\ln\omega < 1$.

Утверждение проверяется непосредственным исследованием указанных функций.

Введем следующие обозначения:

$$\gamma = C_{\sigma+2}^{\sigma+2},\ \omega = 1 - \Psi_0^{-\sigma/2-1}\|u_0\|_{L_{\sigma+2}}^{\sigma+2}/\gamma$$

Лемма 7. *Рассмотрим следующие условия:*

$$\Psi_0 > (\mu\gamma/a)^{-2/\sigma} \tag{14}$$

$$\Psi_0^{-\sigma/2} \leqslant (1-\omega+\omega\ln\omega)\cdot\mu\gamma/a \tag{15}$$

$$\mu\|u_0\|_{L_{\sigma+2}}^{2\sigma+4} \geqslant 2a\gamma\Psi_0^{\sigma/2+2} \tag{16}$$

Из (16) следует (15), а из (15) следует (14).

С учетом определения величины ω перепишем (16) в следующем виде

$$\Psi_0^{-\sigma/2} \leqslant \frac{\mu\gamma}{a}\cdot\frac{(1-\omega)^2}{2}$$

Оценивая правую часть с помощью леммы 6, придем к утверждению, что (16)⇒(15).

Оценивая правую часть (15) с помощью леммы 6, получим вторую часть леммы. **Лемма доказана.**

Теорема 5. *Пусть $F(u) = F_2(u)$. Пусть выполняется условие (16). Тогда решение разрушается за конечное время $T \in [T_1;\ T_2] \subseteq [T_1;\ T_3]$, где*

$$T_1 = \frac{1}{\sigma a}\ln\frac{\mu\gamma}{\mu\gamma - a\Psi_0^{-\sigma/2}},$$

T_2 – единственное решение уравнения $A + Bt + Ce^{-\sigma a t} = 0$, где

$$A = \Psi_0^{-\sigma/2} - \frac{\gamma\mu}{a},\ B = \mu\sigma\left(\gamma - \frac{\|u_0\|_{L_{\sigma+2}}^{\sigma+2}}{\Psi_0^{\sigma/2+1}}\right),\ C = \frac{\gamma\mu}{a},$$

$$T_3 = \frac{\Psi_0^{-\sigma/2-1}\|u_0\|_{L_{\sigma+2}}^{\sigma+2} - \sqrt{\Psi_0^{-\sigma-2}\|u_0\|_{L_{\sigma+2}}^{2\sigma+4} - 2a\mu^{-1}\gamma\Psi_0^{-\sigma/2}}}{\gamma a\sigma}$$

Если $\mu \leqslant 0$ или $\Psi_0 \leqslant (\mu\gamma/a)^{-2/\sigma}$, то решение существует глобально по времени и ограничено.

Таким образом, теорема дает грубую верхнюю оценку времени разрушения решения в виде формулы и более точную оценку в неявном виде.

Замечание. Из доказательства будет видно, что в первой части теоремы можно ослабить требования, потребовав (15) вместо (16). Тогда будет гарантировано выполнение оценки $T \in [T_1;\ T_2]$, но не $T \in [T_1;\ T_3]$.

Замечание. Из условия (16) (или (15)) следует, что $\mu > 0$. Кроме того, из (16) (или (15)) следует (14) в силу леммы 7. Значит, названные оценки времени разрушения не противоречат второй части теоремы, утверждающей, что решение глобально ограничено при $\mu \leqslant 0$ или $\Psi_0 \leqslant (\mu\gamma/a)^{-2/\sigma}$.

Как и в доказательстве теоремы 4, положим в (6) $w = v$ и $w = v'$. Получим следующие энергетические равенства:

$$I.\ \frac{1}{2}\frac{d\Phi}{dt} = \mu e^{-\sigma a t}\|v\|_{L_{\sigma+2}}^{\sigma+2},$$

$$II.\ W = \mu e^{-\sigma a t}\int\limits_{\Omega}|v|^{\sigma}vv'dx$$

Из I энергетического равенства видно, что при $\mu \leqslant 0$ решение существует глобально по времени и ограничено в смысле нормы H_0^1 для всех начальных данных. В дальнейшем будем считать, что $\mu > 0$.

Найдем оценку сверху для времени разрушения.

Из I энергетического равенства следует, что

$$\Phi'' = -\sigma a e^{-\sigma a t} \cdot 2\mu \int\limits_{\Omega} |v|^{\sigma+2}\, dx + 2\mu e^{-\sigma a t} \int\limits_{\Omega} (\sigma+2)\, |v|^{\sigma}\, v v' dx$$

Преобразуем это соотношение с помощью II энергетического равенства:

$$\Phi'' = 2\,(\sigma+2)\, W - 2\mu\sigma a e^{-\sigma a t} \|v\|^{\sigma+2}_{L_{\sigma+2}}$$

Продолжим преобразования с учетом (12) и теоремы вложения:

$$\Phi\Phi'' - \left(1 + \frac{\sigma}{2}\right) \Phi'^2 + 2\mu\sigma a\gamma e^{-\sigma a t} \Phi^{\sigma/2+2} \geqslant 0$$

Положим $\Phi = z^{-2/\sigma}$. После упрощений получим:

$$z'' \leqslant \gamma\mu\sigma^2 a e^{-\sigma a t} \tag{17}$$

Дважды проинтегрируем это неравенство от 0 до t:

$$z \leqslant f\,(t) \equiv A + Bt + Ce^{-\sigma a t},$$

где

$$A = \Phi_0^{-\sigma/2} - \frac{\gamma\mu}{a},\ B = \mu\sigma \left(\gamma - \frac{\|v_0\|^{\sigma+2}_{L_{\sigma+2}}}{\Phi_0^{\sigma/2+1}}\right),\ C = \frac{\gamma\mu}{a}$$

Несложно убедиться, что $f\,(\cdot)$ достигает минимума при $t = (\sigma a)^{-1} \ln\,(C\sigma a/B)$, причем это значение t положительно. Соответствующее значение $f\,(t) = \Phi_0^{-\sigma/2} - \gamma\mu/a + B\,(\sigma a)^{-1}\,(1 + \ln\,(C\sigma a/B))$.

Итак, положительная функция $z\,(t)$ оценивается сверху функцией $f\,(t)$, которая положительна в некоторой окрестности нуля. Значит, решение заведомо будет разрушаться, если $f\,(t)$ положительна в ограниченной окрестности нуля, но не на всей положительной полупрямой, т. е. если ее минимальное значение неположительно.

В силу леммы 6 $0 < 1 - \omega + \omega \ln \omega < 1$. Поэтому условие неположительности минимального значения $f\,(\cdot)$ можно привести к виду

$$\Phi_0^{-\sigma/2} \leqslant C\,(1 - \omega + \omega \ln \omega)$$

В силу леммы 7 это условие следует из предполагаемого условия (16).

Таким образом, T_2 – решение уравнения $f\,(t) = 0$.

Кроме того, можно получить в виде явной формулы более грубую верхнюю оценку для T. Действительно, из (17) следует, что $z'' \leqslant 2D \equiv \gamma\mu\sigma^2 a$. Дважды интегрируя, получим: $z\,(t) \leqslant z_0 + z_0' t + Dt^2$. Правая часть положительна в некоторой окрестности нуля и обращается в 0 при

$$t = T_3 = \frac{\Phi_0^{-\sigma/2-1} \|v_0\|^{\sigma+2}_{L_{\sigma+2}} - \sqrt{\Phi_0^{-\sigma-2} \|v_0\|^{2\sigma+4}_{L_{\sigma+2}} - 2a\mu^{-1}\gamma\Phi_0^{-\sigma/2}}}{\gamma a\sigma}$$

Такое число существует в силу (16).

Несложно убедиться, что $f\,(t) \leqslant z_0 + z_0' t + Dt^2\ \forall t \geqslant 0$, следовательно, $T_2 \leqslant T_3$ (если эти числа существуют).

Найдем оценку снизу для времени разрушения.

Рассмотрим интегральное тождество:

$$\int\limits_{\Omega} \left(\frac{\partial}{\partial t}\,(\Delta u - u) + a\,(\Delta u - u) + \mu\,|u|^{\sigma} u\right) w dx = 0\ \forall w \in H_0^1$$

Положим в нем $w = u$ и после интегрирования по частям получим III энергетическое равенство:

$$\frac{1}{2}\frac{d\Psi}{dt} + a\Psi = \mu\|u\|_{L_{\sigma+2}}^{\sigma+2}$$

Рассуждая по аналогии с доказательством теоремы 4, найдем следующую оценку:

$$\Psi \leqslant \left[\frac{\mu\gamma}{a} - \left(\frac{\mu\gamma}{a} - \Psi_0^{-\sigma/2}\right)e^{\sigma a t}\right]^{-2/\sigma}$$

Выражение в квадратных скобках положительно в некоторой окрестности нуля и обращается в 0 при

$$t = T_1 = \frac{1}{\sigma a}\ln\frac{\mu\gamma}{\mu\gamma - a\Psi_0^{-\sigma/2}},$$

если выполняется условие (14), которое следует из (16) в силу леммы 7. Если же $\Psi_0 \leqslant (\mu\gamma/a)^{-2/\sigma}$, то величина Ψ ограничена глобально по времени, т. е. решение не разрушается. **Теорема доказана.**

Список литературы

[1] А. И. Аристов. О задаче Коши для нелинейного уравнения соболевского типа с переменным коэффициентом. // Сборник тезисов XVII Международной научной конференции студентов, аспирантов и молодых ученых "ЛОМОНОСОВ-2010". Секция "ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И КИБЕРНЕТИКА". С. 40-41. Москва, Макс-Пресс, 2010.

[2] А. И. Аристов. Асимптотика при больших временах решения задачи Коши для уравнения соболевского типа с аналитической нелинейностью. // Сборник статей молодых ученых ф-та ВМК МГУ 2009. С. 17-22. Москва, Макс-Пресс, 2009.

[3] G. N. Watson. A treatise on the theory of Bessel functions. Cambridge University Press, Cambridge, England, 1944.

[4] N. Hayashi, E. Kaikina, P. Naumkin, I. Shishmarev. Asymptotics for Dissipative Nonlinear Equations. Springer-Verlag. 2006.

[5] А. Г. Свешников, А. Б. Альшин, М. О. Корпусов, Ю. Д. Плетнер. Линейные и нелинейные уравнения соболевского типа. М., Физматлит, 2007.

[6] А. Г. Свешников, А. Б. Альшин, М. О. Корпусов. Нелинейный функциональный анализ и его приложения к уравнениям в частных производных. М., Научный мир, 2008.

УДК 519.716

SFE-ЗАМКНУТЫЕ КЛАССЫ ТРЕХЗНАЧНОЙ ЛОГИКИ

© 2010 г. В. С. Федорова

mathcyb@cs.msu.su

Кафедра математической кибернетики

В теории функций многозначной логики одними из важнейших являются вопросы классификационного характера. Чаще всего классификации строятся на основе каких-либо операторов замыкания. Самая известная классификация этого типа — это классификация на основе операции суперпозиции. Для булевых функций она приводит к счетной решетке замкнутых классов [15], [16], но уже для функций трехзначной логики подобная классификация оказывается континуальной [14]. Естественным образом возникает вопрос о построении конечной (или хотя бы счетной) классификации для функций многозначной логики на основе более сильного, чем оператор суперпозиции, оператора замыкания. Существующие классификации используют разнообразные средства расширения традиционного оператора суперпозиции: применение чисто функциональных приемов, обращение к логико-функциональным языкам различных типов, использование идей из программирования и т. д. [4], [2], [12].

В настоящей статье вводится новый оператор замыкания, основывающийся на системах функциональных уравнений, — оператор SFE-замыкания (system of functional equations), который существенно отличается от известных сильных операторов замыкания как по способу задания, так и по порождаемым классификациям функций многозначных логик. По-видимому, оператор SFE-замыкания является наиболее сильным из известных операторов замыкания. В частности, в классе P_2 булевых функций образуется лишь два SFE-замкнутых класса: сам класс P_2 и класс S самодвойственных функций. В данной работе для множества функций многозначной логики приводятся некоторые SFE-полные системы функций и доказывается SFE-замкнутость классов функций определенного типа, а на множестве функций трехзначной логики строится полная решетка SFE-замкнутых классов с указанием порождающих функций для каждого класса.

Пусть $k \geq 2$, $E_k = \{0, 1, \ldots, k-1\}$, P_k — множество всех функций на E_k (множество функций k-значной логики). Для любого $n \geq 1$ и любого множества $Q \subseteq P_k$ обозначим через $Q^{(n)}$ множество всех n-местных функций из Q. Для любых $n \geq 1$ и i, $1 \leq i \leq n$, определим на E_k *селекторную функцию* $e_i^n(x_1, \ldots, x_i, \ldots, x_n) = x_i$. На множестве P_k предполагаем заданной операцию *суперпозиции* [13]. Классы функций из P_k, замкнутые относительно операции суперпозиции, далее называем *замкнутыми классами*. Понятия *порождающей системы* и *базиса* замкнутого класса [13], если специально не оговаривается оператор замыкания, относятся к операции суперпозиции. Замкнутые классы, содержащие все селекторные функции, называем *клонами* [3].

В определении языка функциональных уравнений k-значной логики придерживаемся терминологии работ [8], [10], [11]. Предполагаем, что каждая функция из P_k имеет индивидуальное обозначение. Для обозначения n-местных функций из P_k используем символы $f_i^{(n)}$, которые называем *функциональными константами*. Общепринятые обозначения $0, 1, \ldots, k-1, \overline{x}$ сохраняем за константами и отрицанием Поста ($\overline{x} = x + 1 \pmod{k}$). Иногда для обозначения функциональных констант мы будем использовать символы g, h (с индексами или без них).

Наряду с функциональными константами рассматриваем *функциональные переменные*. Для обозначения n-местных функциональных переменных используем символы $\varphi_i^{(n)}$. Областью значений функциональной переменной $\varphi_i^{(n)}$ служит множество $P_k^{(n)}$. В случае, когда это не приводит к недоразумению, верхние индексы у функциональных переменных будем опускать.

Помимо функциональных переменных используем обычные *индивидные переменные* $x_1, x_2, \ldots$ с областью значений E_k. Иногда для лучшего понимания структуры формулы в качестве индивидных переменных будем использовать переменные y, z.

Пусть $Q \subseteq P_k$. Определим понятие *терма над* Q. Всякая индивидная переменная есть терм над Q. Если $t_1, \ldots, t_n$ — термы над Q, $f_i^{(n)}$ — функциональная константа, являющаяся обозначением функции из Q, $\varphi_j^{(n)}$ — функциональная переменная, то выражения

$$f_i^{(n)}(t_1, \ldots, t_n), \quad \varphi_j^{(n)}(t_1, \ldots, t_n)$$

суть термы над Q.

Равенством над Q называем любое выражение вида $t_1 = t_2$, где t_1, t_2 — термы над Q. Равенства $t_1 = t_2$ и $t_2 = t_1$ в дальнейшем не различаем. Равенства над Q называем также *функциональными уравнениями над* Q.

Пусть T — функциональное уравнение над Q и $\varphi_{i_1}^{(n_1)}, \ldots, \varphi_{i_m}^{(n_m)}$ — все функциональные переменные, входящие в уравнение T. *Решением уравнения* T называем систему функций k-значной логики $\{f_{j_1}^{(n_1)}, \ldots, f_{j_m}^{(n_m)}\}$, которая после замены каждой функциональной переменной $\varphi_{i_s}^{(n_s)}$ соответствующей функциональной константой $f_{j_s}^{(n_s)}$ превращает уравнение T в тождество относительно всех значений входящих в это уравнение индивидных переменных. Отметим, что решением уравнения над Q могут быть функции, не входящие в множество Q.

Пусть Ξ — конечная система уравнений над Q. *Решением системы уравнений* Ξ называем систему функций k-значной логики, которая является решением каждого уравнения, входящего в Ξ.

Для того чтобы определять некоторые функции k-значной логики с помощью решений систем уравнений, выделим одну из функциональных переменных системы уравнений Ξ, которую назовём *главной функциональной переменной* этой системы. Пусть $\varphi_i^{(n)}$ — главная функциональная переменная системы уравнений Ξ и $F \subseteq P_k^{(n)}$. Говорим, что множество функций F *определяется системой уравнений* Ξ, если F является множеством всех тех n-местных функций, которые входят в решения системы Ξ в качестве компоненты по переменной $\varphi_i^{(n)}$.

Пусть $Q \subseteq P_k$. *Замыканием* множества Q относительно систем функциональных уравнений (коротко SFE-*замыканием*) назовем множество всех функций из P_k, которые определяются как одноэлементные множества системами функциональных уравнений над Q. SFE-замыкание множества Q обозначим через SFE$[Q]$. Множество Q назовем SFE-замкнутым, если $Q = \mathrm{SFE}[Q]$. Понятия SFE-*полноты*, SFE-*предполноты* и SFE-*порождающей системы* вводятся по аналогии с соответствующими понятиями для операции суперпозиции [13].

Из введенного определения следует, что SFE-замыкание удовлетворяет трем аксиомам замыкания [3]:

1) $Q \subseteq \mathrm{SFE}[Q]$,

2) если $Q \subseteq R$, то $\mathrm{SFE}[Q] \subseteq \mathrm{SFE}[R]$ (монотонность),

3) $\mathrm{SFE}[\mathrm{SFE}[Q]] = \mathrm{SFE}[Q]$ (идемпотентность),

то есть действительно является замыканием.

Для любого множества Q (в том числе и для $Q = \emptyset$) множеству SFE$[Q]$ принадлежат все селекторные функции, также SFE-замыкание множества Q замкнуто относительно операции суперпозиции (эти свойства для случая булевых функций фактически доказаны в работе [9] и без труда переносятся на случай функций k-значной логики). Таким образом, любое SFE-замкнутое множество является клоном.

Докажем SFE-полноту некоторых систем функций k-значной логики. Для большей наглядности будем опускать фигурные скобки, в которых перечислены функции, внутри квадратных скобок SFE-замыкания.

Пусть для $i \in E_k$

$$j_i(x) = \begin{cases} 1, & \text{если } x = i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Теорема 1. *Справедливы следующие равенства:*

1. $\mathrm{SFE}[0, 1, \ldots, k-1] = P_k$ $(k \geq 2)$.
2. *Для любого* $s \in E_k$ $\mathrm{SFE}[0, 1, \ldots, s-1, s+1, \ldots, k-1] = P_k$ $(k \geq 3)$.
3. $\mathrm{SFE}[j_0, j_1, \ldots, j_{k-1}] = P_k$ $(k \geq 3)$.
4. *Для любого* $s \in E_k$ $\mathrm{SFE}[j_0, j_1, \ldots, j_{s-1}, j_{s+1}, \ldots, j_{k-1}] = P_k$ $(k \geq 3)$.

Доказательство. 1. Пусть $f(x_1, \ldots, x_n)$ — произвольная функция из P_k. Задающая ее система функциональных уравнений с единственной функциональной переменной φ будет состоять из k^n уравнений вида

$$\varphi(a_1, \ldots, a_n) = f(a_1, \ldots, a_n),$$

где $a_1, \ldots, a_n \in E_k$, $f(a_1, \ldots, a_n)$ — константа из E_k, соответствующая значению рассматриваемой функции на наборе $(a_1, \ldots, a_n)$.

2. Сначала получим отрицание Поста $\varphi(x) = \overline{x}$ как единственное решение следующей системы функциональных уравнений над множеством $\{0, 1, \ldots, s-1, s+1, \ldots, k-1\}$:

$$\begin{cases} \varphi^k(x) = x, \\ \varphi(0) = 1, \\ \varphi(1) = 2, \\ \cdots\cdots\cdots \\ \varphi(s-2) = s-1, \\ \varphi(s+1) = s+2, \\ \cdots\cdots\cdots \\ \varphi(k-1) = 0. \end{cases}$$

Здесь, как и везде далее, пользуемся обозначением

$$\varphi^k(x) = \underbrace{\varphi(\varphi(\ldots\varphi}_{k}(x)\ldots)).$$

Последние $k-2$ уравнения задают функцию $\varphi(x)$ на всех наборах, кроме двух: $x = s-1$ и $x = s$. Из первого уравнения системы следует, что $\varphi(x)$ должна принимать все k значений, поэтому для значений функции на пропущенных наборах возможны лишь два варианта: $\varphi(s-1) = s$, $\varphi(s) = s+1$ и $\varphi(s-1) = s+1$, $\varphi(s) = s$. Однако последний случай можно рассматривать как перестановку, содержащую одну неподвижную точку и цикл длины $k-1$, который после возведения этой перестановки в степень k не даст тождественную перестановку, как того требует первое уравнение системы. Таким образом, единственным возможным решением является функция $\varphi(x) = \overline{x}$. Для завершения доказательства получим недостающую константу s как $\overline{s-1}$ и воспользуемся первым пунктом этой теоремы.

3. Поскольку для любого x справедливо равенство $j_2(j_0(x)) = 0$ и $j_0(0) = 1$, при построении системы функциональных уравнений над множеством функций $\{j_0, j_1, \ldots, j_{k-1}\}$ для наглядности будем пользоваться константами 0 и 1, предполагая их заданными соответствующими суперпозициями функций $j_0(x)$ и $j_2(x)$.

Как и в доказательстве предыдущего утверждения теоремы, получим отрицание Поста $\varphi(x) = \overline{x}$ как единственное решение системы функциональных уравнений над множеством $\{j_0, j_1, \ldots, j_{k-1}\}$:

$$\begin{cases} \varphi(0) = 1, \\ j_2(\varphi(1)) = 1, \\ j_3(\varphi^2(1)) = 1, \\ \cdots\cdots\cdots \\ j_{k-1}(\varphi^{k-2}(1)) = 1, \\ j_0(\varphi^{k-1}(1)) = 1. \end{cases}$$

Так как функции $j_i(x)$ принимают значение 1 только при $x = i$, $i \in \{0, 1, \ldots, k-1\}$, i-тая строка этой системы единственным образом задает функциональную переменную $\varphi(i-1) = i \pmod{k}$.

Имея отрицание Поста и константы 0 и 1, получаем полный набор констант и возможность воспользоваться первым утверждением теоремы.

4. Снова получим отрицание Поста с помощью системы функциональных уравнений над множеством $\{j_0, j_1, \ldots, j_{s-1}, j_{s+1}, \ldots, j_{k-1}\}$:

$$\begin{cases} \varphi^k(x) = x, \\ j_0(\varphi(x)) = j_{k-1}(x), \\ j_1(\varphi(x)) = j_0(x), \\ j_2(\varphi(x)) = j_1(x), \\ \cdots\cdots\cdots \\ j_{s-1}(\varphi(x)) = j_{s-2}(x), \\ j_{s+2}(\varphi(x)) = j_{s+1}(x), \\ \cdots\cdots\cdots \\ j_{k-1}(\varphi(x)) = j_{k-2}(x). \end{cases}$$

Уравнения со второго и до последнего однозначно определяют $\overline{x}$ на всех наборах, кроме $x = s-1$ и $x = s$. Проведя рассуждения, аналогичные приведенным в доказательстве второго пункта данной теоремы, получим, что $\varphi(x) = \overline{x}$ есть единственное решение этой системы функциональных уравнений. Поскольку справедливо тождество $j_{s+1}(\overline{x}) = j_s(x)$, имеем $\mathrm{SFE}[j_0, j_1, \ldots, j_{s-1}, j_{s+1}, \ldots, j_{k-1}] = \mathrm{SFE}[j_0, j_1, \ldots, j_{s-1}, j_s, j_{s+1}, \ldots, j_{k-1}] = P_k$, согласно третьему утверждению этой теоремы.

Теорема 1 полностью доказана.

Замечание. *В четвертом пункте теоремы 1 вместо функций $j_0(x), \ldots, j_{k-1}(x)$ можно использовать функции*

$$f_{ijr}(x) = \begin{cases} j, \text{ если } x = i, \\ r \text{ в остальных случаях,} \end{cases}$$

где j, r — различны, j, $r \in E_k$. Доказательство при этом совершенно не изменится.

Пусть $\rho(x_1, \ldots, x_m)$ — предикат на множестве E_k, то есть отображение $\rho : E_k^m \to \{T, F\}$, где T, F — истинностные значения истина и ложь. Говорят, что функция $f(x_1, \ldots, x_n) \in P_k$ *сохраняет предикат* ρ, если для любых n наборов $(a_{11}, a_{21}, \ldots, a_{m1}), \ldots, (a_{1n}, a_{2n}, \ldots, a_{mn})$, удовлетворяющих предикату ρ, набор $(f(a_{11}, a_{12}, \ldots, a_{1n}), \ldots, f(a_{m1}, a_{m2}, \ldots, a_{mn}))$ также удовлетворяет предикату ρ.

Обозначим через $\mathrm{Pol}(\rho)$ множество всех функций из P_k, сохраняющих предикат ρ. Известно [1], что для любого предиката ρ множество $\mathrm{Pol}(\rho)$ является замкнутым относительно суперпозиции классом, содержащим все селекторные функции (то есть клоном).

Напомним [17], что каждый предполный в P_k класс определяется (в смысле функтора Pol) предикатом одного из шести семейств **P**, **O**, **L**, **E**, **C**, **B**. При этом семейство **P** состоит из предикатов, которые являются графиками перестановок на E_k, разлагающихся в произведение циклов одной и той же простой длины, а все одноместные предикаты семейства **C** имеют вид $x \in D$, где $\emptyset \neq D \subset E_k$.

Теорема 2. *Пусть Q — предполный в P_k класс, который определяется предикатом одного из семейств* **O**, **L**, **E**, **C**, **B**. *Тогда* $\mathrm{SFE}[Q] = P_k$.

Доказательство. Как следует из [17], предполный класс Q, который определяется предикатом одного из семейств **O**, **L**, **E**, **B** или неодноместным предикатом семейства **C**, содержит все константы $0, 1, \ldots, k-1$, то есть является SFE-полным по первому утверждению теоремы 1.

Для завершения доказательства осталось рассмотреть случай, когда класс Q определяется одноместным предикатом семейства **C**. При этом возможны два варианта: центр предиката D содержит не меньше двух элементов множества E_k и D состоит в точности из одного элемента $d \in E_k$.

Пусть $|D| \geq 2$, поэтому существуют различные $d_1, d_2 \in D$. Тогда функции $f_{id_1d_2}(x)$, $i \in E_k$, принадлежат классу Q и по замечанию к теореме 1 $\mathrm{SFE}[f_{0d_1d_2}, \ldots, f_{k-1,d_1d_2}] = P_k$.

Пусть $|D| = 1$, то есть $D = \{d\}$. Тогда классу Q принадлежат функции

$$f_{0jd}(x), f_{1jd}(x), \ldots, f_{d-1,jd}(x), f_{d+1,jd}(x), \ldots, f_{k-1,jd}(x),$$

где $j \neq d$, $j \in E_k$ (поскольку $f_{ijd}(d) = d$ для любого i), которые образуют SFE-полную систему, как следует из замечания к теореме 1.

Теорема 2 доказана.

Пусть $f(x_1, \dots, x_n) \in P_k$ и π — перестановка на множестве E_k. Положим

$$f^{\pi}(x_1, \dots, x_n) = \pi^{-1}\big(f(\pi(x_1), \dots, \pi(x_n))\big),$$

где π^{-1} — перестановка, обратная к π. Функция f^{π} называется *двойственной* к функции f относительно перестановки π. Если $f^{\pi} = f$, то говорят, что f *самодвойственна* относительно перестановки π. Множество всех функций из P_k, самодвойственных относительно перестановки π, обозначим через S_{π}. Если G — непустое множество перестановок на E_k, то пусть S_G есть пересечение всех множеств S_{π}, где $\pi \in G$.

Известно [3], что для любого множества перестановок G множество S_G образует клон. Более того, клон S_G совпадает с клоном $S_{G'}$, где G' — группа перестановок, порождаемая множеством G.

Следующее утверждение есть некоторое обобщение Утверждения 2 из работы [11].

Утверждение 1. (Принцип двойственности для SFE-замыкания.) *Пусть система Ξ функциональных уравнений над множеством $\{g_1, \dots, g_s\}$ функций из P_k определяет функцию f и π — перестановка на множестве E_k. Тогда система уравнений Ξ^{π}, полученная из системы Ξ заменой каждой функциональной константы g_i, $i \in \{1, 2, \dots, s\}$, соответствующей функциональной константой g_i^{π}, определяет функцию f^{π}.*

Доказательство. Пусть $\varphi, \varphi_1, \dots, \varphi_r$ — все функциональные переменные системы Ξ и φ — ее главная функциональная переменная. Пусть также $(f, f_1, \dots, f_r)$ — решение системы уравнений Ξ. Согласно основным определениям при замене в системе Ξ функциональных переменных $\varphi, \varphi_1, \dots, \varphi_r$ соответствующими функциональными константами $f, f_1, \dots, f_r$ каждое уравнение системы Ξ превращается в тождество (относительно всех входящих в уравнение индивидных переменных).

Заменим теперь в полученной системе все функциональные константы $f, f_1, \dots, f_r, g_1, \dots, g_s$ функциональными константами $f^{\pi}, f_1^{\pi}, \dots, f_r^{\pi}, g_1^{\pi}, \dots, g_s^{\pi}$ (отдельно стоящие термы вида x_i в уравнениях системы Ξ при этом оставляем без изменений). Нетрудно видеть, что после такой замены полученные уравнения вновь окажутся тождествами. Это следует, например, из известного принципа двойственности для суперпозиции [13]: если

$$h(y_1, \dots, y_n) = h_0(h_1(y_1, \dots, y_n), \dots, h_m(y_1, \dots, y_n)),$$

то

$$h^{\pi}(y_1, \dots, y_n) = h_0^{\pi}(h_1^{\pi}(y_1, \dots, y_n), \dots, h_m^{\pi}(y_1, \dots, y_n)).$$

Таким образом, система функций $(f^{\pi}, f_1^{\pi}, \dots, f_r^{\pi})$ будет решением системы уравнений Ξ^{π}. Единственность решения f^{π} системы Ξ^{π} следует из единственности решения f системы Ξ и возможности обратного перехода от системы Ξ^{π} к системе Ξ с помощью перестановки π^{-1}.

Утверждение 1 доказано.

Следствие. *Для любой группы G перестановок на множестве E_k класс S_G является* SFE-*замкнутым.*

Перейдем к рассмотрению трехзначной логики P_3 и нахождению в ней всех SFE-замкнутых классов.

Обозначим через H_3 класс всех тех функций трехзначной логики, которые самодвойственны относительно любых перестановок множества E_3 (такие функции называют также *однородными*). В терминах групп перестановок имеем $H_3 = S_{S_3}$, где S_3 — полная симметрическая группа перестановок на E_3. Важную роль в дальнейших исследованиях будут играть функции тернарный дискриминатор $p(x, y, z)$ и $r_3(x, y)$, которые определяются следующими соотношениями [7]:

$$p(x, y, z) = \begin{cases} z, & \text{если } x = y, \\ x & \text{в противном случае;} \end{cases}$$

$$r_3(x, y) = \begin{cases} z, & \text{если } \{x, y, z\} = E_k, \\ x & \text{в остальных случаях.} \end{cases}$$

Как было доказано в работе [7], эти две функции составляют базис класса H_3: $[p, r_3] = H_3$. Покажем, что любая однородная функция может быть получена как единственное решение системы функциональных уравнений трехзначной логики без функциональных констант.

Теорема 3. *Имеет место равенство* $\mathrm{SFE}[\emptyset] = H_3$.

Доказательство. Рассмотрим следующую систему функциональных уравнений без функциональных констант:

$$\begin{cases} \varphi_1^3(x) = x, \\ \psi_1(x, x, y) = y, \\ \psi_1(x, y, x) = x, \\ \psi_1(x, y, y) = x, \\ \psi_1(x, \varphi_1(x), y) = x, \end{cases}$$

где $\psi_1(x, y, z)$ — главная функциональная переменная. Первому уравнению системы удовлетворяют три функции: $\varphi_1(x) = x$, $\varphi_1(x) = \overline{x} = x + 1$ и $\varphi_1(x) = \overline{\overline{x}} = x + 2$ (сложение здесь и везде далее производится по модулю 3). Второе, третье и четвертое уравнения задают функцию, совпадающую с функцией $p(x, y, z)$ везде, кроме шести наборов, состоящих из всех различных цифр из E_3. Наконец, последнее уравнение системы, которое должно быть справедливо при произвольных значениях переменных x, y, во-первых, исключает $\varphi_1(x) = x$ из множества решений первого уравнения, а во-вторых, верно доопределяет $\psi_1(x, y, z) = p(x, y, z)$ на оставшихся шести наборах, поскольку с учетом проведенных рассуждений x и $\varphi_1(x)$ — различны, а y — произвольно. Таким образом, единственным решением данной системы является тернарный дискриминатор $p(x, y, z)$.

Теперь построим систему функциональных уравнений без функциональных констант с главной функциональной переменной $\psi_2(x, y)$, единственным решением которой будет функция $r_3(x, y)$, причем для упрощения изложения будем пользоваться и тернарным дискриминатором, предполагая его заданным уже описанной системой уравнений:

$$\begin{cases} \varphi_2^3(x) = x, \\ \psi_2(x, x) = x, \\ \psi_2(x, \varphi_2(x)) = \varphi_2^2(x), \\ p(x, \varphi_2(x), y) = x. \end{cases}$$

Здесь снова первое и последнее уравнения позволяют записать множество значений функциональной переменной $\varphi_2(x)$ как $\{x+1, x+2\}$, откуда следует, что в третьем уравнении x, $\varphi_2(x)$ и $\varphi_2^2(x)$ различны и образуют в объединении E_3. Второе уравнение завершает задание функции $r_3(x, y)$.

Итак, с помощью систем функциональных уравнений без функциональных констант был получен базис по суперпозиции $\{p, r_3\}$ класса всех однородных функций H_3, откуда следует включение $H_3 \subseteq \mathrm{SFE}[\emptyset]$. Однако согласно Следствию из Утверждения 1 класс $S_{S_3} = H_3$ является SFE-замкнутым, а значит, он должен совпадать с минимальным SFE-замкнутым классом $\mathrm{SFE}[\emptyset]$.

Теорема 3 доказана.

Таким образом, любой SFE-замкнутый класс трехзначной логики содержит класс H_3 функций, самодвойственных относительно любых перестановок множества E_3, а значит, и тернарный дискриминатор $p(x, y, z)$. Поэтому в дальнейших рассуждениях будем пользоваться результатами работ [6], где приведено предикатное описание всех классов k-значной логики, содержащих тернарный дискриминатор (так называемые *дискриминаторные классы*), и [5], где в явном виде построены все 144 дискриминаторных класса трехзначной логики. В общем же случае число этих классов достаточно велико и выражается формулой, имеющей вид двойной экспоненты от k.

Итак, нам понадобятся следующие предикаты:

$$\begin{array}{ll} e_a(x) \equiv (x = a), & e_{ab}(x) \equiv (x \in \{a, b\}), \\ \sigma(x, y) \equiv (x + 1 = y), & \sigma_{ab}(x, y) \equiv (x \in \{a, b\}) \mathbin{\&} \sigma(x, y), \\ \sigma^0(x, y) \equiv (2x = y), & \sigma^0_{ab}(x, y) \equiv (x \in \{a, b\}) \mathbin{\&} \sigma^0(x, y), \\ \sigma^1(x, y) \equiv (2x + 2 = y), & \sigma^1_{ab}(x, y) \equiv (x \in \{a, b\}) \mathbin{\&} \sigma^1(x, y), \\ \sigma^2(x, y) \equiv (2x + 1 = y), & \sigma^2_{ab}(x, y) \equiv (x \in \{a, b\}) \mathbin{\&} \sigma^2(x, y), \end{array}$$

где $a \in \{0,1,2\}$, $ab \in \{01,02,12\}$, а сложение и умножение выполняются по модулю 3. Дискриминаторные классы в [5] описываются как классы сохранения предикатов с использованием функтора Pol. Докажем, что некоторые из этих классов являются SFE-предполными.

Теорема 4. *Следующие классы функций являются* SFE-*предполными в* P_3:

$$S_{x+1} = \mathrm{Pol}(\sigma),\ S_{2x} = \mathrm{Pol}(\sigma^0),\ S_{2x+2} = \mathrm{Pol}(\sigma^1),\ S_{2x+1} = \mathrm{Pol}(\sigma^2).$$

Доказательство. Класс S_{x+1} образуют функции, самодвойственные относительно графика перестановки $\pi = (012)$, которая представляет собой один цикл простой длины 3. Это означает, что класс S_{x+1} входит в семейство $\mathbf{P}$ предполных в P_3 классов. Однако этот класс является SFE-замкнутым по Следствию из Утверждения 1 (как и классы S_{2x}, S_{2x+2} и S_{2x+1}) и поэтому SFE-предполным.

Нетрудно заметить, что все функции класса $S_{2x+2i} = \mathrm{Pol}(\sigma^i)$, $i \in E_3$, сохраняют предикат e_i, и в решетке дискриминаторных классов непосредственно над S_{2x+2i} находится класс функций $C_i = \mathrm{Pol}(e_i)$. Классы C_0, C_1 и C_2 SFE-полны согласно Замечанию к Теореме 1, поскольку содержат пары функций $\{j_1(x) = f_{110}(x), j_2(x) = f_{210}(x)\}$, $\{f_{001}(x), f_{201}(x)\}$ и $\{f_{002}(x), f_{102}(x)\}$ соответственно, откуда следует, что классы S_{2x}, S_{2x+2} и S_{2x+1} являются SFE-предполными.

Теорема 4 доказана.

Теорема 5. *Следующие классы функций* SFE-*полны в* P_3:

$$\begin{gathered} S^{01}_{x+1} = \mathrm{Pol}(\sigma_{01}),\ S^{02}_{x+1} = \mathrm{Pol}(\sigma_{02}),\ S^{12}_{x+1} = \mathrm{Pol}(\sigma_{12}),\\ S^{01}_{2x} = \mathrm{Pol}(\sigma^0_{01}),\ S^{01}_{2x+2} = \mathrm{Pol}(\sigma^1_{01}),\ S^{02}_{2x+1} = \mathrm{Pol}(\sigma^2_{02}),\\ S^{12}_{2x} = \mathrm{Pol}(\sigma^0_{12}),\ S^{02}_{2x+2} = \mathrm{Pol}(\sigma^1_{02}),\ S^{01}_{2x+1} = \mathrm{Pol}(\sigma^2_{01}). \end{gathered}$$

Доказательство. Рассмотрим по очереди все девять классов.

1. Легко заметить, что функции $\max(x,y)$ и $\min(x,y)$ принадлежат классу S^{01}_{x+1}, поэтому ему принадлежат и максимум и минимум трех переменных: $\max_3(x,y,z) = \max(x, \max(y,z))$ и $\min_3(x,y,z) = \min(x, \min(y,z))$. Система функциональных уравнений с функциональными константами $p(x,y,z) \in \mathrm{SFE}[\emptyset]$, $\max_3(x,y,z)$, $\min_3(x,y,z)$ и тремя функциональными переменными $\varphi(x)$, $\psi_1(x)$, $\psi_2(x)$

$$\begin{cases} \varphi^3(x) = x,\\ p(x,\varphi(x),y) = x,\\ \psi_1(x) = \max_3(x,\varphi(x),\varphi^2(x)),\\ \psi_2(x) = \min_3(x,\varphi(x),\varphi^2(x)) \end{cases}$$

дает единственное решение $\psi_1(x) \equiv 2$ или $\psi_2(x) \equiv 0$ в зависимости от того, какую из функциональных переменных назначить главной. При этом снова, как и в доказательстве Теоремы 3, используется тот факт, что $x, \varphi(x), \varphi^2(x)$ различны и образуют в объединении E_3. Имея две константы из трех, по второму утверждению Теоремы 1 получаем SFE-полноту класса S^{01}_{x+1}.

2. Функции, заданные столбцами значений, которые для облегчения восприятия будем записывать в строчку, $f_2(x,y) = (010\ 111\ 012)$ и $f'_2(x,y) = (010\ 112\ 022)$ содержатся в классе S^{02}_{x+1}, при этом $f'_2(x,\overline{x}) = (120) = \overline{x}$, а $f'_2(x,\overline{\overline{x}}) = (012) = x$. Поэтому единственным решением системы уравнений над $\mathrm{SFE}[S^{02}_{x+1}]$

$$\begin{cases} \varphi^3(x) = x,\\ p(x,\varphi(x),y) = x,\\ f'_2(x,\varphi(x)) = \varphi(x) \end{cases}$$

будет отрицание Поста $\varphi(x) = \overline{x}$. Однако $f_2(x,\overline{x}) = (110)$ и $f_2(f_2(x,\overline{x}),\overline{\overline{x}}) = (111)$, то есть получена константа 1. Одна константа и $\overline{x}$ дают полный набор констант, и поэтому, согласно первому утверждению Теоремы 1, класс S^{02}_{x+1} является SFE-полным.

3. Функции $f_3(x,y) = (002\ 011\ 212)$ и $f'_3(x,y) = (012\ 111\ 212)$ находятся в классе S^{12}_{x+1} и, как нетрудно заметить, $f_3(x,\overline{x}) = (012) = x$, а $f_3(x,\overline{\overline{x}}) = (201) = \overline{\overline{x}}$. Поэтому следующая система уравнений с функциональными константами $p(x,y,z)$, $f_3(x,y)$

$$\begin{cases} \varphi^3(x) = x,\\ p(x,\varphi(x),y) = x,\\ f_3(x,\varphi(x)) = x \end{cases}$$

имеет единственное решение $\varphi(x) = \overline{x}$. При этом $f_3'(x, \overline{x}) = (112)$ и $f_3'(f_3'(x, \overline{x}), \overline{\overline{x}}) = (111)$, то есть $f_3'(f_3'(x, \overline{x}), \overline{\overline{x}}) \equiv 1$. Используя отрицания Поста и единицу, можно получить все константы, что и означает, согласно первому утверждению Теоремы 1, SFE-полноту класса S_{x+1}^{12}.

4. Класс S_{2x}^{01}, как и класс S_{x+1}^{01} из первого пункта доказательства, содержит функции $\max(x, y)$ и $\min(x, y)$, и поэтому также является SFE-полным.

5. Функция $f_5(x, y) = (002\ 012\ 222) \in S_{2x+2}^{01}$, при этом выполняются следующие тождества: $f_5(x, \overline{x}) = (022)$, $f_5(f_5(x, \overline{x}), \overline{\overline{x}}) = (222)$ и $f_5(x, \overline{\overline{x}}) = (202)$, $f_5(f_5(x, \overline{\overline{x}}), \overline{x}) = (202)$. Значит, система функциональных уравнений над SFE-замыканием класса S_{2x+2}^{01}

$$\begin{cases} \varphi^3(x) = x, \\ p(x, \varphi(x), y) = x, \\ f_5(f_5(x, \varphi(x)), \varphi^2(x)) = f_5(x, \varphi(x)) \end{cases}$$

имеет единственное решение $\varphi(x) = \overline{\overline{x}}$. Константа 1 сохраняет предикат σ_{01}^1, что вместе с двойным отрицанием Поста позволяет получить все константы. SFE-полнота класса S_{2x+2}^{01} доказана.

6. Класс S_{2x+1}^{02} является SFE-полным, так как содержит функции $\max(x, y)$ и $\min(x, y)$.

7, 8, 9. Классы S_{2x}^{12}, S_{2x+2}^{02}, S_{2x+1}^{01} SFE-полны, поскольку содержат пары функций $\{f_{221}(x), f_{121}(x)\}$, $\{f_{220}(x), f_{020}(x)\}$ и $\{j_0(x), j_1(x)\}$ соответственно, которые образуют SFE-полные системы в P_3 согласно замечанию к Теореме 1 и ее же четвертому утверждению.

Теорема 5 полностью доказана.

Докажем теорему о базисах четырех предполных классов S_{x+1}, S_{2x}, S_{2x+2} и S_{2x+1}.

Теорема 6. $S_{x+1} = \mathrm{SFE}[x+1]$, $S_{2x} = \mathrm{SFE}[2x]$, $S_{2x+2} = \mathrm{SFE}[2x+2]$ *и* $S_{2x+1} = \mathrm{SFE}[2x+1]$.

Доказательство. Нетрудно видеть, что любая функция f от n переменных, принадлежащая классу S_{x+1}, может быть полностью задана системой из 3^n уравнений вида

$$\varphi(x + i_1, \ldots, x + i_n) = x + i_0,$$

где $i_j \in E_3$, $j \in \{0, 1, \ldots, n\}$, и любая из функций $x + i_j$ порождается единственной функцией $x + 1$. Действительно, каждое из таких уравнений определяет значения функции f на трех наборах вида $(a_1, \ldots, a_n)$, $(a_1 + 1, \ldots, a_n + 1)$ и $(a_1 + 2, \ldots, a_n + 2)$, причем поскольку f сохраняет перестановку $\pi(x) = x + 1$, соответствующие значения функции $f(a_1, \ldots, a_n) = b_1$, $f(a_1 + 1, \ldots, a_n + 1) = b_2$ и $f(a_1 + 2, \ldots, a_n + 2) = b_3$ будут иметь вид $b_2 = b_1 + 1$, $b_3 = b_1 + 2$. При этом из построенной системы уравнений, очевидно, можно оставить лишь треть, беря только одно из трех уравнений вида

$$\begin{gathered} \varphi(x + i_1, \ldots, x + i_n) = x + i_0, \\ \varphi(x + (i_1 + 1), \ldots, x + (i_n + 1)) = x + (i_0 + 1), \\ \varphi(x + (i_1 + 2), \ldots, x + (i_n + 2)) = x + (i_0 + 2). \end{gathered}$$

Каждый из классов S_{2x+2i}, $i \in E_3$, содержит константу i. Функциональное уравнение

$$\varphi(x) = 2\varphi(x) + 2i$$

с функциональной константой $2x + 2i \in \{2x, 2x + 2, 2x + 1\}$для каждого i имеет единственное решение $\varphi(x) = i$.

Произвольная функция $g(x_1, \ldots, x_n)$ из класса S_{2x+2i} определяется системой функциональных уравнений, состоящей из 3^n уравнений вида

$$\psi(g_1(x), \ldots, g_n(x)) = g_0(x),$$

где $g_0, g_1, \ldots, g_n \in \{i, x, 2x + 2i\}$. Любое такое уравнение задает функцию g на тройке наборов, один из которых $(i, \ldots, i)$, на котором g должна принимать значение i, поскольку функции класса S_{2x+2i} сохраняют i. Поэтому в построенной системе также есть „лишние“ уравнения, которые можно вычеркнуть, — это, во-первых, $g(i, \ldots, i) = i$, а во-вторых, любое из пары уравнений, получающихся одно из другого заменой всех вхождений x на $2x+2i$ и одновременно $2x+2i$ на x.

Теорема 6 доказана.

Теорема 7. *Пусть*

$$R = \mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma_{01}, \sigma_{02}, \sigma_{12}, \sigma^0_{01}, \sigma^0_{02}, \sigma^0_{12}, \sigma^1_{01}, \sigma^1_{02}, \sigma^1_{12}, \sigma^2_{01}, \sigma^2_{02}, \sigma^2_{12}).$$

Тогда $\mathrm{SFE}[R] = P_3$.

Доказательство. Определим в классе R функцию $h(x, y_1, y_2, z_1, z_2, z_3)$ следующим образом:

$$\begin{cases} h(x, y_1, y_2, x, x+1, x+2) = y_1, \\ h(x, y_1, y_2, z_1, z_2, z_3) = y_2 \text{ для остальных значений } z_1, z_2, z_3. \end{cases}$$

Очевидно, что h сохраняет любое подмножество множества E_3, а значит, удовлетворяет предикатам $e_0, e_1, e_2, e_{01}, e_{02}, e_{12}$. Кроме того, h сохраняет и предикаты $\sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab}$, где $ab \in \{01, 02, 12\}$. В самом деле, любому такому предикату удовлетворяют ровно два набора. Поэтому при выборе шести столбцов из матрицы этого предиката будет получена матрица, каждая строка которой содержит не более двух различных элементов. Согласно определению функции h значение этой функции на каждой из строк матрицы будет совпадать со значением переменной y_2. То есть будет выбран третий столбец матрицы, откуда следует сохранение указанного предиката функцией h.

Таким образом, если при фиксированных x, z_1, z_2, z_3 и всех y_1, y_2 выполняется равенство

$$h(x, y_1, y_2, z_1, z_2, z_3) = y_1,$$

то обязательно $z_1 = x$, $z_2 = x+1$, $z_3 = x+2$. С учетом этого обстоятельства нетрудно построить систему функциональных уравнений над h, определяющую отрицание Поста $x+1$:

$$\begin{cases} \varphi^3(x) = x, \\ h(x, y_1, y_2, x, \varphi(x), \varphi^2(x)) = y_1. \end{cases}$$

Значит, SFE-замыкание класса R содержит функцию $x+1$, однако по Теореме 6 $\mathrm{SFE}[x+1] = S_{x+1}$, и поэтому $S_{x+1} \subseteq \mathrm{SFE}[R]$. Класс S_{x+1} является предполным по Теореме 4, а классу R принадлежат также функции, не сохраняющие предикат σ, например, функция $g(x, y, z)$:

$$g(x, y, z) = \begin{cases} 0, \text{ если } x, y, z \text{ различны}, \\ x \text{ в противном случае}. \end{cases}$$

Откуда получаем, по определению предполного класса, $\mathrm{SFE}[R] = P_3$.

Теорема 7 доказана.

Поскольку R есть минимальный дискриминаторный класс функций, не сохраняющих все четыре предиката σ, σ^0, σ^1, σ^2, а классы S_{x+1}, S_{2x}, S_{2x+2}, S_{2x+1} являются предполными в нем [5], получаем следующее важное утверждение.

Следствие. *В* P_3 *существует ровно четыре* SFE-*предполных класса* S_{x+1}, S_{2x}, S_{2x+2}, S_{2x+1}.

Теорема 8. *Класс* H_3 SFE-*предполон в классах* S_{x+1}, S_{2x}, S_{2x+2}, S_{2x+1}.

Доказательство. Согласно [5] наименьший дискриминаторный класс функций есть класс функций, сохраняющих все перечисленные выше предикаты — $\mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma, \sigma^0, \sigma^1, \sigma^2, \sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab})$, где $ab \in \{01, 02, 12\}$. По определению, это класс функций, сохраняющих любое подмножество E_3 и самодвойственных относительно любой перестановки на E_3. Согласно обозначениям статьи [7], это класс S_3^*, порождаемый тернарным дискриминатором $p(x, y, z)$, $S_3^* \subset H_3$. Значит, по Теореме 3, $\mathrm{SFE}[S_3^*] = H_3$.

Класс S_3^* является предполным в следующих дискриминаторных классах [5]:

$$\begin{aligned}
T &= \mathrm{Pol}(e_0, e_1, e_2, \sigma, \sigma^0, \sigma^1, \sigma^2), \\
T_\sigma &= \mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma, \sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab}), \\
T_{\sigma^0} &= \mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma^0, \sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab}), \\
T_{\sigma^1} &= \mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma^1, \sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab}), \\
T_{\sigma^2} &= \mathrm{Pol}(e_0, e_1, e_2, e_{01}, e_{02}, e_{12}, \sigma^2, \sigma_{ab}, \sigma^0_{ab}, \sigma^1_{ab}, \sigma^2_{ab}).
\end{aligned}$$

Построим их SFE-замыкания.

Класс T состоит из функций трехзначной логики, самодвойственных относительно любых перестановок на множестве E_3, то есть это класс всех однородных функций H_3 и $\mathrm{SFE}[T] = \mathrm{SFE}[H_3] = H_3$.

Рассмотрим функцию $g(x, y, z)$ из класса T_σ, такую что $g(x, y, z) = x$ везде, кроме наборов, состоящих из трех различных значений: $g(0, 1, 2) = 1$, $g(0, 2, 1) = 0$, $g(1, 0, 2) = 1$, $g(1, 2, 0) = 2$, $g(2, 0, 1) = 0$, $g(2, 1, 0) = 2$. Нетрудно заметить, что g принадлежит классу T_σ, но не принадлежит классу H_3. Единственным решением системы функциональных уравнений

$$\begin{cases} \varphi^3(x) = x, \\ p(x, \varphi(x), y) = x, \\ g(x, \varphi(x), \varphi^2(x)) = \varphi(x) \end{cases}$$

над функциональными константами из T_σ будет отрицание Поста $\varphi(x) = x+1$. Значит, $\mathrm{SFE}[x+1] = S_{x+1} \subseteq \mathrm{SFE}[T_\sigma]$, а поскольку в T_σ все функции сохраняют предикат σ, то $\mathrm{SFE}[T_\sigma] \subseteq S_{x+1}$, и поэтому $\mathrm{SFE}[T_\sigma] = S_{x+1}$.

Класс T предполон в классе $T' = \mathrm{Pol}(e_0, e_1, e_2, \sigma)$, причем $T_\sigma \subset T' \subset S_{x+1}$ [5], поэтому SFE-замыкание T' есть S_{x+1}.

Введем три функции $g_{\sigma^0}(x, y, z)$, $g_{\sigma^1}(x, y, z)$, $g_{\sigma^2}(x, y, z)$, равные своей третьей переменной z на всех наборах, кроме следующих:

$$g_{\sigma^0}(0,1,2) = 2, g_{\sigma^0}(0,2,1) = 1, g_{\sigma^0}(1,0,2) = g_{\sigma^0}(1,2,0) = g_{\sigma^0}(2,0,1) = g_{\sigma^0}(2,1,0) = 0,$$
$$g_{\sigma^1}(1,0,2) = 2, g_{\sigma^1}(1,2,0) = 0, g_{\sigma^1}(0,1,2) = g_{\sigma^1}(0,2,1) = g_{\sigma^1}(2,0,1) = g_{\sigma^1}(2,1,0) = 1,$$
$$g_{\sigma^2}(0,1,2) = g_{\sigma^0}(0,2,1) = g_{\sigma^0}(1,0,2) = g_{\sigma^0}(1,2,0) = 2, g_{\sigma^0}(2,0,1) = 1, g_{\sigma^0}(2,1,0) = 0.$$

Несложно показать, что $g_{\sigma^i} \in T_{\sigma^i}$, $i \in E_3$, и одновременно $g_{\sigma^i} \notin H_3$. В зависимости от значения i система функциональных уравнений над T_{σ^i} с главной функциональной переменной $\varphi(x)$

$$\begin{cases} \varphi^2(x) = x, \\ \psi(y) = \psi(z), \\ r_3(x, \varphi(x)) = \psi(y), \\ g_{\sigma^i}(z, x, \varphi(x)) = \varphi(x) \end{cases}$$

будет иметь единственное решение $\varphi(x) = 2x$ при $i = 0$, $\varphi(x) = 2x + 2$ при $i = 1$ и $\varphi(x) = 2x + 1$ при $i = 2$. В самом деле, множество решений первого уравнения системы — это $\{x, 2x, 2x + 1, 2x + 2\}$; второе уравнение вводит вспомогательную функцию $\psi(x)$, являющуюся константой, а третье — исключает тождественную функцию из множества значений главной функциональной переменной $\varphi(x)$. Далее в зависимости от используемой функциональной константы g_{σ^i} множество $\{2x, 2x + 1, 2x + 2\}$ решений рассматриваемой системы сокращается до единственной функции. Таким образом, $\mathrm{SFE}[2x + 2i] = S_{2x+2i} \subseteq \mathrm{SFE}[T_{\sigma^i}]$ и, аналогично сказанному выше, $S_{2x+2i} = \mathrm{SFE}[T_{\sigma^i}]$.

Теорема 8 доказана.

Построение решетки SFE-замкнутых классов в P_3 завершено.

Следствие. *В P_3 существует ровно шесть* SFE-*замкнутых классов:*

$$P_3,\ S_{x+1},\ S_{2x},\ S_{2x+2},\ S_{2x+1},\ H_3.$$

Причем S_{x+1}, S_{2x}, S_{2x+2}, S_{2x+1} являются SFE-*предполными, а класс H_3* SFE-*предполон во всех этих четырех классах.*

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект 09-01-00701).

Список литературы

[1] Боднарчук В. Г., Калужнин Л. А., Котов В. Н., Ромов Б. А. *Теория Галуа для алгебр Поста.* Кибернетика. 1969. № 3, с. 1–10. № 5, с. 1–9.

[2] Голунков Ю. В. *Полнота систем функций в операторных алгоритмах, реализующих функции k-значной логики.* Вероятностные методы и кибернетика. 1980. № 17. С. 23–34.

[3] Кон П. *Универсальная алгебра.* М.: Мир. 1968.

[4] Марченков С. С. *S-классификация функций многозначной логики.* Дискретная математика. 1997. Т. 9, № 3. С. 125–152.

[5] Марченков С. С. *Дискриминаторные классы трехзначной логики.* Математические вопросы кибернетики. Выпуск 12. 2003. С. 15–26.

[6] Марченков С. С. *Клоновая классификация дуально дискриминаторных алгебр с конечным носителем.* Математические заметки. 1997. Т. 61, № 3. С. 359–366.

[7] Марченков С. С. *Однородные алгебры.* Проблемы кибернетики. 1982. М.: Наука. Выпуск 39. С. 85–106.

[8] Марченков С. С. *Эквациональное замыкание.* Дискретная математика. 2005. Т. 17, вып. 2. С. 117–126.

[9] Марченков С. С., Федорова В. С. *О решениях систем функциональных булевых уравнений.* Дискретный анализ и исследование операций. 2008. Т. 15, № 6. С. 48–57.

[10] Марченков С. С., Федорова В. С. *О решениях систем функциональных уравнений многозначной логики.* Доклады РАН. 2009. Т. 426, № 4. С. 448–449.

[11] Марченков С. С., Федорова В. С. *Решения систем функциональных уравнений многозначной логики.* Вестник МГУ. Серия 15. Вычислительная математика и кибернетика. 2009. № 4. С. 29–33.

[12] Тайманов В. А. *О функциональных системах k-значной логики с операциями программного типа.* Доклады АН СССР. 1983. Т. 268, № 6. С. 1307–1310.

[13] Яблонский С. В. *Введение в дискретную математику.* М.: Наука. 1986.

[14] Янов Ю. И., Мучник А. А. *О существовании k-значных замкнутых классов, не имеющих конечного базиса.* Доклады АН СССР. 1959. Т. 127, № 1. С. 44–46.

[15] Post E. L. *Introduction to a general theory of elementary propositions.* Amer. J. Math. 1921. V. 43. P. 163–185.

[16] Post E. L. *The two-valued iterative systems of mathematical logic.* Princeton Univ. Press, Princeton, NJ. 1941.

[17] Rosenberg I. G. *Über die funktionale Vollständigkeit in den mehrwertigen Logiken.* Rozpravy Československe Acad. Věd., Řada Mat. Přir. Věd. Praha. 1970. Vol. 80, № 4. P. 3–93.

УДК 004.946

РЕКОНСТРУКЦИЯ ФОРМЫ И ТЕКСТУРЫ ГОЛОВЫ ЧЕЛОВЕКА ПО НАБОРУ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ СИНТЕТИЧЕСКИХ ГИБКИХ МОДЕЛЕЙ

© 2010 г. М. А. Федюков

mfedyukov@graphics.cs.msu.ru

Кафедра автоматизации систем вычислительных комплексов,
Лаборатория компьютерной графики и мультимедиа

1 Введение

Задача моделирования головы человека является актуальной во многих областях: идентификации, отслеживании, видеокодировании на базе трехмерных моделей, моделировании виртуального присутствия и других.

В зависимости от области применения моделирование может производиться с помощью трехмерного лазерного сканера [1, 2], по стереопаре [3, 4], по одной фотографии [5, 6], либо по набору изображений: фотографиям или кадрам видеопоследовательности [7, 8]. При постановке задачи реконструкции всей модели головы (но не только модели лица) с минимальными требованиями к входным данным оптимальным является последний подход.

Также методы моделирования можно классифицировать по типу выходных данных. В некоторых работах ими является облако точек [9]. Во многих работах реконструированная форма представляется в виде полигональной модели [3, 5, 6, 8].

Отдельной задачей, ставшей актуальной с развитием многопользовательских трехмерный сред является реконструкция *гибких моделей* [11], описывающих форму головы компактным набором параметров. Широко используемыми гибкими моделями являются *модели активной формы* [12], *модели активной внешности* [13], *морфируемые модели* [14], однако все они получены из реальных статистических данных (баз данных фотографий лиц, баз данных трехмерных отсканированных моделей лиц). В данной же работе рассматривается подход на основе синтетических гибких моделей головы. Положительной стороной таких моделей является четкое соответствие морфологическим характеристикам головы человека, таким как расстояние между глазами, пухлость губ или глубина ямочки подбородка (см., например, модель LAD [15], рис. 1). Как следствие, в зависимости от приложения преимуществами могут являться простота ее редактирования неподготовленным пользователем, возможность работы с ней на уровне возрастных, гендерных или этнотерриториальных морфологических характеристик. Другим преимуществом таких моделей является их компактное представление: в клиент-серверной архитектуре достаточно передать один раз гибкую модель, а впоследствии передавать только набор ее параметров, что ускоряет отправку клиенту нескольких моделей и их анимацию. Отдельной же сложностью работы с такими моделями является невозможность для некоторых входных фотографий реконструировать ряд черт лица с достаточной точностью из-за отсутствия соответствующих параметров в модели.

2 Постановка задачи

Везде, где не оговорено обратное, будем понимать под *полигональной моделью и* вектор $(u_1, u_2, ..., u_N)$, $u_i \in \mathbb{R}^3, i = 1..N$, состоящий из радиус-векторов вершин. Множество граней полигональной модели фиксировано, и потому не рассматривается с целью упрощения описания.

На отдельных же этапах, касающихся получения и работы с текстурой, будем рассматривать *текстурированную полигональную модель* $\tilde{u} = (\tilde{u}_1, \tilde{u}_2, ..., \tilde{u}_N)$, $\tilde{u}_i \in \mathbb{R}^5$, $\tilde{u}_i = (\tilde{u}_{i,x}, \tilde{u}_{i,y}, \tilde{u}_{i,z}, \tilde{u}_{i,u}, \tilde{u}_{i,v})$, $\tilde{u}_{i,x}, \tilde{u}_{i,y}, \tilde{u}_{i,z}, i = 1..N$ — пространственные координаты вершины, $\tilde{u}_{i,u}, \tilde{u}_{i,v}$ — текстурные координаты вершины.

Под *параметризованной полигональной моделью* a будем подразумевать синтетическую гибкую модель $a = b + Sp$, где $b = (b_1, b_2, ..., b_N)$ — *базовая* полигональная модель, задающая координаты вершин по умолчанию, $S = [s_{ij}]$ ($j = 1..K$; $s_{ij} \in \mathbb{R}^3 i = 1..N$) — матрица *смещений*, $p = (p_1, p_2, ..., p_K)$ ($p_j \in \mathbb{R}, -1 \leq p_j \leq 1, j = 1..K$) — вектор *параметров* модели, N — число вершин полигональной модели, K — число параметров.

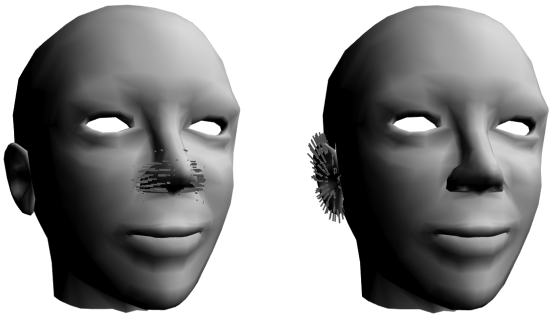

Рис. 1: Визуализация наборов векторов смещений, отвечающих за ширину носа (слева) и размер ушей (справа).

Решаемая задача заключается в разработке системы, определяющей значения всех параметров p и текстуру по набору из 1—4 фотографий (анфас, в профиль слева, в профиль справа и сзади).

3 Предлагаемый метод

Схема работы системы представлена на рис. 2.

Базовая полигональная модель предварительно размечена. *Универсальная разметка* $T = (t_1, t_2, ..., t_M)$ представляет собой набор векторов $t_i = (t_1^i, t_2^i, ..., t_{L_i}^i)$ ($t_j^i \in \mathbb{N}, t_j^i \leq N$; $i = 1..M$; $j = 1..L_i$; M — число векторов, L_i — число элементов векторов, N — число вершин полигональной модели), элементами которых являются индексы t_i^j вершин базовой полигональной модели. Таким образом, вектор $q_i = (q_1^i, q_2^i, ..., q_{L_i}^i)$, где $q_j^i = b_{t_j^i}$ задает ломаную, узлами которой являются вершины базовой полигональной модели. Далее $(q_1, q_2, ..., q_M)$ будем обозначать как Q. Каждая ломаная описывает одну характерную черту лица человека, такую как контур брови, носа, или всей головы.

Разметка (ломаные) заданы для каждой проекции — анфас, в профиль слева, в профиль справа и сзади: T_f (Q_f), T_l (Q_l), T_r (Q_r) и T_b (Q_b), соответственно (см. рис. 3). Однако, в силу того, что работа с разметкой и ломаными на большинстве этапов происходит одинаково для каждой проекции, для упрощения записи, где это возможно, будем писать T и Q, не уточняя индекс проекции.

3.1 Выделение черт лица

На каждой фотографии с помощью активных моделей формы находятся характерные точки лица, такие как уголки глаз и кончик носа. На каждую фотографию проецируется соответствующий набор ломаных: $\Omega = \mathbf{P}Q$, где $\mathbf{P} : \mathbb{R}^3 \to \mathbb{R}^2$ — оператор ортогонального проецирования на координатную плоскость. Узлы спроецированных ломаных, соответствующие найденным характерным точкам, перемещаются в соответствующие координаты. По ним вычисляется матрица аффинного преобразования, с помощью которой перемещаются все остальные

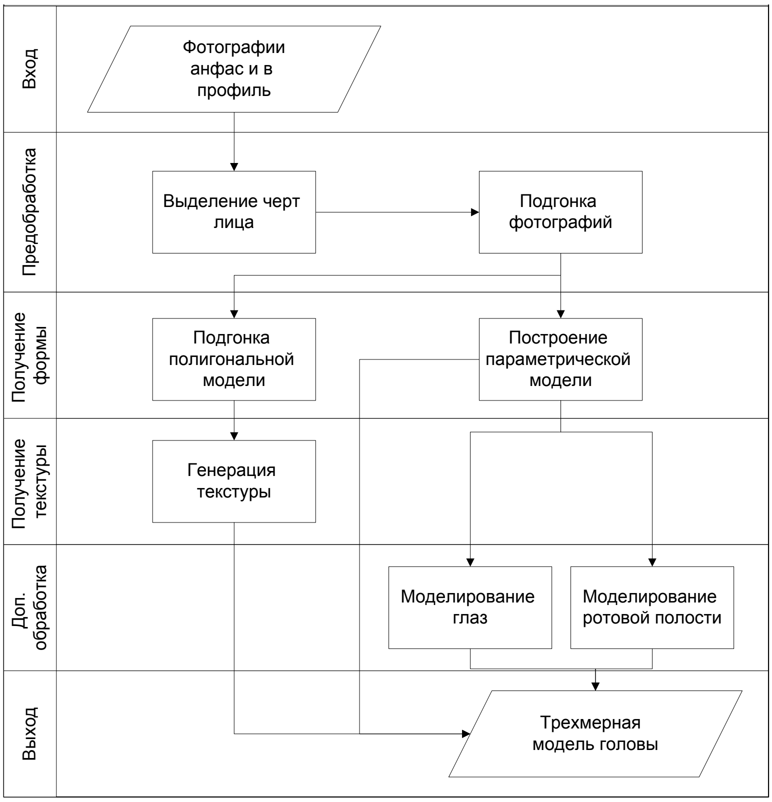

Рис. 2: Схема работы системы реконструкции головы человека по набору изображений.

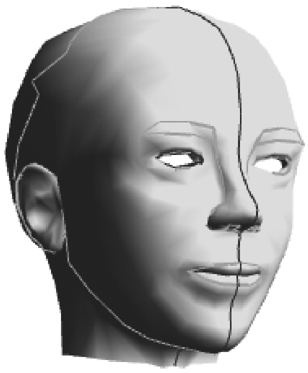

Рис. 3: Разметка для фотографий анфас (светло-серая) и в профиль справа (темно-серая).

узлы каждой ломаной. После этого дополнительно скорректировать положение узлов спроецированных ломаных Ω можно вручную.

В дальнейшем, не ограничивая общности, будем полагать, что на вход системе подаются 2 фотографии: анфас и в профиль справа. Работа системы с тремя или четырьмя фотографиями производится аналогично работе с двумя. В случае же подачи системе на вход только фотографии анфас, в качестве Ω используются контуры, описывающие среднестатистическую модель головы [16, 17].

3.2 Подгонка фотографий

На данном этапе фотографии, как текстурированные прямоугольники, располагаются в $\mathbb{R}^3$, корректируются их положение, поворот и масштаб.

Полагаем, что плоскость Oxy расположена горизонтально, ось Y направлена на наблюдателя, ось X — вправо, ось Z — вверх.

Сначала определяется положение фотографии анфас. Она располагается в плоскости Oxz и поворачивается вдоль оси Y таким образом, чтобы ось симметрии ломаных Ω_f на ней совпадала с осью Z.

Затем фотография в профиль располагается в плоскости Oyz и ее положение, поворот, и масштаб определяются при помощи минимизации среднеквадратичного отклонения вершин, входящих одновременно как в набор ломаных для фотографии анфас Q_f, так и в набор ломаных для фотографии анфас Q_r.

3.3 Построение параметризованной модели

На данном этапе положение узлов ломаных Ω считается фиксированным.

Цель этапа заключается в подгонке проекций ломаных, задаваемых параметризованной моделью, к ломаным, полученным на предыдущем этапе. Формально она может быть сформулирована как нахождение значений параметров p, при которых функционал $\mathcal{F} = \sum_{i=1}^{M} \sum_{j=1}^{L_i} (\omega_j^i - \omega'^i_j)^2$ принимает минимальное значение [18]. Здесь ω'^i_j — узлы ломаных $\omega' = \mathbf{P}Q$, а так как ломаные Q проходят по вершинам параметризованной модели $a(p)$, $\omega'^i_j = \omega'^i_j(p)$ (см. рис. 4).

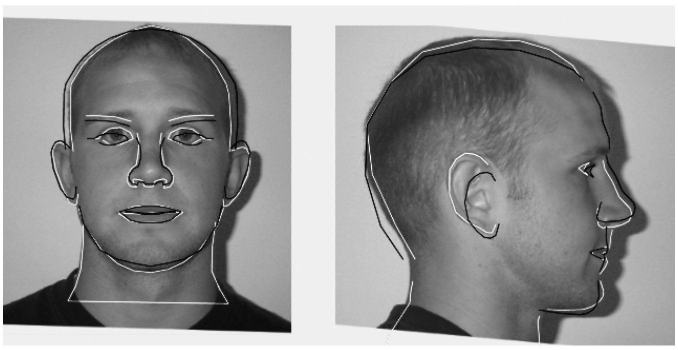

Рис. 4: Ломаные Ω, описывающие контуры характерных черт лица (белые), и Ω', подбираемые на данном этапе (черные).

Однако, в таком решении есть следующая проблема. Матрица смещений S может быть задана так, что множество всех возможных параметризованных полигональных моделей $\{a(p); p \in \mathbb{R}^K\}$ (K — число параметров) содержится в множестве всех возможных базовых полигональных моделей $\{b = (b_1, b_2, ..., b_N) | b_i \in \mathbb{R}^3\}$, но не равно ему. Т. е., на вход могут быть поданы такие фотографии и, в свою очередь, по ним будут подобран такой набор ломаных Ω, что не найдется параметров, которые преобразуют базовую модель к требуемому виду.

Так как психофизический процесс распознавания лиц человеком состоит из нескольких этапов, обработка форм отдельных черт лица и обработка композиции всех черт происходят на разных этапах и вносят разный вклад в узнаваемость лиц [10, 19, 20], целесообразно разбить набор ломаных Q на группы Q^m $(m = 1, 2, ..., H)$ таким образом, чтобы в группах $Q_1, Q_2, ..., Q_{H-1}$ содержался контур только одной черты лица (для подбора ее формы), а в группе Q_H — контуры всего лица и минимальный набор характерных точек всех остальных черт лица. Таким образом, задача сводится к независимому поиску значений из наборов p^m, минимизирующих функционалы $G_m = \sum_{i=l_m}^{l_{m+1}-1} \sum_{j=1}^{L_i} (\omega_j^{k_i} - {\omega'}_j^{k_i})^2$.

Минимизация осуществляется методом градиентного спуска.

3.4 Подгонка полигональной модели

Итак, получен набор параметров p, минимизирующий отклонение контуров характерных черт лица на модели a от контуров на входных фотографиях. Однако, в зависимости от заданной матрицы смещений S значения функционалов невязки G_m могут быть достаточно большими и получаемая полигональная модель a может быть недостаточно точной для генерации по ней текстуры.

Для решения этой проблемы создается текстурированная полигональная модель $\tilde{c}$ — копия базовой текстурированной модели $\tilde{b}$. Затем ломаные на фотографиях Ω проецируются оператором $\mathbf{P}^{-1}$ из $\mathbb{R}^2$ в $\mathbb{R}^3$ (получаемые ломаные будем обозначать $\hat{Q}$): $\hat{q}^i_{j,x} = \omega^i_{j,x}$, $\hat{q}^i_{j,y} = \omega^i_{j,y}$, а значение z-компоненты восстанавливается из исходной модели: $\hat{q}^i_{j,z} = q^i_{j,z}$.

Подгонка полигональной модели $\tilde{c}$ происходит итеративно. На каждой итерации i вершины, принадлежащие одной черте лица (заданной универсальной разметкой T), перемещаются в соответствующие координаты $\hat{Q}$: $\tilde{c}_{t^i_j} = \hat{q}^i_j$. Кроме того, на каждой итерации перемещаются вершины, лежащие в окрестности $\tilde{c}_{t^i_j}$. Окрестность также находится итеративно. Каждой вершине модели $\tilde{c}$ присваивается вес ν_k $(k = 1..N$, N — число вершин полигональной модели). Изначально веса всех вершин равны 0. На первой итерации веса вершин, принадлежащих текущей черте лица, $\tilde{c}_{t^i_j}$ принимаются равными 1. На очередной итерации j перебираются все вершины модели $\tilde{c}$, соединенные ребрами с вершинами, участвовавшими в предыдущей итерации, за исключением вершин $\tilde{c}_{t^m_j}|_{m \neq i}$, не принадлежащих черте лица итерации i. Веса перебираемых вершин вычисляются следующим образом: $\nu^k_{\tilde{c}_j} = max(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\lambda j}{2\sigma^2}\right), \nu^k_{\tilde{c}_j})$, где σ — коэффициент масштаба, λ — нормировочный коэффициент.

3.5 Генерация текстуры

Для генерации текстуры создается текстурированная полигональная модель $\tilde{d}$.

В случае фотографии анфас, значения элементов $\tilde{d}$ задаются следующим образом: $\tilde{d}_x = 0$, $\tilde{d}_y = \tilde{c}_u$, $\tilde{d}_z = \tilde{c}_v$, $\tilde{d}_u = \mathbf{P}\tilde{c}_y$, $\tilde{d}_v = \mathbf{P}\tilde{c}_z$, где $\mathbf{P}$ — оператор ортогонального проецирования. Полученная плоская полигональная модель визуализируется в текстуру (в данном случае, фронтальную).

Аналогичным образом генерируется текстура по фотографии в профиль. Для использования на этапе смешивания текстур генерируется карта нормалей. Значения в каждой точке маски смешивания вычисляются как косинус угла между осью Y и нормалью в данной точке (см. рис. 5).

3.6 Дополнительная обработка

Если система, в которой будет использоваться полученная модель головы, предполагает анимацию (в частности, моргание глаз и открывание рта), то исходная базовая модель головы не включает себя модели глаз, зубов, языка и полости рта. Помимо них на данном этапе могут моделироваться дополнительные элементы модели, такие как очки, головные уборы, прическа, одежда и т. п. (см. рис. 7).

Модель каждого глаза, представляющая собой объединение двух сфер (глазное яблоко и зрачок) переносится и масштабируется в соответствии с положением ломаной $\hat{Q}$, описывающей контур данного глаза. Усредненный цвет радужки глаза определяется по фотографии анфас.

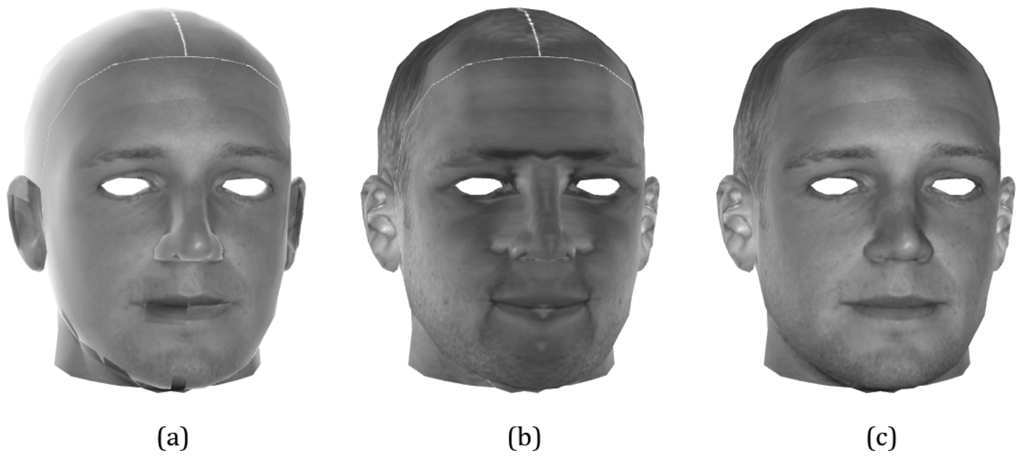

Рис. 5: (a) — текстура, полученная по фотографии анфас, с маской, (b) — текстура, полученная по фотографии в профиль, (c) — смешанная текстура.

Затем соответственно изменяется цвет радужки на предварительно подготовленной развертке фотографии глаза.

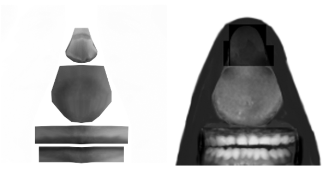

Рис. 6: Фрагменты текстуры, полученной с фотографий и предварительно подготовленной текстуры ротовой полости.

Модели полости рта, а также языка и зубов, подгоняются аналогично моделям глаз. Текстура, полученная с фотографий, смешивается с предварительно полученной разверткой фотографий ротовой полости, включающей зубы и язык (см. рис. 6).

В случае ограничения на размер текстуры она может обрабатываться билатеральным фильтром, для сглаживания крупной зернистости на участках кожи, с сохранением при этом мелких деталей на участках с высокой локальной контрастностью.

4 Заключение и предстоящая работа

В статье представлен разработанный метод построения параметризованной модели головы человека по набору изображений. Данный подход применим к любым параметризованным моделям, обладающим описанными свойствами. В рамках дальнейшей работы по данному проекту предполагается как улучшение предложенных алгоритмов, так и добавление новых этапов, таких как фильтрация текстуры и моделирование прически в соответствии с входными фотографиями.

Список литературы

[1] C. Xu, L. Quan, Y. Wang, T. Tan, and M. Lhuillier. Adaptive multi-resolution fitting and its application to realistic head modeling. In *Proceedings of Geometric Modeling and Processing*, 2004.

Рис. 7: Итоговые текстурированные модели.

[2] Frank Hülsken, Christian Eckes, Roland Kuck, Jörg Unterberg, and Sophie Jörg. Modeling and animating virtual humans for real-time applications. In *The International Journal of Virtual Reality, 2007, 6(4):11-20*, 2007.

[3] G. Galicia and A. Zakhor. Depth based recovery of human facial features from video sequences. In *Image Processing, 1995. Proceedings., International Conference on*, volume 2, 1995.

[4] R. Lengagne, P. Fua, and O. Monga. 3D stereo reconstruction of human faces driven by differential constraints. *Image and Vision Computing*, 18(4):337–343, 2000.

[5] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3D structure recovery of faces. *Lecture Notes in Computer Science*, pages 99–113, 2004.

[6] W.A.P. Smith and E.R. Hancock. A Model-Based Method for Face Shape Recovery. In *Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA*, pages 268–276. Springer, 2005.

[7] P. Fua. Regularized bundle-adjustment to model heads from image sequences without calibration data. *International Journal of Computer Vision*, 38(2):153–171, 2000.

[8] VG Zhislina, DV Ivanov, VF Kuriakin, VS Lempitskii, EM Martinova, KV Rodyushkin, TV Firsova, AA Khropov, and AV Shokurov. Creating and Animating Personalized Head Models from Digital Photographs and Video. *Programming and Computer Software*, 30(5):242–257, 2004.

[9] Y. Xu, C.S. Xu, Y.L. Tian, S.D. Ma, and M.L. Luo. 3D face image acquisition and reconstruction system. In *IEEE Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings*, volume 2, 1998.

[10] M. Minear, D.C. Park. A lifespan database of adult facial stimuli. In *Behavior Research Methods, Instruments and Computers.*, 36, 2004.

[11] Lanitis, A., Taylor, C., and Cootes, T. Automatic face identification system using flexible appearance models. *Image and Vision Computing 13*, 5, 393–401, 1995.

[12] Cootes, T., Taylor, C., Cooper, D., Graham, J., et al. Active shape models-their training and application. *Computer vision and image understanding 61*, 1, 38–59, 1995.

[13] Cootes, T., Edwards, G., and Taylor, C. Active appearance models. *Computer Vision—ECCV'98*, 484, 1998.

[14] Blanz, V., and Vetter, T. A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 187–194, 1999.

[15] Agent appearance definition. http://lib.openmetaverse.org/wiki/AgentSetAppearance, as of 2010.

[16] Leslie G. Farkas. *Anthropometry of the head and face in medicine.* New York : Elsevier, 1981.

[17] Leslie G. Farkas and Ian R. Munro. *Anthropometric facial proportions in medicine*, chapter Facial Proportions in Medical Illustration. Springfield: Charles C. Thomas, 1987.

[18] М. А. Федюков. Аппроксимация полигональной сетки головы человека с помощью параметризованной антропометрической модели. In *Труды международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов–2009»*, 2009.

[19] A. Schwaninger, C.C. Carbon, and H. Leder. Expert face processing: Specialization and constraints. *Development of face processing*, pages 81–97, 2003.

[20] C. Wallraven, A. Schwaninger, and H.H. Bülthoff. Learning from humans: computational modeling of face recognition. *Network: Computation in Neural Systems*, 16(4):401–418, 2005.

УДК 512.57

О НАДСТРУКТУРЕ КЛАССОВ МОНОТОННЫХ ФУНКЦИЙ В МНОГОЗНАЧНЫХ ЛОГИКАХ

© 2010 г. В. Б. Ларионов

Кафедра Математической кибернетики

1 Основные понятия

Обозначим множество $E_k = \{0, 1, \ldots, k-1\}$.

Определение 1. Функция $f(x_1, \ldots, x_n)$ называется функцией k-значной логики ($k \geqslant 2$), если она определена на E_k^n и все ее значения принадлежат E_k.

Будем использовать следующие стандартные обозначения. Множество всех функций k-значной логики обозначим P_k. Для любого подмножества A из P_k через $[A]$ будем обозначать замыкание относительно операции суперпозиции (для функций далее везде будет идти речь именно об этом типе замыкания).

Пусть на E_k задано некоторое отношение частичного порядка r. Возьмем два произвольных набора $\tilde{a} = (a_1, \ldots, a_n)$ и $\tilde{b} = (b_1, \ldots, b_n)$ из E_k^n. Будем говорить, что $\tilde{a}$ не превосходит $\tilde{b}$ относительно частичного порядка r и записывать $\tilde{a} \leqslant_r \tilde{b}$, если для любого $1 \leqslant i \leqslant n$ справедливо неравенство $a_i \leqslant_r b_i$.

Определение 2. Функция $f(x_1, \ldots, x_n)$ называется монотонной относительно частичного порядка r, если для любых двух наборов $\tilde{a}, \tilde{b} \in E_k^n$ таких, что $\tilde{a} \leqslant_r \tilde{b}$, выполнено $f(\tilde{a}) \leqslant_r f(\tilde{b})$. Множество всех функций из P_k, монотонных относительно r, называется монотонным классом M_r.

Для краткости мы будем задавать частичный порядок r частично упорядоченным множеством (ЧУМ) H из элементов E_k, а соответствующий монотонный класс обозначать M_H.

Определение 3. Пусть $p(x_1, \ldots, x_m)$ – некоторый предикат, определенный на E_k^m, $f(y_1, \ldots, y_n)$ – функция из P_k. Говорят, что функция $f(y_1, \ldots, y_n)$ сохраняет предикат $p(x_1, \ldots, x_m)$, если для любых n наборов $\tilde{a_i} = (a_{i1}, \ldots, a_{im})$ $(i = 1, \ldots, n)$, удовлетворяющих предикату p, набор $f(a_{11}, \ldots, a_{n1}), \ldots, f(a_{1m}, \ldots, a_{nm})$ также удовлетворяет предикату p. По определению будем считать, что тождественно ложный предикат сохраняет любая функция.

Будем обозначать через $\mathrm{Pol}(p)$ множество функций, сохраняющих предикат p.

Класс M_H является замкнутым классом функций, сохраняющих предикат $R(x,y) = TRUE \iff x \leqslant_r y$ ([3]). Везде далее, когда мы будем писать, что монотонный класс задается предикатом R, мы будем подразумевать именно описанный предикат $R(x,y)$.

Одним из семейств предполных классов в P_k при $k \geqslant 3$ является семейство $\mathbf{M}$ ([4]) — подмножество множества всех классов монотонных функций. Известно ([5]), что монотонный класс является предполным (принадлежит множеству $\mathbf{M}$) тогда и только тогда, когда частичный порядок, сохраняемый им, обладает в точности одним минимальным и одним максимальным элементом.

Ранее автором изучался вопрос о строении надструктуры монотонных классов, не входящих в семейство $\mathbf{M}$ ([1], [2]). Было показано, что существуют монотонные классы с бесконечной надструктурой, минимальной логикой с подобным классом является P_4.

В данной статье представлены новые методы работы с формулами предикатов, на основе которых доказываются некоторые общие результаты о надструктуре классов монотонных функций. Также описывается критерий для наличия бесконечной надструктуры для некоторого семейства классов монотонных функций.

Далее мы приведем некоторые результаты о соответствии Галуа между замкнутыми классами предикатов и функций.

Пусть T – некоторая система предикатов.

Определение 4.

1. Если p – обозначение m-местного предиката из T, $x_1, \ldots, x_m$ – некоторые различные символы переменных, то $p(x_1, \ldots, x_m)$ – формула над T, реализующая предикат $p(x_1, \ldots, x_m)$. В этой формуле все переменные $x_1, \ldots, x_m$ – свободные, связанные переменные отсутствуют.

2. Пусть $A(y_1, \ldots, y_l)$ – формула над T со свободными переменными $y_1, \ldots, y_l$, реализующая предикат $p(y_1, \ldots, y_l)$. Тогда $(\exists y_i)A(y_1, \ldots, y_i, \ldots, y_l)$ – формула над T со свободными переменными $y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_l$ и связанными переменными такими же, как у $A(y_1, \ldots, y_l)$, с добавлением y_i, реализующая предикат $\exists y_i p(y_1, \ldots, y_i, \ldots, y_l)$.

3. Пусть опять $A(y_1, \ldots, y_l)$ – формула над T со свободными переменными $y_1, \ldots, y_l$, реализующая предикат $p(y_1, \ldots, y_l)$, $x_{i_1}, \ldots, x_{i_l}$ – не обязательно различные символы переменных, отличающиеся от символов связанных переменных формулы $A(y_1, \ldots, y_l)$. Тогда $A(x_{i_1}, \ldots, x_{i_l})$ – формула над T со свободными переменными $x_{i_1}, \ldots, x_{i_l}$ (тут имеются в виду все различные переменные из списка $x_{i_1}, \ldots, x_{i_l}$) и связанными переменными такими же, как у $A(y_1, \ldots, y_l)$, реализующая предикат $p(x_{i_1}, \ldots, x_{i_l})$.

4. Пусть $A(x_1, \ldots, x_l)$ и $B(y_1, \ldots, y_m)$ – формулы над T с множествами свободных переменных соответственно $x_1, \ldots, x_l$ и $y_1, \ldots, y_m$, реализующие предикаты $p_1(x_1, \ldots, x_l)$ и $p_2(y_1, \ldots, y_m)$, причем множества связанных переменных этих формул не пересекаются. Тогда $A(x_1, \ldots, x_l)\&B(y_1, \ldots, y_m)$ – формула над T со свободными переменными $x_1, \ldots, x_l, y_1, \ldots, y_m$ (опять мы подразумеваем все различные переменные из списка $x_1, \ldots, x_l, y_1, \ldots, y_m$), реализующая предикат $p_1(x_1, \ldots, x_l)\&p_2(y_1, \ldots, y_m)$. Множество связанных переменных этой формулы – объединение множеств связанных переменных формул $A(x_1, \ldots, x_l)$ и $B(y_1, \ldots, y_m)$.

Формулами над системой T являются только объекты, которые можно построить за конечное число шагов по пунктам 1-4. Других формул над системой предикатов T не существует.

Замыканием $[T]$ системы T называется множество предикатов, реализуемых всевозможными формулами над T.

Далее для упрощения изложения будем называть операции над предикатами следующим образом: добавление квантора существования (пункт 2 определения) – проекцией по соответствующей переменной, подстановку вместо $y_{j_1}, \ldots, y_{j_h}$ одного и того же символа переменной (пункт 3 определения) – отождествлением переменных $y_{j_1}, \ldots, y_{j_h}$, операцию, описанную в пункте 4 – произведением предикатов p_1 и p_2 (мы явно будем указывать общие свободные переменные предикатов p_1 и p_2 там, где это необходимо).

Определение 5. Предикаты R_1 и R_2 называются эквивалентными, если $\mathrm{Pol}(R_1) = \mathrm{Pol}(R_2)$. Под $R_1 = R_2$ будем подразумевать равенство предикатов как функций.

Определение 6. Пусть на множестве $\{1, 2, \ldots, m\}$ задано отношение эквивалентности ϵ, то есть бинарное рефлексивное, симметричное и транзитивное отношение. Диагональю, соответствующей отношению ϵ, называется такой предикат d_ϵ, что

$$d_\epsilon(x_1, \ldots, x_m) = \mathop{\&}_{\epsilon(i,j)} (x_i = x_j),$$

где конъюнкия берется по всем числам i, j из $\{1, 2, \ldots, m\}$, эквивалентным в смысле отношения ϵ.

Множество всех диагоналей будем обозначать через D. Двухместную диагональ, соответствующую полному отношению ϵ, обозначим через $d(x_1, x_2)$, то есть $d(x_1, x_2) = (x_1 = x_2)$.

Отметим, что в случае единичного отношения ϵ, соответствующая ему диагональ представляет собой всюду истинный предикат.

Лемма 1([6]). Если $R_1 \in [R_2]$, то $\mathrm{Pol}(R_2) \subseteq \mathrm{Pol}(R_1)$.

Обозначим через $\mathrm{Inv}(f)$ множество всех предикатов, каждый из которых сохраняется функцией f. Для произвольных множества функций A и множества предикатов T положим $\mathrm{Inv}(A) = \cap_{f \in A}\mathrm{Inv}(f)$, $\mathrm{Pol}(T) = \cap_{p \in T}\mathrm{Pol}(p)$.

Основными результатами теории Галуа для замкнутых классов являются соотношения ([7]):

$$A = \mathrm{Pol}(\mathrm{Inv}(A)), \quad T = \mathrm{Inv}(\mathrm{Pol}(T)), \tag{1}$$

где A – замкнутый класс функций, содержащий все селекторные функции (т.е. функции вида $f(x_1, \ldots, x_n) = x_i$), а T – замкнутый класс предикатов, содержащий все диагонали.

Лемма 2([7]). $D = [d(x_1, x_2)]$.

Лемма 3. Пусть $p(x) \in [R]$ — одноместный предикат. Тогда p является тождественно истинным предикатом.

Пусть предикат $p \in [R]$, где R – предикат, задающий монотонный класс. Запишем p в виде формулы F. Заметим, что в формуле F можно вынести вперед все кванторы существования, не изменив предикат p.

Сопоставим F ориентированный граф G_F по следующему правилу: между множеством вершин G_F и множеством переменных F (учитываем и свободные и связанные) существует взаимно однозначное соответствие. Вершину, соответствующую переменной x, пометим символом "x" , если переменная x свободная и "$\exists x$" , если связанная. Данную вершину для краткости изложения будем обозначать v_x. В графе G_F есть ориентированное ребро (v_x, v_y) тогда и только тогда, когда в формуле F содержится запись $R(y, x)$.

Лемма 4([1]). Для любого $p \in [R]$ существует эквивалентный предикат p', задаваемый формулой F', такой, что в графе $G_{F'}$ отсутствуют ориентированные циклы. Если при этом существует формула F, реализующая p такая, что в графе G_F нет ориентированного цикла, в котором хотя бы две вершины отвечают свободным переменным, то $p' = p$.

Согласно лемме 4, формуле F' с графом $G_{F'}$ без ориентированных циклов можно следующим образом сопоставить частично упорядоченное множество $L_{F'}$. Между множеством элементов ЧУМ $L_{F'}$ и множеством всех переменных в формуле F' существует взаимно однозначное соответствие (опять обозначим это так: элемент $v_x \in L_{F'}$ соответствует переменной $x \in F'$). Для всех переменных полагаем $v_x \leqslant v_x$, и, если формула F' содержит запись $R(x, y)$, то $v_x \leqslant v_y$. Замыканием по транзитивности получаем искомое ЧУМ $L_{F'}$ (поскольку в графе $G_{F'}$ отсутствуют ориентированные циклы).

Определение 7. Через $\overline{F}_1$ обозначим множество формул F над $\{R\}$, задающих предикаты $p \in [R]$, для которых в графе G_F отсутствуют ориентированные циклы, а также компоненты связности, в которых отсутствуют вершины, соответствующие свободным переменным.

Непосредственным следствием леммы 4 является следующая лемма.

Лемма 5. Пусть предикат $p \in [R]$. Тогда существует предикат $p' \in [R]$, задаваемый формулой $F \in \overline{F}_1$, такой что p и p' эквивалентны.

Определение 8. Формулу F, реализующую некоторый предикат $p \in [R]$, назовем несвязной, если граф G_F несвязен.

Пусть для некоторого предиката $p \in [R]$ граф G_F его формулы F имеет t компонент связности $G_{F_1}, \ldots, G_{F_t}$. Каждой компоненте связности G_{F_i} соответствует некоторая подформула F_i формулы F, задающая свой предикат p_i, $i = 1, \ldots, t$. По определению графа G_F получаем, что $p = p_1 \& p_2 \& \ldots \& p_t$, причем в указанной конъюнкции разные предикаты-сомножители p_i и p_j не имеют общих переменных, $p_i \in [R]$, где $i = 1, \ldots, t$.

Лемма 6([3]). Для произвольного предиката R из представления $R = R_1 \& R_2 \& \ldots R_t$, где предикаты $R_1, \ldots, R_t$ не имеют общих переменных и не являются тождественно ложными, следует $\mathrm{Pol}\, R = \mathrm{Pol}\, R_1 \cap \mathrm{Pol}\, R_2 \cap \ldots \cap \mathrm{Pol}\, R_t$.

Таким образом, предикат p задает пересечение классов, задаваемых предикатами p_i, $i = 1, \ldots, t$.

Следующая лемма является следствием соотношений Галуа и свойств формул, задающих предикаты $p \in [R]$.

Лемма 7. Пусть $p_1, p_2 \in [R]$, где R – предикат, задающий монотонный класс.

1. Пусть предикат p_2 задается формулой $F \in \overline{F}_1$. Если $\mathrm{Pol}\, p_1 \subseteq \mathrm{Pol}\, p_2$, то выполнено $p_2 \in [p_1]$.

2. Если для некоторого предиката p_3 выполнено $\mathrm{Pol}\, R \subseteq \mathrm{Pol}\, p_3$, то $p_3 \in [R]$.

Лемма 8([1]). Пусть предикат $p \in [R]$.

1. Если найдется формула $F \in \overline{F}_1$, задающая предикат p, такая, что в ЧУМ L_F существует пара сравнимых элементов v_{x_1}, v_{x_2}, то справедливо равенство $\mathrm{Pol}\, p = \mathrm{Pol}\, R$.

2. Если выполнено соотношение $\mathrm{Pol}\, p = \mathrm{Pol}\, R$, то для любой формулы $F \in \overline{F}_1$, задающей предикат p, в ЧУМ L_F найдется пара сравнимых элементов, соответствующих свободным переменным предиката p.

2 Невырожденные предикаты

Согласно лемме 8, изучение надструктуры класса монотонных функций M_H сводится к изучению всевозможных формул, реализующих предикаты над множеством $\{R\}$. К сожалению таких предикатов и формул очень много. В этом разделе доказываются результаты, позволяющие ограничить количество рассматриваемых формул. В конце раздела доказывается основная лемма, устанавливающая ограничения на выразимость формулами друг из друга различных предикатов, задающих замкнутые классы, содержащие класс M_H.

Определение 9. Аппроксимацией предиката $p(x_1, \dots, x_n)$, где $n \geqslant 2$, назовем следующий n-местный предикат

$$\Psi(p)(x_1, \dots, x_n) = \exists y_1, \dots, y_n p(y_1, x_2, x_3, \dots, x_{n-1}, x_n) \&$$

$$\& p(x_1, y_2, x_3, \dots, x_{n-1}, x_n) \& \dots \& p(x_1, x_2, x_3, \dots, y_{n-1}, x_n) \&$$

$$\& p(x_1, x_2, x_3, \dots, x_{n-1}, y_n).$$

Аппроксимацией одноместного предиката $p(x)$ назовем одноместный предикат $\Psi(p)(x) = \exists y p(y)$ (x — несущественная переменная предиката $\Psi(p)(x)$).

Определение 10. Предикат $p(x_1, \dots, x_n)$, где $n \geqslant 2$, назовем невырожденным, если существует набор $\tilde{a} \in E_k^n$ такой, что $p(\tilde{a}) = FALSE$, но для любого номера $i \in \{1, \dots, n\}$ существует элемент $b_i \in E_k$ такой, что $p(a_1, \dots, a_{i-1}, b_i, a_{i+1}, \dots, a_n) = TRUE$. В случае одноместного предиката, $p(x)$ — невырожден тогда и только тогда, когда он отличен от тождественно истинного или ложного предиката. В противном случае предикат назовем вырожденным.

Отметим некоторые простые свойства введенных понятий.

Если $\Psi(p)(\tilde{a}) = FALSE$, то $p(\tilde{a}) = FALSE$.

Если предикат p имеет хотя бы одну несущественную переменную, то он вырожден.

Лемма 9. Предикат равен своей аппроксимации тогда и только тогда, когда он вырожден.

Доказательство Пусть p — одноместный предикат. По определению p вырожден тогда и только тогда, когда он является тождественно истинным или ложным предикатом. Равенство $p(x) = \Psi(p)(x)$ также имеет место в том и только том случае, когда p — тождественно истинен или ложен.

Пусть теперь p имеет местность $n > 1$. Предположим вначале, что предикат вырожден. Это означает, что для любого набора $\tilde{a}$ такого, что $p(\tilde{a}) = FALSE$, найдется номер i такой, что $\exists y_i p(a_1, \dots, a_{i-1}, y_i, a_{i+1}, \dots, a_n) = FALSE$, откуда $\Psi(p)(\tilde{a}) = FALSE$. С учетом сказанного выше получаем, что $\Psi(p)(\tilde{x}) = p(\tilde{x})$.

Если же предикат p невырожден, то для набора $\tilde{a}$ из определения 10 справедливо, что для любого $i \in \{1, \dots, n\}$ $\exists y_i p(a_1, \dots, a_{i-1}, y_i, a_{i+1}, \dots, a_n) = TRUE$, откуда $\Psi(p)(\tilde{a}) = TRUE$, но $p(\tilde{a}) = FALSE$, то есть предикат p не равен своей аппроксимации.

Лемма доказана.

Лемма 10. Если для некоторого предиката p справедливо представление $p(\tilde{x}) = p_1(\tilde{x}) \& \dots \& p_l(\tilde{x})$, где предикаты-сомножители $p_1, \dots, p_l$ вырождены, то предикат $p(\tilde{x})$ также вырожден.

Доказательство Возьмем любой набор $\tilde{a}$ такой, что $p(\tilde{a}) = FALSE$. Это означает, что найдется предикат p_i, где $i = 1, \dots, s$ такой, что $p_i(\tilde{a}) = FALSE$. Поскольку p_i вырожден, то либо p_i — одноместный предикат, но тогда он тождественно ложный, и предикат p — вырожден, либо местность p_i больше единицы, и найдется номер j такой, что предикат $\exists y_j p_i(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)$ ложен на наборе $(a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n)$. Отсюда получаем, что на указанном наборе ложен и предикат $\exists y_j p(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n)$. В самом деле, $\exists y_i p(\tilde{u}) = (\exists y_i)(p_1(\tilde{u})\& \ldots \& p_i(\tilde{u})\& \ldots \& p_l(\tilde{u}))$, где $\tilde{u} = (x_1, \ldots, x_{j-1}, y_i, x_{j+1}, \ldots, x_n)$, но предикат $p_i(a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n) = FALSE$ при любом b.

Получаем, что $\Psi(p)(\tilde{a}) = FALSE$. Поскольку $\tilde{a}$ — произвольный набор, то $\Psi(p) = p$, и по лемме 9 p — вырожденный предикат.

Лемма доказана.

Лемма 11. Если предикат $p(\tilde{x}^n)$, где $n > 1$, не являющийся тождественно ложным, вырожден, то

$$\operatorname{Pol}(p(x_1, \ldots, x_n)) = \bigcap_{i=1}^{n} \operatorname{Pol}(\exists y_i p(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)).$$

Доказательство Сделаем следующее обозначение: $p_i = \exists y_i p(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)$. Из указанного соотношения следует, что $p_i \in [p]$. Но тогда множеству $[p]$ принадлежит и следующий предикат: $p'(x_1, \ldots, x_{n(n-1)}) = p_1(x_1, \ldots, x_{n-1})\& \ldots \& p_n(x_{n(n-1)-n+2}, \ldots, x_{n(n-1)})$. Если бы какой-то из предикатов p_i являлся тождественно ложным, то p также бы был тождественно ложен, что невозможно в силу условия леммы. По лемме 6 получаем, что $\operatorname{Pol} p' = \bigcap_{i=1}^{n} \operatorname{Pol} p_i$. Итак, по лемме 1 имеем, что $\operatorname{Pol} p \subseteq \operatorname{Pol} p' = \bigcap_{i=1}^{n} \operatorname{Pol} p_i$.

С другой стороны, если предикат p вырожден, то $p = \Psi(p)$, следовательно $p \in [p_1, \ldots, p_n]$. Поскольку $p_i \in [p']$ (предикат p_i получается из p' проекцией по всем переменным, кроме $x_{i(n-1)-n+2}, \ldots, x_{i(n-1)}$), получаем, что $p \in [p']$, откуда по лемме 1 имеем $\operatorname{Pol} p' \subseteq \operatorname{Pol} p$.

Объединяя полученные результаты, имеем $\operatorname{Pol} p = \operatorname{Pol} p' = \bigcap_{i=1}^{n} \operatorname{Pol} p_i$.

Лемма доказана.

Определение 11. Обозначим через T_1 множество предикатов $p(\tilde{x}^n) \in [R]$, $n > 1$, обладающих следующими свойствами:

1. $\operatorname{Pol} p \neq \operatorname{Pol} R$.
2. Предикат p невырожден.
3. Для любого $i \in \{1, \ldots, n\}$ выполнено $\operatorname{Pol} p(x_1, \ldots, x_n) \neq \operatorname{Pol}(\exists y p(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n))$.

Рассмотрим важное свойство семейства T_1.

Лемма 12. Любой класс $A = \operatorname{Pol} p$, где $p \in [R]$, $\operatorname{Pol} p \neq \operatorname{Pol} R$, $A \neq P_k$, представим конечным пересечением: $A = \bigcap_i \operatorname{Pol} p_i$, где все p_i — предикаты из семейства T_1.

Доказательство Возьмем произвольный класс $A = \operatorname{Pol} p$, $A \neq P_k$. Если предикат $p \in T_1$, то искомое представление тривиально: $A = \operatorname{Pol} p$.

Заметим, что предикат p не может быть тождественно истинным или ложным в силу $A \neq P_k$, а также не может быть одноместным в силу леммы 3.

Пусть $p \notin T_1$. Тогда либо существует проекция p_i предиката $p(x_1, \ldots, x_n)$ по переменной x_i, такая что $\operatorname{Pol} p_i = \operatorname{Pol} p$, либо p вырожден ($\operatorname{Pol} p \neq \operatorname{Pol} R$ выполнено по условию леммы). В первом случае вместо p рассмотрим предикат p_i и повторим рассуждения сначала. Во втором случае применив к предикату p лемму 11, получим $A = \bigcap_{i=1}^{n} \operatorname{Pol} p_i$, причем для любого $i \in \{1, \ldots, n\}$ выполнено $A \neq \operatorname{Pol} p_i$ (следует из рассмотрения первого случая). Рассмотрим теперь аналогичным образом все предикаты p_i. Если $p_i \in T_1$, оставим множество $\operatorname{Pol} p_i$ в пересечении, если p_i — тождественно истинный или ложный предикат, то вычеркнем множество $\operatorname{Pol} p_i$ (оно равно P_k) из пересечения (при этом, в силу сказанного выше и $p_i \in [R]$, у нас не останется одноместных предикатов), иначе проделаем описанную процедуру разложения с классом $\operatorname{Pol} p_i$ (опять, поскольку $A \neq \operatorname{Pol} R$, то $\operatorname{Pol} p_i \neq \operatorname{Pol} R$, p_i отличен от тождественно истинного или ложного предиката). Поскольку на каждом шаге местность рассматриваемых предикатов уменьшается на единицу, процесс завершится. В конце мы получим представление:

$$A = \bigcap_{i=1}^{r} \operatorname{Pol} p_i,$$

где все $p_i \in T_1$.

Лемма доказана.

Определение 12. Замкнутый класс A называется предикатно-описуемым, если существует предикат p, такой что справедливо представление $A = \mathrm{Pol}\, p$.

Через $A(T_1)$ обозначим множество различных классов $\mathrm{Pol}\, p$, где $p \in T_1$.

Лемма 13. Число замкнутых классов над M_R бесконечно тогда и только тогда, когда множество $A(T_1)$ содержит бесконечное число классов.

Доказательство Если множество $A(T_1)$ бесконечно, то достаточно учесть, что по построению любой $p \in T_1$ реализуется формулой над R, откуда по лемме 1 получаем, что $\mathrm{Pol}\, p \supseteq M_R$.

Пусть теперь над M_R существует бесконечное число классов.

Рассмотрим вначале случай, когда существует класс A, содержащий M_R такой, что A не является предикатно-описуемым. Рассмотрим множество $\mathrm{Inv}\, A$. Обозначим через H множество различных классов $B = \mathrm{Pol}\, p$, где $p \in \mathrm{Inv}\, A$.

Возьмем произвольный класс B из множества H. По определению H найдется предикат $p \in \mathrm{Inv}\, A$ такой, что $B = \mathrm{Pol}\, p$. По определению множества $\mathrm{Inv}\, A$ любая функция $f \in A$ сохраняет предикат p, следовательно $f \in B$. Получаем, что $A \subseteq B$.

Покажем теперь, что H — бесконечное множество. Предположим обратное, пусть H состоит из замкнутых классов $B_1, \ldots, B_N$. По определению H для каждого элемента B_i справедливо представление $B_i = \mathrm{Pol}\, p_i$, где $p_i \in \mathrm{Inv}\, A$, $i \in \{1, \ldots, N\}$.

Покажем, что $A = \bigcap\limits_{i=1}^{N} B_i$.

Выше было показано, что для любого $B \in H$ справедливо $A \subseteq B$, откуда $A \subseteq \bigcap\limits_{i=1}^{N} B_i$. Пусть теперь $f \in \bigcap\limits_{i=1}^{N} B_i$. Предположим, что $f \notin A$. Это означает, что в $\mathrm{Inv}\, A$ найдется предикат p' такой, что f не сохраняет p'. По построению множества H в нем должен содержатся класс, равный $\mathrm{Pol}\, p'$. Следовательно, указанный класс есть во множестве $\{B_1, \ldots, B_N\}$, но функция f принадлежит всем классам данного множества, а значит сохраняет предикат p'. Полученное противоречие доказывает, что $f \in A$. Итак, $\bigcap\limits_{i=1}^{N} B_i \subseteq A$.

Заметим, что в класс $\bigcap\limits_{i=1}^{N} B_i$ входят все функции, сохраняющие предикаты $p_1, \ldots, p_N$ и только они. Получаем, что $\bigcap\limits_{i=1}^{N} B_i = \mathrm{Pol} \mathop{\&}\limits_{i=1}^{N} p_i$, откуда

$$A = \mathrm{Pol}\left(\mathop{\&}\limits_{i=1}^{N} p_i\right).$$

Но мы предположили, что класс A не представим в виде $A = \mathrm{Pol}\, p$. Полученное противоречие показывает, что множество H содержит бесконечное число классов.

Итак, мы доказали, что в рассматриваемом случае существует бесконечное число различных классов, представимых в виде $\mathrm{Pol}\, p$, содержащих класс A, а, следовательно, и класс M_R.

Если же все классы A, содержащие M_R, являются предикатно-описуемыми, то мы опять получаем, что существует бесконечное число различных классов, представимых в виде $\mathrm{Pol}\, p$, содержащих класс M_R.

Предположим, что множество $A(T_1)$ содержит конечное число классов: $A_1, \ldots, A_N$, $A_i = \mathrm{Pol}\, p_i$, $p_i \in T_1$. Рассмотрим произвольный класс A такой, что $M_R \subset A \subset P_k$, $A = \mathrm{Pol}\, p$. По второму пункту леммы 7 получаем, что $p \in [R]$. Из леммы 12 следует, что замкнутый класс A представим пересечением классов $\mathrm{Pol}\, p_i$. Поскольку классов, представимых указанным пересечением конечное число, а всего классов вида $\mathrm{Pol}\, p$, содержащих M_R, бесконечно, приходим к противоречию, то есть множество $A(T_1)$ бесконечно.

Лемма доказана.

Докажем еще несколько свойств невырожденных предикатов, которые нам понадобятся в дальнейшем.

Лемма 14. Любой невырожденный предикат $p \in [R]$ такой, что $\mathrm{Pol}\, p \neq P_k$, можно задать формулой F из семейства $\overline{F}_1$.

Доказательство Пусть $p \in [R]$ — невырожденный предикат. Предположим, что формула $F_1 \notin \overline{F}_1$ задает p. Согласно лемме 5, мы можем построить эквивалентный предикат p', задаваемый формулой $F_2 \in \overline{F}_1$. При построении p' мы сначала используем лемму 4. При этом предикат меняется, если существует пара вершин, отвечающих свободным переменным и входящих в один ориентированный цикл графа G_{F_1}. После этого применяется лемма 5, в которой из графа формулы удаляются компоненты связности, в которых нет ни одной вершины, соответствующей свободной переменной. При этой процедуре предикат не меняется.

Итак, предикат может изменится только при отождествлении вершин ориентированного цикла, в который входят две вершины v_{x_1}, v_{x_2} графа G_{F_1}, соответствующие свободным переменным x_1, x_2 формулы F_1. Отметим, что в любом наборе $\tilde{a} = (a_1, \dots, a_n)$ таком, что $p(\tilde{a}) = TRUE$ будет выполнено $a_1 = a_2$ (мы считаем, что $x_i = a_i$).

Если p — двухместный предикат, то с учетом рассуждений выше получаем, что p — диагональ, откуда $\mathrm{Pol}\, p = P_k$, что противоречит условию леммы.

Пусть местность предиката p больше 2, $\tilde{a}$ набор такой, что $p(\tilde{a}) = FALSE$, $(\exists x_i p)(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) = TRUE$ для любого $i \in \{1, \dots, n\}$ (по определению невырожденного предиката). Если $a_1 = a_2$, то с учетом $p(\tilde{a}) = FALSE$ получаем, что для любого $b \neq a_2$ справедливо $p(b, a_2, \dots, a_n) = FALSE$, откуда $(\exists x_1 p)(a_2, \dots, a_n) = FALSE$. Приходим к противоречию. Если же $a_1 \neq a_2$, то $(\exists x_3 p)(a_1, a_2, a_4, \dots, a_n) = FALSE$ (тут мы учитываем, что местность предиката p больше 2). Получаем, что невырожденный предикат местности больше 2 не может иметь переменных, связанных отношением диагонали. Отсюда получаем, что при применении лемм 4 и 5 предикат p не изменится.

Лемма доказана.

Лемма 15. Если для невырожденного предиката p справедливо представление

Доказательство

$$p(x_1, \dots, x_n) = p_1(x_1^1, \dots, x_{i_1}^1) \& \dots \& p_l(x_1^s, \dots, x_{i_s}^s), \tag{2}$$

X_j — множество различных переменных из списка $x_1^j, \dots, x_{i_j}^j$, $X = \{x_1, \dots, x_n\}$. Тогда найдется номер $j \in \{1, \dots, s\}$ такой, что $X_j = X$ и предикат p_j является невырожденным.

Доказательство Введем предикаты $h_1, \dots, h_s$ такие, что $h_j(x_1, \dots, x_n) = p_j(x_1^j, \dots, x_{i_j}^j)$ (то есть мы дополняем p_j несущественными переменными, если необходимо). Равенство (2) принимает вид $p(x_1, \dots, x_n) = h_1(x_1, \dots, x_n) \& \dots \& h_s(x_1, \dots, x_n)$.

По условию леммы предикат p невырожден. Следовательно, если все предикаты $h_i(\tilde{x})$ вырождены, по лемме 10 мы приходим к противоречию. Значит найдется невырожденный предикат $h_m(\tilde{x}) = p_m(x_1^m, \dots, x_{i_m}^m)$. Если бы $X_m \neq X$, то предикат h_m оказался вырожденным (по описанным выше свойствам невырожденных предикатов).

Лемма доказана.

Далее будем рассматривать такие ЧУМ H, задающие монотонные классы, которые обладают единственным минимальным элементом m. Случай, когда указанные ЧУМ имеют единственный максимальный элемент, рассматривается аналогично ([5]).

Лемма 16. Пусть $p \in [R]$ задается формулой F. Тогда можно так преобразовать формулу F, чтобы в ЧУМ L_F соответствующие свободным переменным вершины и только они являлись минимумами, и при этом предикат p не изменится. Если F принадлежит семейству $\overline{F}_1$, то указанное преобразование можно провести так, чтобы преобразованная формула осталась в семействе $\overline{F}_1$.

Доказательство Рассмотрим ЧУМ L_F. Пусть связанные переменные $y_1, \dots, y_s$ формулы F соответствуют всем элементам множества L_F, меньше которых не существует элементов L_F, соответствующих свободным переменным формулы F. Удалим из формулы F все вхождения $R(\cdot, \cdot)$, в которых встречается переменная из множества $\{y_1, \dots, y_s\}$, а также все кванторы существования по переменным $\{y_1, \dots, y_s\}$. Полученную формулу обозначим F', а соответствующий ей предикат — r. Покажем, что $r = p$.

Если $p(\tilde{a}) = TRUE$, очевидно, что $r(\tilde{a}) = TRUE$. Обратно, пусть $r(\tilde{a}) = TRUE$. Придадим удаленным связанным переменным в F значение минимума m, остальным — те же значения, что и в F'. При этом получим, что $p(\tilde{a}) = TRUE$.

Теперь предположим, что $F \in \overline{F}_1$ и покажем, что $F' \in \overline{F}_1$.

При проведенной процедуре удаления минимальных вершин, в графе $G_{F'}$ формулы F' не могут появиться ориентированные циклы. Предположим, что в указанном графе возникнет компонента связности, в которой нет ни одной вершины, соответствующей свободной переменной. Но по определению ЧУМ $L_{F'}$ формулы F' два элемента $v_{y_1}, v_{y_2} \in L_{F'}$ сравнимы тогда и только тогда, когда в графе $G_{F'}$ найдется ориентированный путь между вершинами v_{y_1}, v_{y_2} (направление указанного пути для нас сейчас не важно). Мы для удобства обозначили одинаково два элемента множества $L_{F'}$ и две вершины графа $G_{F'}$, соответствующие переменным y_1, y_2 формулы F'. Из этого следует, что в ЧУМ $L_{F'}$ будет содержаться минимальный элемент v_y, соответствующий некоторой переменной y, которой в графе $G_{F'}$ отвечает вершина v_y из рассматриваемой компоненты связности. Однако по рассуждениям выше таких элементов в ЧУМ $L_{F'}$ быть не может.

Лемма доказана.

Докажем следующую основную лемму данного раздела.

Лемма 17. Пусть $p_1(x_1, \ldots, x_{n_1}), \ldots, p_l(x_1, \ldots, x_{n_l}) \in [R]$ — невырожденные предикаты местности соответственно $n_1, \ldots, n_l$, задаваемые формулами $F_1, \ldots, F_l \in \overline{F}_1$, $n = \max(n_1, \ldots, n_l)$, $\mathrm{Pol}\, p_i \neq \mathrm{Pol}\, R$. Тогда любой невырожденный предикат p' из множества $[p_1, \ldots, p_l]$ имеет местность $r \leqslant n$.

Доказательство Пусть невырожденный предикат p' реализуется некоторой формулой над $[p_1, \ldots, p_l]$:

$$p'(x_1, \ldots, x_r) = \exists y_1, \ldots, y_s R_1(u_1^1, \ldots, u_{s_1}^1) \& \ldots \& R_t(u_1^t, \ldots, u_{s_t}^t), \tag{3}$$

все предикаты $R_i \in \{p_1, \ldots, p_l\}$, все переменные $u_j^i \in \{x_1, \ldots, x_r, y_1, \ldots, y_s\}$.

Рассмотрим формулы $F_i \in \overline{F}_1$ из условия леммы, задающие предикаты $p_i \in [R]$ с ЧУМ L_{F_i}. Применим к формулам F_i лемму 16. Далее будем считать, что в каждом ЧУМ L_{F_i} все минимумы соответствуют свободным переменным формулы F_i, и все формулы $F_i \in \overline{F}_1$.

Обозначим через F' следующую формулу, реализующую предикат p':

$$F'(x_1, \ldots, x_r) = \exists y_1, \ldots, y_s F_{R_1}(u_1^1, \ldots, u_{s_1}^1) \& \ldots \& F_{R_t}(u_1^t, \ldots, u_{s_t}^t),$$

полученную заменой в выражении (3) предикатов R_j на формулы F_{R_j} следующим образом: если предикат R_j является p_i, то в выражении (3) мы заменяем предикат R_j на формулу F_i, реализующую p_i, переобозначая ее через F_{R_j}.

Построим новую формулу F'', реализующую предикат p', из формул $F_1, \ldots, F_l$.

Сначала мы делаем конъюнкцию t подходящих экземпляров формул $F_1 \ldots, F_l$ (эти формулы будут такими же, как и в последнем выражении для F'):

$$\tilde{F}_1(x_1, \ldots, x_{s_1 + \ldots + s_t}) = F_{R_1}(x_1, \ldots, x_{s_1}) \& \ldots \& F_{R_t}(x_{s_t - s_{t-1} + 1}, \ldots, x_{s_t}),$$

где все переменные x_j^i различны. При этом получаем формулу $\tilde{F}_1$ с графом $G_{\tilde{F}_1}$, состоящим из t не связанных между собой компонент, каждая из которых с точностью до пометок совпадает с некоторым графом из множества $\{G_{F_1}, \ldots, G_{F_l}\}$. Все минимальные элементы ЧУМ $L_{\tilde{F}_1}$ и только они соответствуют свободным переменным.

Далее проводим некоторое отождествление переменных в формуле $\tilde{F}_1$ (соответствующее выражению (3)). Полученную формулу обозначим $\tilde{F}_2$. При этом мы получим граф $G_{\tilde{F}_2}$ отождествлением некоторых вершин, соответствующих свободным переменным графа $G_{\tilde{F}_1}$.

Наконец, берем проекцию формулы $\tilde{F}_2$ по переменным $y_1, \ldots, y_s$, чтобы получить формулу $\tilde{F}_3$, реализующую предикат p'. Поскольку у формулы $\tilde{F}_2$ указанные переменные были свободными, то в ЧУМ $L_{\tilde{F}_3}$ элементы, соответствующие переменным $y_1, \ldots, y_s$, являются минимальными.

Пусть переменная y_1 получена на втором этапе некоторым отождествлением переменных $x_{i_1}, \ldots, x_{i_s}$. Выполним следующую процедуру: вместо указанного отождествления и последующей проекции по переменной y_1 мы независимо возьмем проекции по каждой переменной $x_{i_1}, \ldots, x_{i_s}$. Полученную формулу, реализующую некоторый предикат p'', обозначим через F''.

Если описывать проведенную операцию на содержательном уровне, то можно сказать, что мы разорвали связи, установленные на втором этапе построения формулы F'' между компонентами связности графа $G_{\tilde{F}_1}$, полученные отождествлением вершин, по которым на третьем этапе берется проекция. В результате мы получили граф $G_{F''}$, где указанные компоненты связности соединены только по вершинам, соответствующим свободным переменным формулы F''.

Итак, мы получили следующую формулу:

$$F''(x_1, \ldots, x_r) = \exists y_1^1, \ldots, y_s^t F_{R_1}(u_1^1, \ldots, u_{s_1}^1) \& \ldots \& F_{R_t}(u_1^t, \ldots, u_{s_t}^t), \tag{4}$$

где переменные $u_j^i \in \{x_1, \ldots, x_r, y_1^i, \ldots, y_s^i\}$, то есть каждая переменная y_j^i ($i \in \{1, \ldots, t\}$, $j \in \{1, \ldots, s\}$) входит только в один сомножитель $F_{R_i}(u_1^i, \ldots, u_{s_i}^i)$ конъюнкции в правой части (4). В указанной формуле мы разорвали все связи (то есть общие переменные) между разными сомножителями конъюнкции, осуществляемые по связанным переменным. То есть в данной формуле сомножители могут иметь только общие переменные, являющиеся свободными.

Покажем, что $p' = p''$. Это непосредственно следует из леммы 16. Мы можем удалить из формул F' и F'' все связанные переменные, соответствующие минимумам соответственно в $L_{F'}$ и $L_{F''}$. При этом мы получим равные формулы, задающие равные предикаты ($G_{F'}$ и $G_{F''}$ отличаются друг от друга лишь расклейкой минимальных вершин, соответствующих связанным переменным).

Поскольку $F_1, \ldots, F_l \in \overline{F}_1$, то $F'' \in \overline{F}_1$. Перепишем (4) следующим образом:

$$p''(x_1, \ldots, x_r) = \hat{p}_1(w_1^1, \ldots, w_{i_1}^1) \& \ldots \& \hat{p}_t(w_1^t, \ldots, w_{i_t}^t), \tag{5}$$

где все $w_j^i \in \{x_1, \ldots, x_n\}$, $\hat{p}_j$ — предикат, задаваемый формулой $\exists y_1^j, \ldots, y_s^j F_{R_j}(u_1^j, \ldots, u_{s_j}^j)$ в выражении (4), местность которого $i_j \leqslant s_j$, то есть для некоторого $h \in \{1, \ldots, l\}$ местность $\hat{p}_j$ не превосходит местности некоторого p_h или $i_j \leqslant n_h \leqslant n$. Еще раз отметим, что каждая свободная переменная y_j^i входит только в один сомножитель F в конъюнкции в правой части (4), поэтому и возможен переход к произведению предикатов.

По условию p' — невырожденный предикат. Значит, p'' также невырожден. Применим лемму 15 для предиката p'' и выражения (5). Получаем, что найдется невырожденный предикат $\hat{p}_j(w_1^j, \ldots, w_{i_j}^j)$, такой, что $\{x_1, \ldots, x_r\} = \{w_1^j, \ldots w_{i_j}^j\}$, откуда $r = i_j \leqslant n$.

Лемма доказана.

3 Классы монотонных функций, сохраняющих ЧП с одним минимальным элементом.

Продолжим рассмотрение ЧУМ H, задающих монотонный класс, имеющих единственный минимальный элемент. В данном разделе мы применим полученные ранее свойства невырожденных предикатов к изучению надструктуры класса M_H.

Нам в дальнейшем понадобится семейство предполных классов **C**.

Определение 13. Класс функций B принадлежит семейству **C**, если B — множество функций, сохраняющих центральный предикат p, т.е. предикат, обладающий следующими свойствами:

1. Абсолютная симметричность. Для любой подстановки σ на множестве $\{1, \ldots, n\}$ и любого набора $\tilde{a} \in E_k^n$ справедливо $p(a_1, \ldots, a_n) = TRUE \Longleftrightarrow p(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) = TRUE$.

2. Абсолютная рефлексивность. Предикат p истинен на любом наборе $\tilde{a}$, таком что найдутся различные номера $1 \leqslant i, j \leqslant m$ (m — местность p), удовлетворяющие $a_i = a_j$ (в $\tilde{a}$ присутствуют хотя бы две равные компоненты).

3. Существует центр предиката p, то есть существует элемент $h \in E_k$ такой, что для любого набора $\tilde{a} \in E_k^n$, у которого есть компонента, равная h, справедливо $p(\tilde{a}) = TRUE$.

На основе полученных выше результатов докажем следующую теорему, описывающую некоторые важные свойства надрешеток классов монотонных функций, сохраняющих ЧУМ H, обладающее единственным минимальным элементом.

Теорема 1. Пусть $M_H = \mathrm{Pol}\, R$ — монотонный класс, сохраняющий ЧУМ H с единственным минимальным элементом. Тогда справедливы следующие утверждения:

1. Для любого $p \in [R]$, $\mathrm{Pol}\, p \neq \mathrm{Pol}\, R$ число замкнутых классов, содержащих $\mathrm{Pol}\, p$, конечно.
2. Замкнутый класс M_H предполон в единственном классе $A = \bigcap\limits_{p \in T(R)} \mathrm{Pol}\, p$, где $T(R)$ — множество предикатов $p \in [R]$ таких, что $\mathrm{Pol}\, p \neq \mathrm{Pol}\, R$. Причем класс A предикатно-описуем тогда и только тогда, когда надрешетка M_H конечна.
3. Замкнутый класс M_H содержится в единственном предполном классе C семейства $\mathbf{C}$, представимом в виде $\mathrm{Pol}\, (\exists y R(x_1, y) \& R(x_2, y))$.

Доказательство Докажем вначале первый пункт. Пусть $p \in [R]$, $\mathrm{Pol}\, p \neq \mathrm{Pol}\, R$. Предположим, что $\mathrm{Pol}\, p \neq P_k$ (в ином случае утверждение очевидно). По лемме 12 класс $\mathrm{Pol}\, p$ представим конечным пересечением классов $\bigcap_i \mathrm{Pol}\, p_i$, где все предикаты $p_i \in T_1$. Обозначим через P систему предикатов, состоящую из всех предикатов p_i, участвующих в разложении $\mathrm{Pol}\, p$. Пользуясь леммой 14, будем считать, что все предикаты p_i задаются формулами из множества $\overline{F}_1$. Отметим, что p выражается формулой над системой P. Обозначим через m максимальную местность предикатов из P.

Рассмотрим невырожденный предикат $p' \in [p]$. Указанный предикат реализуется формулой над системой P. По лемме 17 местность предиката p' не превосходит m. Получаем, что число невырожденных предикатов p' из множества $[p]$ конечно. Предположим, что замкнутый класс A, содержащий $\mathrm{Pol}\, p$, представим в виде $\mathrm{Pol}\, p''$, где p'' — невырожденный предикат. По лемме 14 его можно задать формулой $F'' \in \overline{F}_1$, а по первому пункту леммы 7 мы получим, что $p'' \in [p]$. А таких предикатов, как мы уже установили, конечное число.

Итак, замкнутых классов вида $\mathrm{Pol}\, p'$, где p' — невырожденный предикат, содержащих $\mathrm{Pol}\, p$, конечное число.

Рассмотрим теперь произвольный предикатно-описуемый класс вида $\mathrm{Pol}\, p'$, содержащий $\mathrm{Pol}\, p$, где предикат p' вырожден, $\mathrm{Pol}\, p' \neq P_k$. По лемме 7 $p' \in [R]$. По лемме 12 класс $\mathrm{Pol}\, p'$ представим конечным пересечением классов $\bigcap_i \mathrm{Pol}\, p'_i$, где все предикаты $p'_i \in T_1$, то есть невырождены. Поскольку $\mathrm{Pol}\, p \subseteq \mathrm{Pol}\, p'$, то $\mathrm{Pol}\, p \subseteq \mathrm{Pol}\, p'_i$. Но выше мы показали, что предикатно-описуемых классов, содержащих $\mathrm{Pol}\, p$ и задаваемых невырожденным предикатом конечное число. Следовательно, их пересечений также будет конечное число.

Итак, число предикатно-описуемых классов, содержащих $\mathrm{Pol}\, p$, конечно.

Если в решетке над $\mathrm{Pol}\, p$ существует замкнутый класс A, не являющийся предикатно-описуемым, то мы получим бесконечное число различных классов вида $\mathrm{Pol}\, p'$, где $p' \in \mathrm{Inv}\, A$ (см. доказательство леммы 13), содержащих $\mathrm{Pol}\, p$. Это противоречит доказанному выше.

Докажем теперь второй пункт.

Рассмотрим класс A из условия теоремы: $A = \bigcap\limits_{p \in T(R)} \mathrm{Pol}\, p$, где $T(R) = \{p \in [R],\ \ \mathrm{Pol}\, p \neq \mathrm{Pol}\, R\}$. Заметим, что $M_H \subseteq A$. В самом деле, из леммы 1 следует, что любая функция $f \in M_H$ сохраняет все предикаты $p \in [R]$, откуда получаем, что $f \in \bigcap\limits_{p \in [R]} \mathrm{Pol}\, p \subseteq \bigcap\limits_{p \in T(R)} \mathrm{Pol}\, p = A$.

Рассмотрим произвольный класс B такой, что $M_H \subseteq B$, $M_H \neq B$. Из [7] получаем, что $\mathrm{Inv}\, B \subseteq \mathrm{Inv}\, M_H$. Из очевидного равенства $d(x_1, x_2) = R(x_1, x_2) \& R(x_2, x_1)$ следует $D \subseteq [R]$. С учетом 1 имеем $\mathrm{Inv}\, M_H = \mathrm{Inv}\, \mathrm{Pol}\, R = \mathrm{Inv}\, \mathrm{Pol}\, [R] = [R]$. Объединяя полученные результаты, получаем

$$\mathrm{Inv}\, B \subseteq [R].$$

Предположим, что во множестве $\mathrm{Inv}\, B$ есть предикат p такой, что $\mathrm{Pol}\, p = \mathrm{Pol}\, R$. По уже доказанному $p \in [R]$, откуда по первому пункту леммы 7 (полагая $p_2 = R$, $p_1 = p$) получаем, что $R \in [p]$, откуда $[R] \subseteq [p] \subseteq \mathrm{Inv}\, B$, Получаем, что все функции класса B сохраняют множество предикатов $[R]$. Откуда имеем, что $B \subseteq M_H$. Из данного противоречия получаем, что во множестве $\mathrm{Inv}\, B$ нет предиката p такого, что $\mathrm{Pol}\, p = M_H$. Откуда $\mathrm{Inv}\, B \subseteq T(R) = \mathrm{Inv}\, A$. Из [7] получаем, что $A \subseteq B$.

Предположим, что надструктура класса монотонных функций M_H бесконечна. По лемме 13 множество $A(T_1)$ будет содержать бесконечное число классов. По доказанному выше каждый из классов множества $A(T_1)$ содержит класс A. Предположим, что класс A предикатно-описуем. Пусть $A = \mathrm{Pol}\, p$. По второму пункту леммы 7 $p \in [R]$. Из уже доказанного первого пункта данной теоремы следует, что число различных замкнутых классов, содержащих $\mathrm{Pol}\, p = A$, конечно. Полученное противоречие показывает, что класс A не является предикатно-описуемым.

С другой стороны, если класс A не является предикатно-описуемым, то рассмотрим множество различных классов $\mathrm{Pol}\, p$, где $p \in \mathrm{Inv}\, A$. Аналогично доказательству леммы 13 можно доказать, что указанные классы образуют бесконечню надструктуру класса A, а значит, и M_H.

Итак, A является предикатно-описуемым классом тогда и только тогда, когда надструктура класса M_H конечна.

Покажем теперь, что $M_H \neq A$.

Предположим, что $M_H = A$. Поскольку $M_H = \mathrm{Pol}\, R$, то в рассматриваемом случае класс A является предикатно-описуемым, откуда по уже доказанному получаем, что число замкнутых классов, строго содержащих M_H конечно. Пусть $A_1, \ldots, A_N$ — все указанные классы. Все они предикатно-описуемы (иначе бы мы опять пришли к бесконечной надструктуре). Пусть $A_i = \mathrm{Pol}\, p_i$, $\mathrm{Pol}\, p_i \neq \mathrm{Pol}\, R$, $i \in \{1, \ldots, N\}$. Тогда по построению класса A получаем, что

$$A = \bigcap_{i=1}^{N} \mathrm{Pol}\, p_i = \mathrm{Pol}\,(p_1 \& \ldots \& p_N),$$

где предикаты в конъюнкции справа не имеют общих переменных. Обозначив $p' = p_1 \& \ldots \& p_N$, получаем $A = \mathrm{Pol}\, p'$. По лемме 7 получаем, что $R \in [p']$ (в первом пункте леммы 7 полагаем $p_2 = R$, $p_1 = p'$). Однако, по лемме 8, получаем, что для любого $i \in \{1, \ldots, N\}$, для любой формулы $F_i \in \overline{F_1}$, задающей предикат p_i, в ЧУМ L_F не будет пары сравнимых элементов, соответствующих свободным переменным F_i.

Предположим, что найдется формула F', задающая предикат p' такая, что в ЧУМ $L_{F'}$ найдется пара сравнимых элементов $v_{x_1} \leqslant v_{x_2}$, соответствующих свободным переменным x_1, x_2 формулы F'. Пусть эти переменные относятся к разным предикатам p_i, p_j из конъюнкции, задающей p'. Рассмотрим два элемента $a, b \in E_k$ таких, что не выполнено $a \leqslant b$. По лемме 3 выполнено $p_j(b, \ldots, b) = TRUE$, $p_l(a, \ldots, a) = TRUE$ для всех $l \in \{1, \ldots, N\} \setminus \{j\}$. Следовательно, поскольку предикаты в конъюнкции, задающей p', не имеют общих переменных, мы подобрали набор, на котором предикат p' будет истинен. Однако при этом по построению указанного набора $x_1 = a$, $x_2 = b$. Получаем, что предикат, задаваемый формулой F' ложен на рассматриваемом наборе. Полученное противоречие показывает, что переменные x_1, x_2 относятся к одному сомножителю конъюнкции p_j. Взяв проекции по всем свободным переменным формулы F', соответствующим предикатам $p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_N$, мы получим формулу, задающую предикат p_j, в ЧУМ которой есть сравнимые элементы v_{x_1}, v_{x_2}, соответствующие переменным предиката p_j. Опять приходим к противоречию.

Итак, для любой формулы F', задающей предикат p', в ЧУМ $L_{F'}$ отсутствуют сравнимые элементы, соответствующие свободным переменным F'. По лемме 8 получаем, что $\mathrm{Pol}\, p' \neq \mathrm{Pol}\, R$, или $M_H \neq A$. Полученной противоречие завершает доказательство того, что $M_H \neq A$.

Итак, возьмем произвольную функцию $f \notin M_H$, и рассмотрим замкнутый класс $F = [M_H \bigcup \{f\}]$. Очевидно, что $M_H \subset F$, откуда по уже доказанному $A \subseteq F$. Получаем, что класс монотонных функций M_H предполон в A. Пусть M_H предполон еще в каком-то классе $G \neq A$, но тогда $A \subset G$. Приходим к противоречию.

Перейдем к третьему пункту.

Введем следующее обозначение: $p_{max}(x_1, x_2) = \exists y R(x_1, y) \& R(x_2, y)$. Обозначим $C = \mathrm{Pol}\, p_{max}$.

Покажем, что C — предполный класс из семейства $\mathbf{C}$.

Симметричность предиката p_{max} очевидна.

Для любого $a \in E_k$ выполнено $p_{max}(a, a) = TRUE$, так как в формуле выше, задающей предикат p_{max}, мы всегда можем положить $y = a$. Получаем, что p_{max} — абсолютно рефлексивный предикат. Напомним, что ЧУМ H обладает единственным минимальным элементом, который мы обозначили m. Из того факта, что $R(m,a) = TRUE$ для любого $a \in E_k$, следует, что для любого $b \in E_k$ справедливо $p_{max}(b,m) = \exists y R(b,y) \& R(m,y) = \exists y R(b,y)$. Принимая в последнем выражении $y = b$, получаем, что для любого $b \in E_k$ выполнено $p_{max}(b,m) = TRUE$. Из симметричности p_{max} получаем, что $p_{max}(m,d) = TRUE$ для любого $d \in E_k$. Итак, центру p_{max} принадлежит, по крайней мере, элемент $m \in E_k$.

Итак, замкнутый класс C является предполным из семейства $\mathbf{C}$.

Покажем теперь, что для любого замкнутого класса B такого, что $M_H \subseteq B \subset P_k$ выполнено $B \subseteq C$.

Предположим вначале, что класс B является предикатно-описуемым. Пусть $B = \mathrm{Pol}\, p$. По лемме 7 $p \in [R]$. Обозначим через F формулу, реализующую p над $\{R\}$. Согласно лемме 5, мы можем считать, что $F \in \overline{F}_1$. Применим теперь к формуле F лемму 16. Далее будем считать, что в ЧУМ L_F формулы F все минимальные элементы и только они соответствуют свободным переменным F. Пусть G_F — граф формулы F.

Покажем, что в G_F найдется компонента связности с по крайней мере двумя вершинами, соответствующими свободным переменным F. Предположим противное. Пусть $G_1, \ldots, G_n$ — все компоненты связности графа G_F. В каждом графе G_i только одна вершина v_{x_i}, отвечающая свободной переменной x_i формулы F $(i \in \{1, \ldots, n\})$. Каждый граф G_i очевидным образом задает свой одноместный предикат $p_i(x_i)$. По графу G_i мы можем построить формулу F_i, которая реализует p_i над $\{R\}$. Получаем, что $p_i \in [R]$ для всех $i \in \{1, \ldots, n\}$. По лемме 3 все предикаты $p_1, \ldots, p_n$ являются тождественно истинными. По построению рассматриваемых предикатов будет справедлива формула $p = p_1 \& \ldots \& p_n$. Получаем, что p — тождественно истинный предикат, откуда $B = P_k$. Полученное противоречие доказывает, что в предикате G найдется компонента связности G_i, в которой есть по крайней мере две вершины, соответствующие свободным переменным формулы F.

Пусть v_{x_1} — некоторый элемент ЧУМ L_{G_i} графа G_i, соответствующий свободной переменной x_1 формулы F. Рассмотрим всевозможные максимумы $v_{y_1}, \ldots, v_{y_s}$ ЧУМ L_{G_i} такие, что $v_{x_1} \leqslant v_{y_i}$ для всех $i \in \{1, \ldots, s\}$. Пусть элементы $v_{x_1}, v_{x_2}, \ldots, v_{x_r} \in L_{G_i}$ — все минимальные элементы, каждый из которых меньше некоторого v_{y_j} в L_{G_i}. Если $r = 1$, то есть кроме v_{x_1} не найдется ни одного минимума, меньшего некорого v_{y_j}, $j \in \{1, \ldots, s\}$, то, принимая во внимание тот факт, что все минимальные элементы L_{G_i} соответствуют свободным переменным F, получим, что в рассматриваемой компоненте связности G_i только одна вершина, соответствующая свободной переменной F. Полученное противоречие показывает, что найдется элемент $v_{y_j} \in L_{G_i}$ такой, что $v_{y_j} \geqslant v_{x_1}, v_{x_2}$, где x_1, x_2 — некоторые различные свободные переменные F.

Возьмем в предикате p проекции по всем переменным, кроме указанных, и обозначим полученный предикат $p'(x_1, x_2)$. Покажем, что $p' = p_{max}$.

Пусть элементы $a, b \in E_k$ такие, что $p'(a,b) = TRUE$. Пусть при этом переменная y_j формулы F приняла значение c. Присвоив единственной связанной переменной y в формуле, задающей предикат p_{max}, значение c, получим $p_{max}(a,b) = TRUE$.

Пусть теперь $p_{max}(a,b) = TRUE$, пусть при этом единственная связанная переменная y в формуле, задающей p_{max}, приняла значение $c \in E_k$.

Пусть F' — формула, полученная из F выбрасыванием всех кванторов существования. Рассмотрим произвольный элемент v_z ЧУМ $L_{F'}$, где z — произвольная переменная формулы F'. Осуществим следующее присвоение: если $z = x_1$, положим $z = a$, если $z = x_2$, положим $z = b$, в ином случае положим $z = c$. Поскольку элементы v_{x_1}, v_{x_2} являются минимальными в L_F, а значит в $L_{F'}$, и $a \leqslant c$, $b \leqslant c$, то указанное присвоение непротиворечиво. Обозначим через p'' предикат, реализуемый формулой F'. Отметим, что предикат p', получается из p'' проекциями по всем переменным, кроме x_1, x_2. Мы построили набор, на котором предикат p'' истинен, при этом $x_1 = a$, $x_2 = b$. Получаем, что $p'(a,b) = TRUE$.

Итак, $p' = p_{max}$. Получаем, что $p_{max} \in [p]$. По лемме 1 получаем, что $B \subseteq C$.

Если бы существовал предполный класс D отличный от C, такой что $M_H \subset D$, то по доказанному мы бы получили $D \subseteq C$. Полученное противоречие доказывает единственность предполного класса C.

Теорема доказана.

Из доказанной теоремы следует, что в случае бесконечной надструктуры класса M_H (напомним, H имеет единственный минимальный элемент), все предикатно-описуемые классы находятся вверху надрешетки M_H, а не являющиеся предикатно-описуемыми — под ними. Причем надрешетка M_H, если не рассматривать сам класс M_H и все P_k, всегда имеет единственные минимальный и максимальный элементы — соответственно классы A и C.

4 Необходимые и достаточные условия наличия бесконечной надструктуры у класса M_H

Теперь докажем теорему, описывающую некоторое семейство классов монотонных функций, надструктура которых заведомо конечна.

Через $L_{i,j}$ будем обозначать множество ЧУМ, имеющих i максимальных элементов и j минимальных.

Рассмотрим произвольное ЧУМ H из элементов множества E_k, принадлежащее множеству $L_{s,1}$. Пусть $h_1, \ldots, h_s$ — все максимальные элементы множества H.

Пусть W_i — некоторое непустое подмножество множества $W = \{1, \ldots, s\}$. Обозначим через H_{W_i} частично упорядоченное множество, состоящее из элементов $d \in H$ таких, что $d \leqslant h_j$ для любого $j \in W_i$ и $d \nleqslant h_j$, если $j \notin W_i$. Для любых двух элементов $b_1, b_2 \in H_{W_i}$ выполнено $b_1 \leqslant b_2$ тогда и только тогда, когда $b_1 \leqslant b_2$ в H.

Определение 14. Множество $H \in L_{s,1}$ назовем простым, если выполнены следующие условия:

1. Для любого непустого подмножества W_i множества W множество H_{W_i} непусто.

2. Каждое множество H_{W_i} имеет единственный максимальный элемент M_{W_i}.

3. Если $W_i \subseteq W_j$, то $M_{W_i} \geqslant M_{W_j}$ для любых подмножеств W_i, W_j множества W.

Определение 15. Множество элементов $\{a_1, \ldots, a_r\}$, не имеющих общего максимума в H, будем называть тупиковым, если при вычеркивании из него любого элемента оставшиеся обладают общим максимумом в H.

Теорема 2.

1. Любой класс монотонных функций, сохраняющих простое ЧУМ, имеет конечную надструктуру.

2. Местность произвольного невырожденного предиката $p \in [R_H]$, где H — простое множество, не превосходит максимальной мощности тупикового подмножества элементов из H, не имеющих общего максимума.

Доказательство Пусть $H \in L_{s,1}$ — простое ЧУМ. Рассмотрим монотонный класс $M_H = \operatorname{Pol} R_H$. Возьмем произвольный предикат $p \in [R_H]$ такой, что $\operatorname{Pol} p \neq M_H$ и $\operatorname{Pol} p \neq P_k$. Следовательно, предикат p не может быть тождественно истинным. Из леммы 3 следует, что местность p больше единицы. Существует набор $\tilde{a}$ такой, что $p(\tilde{a}) = FALSE$.

Рассмотрим некоторую формулу F, реализующую предикат p над $\{R_H\}$. Обозначим через L_F ЧУМ формулы F. Согласно лемме 16 можно считать, что все минимальные элементы L_F и только они соответствуют свободным переменным формулы F. Обозначим указанные элементы через $v_{x_1}, \ldots, v_{x_n}$.

Далее присвоим всем переменным формулы F значения так, чтобы для любых двух переменных z_1, z_2, принявших значения соответственно b_1, b_2, и таких, что $v_{z_1} \leqslant v_{z_2}$ в L_F, выполнялось $b_1 \leqslant b_2$ в H.

Всем свободным переменным x_i присвоим значения a_i, которые они принимают в $p(\tilde{a})$. Рассмотрим теперь произвольный элемент $v_y \in L_F$, соответствующий связанной переменной y. Пусть $x_{i_1}, \ldots, x_{i_r}$ — все свободные переменные формулы F такие, что $v_{x_{i_j}} < v_y$ (то есть, иными словами, переменной y требуется присвоить значение, превосходящее значения свободных переменных $x_{i_1}, \ldots, x_{i_r}$). Каждый компонент a_{i_j} набора $\tilde{a}$ принадлежит какому-то одному множеству $H_{W_{i_j}}$ (это множество определяется тем, с какими максимальными элементами ЧУМ H сравним элемент a_{i_j}).

Рассмотрим множества $W_{i_1}, \ldots, W_{i_r}$. Если их пересечение $W'_y = \bigcap_{j=1}^{r} W_{i_j}$ непусто, то присвоим переменной y значение $M_{W'_y}$ (по определению простого множества уканный максимальный элемент множества $H_{W'_y}$ найдется для любого W'_y). Отметим, что W'_y — множество номеров максимальных элементов ЧУМ H, которые больше каждого из элементов множества $\{a_{i_1}, \ldots, a_{i_r}\}$.

Предположим сначала, что для всех переменных y пересечение W'_y оказалось непустым. В этом случае присвоим значения $M_{W'_y}$ всем связанным переменным формулы F.

Рассмотрим две произвольные связанные переменные y_1, y_2 такие, что $v_{y_1} \leqslant v_{y_2}$ в ЧУМ L_F. Пусть указанные две переменные приняли значения соответственно b_1 и b_2. Пусть $x_{i_1}, \ldots, x_{i_r}$ — все свободные переменные, элементы L_F соответствующие которым меньше элемента v_{y_2}. Если мы рассмотрим аналогичное множество свободных переменных для v_{y_1}, то оно будет вложено в множество $\{x_{i_1}, \ldots, x_{i_r}\}$. Отсюда получаем, что $W'_{y_2} \subseteq W'_{y_1}$. По определению простого множества получим, что $M_{W'_{y_2}} \geqslant M_{W'_{y_1}}$, то есть для значений переменных выполнено $d_1 \leqslant d_2$.

Возьмем теперь произвольную свободную переменную x_i и связанную переменную y такую, что $v_{x_i} < v_y$. Пусть в результате описанного выше присвоения переменная y приняла значение d. Это значение является единственным максимальным элементом некоторого множества $H_{W'_y}$, где W'_y — подмножество множества W. Пусть элемент $a_i \in H$ (это значение переменной x_i) принадлежит некоторому множеству H_{W_i}, где W_i — подмножество W. По построению множество W'_y было получено пересечением, в которое входит множество W_i. Отсюда получаем, что $W'_y \subseteq W_i$. Это означает, что $M_{W'_y} \geqslant M_{W_i}$, то есть $d \geqslant a_i$.

Итак, мы присвоили всем переменным формулы F значения так, что $p(\tilde{a}) = TRUE$ (поскольку мы присвоили свободным переменным подходящие компоненты набора $\tilde{a}$ и показали, что существуют подходящие значения для связанных переменных) при условии, что для каждой связанной переменной формулы F подмножество W'_y непусто.

Поскольку мы предположили $p(\tilde{a}) = FALSE$, получаем, что обязательно должна найтись связанная переменная y такая, что $W'_y = \emptyset$. Но по определению W'_y из этого следует существование элементов $a_{i_1}, \ldots, a_{i_r}$, которые не имеют общего максимума в ЧУМ H, а переменной y требуется присвоить значение, которое является указанным максимумом. Из множества $a_{i_1}, \ldots, a_{i_r}$ можно выбрать некоторое тупиковое подмножество элементов $a_{j_1}, \ldots, a_{j_q}$, не имеющих максимума в H. Заметим, что, поскольку все элементы $a_{j_1}, \ldots, a_{j_q}$ должны быть различны (иначе это будет не тупиковое множество), то мощность q множества $\{a_{j_1}, \ldots, a_{j_q}\}$ не превосходит числа элементов ЧУМ H, то есть k.

Получаем, что при проекции предиката p по всем переменным, кроме $x_{j_1}, \ldots, x_{j_q}$, получим новый предикат p', ложный на наборе $(a_{j_1}, \ldots, a_{j_q})$, поскольку в формуле F', полученной указанной проекцией из F, также будет требоваться элемент, являющийся максимумом $a_{j_1}, \ldots, a_{j_q}$. В силу произвольности набора $\tilde{a}$ отсюда следует, что во множестве $[R_H]$ не может быть ни одного невырожденного предиката местности, превосходящей мощность рассматриваемого тупикового множества, то есть k (для невырожденного предиката p требуется, чтобы для его проекции по любой переменной выполнялось $\exists y p(a_1, \ldots, a_{i-1}, y, a_{i+1}, \ldots, a_n) = TRUE$ для некоторого набора $\tilde{a}$ такого, что $p(\tilde{a}) = FALSE$). Этим мы доказали второй пункт.

Поскольку число k фиксировано, то число невырожденных предикатов $p \in [R]$ конечно. Но тогда и конечно число классов $A(T_1)$ (см. определение 11).

По лемме 13 получаем, что число классов, содержащих M_H, конечно.

Теорема доказана.

Далее опишем некоторые семейства классов монотонных функций, для которых удается установить необходимое и достаточное условие наличия бесконечной надструктуры.

Вначале рассмотрим классы монотонных функций, наиболее близкие к предполным, — сохраняющие ЧУМ с двумя максимальными элементами и единственным минимальным. Такое семейство классов уже обладает достаточно интересными свойствами.

Через $\overline{H}_1$ будем обозначать четырехэлементное ЧУМ, множество элементов которого $\{0, 1, 2, 3\}$. В $\overline{H}_1$ справедливы следующие отношения: $0 \geqslant 1$, $0 \geqslant 2$, $3 \geqslant 1$, $3 \geqslant 2$, элементы $0, 3$ несравнимы, элементы $1, 2$ также несравнимы.

Определение 16. Для произвольного k обозначим через Q^1_k множество ЧУМ $H \in L_{21}$, составленных из элементов E_k таких, что ЧУМ H содержит подмножество $\overline{H}_1$, причем во множестве H пары элементов $0,3$ и $1,2$ остаются несравнимыми, не появляется элемента a такого, что $a \leqslant 0,3$ и $a \geqslant 1,2$, и элементы $0,3$ являются максимальными в H.

Далее для простоты будем обозначать через $0,3,1,2$ элементы как множества $\overline{H}_1$, так и множества H (имея в виду те элементы, которые получились при вложении множества $\overline{H}_1$ из одноименных элементов указанного множества). Из контекста будет понятно, о каких именно элементах идет речь.

Введем следующее обозначение: $p_{max}(x_1,x_2) = \exists y R(x_1,y) \& R(x_2,y)$.

Теорема 3. Для того, чтобы в P_k класс монотонных функций M_H, где $H \in L_{2,1}$, обладал бесконечной надструктурой, необходимо и достаточно, чтобы ЧУМ $H \in Q_k^1$. В случае конечной надструктуры класс M_H содержится только в классах $\mathrm{Pol}\, p_{max} \in \mathbf{C}$ и P_k.

Доказательство Если $H \in Q_k^1$, то есть подмножество $\overline{H}_1$ присутствует в множестве H и выполнены условия определения 16, то из [1] получаем, что надструктура класса M_H бесконечна.

Пусть теперь $H \notin Q_k^1$.

Обозначим через h_1, h_2 максимальные элементы множества H. Пусть $H_{\{1,2\}}$ — подмножество множества H, состоящее из элементов b, таких что $b \leqslant h_1$ и $b \leqslant h_2$. Будем считать, что для любых двух элементов b_1, b_2 в $H_{\{1,2\}}$ выполнено $b_1 \leqslant b_2$ тогда и только тогда, когда $b_1 \leqslant b_2$ в H.

Предположим, что множество $H_{\{1,2\}}$ имеет более двух максимальных элементов. Пусть c_1, c_2 — два различных максимальных элемента $H_{\{1,2\}}$. Рассмотрим четверку h_1, h_2, c_1, c_2. Поскольку $c_1, c_2 \in H_{\{1,2\}}$, то $c_1, c_2 \leqslant h_1$ и $c_1, c_2 \leqslant h_2$. По определению элементы c_1, c_2 несравнимы, элементы h_1, h_2 также несравнимы. Предположим, что в ЧУМ H найдется элемент a такой, что $a \leqslant h_1, h_2$ и $a \geqslant c_1, c_2$. Но тогда $a \in H_{\{1,2\}}$, и получаем противоречие с тем, что c_1, c_2 — максимальные элементы множества $H_{\{1,2\}}$. Итак, рассматриваемая четверка образует множество $\overline{H}_1$ (с точностью до пометок), и $H \in Q_k^1$. Полученное противоречие показывает, что множество $H_{\{1,2\}}$ имеет единственный максимальный элемент. Обозначим его через $M_{\{1,2\}}$.

Заметим, что множество $H_{\{1,2\}}$ непусто, ему по крайней мере принадлежит единственный минимальный элемент ЧУМ H. Любой элемент $b \in H \setminus H_{\{1,2\}}$ либо меньше h_1 — множество таких элементов обозначим через $H_{\{1\}}$, либо меньше h_2 — соответственно множество этих элементов обозначим через $H_{\{2\}}$ (опять будем считать, что в $H_{\{1\}}$ и $H_{\{2\}}$ сохранились все отношения между элементами, которые были в H).

Покажем теперь, что множество H — простое. Поскольку H имеет два максимальных элемента, то множество W имеет вид $\{1,2\}$. Все его непустые подмножества суть $W_1 = \{1\}$, $W_2 = \{2\}$, $W_3 = \{1,2\}$. Каждое из множеств H_{W_i} непусто ($H_{\{1,2\}}$ содержит минимальный элемент ЧУМ H, а $H_{\{1\}}, H_{\{2\}}$ содержат по крайней мере максимальные элементы множества H), имеет единственный максимальный элемент (для $H_{\{1,2\}}$ это $M_{\{1,2\}}$, а для $H_{\{1\}}$, $H_{\{2\}}$ — соответственно $M_{\{1\}} = h_1$ и $M_{\{2\}} = h_2$). В рассматриваемом случае $W_1, W_2 \subseteq W_3$. Но, как уже было показано, $M_{\{1,2\}} \leqslant M_{\{1\}}$, $M_{\{1,2\}} \leqslant M_{\{2\}}$. По определению получаем, что множество H — простое. Из теоремы 2 следует, что класс M_H обладает конечной надструктурой.

В теореме 2 было доказано, что местность любого невырожденного предиката $p \in [R_H]$ не превосходит максимальную мощность множества $R_s = \{a_1, \ldots, a_s\}$ элементов $a_i \in H$ такого, что $a_1, \ldots, a_s$ не имеют общего максимума в H, но при вычеркивании любого элемента из R_s оставшееся множество имеет общий максимум в H.

Заметим, что, если множество R_s имеет общий максимум, то оно имеет общий максимум, являющийся максимальным элементом ЧУМ H.

Изучим теперь вопрос, каким может быть число s в рассматриваемом случае.

Предположим, что множество R_s содержит элементы множества $H_{\{1,2\}}$. Выбросим их из R_s, обозначим полученное множество через R'_s. По нашему предположению все элементы множества R'_s должны иметь общий максимум в H. В силу сказанного выше этим общим максимумом в частности является элемент h_1 или h_2. Но, поскольку все элементы множества $H_{\{1,2\}}$ меньше обоих элементов h_1, h_2, то все элементы множества R_s имеют тот же общий максимум, что и элементы множества R'_s. Полученное противоречие показывает, что множество R_s не содержит элементов множества $H_{\{1,2\}}$.

Предположим теперь, что R_s содержит более одного элемента из множества $H_{\{1\}}$ ($H_{\{2\}}$). Удалим указанные элементы из множества R_s, оставив единственный элемент из множества $H_{\{1\}}$ ($H_{\{2\}}$). Снова обозначим полученное множество через R'_s. По нашему предположению все элементы множества R'_s должны иметь общий максимум в H. Как показано выше, в этом случае мы можем взять максимум из числа максимальных элементов множества H, то есть из множества $\{h_1, h_2\}$. Но множество R'_s содержит элемент из $H_{\{1\}}$ ($H_{\{2\}}$), несравнимый с h_2 (h_1), поэтому все элементы множества R'_s имеют общий максимум h_1 (соответственно, h_2). Но тогда указанный максимум имеют и все элементы множества R_s. Полученное противоречие показывает, что множество R_s состоит из двух элементов b_1, b_2 таких, что $b_1 \in H_{\{1\}}$, $b_2 \in H_{\{2\}}$.

Получаем, что произвольный невырожденный предикат $p \in [R_H]$ имеет местность 2 (по лемме 3 одноместные предикаты $p \in [R_H]$ всегда вырождены), и в его формуле F есть переменная y, которая должна принять значение, являющееся максимумом значений, которые принимают переменные предиката. По лемме 16 будем считать, что в ЧУМ L_F все минимальные элементы и только они соответствуют свободным переменным F.

Покажем, что $p(x_1, x_2) = p_{max}(x_1, x_2) = \exists y R_H(x_1, y) \& R_H(x_2, y)$.

Если $p(a_1, a_2) = TRUE$, то элементы a_1, a_2 имеют максимум b в ЧУМ H, следовательно, $p_{max}(a_1, a_2) = TRUE$. Если же $p_{max}(a_1, a_2) = TRUE$, то опять получаем, что элементы a_1, a_2 имеют максимум b в ЧУМ H. Придадим всем связанным переменным формулы F значение b. Поскольку все элементы L_F, соответствующие связанным переменным, больше элементов, соответствующих свободным, то получим $p(a_1, a_2) = TRUE$.

Итак, во множестве $[R_H]$ есть только один невырожденный предикат p_{max}.

Рассмотрим произвольный класс $A \supset M_H$, $A \neq P_k$. Поскольку, как мы уже доказали, надструктура M_H конечна, то класс A предикатно-описуем. Пусть $A = \mathrm{Pol}\, p$. По лемме 12 класс A представим пересечением классов вида $\mathrm{Pol}\, p_i$, где p_i — предикаты из семейства T_1, то есть невырожденные предикаты. Отсюда получаем, что $A = \mathrm{Pol}\, p_{max}$.

Теорема доказана.

Рассмотрим вопрос о неограниченности конечной надструктуры монотонных классов.

Теорема 4. Для любого числа $n > 0$ существует k-значная логика P_k с монотонным классом M_H таким, что число различных классов B, удовлетворяющих $M_H \subset B \subset P_k$, конечно, но превосходит число n.

Доказательство Зафиксируем число $n > 0$. Возьмем $s = n+1$. Пусть H_s — простое ЧУМ, принадлежащее семейству $L_{s,1}$, такое, что для любого непустого подмножества W_i множества $W = \{0, 1, \dots, s\}$ соответствующее множество H_{W_i} состоит из единственного элемента M_{W_i}.

Получаем, что множество H_s содержит s максимальных элементов, $\binom{s}{2}$ элемента M_{W_i}, соответствующих двухэлементным множествам W_i, $\binom{s}{3}$ элемента M_{W_i}, соответствующих трехэлементным множествам W_i и так далее. Единственным минимальным элементом H_s является M_W, то есть элемент, меньший всех максимумов $h_1, \dots, h_s$. Итак, число элементов в H_s равно $s + \binom{s}{2} + \binom{s}{3} + \ldots + \binom{s}{s-1} + \binom{s}{s} = 2^s - 1$. ЧУМ H_s представляет собой s-мерный булев куб с удаленным максимальным элементом.

Рассмотрим k-значную логику, где $k = 2^s - 1 = 2^{n+1} - 1$, и класс монотонных функций M_{H_s}, сохраняющих ЧУМ H_s. Поскольку ЧУМ H_s — простое, то по теореме 2 получаем, что надструктура M_{H_s} конечна. Как и раньше, через R будем обозначать двухместный предикат, задающий класс M_{H_s}.

Рассмотрим предикат $p^l_{max}(x_1, \dots, x_l)$ для $l > 1$. Положим $p^l_{max}(a_1, \dots, a_l) = TRUE$ тогда и только тогда, когда элементы $a_1, \dots, a_l$ имеют общий максимум в H.

Указанные предикаты можно реализовать формулой над $\{R\}$ следующим образом:

$$p^l_{max}(x_1, \dots, x_l) = \exists y R(x_1, y) \& R(x_2, y) \& \ldots \& R(x_l, y). \tag{6}$$

Получаем, что $p^l_{max}(\tilde{x}) \in [R]$ для любого $l > 1$.

Рассмотрим предикат $p'(x_1, \dots, x_{l-1}) = \exists y p^l_{max}(x_1, \dots, x_{l-1}, y)$, где $l > 2$. Покажем, что $p' = p^{l-1}_{max}$.

Пусть для некоторого набора $\tilde{b} = (b_1, \dots, b_{l-1})$ справедливо $p^{l-1}_{max}(b_1, \dots, b_{l-1}) = TRUE$. По определению предиката p^l_{max} это означает, что элементы $b_1, \dots, b_{l-1}$ имеют общий максимум

d в H_s. Но этот же элемент d является общим максимумом элементов $b_1, \dots, b_{l-1}, M_W$, откуда получаем, что $p^l_{max}(b_1, \dots, b_{l-1}, M_W) = TRUE$. Следовательно $p'(b_1, \dots, b_{l-1}) = TRUE$.

Если же $p'(b_1, \dots, b_{l-1}) = TRUE$, то существует элемент $a \in H_s$ такой, что система $b_1, \dots, b_{l-1}, a$ имеет общий максимум d в ЧУМ H_s. Но тогда элемент d является и общим максимумом системы $b_1, \dots, b_{l-1}$. Следовательно $p^{l-1}_{max}(b_1, \dots, b_{l-1}) = TRUE$.

Итак, $p' = p^{l-1}_{max}$. Поскольку все переменные предиката p^l_{max} симметричны (то есть предикат p^l_{max} не меняется при произвольной перестановке переменных), то любая проекция p^l_{max} по одной переменной дает предикат p^{l-1}_{max}. Отсюда следует, что аппроксимация предиката $\Psi(p^l_{max})(x_1, \dots, x_l)$ является конъюнкцией l предикатов $p^{l-1}_{max}(z_1, \dots, z_{l-1})$, где переменные $z_1, \dots, z_{l-1}$ пробегают всевозможные подмножества из $l-1$ переменных из списка $\{x_1, x_2, \dots, x_l\}$. Получаем, что условиям $p^l_{max}(\tilde{a}) = FALSE$, $\Psi(p^l_{max})(\tilde{a}) = TRUE$ удовлетворяют все наборы $\tilde{a} = (a_1, \dots, a_l)$ такие, что множество элементов $\{a_1, \dots, a_l\}$ не имеет общего максимума в ЧУМ H_s, а любое его подмножество из $l-1$ элемента имеет общий максимум.

Поскольку ЧУМ формулы (6), задающей предикат p^l_{max}, не содержит сравнимых элементов, соответствующих свободным переменным, то, согласно лемме 8, классы $\mathrm{Pol}\, p^l_{max}$ отличны от M_{H_s}. С другой стороны, выше было показано, что $p^{l-1}_{max} \in [p^l_{max}]$ для любого $l > 2$. Получаем, что для любого $l > 2$ предикат $p^2_{max} \in [p^l_{max}]$. По лемме 1 получаем, что $\mathrm{Pol}\, p^l_{max} \subseteq \mathrm{Pol}\, p^2_{max}$. Поскольку ЧУМ H_s имеет единственный минимальный элемент, то по теореме 1 получаем, что класс M_{H_s} содержится в единственном предполном классе $\mathrm{Pol}\, p^2_{max}$. Итак, получаем

$$\mathrm{Pol}\, p^l_{max} \subseteq \mathrm{Pol}\, p^{l-1}_{max} \subseteq \ldots \subseteq \mathrm{Pol}\, p^2_{max} \subset P_k.$$

Получаем, что все классы $\mathrm{Pol}\, p^l_{max}$ отличны от P_k.

Покажем далее, что классы $Polp^r_{max}$ различны для $2 \leqslant r \leqslant s$.

Зафиксируем произвольное число r такое, что $2 \leqslant r \leqslant s$. Рассмотрим r различных максимальных элементов ЧУМ H_s. Пусть W' — множество номеров указанных максимальных элементов. Выполняется соотношение $W' \subseteq W$. Возьмем всевозможные $(r-1)$-элементные подмножества W'_i множества W'. Всего их $\binom{r}{r-1} = r$ штук. Соответствующие единственные элементы множеств $H_{W'_i}$ обозначим через $b_1, \dots, b_r$.

Предположим, что в H_s существует общий максимум элементов $b_1, \dots, b_r$. Это означает, что у элементов $b_1, \dots, b_r$ существует общий максимум, являющийся максимальным элементом ЧУМ H_s. По определению H_s каждый элемент $a \in H_{W_i}$ меньше или равен максимальным элементам с номерами из множества W_i и несравним с остальными максимумами H_s. Согласно выбору элементов $b_1, \dots, b_r$, соответствующие им множества W'_i (напомним, это всевозможные $(r-1)$-элементные подмножества r-элементного множества W') не имеют элемента, принадлежащего всем множествам $W'_1, \dots, W'_r$. Получаем, что элементы $b_1, \dots, b_r$ не имеют общего максимума в H_s.

Рассмотрим любое подмножество $\{b_1, \dots, b_r\}$ из $(r-1)$ элемента. Не ограничивая общности будем рассматривать подмножество $\{b_1, \dots, b_{r-1}\}$. Напомним, что элементу b_r соответствует некоторое $(r-1)$-элементное подмножество номеров максимальных элементов $W'_h \subset W'$. Следовательно, существует элемент $j \in W' \setminus W'_h$. Элемент j должен войти в $(r-1)$ подмножество вида W'_i. Поскольку всего указанных подмножеств r штук, то j принадлежит всем подмножествам $W_1, W_2, \dots, W_{h-1}, W_{h+1}, \dots, W_r$. Указанные подмножества соответствуют элементам $b_1, \dots, b_{r-1}$. Получаем, что элементы $b_1, \dots, b_{r-1}$ имеют общий максимум h_j (максимальный элемент ЧУМ H_s с номером j). Итак, любое $(r-1)$-элементное подмножество множества $\{b_1, \dots, b_r\}$ имеет общий максимум в H_s.

Обозначим набор $\tilde{b} = (b_1, \dots, b_r)$. В силу доказанных выше свойств предикатов p^l_{max}, получаем $p^r_{max}(\tilde{b}) = FALSE$ и $\Psi(p^r_{max})(\tilde{b}) = TRUE$. Получаем, что предикаты p^r_{max} невырождены $(2 \leqslant r \leqslant s)$.

Предположим, что для двух различных чисел r_1, r_2 таких, что $2 \leqslant r_1, r_2 \leqslant s$, выполнено $\mathrm{Pol}\, p^{r_1}_{max} = \mathrm{Pol}\, p^{r_2}_{max}$. Отметим, что формулы вида (6), задающие предикаты p^l_{max} принадлежат множеству $\overline{F}_1$. По первому пункту леммы 7 получаем, что $p^{r_1}_{max} \in [p^{r_2}_{max}]$ и $p^{r_2}_{max} \in [p^{r_1}_{max}]$. Но по лемме 17 мы не можем реализовать формулой невырожденный предикат меньший местности из невырожденного предиката большей местности. Полученное противоречие доказывает, что классы $\mathrm{Pol}\, p^s_{max}, \mathrm{Pol}\, p^{s-1}_{max}, \dots, \mathrm{Pol}\, p^2_{max}$ различны.

Итак, мы построили $s-1=n$ различных классов, отличных от P_k и строго содержащих класс монотонных функций M_{H_s}.

Теорема доказана.

Следующая теорема показывает сложность бесконечной надструктуры классов монотонных функций, сохраняющих ЧУМ с единственным минимальным и двумя максимальными элементами.

Теорема 5. Если для некоторого $H \in L_{2,1}$ надструктура класса M_H бесконечна, то существует бесконечное число различных классов, содержащих M_H и не являющихся предикатно-описуемыми.

Определение 17. Для произвольного k обозначим через Q_k^2 множество ЧУМ $H \in L_{22}$, составленных из элементов E_k, таких, что ЧУМ H содержит подмножество $\overline{H}_1$, причем во множестве H не появляется элемента a такого, что $a \leqslant 0,3$ и $a \geqslant 1,2$; элементы $0,3$ являются максимальными в H или элементы $1,2$ — минимальными.

Теорему 3 можно обобщить следующим образом.

Теорема 6. Для того, чтобы в P_k класс монотонных функций M_H, где $H \in L_{22}$, обладал бесконечной надструктурой необходимо и достаточно, чтобы ЧУМ $H \in Q_k^2$.

Работа выполнена при поддержке РФФИ, грант 09-01-00701.

Список литературы

[1] В. Б. Ларионов. *О положении некоторых классов монотонных k-значных функций в решетке замкнутых классов.* Дискретная математика, Т. 21, вып. 5, 111-116, Москва, 2009.

[2] В. Б. Ларионов. *О монотонных замкнутых классах функций многозначной логики с бесконечной надструктурой.* Материалы VII молодежной научной школы по дискретной математике и ее приложениям (18-23 мая 2009 г.), М.: ИПМ им. М. В. Келдыша РАН, т. 2, 7-12, 2009.

[3] С. В. Яблонский, Г. П. Гаврилов, А. А. Набебин. *Предполные классы в многозначных логиках.* Издательский дом МЭИ, Москва, 1997.

[4] I. Rosenberg. *La structure des fonctions de plusieur variables sur un ensemble fini.* Comptes Rendus, de l'Academ, 260, 3817-3819, Paris, 1965.

[5] В. В. Мартынюк. *Исследование некоторых классов функций в многозначных логиках.* Проблемы кибернетики, вып. 3, М.: Наука, 49-61, 1960.

[6] С. С. Марченков. *Замкнутые классы булевых функций.* Физматлит, Москва, 2000.

[7] В. Г. Боднарчук, В. А. Калужнин, В. Н. Котов, Б. А. Ромов. *Теория Галуа для алгебр Поста.* Кибернетика, №3, 1-10, №5, 1-9, 1969.

УДК 519.7

ИССЛЕДОВАНИЕ МЕТРИЧЕСКОГО АНАЛОГА МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

© 2010 г. Н. А. Найденов

Кафедра математических методов прогнозирования

1 Введение

В интеллектуальном анализе данных классическим является представление информации об объектах в виде признаковых описаний. Общеизвестны проблемы, которые возникают при плохом наборе признаков или при достаточно большом их количестве. Чтобы справиться с указанными проблемами, предложены различные подходы, в числе которых выделяют методы понижения размерности. Один из наиболее известных — это метод главных компонент (Principal components analysis, PCA). Классический МГК [1] может быть сформулирован как задача аппроксимации числовых признаковых описаний объектов, причём на признаки не накладываются никакие дополнительные ограничения.

В настоящее время всё больше прикладных и теоретических работ ориентировано на описание набора объектов не признаками, а попарными расстояниями между объектами. Такие описания называют метрическими. При этом возникают и требуют решения проблемы, аналогичные проблемам признаковых описаний [2]. В работе предложен и исследован аналог метода главных компонент для метрических описаний. На получаемое новое описание малой размерности накладывается дополнительное ограничение в виде условия, что каждая новая компонента описания должна удовлетворять аксиомам полуметрики. Рассматриваемый метод назван метрическим методом главных компонент.

Доказана теорема о сохранении качества аппроксимации при добавлении метрических условий, т.е. дополнительные требования не снижают качества аппроксимации. Показано, что «привлекательные» свойства классического метода главных компонент сохраняются: главные компоненты можно вычислять частично, и они годятся сразу для всех размерностей, главные компоненты ортогональны и можно удобно переходить от старых описаний к новым.

Проведены эксперименты на модельных и реальных метрических данных для анализа необходимости свободного члена в предложенном методе главных компонент.

2 Определения и обозначения

В работе используется стандартное определение полуметрики.

Определение 2.1 *Для данного множества M функция $d : M \times M \to \mathbb{R}$ называется полуметрикой на M, если для любых точек из M она удовлетворяет следующим условиям:*

- $d(x,x) = 0$
- $d(x,y) = d(x,y)$
- $d(x,z) \leqslant d(x,y) + d(y,z)$

То есть, в отличие от метрики, различные точки в M могут находиться на нулевом расстоянии. Тот факт, что функция $d : M \times M \to \mathbb{R}$ является полуметрикой будем обозначать через $d \in MET$.

При работе с классическими признаковыми описаниями описанием объекта является вектор значений разных признаков, при работе с метрическими описаниями пара объектов описывается вектором различных расстояний [2, 6, 7].

Определение 2.2 *Метрическая конфигурация — набор попарных расстояний в одной метрике на выборке.*

Определение 2.3 *Метрическое описание — набор объектов, которым соответствует набор метрических конфигураций.*

Определение 2.4 $\hat{1}$ *— метрика пространства изолированных точек (расстояние между любыми объектами равно 1).*

3 Классический метод главных компонент

Метод главных компонент (англ. Principal components analysis, PCA) — один из самых известных способов сократить размерность данных, потеряв наименьшее количество информации [1, 3, 5]. Он был предложен Карлом Пирсоном (англ. Karl Pearson) в 1901 г. и применяется во многих областях, таких как распознавание образов, компьютерное зрение, сжатие данных и т. п. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных или к сингулярному разложению матрицы данных. Иногда в статистике совместно упоминают метод главных компонент (МГК) и преобразование Кархунена-Лоэва (англ. Karhunen-Loeve) или преобразование Хотеллинга (англ. Hotelling transform). Основное отличие заключается в том, что МГК - метод обработки чисел без привлечения теории вероятностей.

Цель МГК — аппроксимировать исходный набор векторов в пространстве меньшей размерности. Т.е. в линейном векторном пространстве для заданного набора точек требуется найти такое линейное многообразие заданной меньшей размерности, проекции исходных точек на которое в среднеквадратичном смысле ближе всего к исходным точкам.

На практике часто удаётся упростить данные на порядки: от 1000 переменных перейти всего к двум. При этом ничего не выбрасывается: все переменные учитываются. Нередко исследователи интерпретируют главные компоненты как скрытые переменные, управляющие устройством данных.

Формальное описание Одно из стандартных представлений модели МГК следующее. Исходные данные представлены в виде матрицы данных X.

$$X = (f_1 \dots f_n) = (x_1 \dots x_q)' = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & & \vdots \\ x_{q1} & \dots & x_{qn} \end{pmatrix}$$

где признак $f_j \in \mathbb{R}^q$ и признаковое описание объекта $x_i \in \mathbb{R}^n$

Матрица X приближается произведением матриц. По-иному говоря, проводится факторизация матрицы данных. Факторизация записывается следующим образом:

$$X \approx T \cdot P^T$$

где T — матрица счетов или новых описаний объектов, P^T — матрица нагрузок или матрица главных компонент. Условимся символом $\approx$ обозначать приближение в среднеквадратичном смысле. Отметим, что размерность матрицы $T - q \times k$, а матрицы $P^T - k \times n$, $k \leqslant n$.

Среднеквадратичное приближение — это способ аппроксимации, при котором критерий близости матрицы исходных данных и произведения TP^T определяется как сумма квадратов отклонений соответствующих элементов и эта сумма минимальна.

$$Q = \left\| X - TP' \right\|_2^2 \to min$$

Также модель МГК можно представить в векторной записи. Отметим, что у МГК имеется две модели: со свободным членом и без него. Представленные здесь модели — со свободным членом.

$$X = (f_1 \dots f_n) \approx \left(\vec{1}, g_1, \dots, g_k\right) \times (a_0, a_1, \dots, a_k)^T$$

где вектор $a_0 \in \mathbb{R}^n$ — свободный член модели МГК. $a_1, \dots, a_k \in \mathbb{R}^n$ — главные компоненты. $g_1, ..., g_k \in \mathbb{R}^q$ — новые описания объектов.

Свойства метода главных компонент Свою популярность МГК завоевал благодаря ряду полезных свойств:

- **Главные компоненты ортогональны**

  Это позволяет получать новые описания из старых. Нет никакой необходимости вычислять и запоминать все компоненты при переходе к новым описаниям, достаточно использовать лишь столько, сколько удовлетворяет размерности приближения.

- **Главные компоненты не зависят от размерности приближения**

  Сразу находятся все главные компоненты, которые не меняются при изменении размерности приближения.

- **Упорядоченность по вкладу в объяснение исходных данных**

  Главные компоненты упорядочиваются по их вкладу в качество аппроксимации и оптимальным решением для любой размерности будет выбор первых компонент.

В работе предлагается исследовать аналог МГК для метрических описаний.

4 Метрический метод главных компонент

4.1 Метрические описания

При работе с классическими признаковыми описаниями один объект описывается вектором значений разных признаков. При работе с метрическими описаниями пара объектов описывается вектором различных расстояний [7]. Через n обозначим число признаков в признаковых описаниях или число конфигураций сходства в метрических описаниях. Признаковые описания выборки представляются матрицей данных вида

$$X = (f_1 \dots f_n) = (x_1 \dots x_q)' = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & & \vdots \\ x_{q1} & \dots & x_{qn} \end{pmatrix}$$

где признак $f_j \in \mathbb{R}^q$ и признаковое описание объекта $x_i \in \mathbb{R}^n$. Метрические описания выборки представляются матрицей данных вида

$$Y = (\hat{\rho}_1 \dots \hat{\rho}_n) = (y_1 \dots y_t)' = \begin{pmatrix} y_{11} & \dots & y_{1n} \\ \vdots & & \vdots \\ y_{t1} & \dots & y_{tn} \end{pmatrix}$$

где $\hat{\rho}_j \in \mathbb{R}^t$ и метрическое описание пары объектов $y_i \in \mathbb{R}^n$.

4.2 Задача метрического метода главных компонент

Чтобы ниже было легче проводить аналогии,напомним задачу классического МГК, которую сформулируем как задачу аппроксимации набора векторов в линейном многообразии меньшей размерности. В линейном векторном пространстве $\mathbb{R}^n$ для заданного набора точек $x_1, \dots, x_q$ требуется найти такое линейное многообразие L заданной меньшей размерности k, проекции исходных точек на которое в среднеквадратичном смысле ближе всего к исходным точкам. Иными словами, для заданной матрицы признаковых описаний X требуется найти такие $a_0, \ldots, a_k \in \mathbb{R}^n$, определяющие

$$L_k = \{a_0 + g_1 a_1 + \cdots + g_k a_k | g_1, \ldots, g_k \in \mathbb{R}\},$$

и такие $g_1, \ldots, g_k \in \mathbb{R}^q$, которые доставляют минимум функционалу ошибки аппроксимации

$$Q = \sum_{i=1}^{q} dist^2_{euclid}\left(x_i, a_0 + g_{1i} a_1 + \cdots + g_{ki} a_k\right) = \left\|X - (1 g_1 \ldots g_k) \times (a_0 a_1 \ldots a_k)'\right\|_2^2$$

Это задача безусловной оптимизации. Вектор a_0 называется свободным членом модели МГК, векторы $a_1, \ldots, a_k$ называются главными компонентами, векторы $g_1, \ldots, g_k$ называются новыми признаками и составляют новое признаковое описание.

Задачу метрического МГК сформулируем как задачу условной оптимизации.

Задача Для заданной матрицы метрических описаний Y требуется найти такие $a_0, \ldots, a_k \in \mathbb{R}^n$, определяющие

$$L_k = \{a_0 + g_1 a_1 + \cdots + g_k a_k | g_1, \ldots, g_k \in \mathbb{R}\},$$

и такие $\hat{r}_1, \ldots, \hat{r}_k \in MET$, которые доставляют минимум функционалу ошибки аппроксимации

$$Q = \left\|(\hat{\rho}_1 \ldots \hat{\rho}_n) - \left(\hat{1} \hat{r}_1 \ldots \hat{r}_k\right) \times (a_0 a_1 \ldots a_k)'\right\|_2^2$$

Конфигурации $\hat{r}_1, \ldots, \hat{r}_k$ называются новыми метрическими конфигурациями сходства и составляют новое метрическое описание. Таким образом, единственным существенным различием задач являются дополнительные метрические требования на новые метрические описания.

5 Анализ необходимости свободного члена

Как известно, для метода главных компонент распространены две основные модели: со свободным членом и без него. Модель МГК со свободным членом имеет вид:

$$(\hat{\rho}_1, \ldots, \hat{\rho}_n) \approx \left(\hat{1}, \hat{r}_1, \ldots, \hat{r}_k\right) \times (a_0, a_1, \ldots, a_k)^T$$

а без свободного члена —

$$(\hat{\rho}_1, \ldots, \hat{\rho}_n) \approx (\hat{r}_1, \ldots, \hat{r}_k) \times (a_1, \ldots, a_k)^T$$

Напомним, что свободный член модели МГК $\vec{a}_0$ — это точка исходного пространства метрических описаний, соответствующая началу координат в новом пространстве метрических описаний.

В классическом МГК для произвольных признаковых описаний свободный член обычно позволяет существенно улучшить качество аппроксимации при малых размерностях. Действительно, в общем случае нет никаких оснований для того, чтобы выделять объект с нулевым признаковым описанием. Однако при работе с метрическими описаниями возникает гипотеза, что свободный член не оказывает существенного влияния на результат факторизации. Действительно, нулевое метрическое описание можно считать типичным, поскольку оно соответствует парам совпадающих объектов. Поэтому представляется разумным исследовать вопрос о том, не будет ли характерным для метрических описаний всегда иметь практически нулевой свободный член при использовании классического МГК. Отметим, что в данном разделе требования метричности на новые описания не накладываются. Чтобы разобраться с необходимостью свободного члена в МГК, проведём несколько экспериментов.

Метод Будем применять обе модели МГК (со свободным членом и без него) на модельных и реальных данных и для обеих моделей сравнивать качество аппроксимации Q.

Реализация Вся практическая часть данной работы выполнялась в среде программирования **MATLAB 7.7.0 (R2008b)**. Были написаны функции генерации исходных множеств, функции сингулярного разложения матрицы исходных данных, а также функции, визуализирующие полученное качество аппроксимации данных.

Расчёты (эксперименты) МГК осуществляется при помощи функции

```
[u s v] = svd(X)
```

$T = u \cdot s$, $P = v$, где X — исходная матрица данных, T — матрица новых описаний, P — главные компоненты [4]. Последовательно для каждой размерности k матриц T и P находим качество аппроксимации в среднеквадратичном смысле:

$$Q = \left\| X - TP' \right\|_2^2$$

Приведённые ниже графики показывают зависимость среднеквадратичной ошибки Q от размерности новых описаний k.

5.1 Классические признаковые описания

Чтобы выяснить, какое различие в качестве аппроксимации можно назвать существенным, проведём эксперимент на обычных признаковых описаниях. Выберем в пространстве $\mathbb{R}^6$ 4950 точек, лежащих на гиперплоскости, проходящей вдали от начала координат. Представленный ниже график демонстрирует зависимость качества аппроксимации от размерности приближения для обеих моделей МГК (Рис.1).

Из построенного графика видно, что на модель МГК без свободного члена никак не повлияло расположение данных, и качество аппроксимации ухудшается с уменьшением размерности приближения. Модель МГК со свободным членом определила, что данные лежат на гиперплоскости и, можно сказать, избавилась от лишней размерности, тем самым выигрывая в качестве аппроксимации.

5.2 Эксперимент на модельных метрических данных

Данные Объекты выбирались из пространств $\mathbb{R}^2$, $\mathbb{R}^5$ и $\mathbb{R}^{10}$. В качестве метрик в этих пространствах были использованы l_1 , l_2 ,l_∞ ,$\frac{l_1}{1+l_1}$, $\frac{l_2}{1+l_2}$, $\frac{l_\infty}{1+l_\infty}$.

- $l_1 = \sum_{i=1}^{n} |x_i - y_i|$ — расстояние Хэмминга (городских кварталов)
- $l_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$ — евклидово расстояние
- $l_\infty = max_{i=1..n} |x_i - y_i|$ — расстояние Чебышёва

Генерировались следующие модельные наборы: куб, кубический слой, пара отдалённых шаров, два соприкасающихся множества.

Количество точек в каждом модельном наборе равно 100, следовательно, число пар равно 4950. Таким образом, модельные данные имеют размер 4950×6. Результат эксперимента — значения Q для каждого k от 1 до 6.

Результаты Эксперименты проводились на всех типах данных, и полученные результаты были очень похожи. Поэтому приведём один из графиков, так как остальные аналогичные. Данный график (Рис.2) отображает работу МГК на исходных данных типа: кубический слой в $\mathbb{R}^2$.

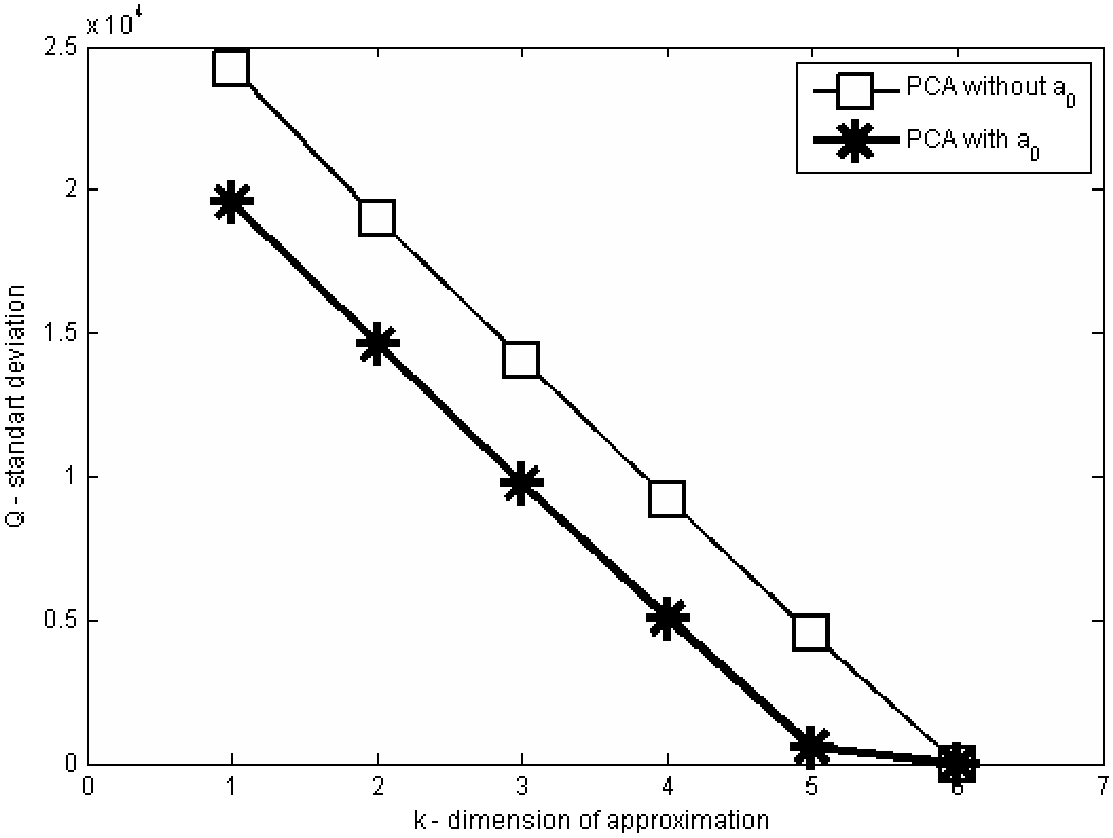

Рис. 1: Зависимость качества аппроксимации от размерности приближения. Линия с квадратами — модель МГК без свободного члена, линия со звёздочками — модель МГК со свободным членом

5.3 Эксперимент на реальных данных

Проведём аналогичный эксперимент на реальных данных. Реальные данные представляют собой биометрическую информацию, а именно берётся геометрия руки человека. Объектом является обработанная фотография ладони. Для эксперимента было выбрано 100 фотографий. Между парой снимков ладоней высчитывается с помощью определённых функций набор из 11 расстояний [8]. Первые 5 расстояний характеризуют сравнение пальцев по ширине, следующие 5 - по изгибу оси пальцев, и 11 расстояние - площадь симметрической разности ладоней при наложении. Таким образом, реальные данные представляют собой матрицу размером 4950×11.

Результаты Ниже представлен график (Рис.3) той же зависимости Q от k.

5.4 Анализ результатов (выводы)

Из результатов эксперимента видно, что для метрических описаний модель со свободным членом уменьшает ошибку аппроксимации незначительно. Поэтому, если бы нашей целью было только повышение качества аппроксимации Q, то свободным членом в обычной модели МГК можно было бы пренебречь.

Также можно заметить, что ошибка аппроксимации на реальных данных падает быстрее, чем на модельных.

Эксперименты на модельных наборах и реальных данных подтвердили, что наличие свободного члена в модели МГК практически не влияет на результаты применения МГК к метрическим описаниям. Однако, при применении модели метода главных компонент к метрическим структурам, свободный член играет очень важную роль.

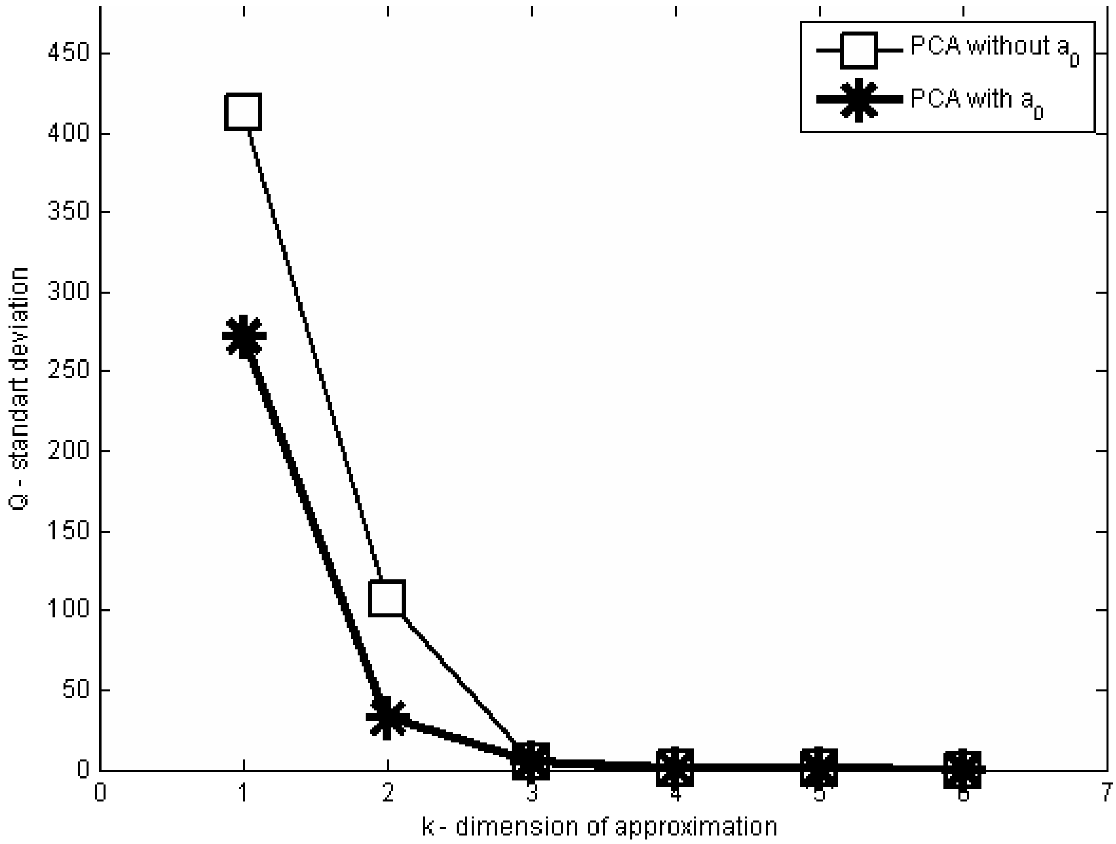

Рис. 2: Зависимость качества аппроксимации от размерности приближения. Линия с квадратами — модель МГК без свободного члена, линия со звёздочками — модель МГК со свободным членом

6 О сохранении качества аппроксимации и свойств МГК

В данном разделе исследуются свойства метрического метода главных компонент. Напомним, что задача метрического МГК отличается от классической задачи МГК дополнительными метрическими требованиями. Вообще говоря, можно было бы ожидать, что выполнение дополнительные требований влечёт ухудшение качества аппроксимации и потерю полезных свойств МГК, перечисленных в пункте 3.

Важнейшим результатом исследования можно считать доказательство того факта, что качество аппроксимации и указанные свойства сохраняются.

Теорема 1 (о сохранении качества аппроксимации) *Метрический МГК даёт ту же самую аппроксимацию исходных метрических описаний, что и классический МГК.*

Доказательство. Идея доказательства состоит в том, что условие метричности достигается благодаря модификации свободного члена.

Применим к исходным метрическим описаниям X классический МГК со свободным членом. Получим новые описания, вообще говоря, не удовлетворяющие метрическим требованиям, и главные компоненты. Размерность приближения примем равной исходной размерности данных $k = n$:

$$X = (f_1 \dots f_n) \approx \left(\vec{1}, g_1, \dots, g_k\right) \times (a_0, a_1, \dots, a_k)'$$

Или в матричной записи

$$X \approx T \cdot P^T$$

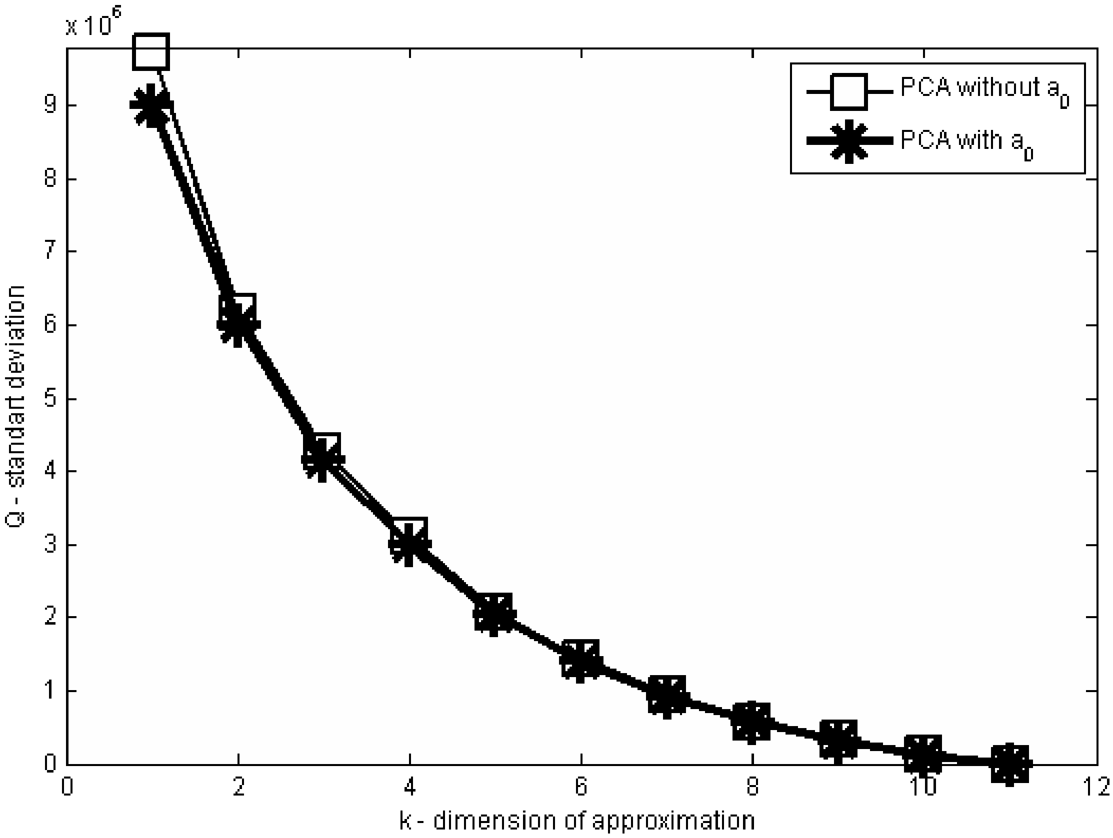

Рис. 3: Зависимость качества аппроксимации от размерности приближения. Линия с квадратами — модель МГК без свободного члена, линия со звёздочками — модель МГК со свободным членом

$$T = \begin{array}{c} \begin{array}{cccc} & g_1 & & g_k \end{array} \\ \begin{pmatrix} 1 & g_1^1 & \dots & g_k^1 \\ \dots & \dots & \dots & \dots \\ 1 & g_1^q & \dots & g_k^q \end{pmatrix} \end{array}, P^T = (a_0, a_1, \dots, a_k)'$$

Столбцы g_j понимаем как новые описания объектов. Повторим, что в общем случае они не удовлетворяют метрическим требованиям. Нашей ближайшей целью будет так модифицировать g_i, чтобы получить полуметрики.

Введём величины $\delta_{i,j,l}^s = g_s^i + g_s^j - g_s^l$ для всех $1 \leqslant i, j, l \leqslant m$ и $1 \leqslant s \leqslant n$ и

$$\delta_s = min_{i,j,l} \delta_{i,j,l}^s.$$

Если $\delta_s \geqslant 0$ для всех $1 \leqslant s \leqslant n$, то все неравенства треугольника выполняются и новые метрические конфигурации являются полуметриками. Результат классического МГК можно вернуть как результат метрического МГК. На самом деле случай $\delta_s \geqslant 0$, для всех $1 \leqslant s \leqslant n$ невозможен. Новые описания в классическом МГК ортогональны. В [7] доказано, что полуметрики не могут быть ортогональными.

Если $\delta_s < 0$, то из всех элементов s-го столбца матрицы новых описаний вычитаем это значение, так называемый *постоянный аддитивный дефект* и полученный результат обозначим $\hat{r}_i$.

$\hat{r}_s = g_s - \vec{\delta_s}$ Здесь $\vec{\delta_s}$ — вектор длины q, все элементы которого равны δ_s.

Полученные новые описания объектов $\hat{r}_s$ удовлетворяют всем свойствам полуметрики (если уже по ним снова вычислить δ_s, то оно будет нулевым). Покажем, что эти описания можно использовать в модели метрического МГК.

Запишем правую часть модели МГК со свободным членом и модифицируем его.

$$T \cdot P^T = \left(\vec{1}, g_1, \dots, g_n\right) \times (a_0, a_1, \dots, a_n)' = a_0 + g_1 a_1 + \dots + g_n a_n =$$

$$= a_0 + \left(g_1 - \vec{\delta}_1\right) a_1 + \dots + \left(g_n - \vec{\delta}_n\right) a_n + \left(\vec{\delta}_1 a_1 + \dots + \vec{\delta}_n a_n\right) =$$

$$= \left(a_0 + \vec{\delta}_1 a_1 + \dots + \vec{\delta}_n a_n\right) + \hat{r}_1 a_1 + \dots + \hat{r}_n a_n = \left(\hat{1}, \hat{r}_1, \dots, \hat{r}_n\right) \times (b_0, a_1, \dots, a_n)',$$

$$\text{где } b_0 = a_0 + \vec{\delta}_1 \cdot a_1 + \dots + \vec{\delta}_n \cdot a_n$$

Таким образом, мы добились выполнения условий метричности новых описаний путём модификации свободного члена. Полученное решение задачи метрического МГК даёт ту же самую аппроксимацию с теми же самыми главными компонентами, что и классический МГК. ■

Опираясь на доказательство **Теоремы 1** можно сформулировать следующую теорему:

Теорема 2 (о сохранении свойств главных компонент) *Решение, даваемое метрическим МГК, содержит те же самые главные компоненты $a_1, \dots, a_k$, что и решение, даваемое классическим МГК.*

Доказательство. Как видно из предыдущего доказательства, главные компоненты в метрическом МГК являются главными компонентами классического МГК.

$$\left(\vec{1}, g_1, \dots, g_n\right) \times (a_0, a_1, \dots, a_n)' = \left(\hat{1}, \hat{r}_1, \dots, \hat{r}_n\right) \times (b_0, a_1, \dots, a_n)$$

Это ведёт к сохранению важных свойств главных компонент. А именно: они ортогональны, независимы от размерности приближения и упорядочены по своему вкладу в объяснение исходных данных. ■

То есть, верны следующие результаты.

- Аппроксимация та же самая, следовательно, качество аппроксимации сохраняется
- Главные компоненты те же самые, следовательно, их полезные свойства сохраняются

Результат работы классического МГК определён с точностью до направления новых описаний. В случае получения результата для одной фиксированной размерности произвол в факторизации существенно возрастает. Представляется целесообразным в будущем использовать такую неоднозначность для смягчения коррекции новых описаний.

7 Заключение

В работе предложен, формализован и исследован аналог метода главных компонент для метрических описаний данных. К классическому методу добавлено дополнительное ограничение в виде условия, что получаемое новое описание исходных данных меньшей размерности удовлетворяет всем аксиомам полуметрики. Доказано, что, несмотря на добавление метрических требований в задачу МГК, качество аппроксимации не ухудшилось, а из всех важных полезных свойств решения классического МГК было утрачено лишь то, которое неминуемо было утратить. Вычислительные затраты возросли.

Главный теоретический результат данной работы — это доказательство теоремы о сохранении качества аппроксимации при добавлении метрических условий. Они достигаются путём модификации свободного члена в классической модели МГК. Показано, что:

- Аппроксимация та же самая, следовательно, качество аппроксимации сохраняется
- Главные компоненты те же самые, следовательно, их полезные свойства сохраняются

Как и в классическом МГК, решение метрического МГК неоднозначно. Это позволяет исследовать возможность введения в задачу метрического МГК других дополнительных ограничений. Таким образом, область применения метода главных компонент была расширена и на метрические структуры данных.

Был проведён ряд экспериментов, анализирующих необходимость использования свободного члена в модели метода главных компонент при работе с метрической информацией. Результаты показали, что на качество аппроксимации метрических данных свободный член практически не влияет.

Список литературы

[1] Jolliffe T. *Principal Component Analysis* (2nd Ed.) Springer-Verlag New York Inc, 2002

[2] Майсурадзе А. И. *О поиске оптимального коллективного слагаемого для набора метрических конфигураций* Искусственный интеллект, №2, с.146-150, 2006

[3] Dunteman G. H. *Principal Component Analysis* no. 07069 in Sage University Paper Series on Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills, 1989.

[4] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. *Singular Value Decomposition* Numerical Recipes in C. 2nd edition. Cambridge University Press

[5] Померанцев А. *Научно-популярная и методическая работа «Метод главных компонент»* http://www.chemometrics.ru/materials/textbooks/pca.htm

[6] Скворцов В. А. *Примеры метрических пространств* Библиотека «Математическое просвещение». Выпуск 9, 2001

[7] Майсурадзе А. И. *О оптимальных разложениях конечных метрических конфигураций* ЖВМ и МФ

[8] Местецкий Л., Бакина И. *Метод сравнения формы ладоней при наличии артефактов* Математические методы распознавания образов (ММРО-14), 2009

УДК 510.2

РАВНОСХОДИМОСТЬ В СМЫСЛЕ ЧЕЗАРОВСКИХ СРЕДНИХ РАЗЛОЖЕНИЙ В РЯДЫ ПО СОБСТВЕННЫМ ФУНКЦИЯМ ОПЕРАТОРОВ ШТУРМА-ЛИУВИЛЛЯ С ПОТЕНЦИАЛАМИ - РАСПРЕДЕЛЕНИЯМИ

© 2010 г. О. А. Швейкина

Кафедра Общей математики

В работе рассматривается оператор Штурма-Лиувилля $Ly = -y'' + q(x)y$ в пространстве $L_2[0;\pi]$ с граничными условиями Дирихле $y(0) = y(\pi) = 0$. Предполагается, что потенциал $q(x)$ — вещественнозначная функция такая, что $q(x) = u'(x)$, $u \in L_2[0;\pi]$ (производная понимается в смысле распределений). Операторы такого вида были впервые определены в работе Савчука, Шкаликова [1]. В работах [1, 2, 3, 4] было показано, что оператор L в случае вещественного потенциала самосопряжен, полуограничен снизу и имеет чисто дискретный спектр. Также были выведены асимптотические формулы для собственных значений и собственных функций оператора.

В настоящей статье изучается вопрос о равносходимости разложения по системе собственных функций оператора L некоторой суммируемой функции f с ее разложением в тригонометрический ряд Фурье (равносходимость рассматривается в смысле суммирования методом чезаровских средних).

Вопрос о равносходимости в случае классического потенциала в настоящее время хорошо изучен, (см., например, работы [5, 6, 7, 8, 9, 10, 11, 12]). В статье Винокурова и Садовничего [13] было доказано, что равносходимость имеет место в случае, когда $f \in L_1$, а потенциал $q(x)$ —вещественный, причем $q(x) = u'(x)$, $u \in BV$. В статье Садовничей [14] были получены достаточные условия на первообразную потенциала u , которые гарантируют равносходимость для произвольной суммируемой функции f. В частности, было показано, что равносходимость будет иметь место, если $u \in L_\infty$ и $\omega_1(u, \frac{1}{n}) = O(\frac{1}{n})$, где ω_1 — интегральный модуль непрерывности функции u. Ниже будет доказано, что в случае, когда суммирование производится методом чезаровских средних, равносходимость частичных сумм будет иметь для произвольной функции $f \in L_1$, при условии, что первообразная потенциала $u \in L_\infty$.

В статье будет использована следующие результаты, полученные в работах [2, 4]:

Лемма А. (Савчук, Шкаликов)

Пусть $q(x) = u'(x)$, $u \in L_2[0;\pi]$ — вещественнозначная функция. Тогда все собственные значения оператора L просты, а ортонормированная система $\{y_n(x)\}_{n=1}^{\infty}$ собственных функций образует базис Рисса в пространстве $L_2[0;\pi]$. Кроме того,

$$y_n(x) = \sqrt{\frac{2}{\pi}}(\sin(nx) + \varphi_n(x)),$$

где $\varphi_n(x) = \varphi_{n,0}(x) + \varphi_{n,1}(x) + \varphi_{n,2}(x)$. При этом

$$\varphi_{n,0}(x) = \sqrt{\frac{2}{\pi}} \left(\int_0^\pi u(t) \left(1 - \frac{t}{\pi}\right) \cos(2nt)\,dt \cdot \sin(nx) + \right.$$
$$\left. + \, x \cdot \cos(nx) \cdot \int_0^\pi \left(-\frac{1}{\pi}\right) u(t) \sin(2nt)\,dt - \int_0^x u(t) \sin(n(x-2t))\,dt \right);$$

$$\varphi_{n,1}(x) = \sqrt{\frac{2}{\pi}}\sin(nx)\cdot\left(-\frac{1}{2n}\int\limits_0^x u^2(t)\sin(2nt)\,dt\right) + \sqrt{\frac{2}{\pi}}\cos(nx)\left[-\frac{x}{\pi}\int\limits_0^\pi u(t)\sin(2nt)\,dt+\right.$$
$$+\frac{x}{2\pi n}\int\limits_0^\pi u^2(t)\cdot(\cos(2nt)-1)\,dt - \frac{2x}{\pi}\int\limits_0^\pi\int\limits_0^t u(t)u(s)\cos(2nt)\sin(2ns)\,ds\,dt+$$
$$\left.+\frac{1}{2n}\int\limits_0^x u^2(t)(1-\cos(2nt))\,dt + 2\int\limits_0^x\int\limits_0^t u(t)u(s)\cos(2nt)\sin(2ns)\,ds\,dt\right];$$

а $\{\|\varphi_{n,2}(x)\|_C\}_{n=1}^\infty \in l_1$.

Переходим к теореме, содержащей основной результат данной работы.

Теорема.

Пусть $Ly = -y'' + q(x)y$ — оператор Штурма-Лиувилля, действующий в пространстве $L_2[0;\pi]$, с граничными условиями Дирихле $y(0) = y(\pi) = 0$. Пусть $q(x) = u'(x)$, где $u \in L_\infty[0;\pi]$ — вещественнозначная функция. Обозначим $\{y_n(x)\}_{n=1}^\infty$ — ортонормированная система собственных функций оператора L. Пусть

$$\sigma_n(x) = \frac{1}{n+1}\sum_{k=1}^n\sum_{j=1}^k (f(t), y_j(t))\cdot y_j(x);$$

$$S_n(x) = \frac{1}{n+1}\sqrt{\frac{2}{\pi}}\sum_{k=1}^n\sum_{j=1}^k (f(t), \sin(jt))\cdot\sin(jx),$$

где $f \in L_1$ — произвольная функция.

Тогда $\|\sigma_n(x) - S_n(x)\|_C \to 0$ при $n\to\infty$.

Доказательство:

Покажем, что $\forall n \in \mathbb{N}$

$$\|\sigma_n(x) - S_n(x)\|_C \leqslant C\cdot\|f(x)\|_{L_1},$$

где C — постоянная, не зависящая от n.

$$\|\sigma_n(x) - S_n(x)\|_C = \frac{1}{n+1}\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^n\sum_{j=1}^k (f(t), \varphi_j(t))\sin(jx)+\right.$$
$$+\sqrt{\frac{2}{\pi}}\sum_{k=1}^n\sum_{j=1}^k (f(t), \sin(jt))\varphi_j(x)+$$
$$\left.+\sum_{k=1}^n\sum_{j=1}^k (f(t), \varphi_j(t))\varphi_j(x)\right\|_C.$$

Рассмотрим каждое слагаемое по отдельности.

1. Применим лемму А к первому слагаемому главного выражения:

$$\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_j(t))\sin(jx) = \sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,0}(t))\sin(jx) + \sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,1}(t))\sin(jx) + \sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,2}(t))\sin(jx)$$

Рассмотрим первое слагаемое правой части последнего выражения:

$$\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,0}(t))\sin(jx) =$$

$$= \sqrt{\frac{2}{\pi}}\left(\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)(1-\frac{s}{\pi})\cos(2js)\cdot\sin(jt)\,ds\,dt\cdot\sin(jx)+\right. \tag{1}$$

$$+ \sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^\pi(-\frac{t}{\pi})f(t)u(s)\sin(2js)\cdot\cos(jt)\,ds\,dt\cdot\sin(jx)+ \tag{2}$$

$$+ \sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^t f(t)u(s)\sin(2js)\cdot\cos(jt)\,ds\,dt\cdot\sin(jx)- \tag{3}$$

$$\left. - \sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^t f(t)u(s)\cos(2js)\cdot\sin(jt)\,ds\,dt\cdot\sin(jx)\right). \tag{4}$$

1.1 Рассмотрим слагаемое (1). Так как $s\in[0;\pi]$, то $|1-\frac{s}{\pi}|\leqslant 1$.
Заметим, что

$$\cos(2js)\cdot\sin(jt)\cdot\sin(jx) = \cos(2js)\cdot\frac{1}{2}(\cos(j(t-x))-\cos(j(t+x))) =$$
$$= \frac{1}{4}\left(\cos(j(2s+t-x))+\cos(j(2s-t+x))-\cos(j(2s+t+x))-\cos(j(2s-t-x))\right).$$

Отсюда получим, что исходное выражение (1)

$$\left|\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)(1-\frac{s}{\pi})\cos(2js)\cdot\sin(jt)\,ds\,dt\cdot\sin(jx)\right|\leqslant$$

$$\leqslant\left|\frac{1}{4}\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)\sum_{k=1}^{n}\left[D_k(2s+t-x)+D_k(2s-t+x)-D_k(2s+t+x)-D_k(2s-t-x)\right]ds\,dt\right|,$$

где $D_n(x)=\frac{1}{2}+\sum_{k=1}^{n}\cos(kx)$ — ядро Дирихле.
Рассмотрим одно из получившихся четырех слагаемых:

$$\left|\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)\sum_{k=1}^{n}D_k(2s+t-x)\,ds\,dt\right|.$$

Принимая во внимание, что $\frac{1}{n+1}\sum_{k=1}^{n} D_k(x) = K_n(x)$, где $K_n(x)$ — ядро Фейера, имеем:

$$\left|\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)\sum_{k=1}^{n} D_k(2s+t-x)\,ds\,dt\right| = \left|\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)(n+1)K_n(2s+t-x)\,ds\,dt\right|.$$

Т.к. $u \in L_\infty$, то по теореме Фейера (см. [15], том 1, стр. 149) $|\int_{-\pi}^{\pi} u(x+t)K_n(t)\,dt| \leqslant M$, где M — константа, не зависящая от n.

Продолжим функцию u периодически за отрезок $[0;\pi]$ и сделаем замену $2s+t-x=\xi$. Получим:

$$\left|\int\limits_0^\pi\int\limits_0^\pi f(t)u(s)(n+1)K_n(2s+t-x)\,ds\,dt\right| =$$
$$=\left|\frac{1}{2}\int\limits_0^\pi f(t)\int\limits_{t-x}^{2\pi+t-x} u(\frac{\xi-t+x}{2})(n+1)K_n(\xi)d\xi\,dt\right| \leqslant$$
$$\leqslant (n+1)\cdot M\int\limits_0^\pi |f(t)|\,dt = M(n+1)\|f(t)\|_{L_1}.$$

Проделывая аналогичные преобразования для трех оставшихся слагаемых из (10), в результате получаем, что слагаемое (1) не превосходит величины $C_1\|f(x)\|_{L_1}(n+1)$, где C_1 — константа, не зависящая от n.

1.2 Следующее слагаемое (2) рассматривается так же, как слагаемое (1) в пункте 1.1, так как функция $\left(-\frac{t}{\pi}\right)\cdot f(t) \in L_1$ и $\left\|\left(-\frac{t}{\pi}\right)\cdot f(t)\right\|_{L_1} \leqslant \|f(t)\|_{L_1}$. В итоге получим, что (2) не превосходит величины $C_2\|f(x)\|_{L_1}(n+1)$.

1.3 Переходим к слагаемому (3). Введем функцию $u_1(s)$ такую, что

$$\begin{cases} u_1(s) = u(s), & s \in [0;t] \\ u_1(s) = 0, & s \in (t;\pi] \end{cases}$$

и периодически продолжим ее за отрезок $[0;\pi]$.

Далее аналогично получим, что

$$\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^t f(t)u(s)\sin(2js)\cdot\cos(jt)\,ds\,dt\cdot\sin(jx)\right\|_C =$$
$$=\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^\pi f(t)u_1(s)\sin(2js)\cdot\cos(jt)\,ds\,dt\cdot\sin(jx)\right\|_C \leqslant C_3\|f(x)\|_{L_1}(n+1)$$

так как $u_1 \in L_\infty[0;\pi]$.

1.4 Слагаемое (4) рассматривается по схеме слагаемого (3) из пункта 1.3. Получаем, что

$$\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^\pi\int\limits_0^t f(t)u(s)\cos(2js)\cdot\sin(jt)\,ds\,dt\cdot\sin(jx)\right\|_C \leqslant C_4\|f(x)\|_{L_1}(n+1).$$

Таким образом,

$$\sqrt{\frac{2}{\pi}}\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,0}(t))\sin(jx)\right\|_C \leqslant$$

$$\sqrt{\frac{2}{\pi}}(C_1\|f(x)\|_{L_1}(n+1)+C_2\|f(x)\|_{L_1}(n+1)+C_3\|f(x)\|_{L_1}(n+1)+C_4\|f(x)\|_{L_1}(n+1)) = \\ = C_1^1\|f(x)\|_{L_1}(n+1),$$

где C_1^1 — константа, не зависящая от n.
1.5 Перейдем ко второму слагаемому правой части (4).

$$\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,1}(t))\sin(jx)\right\|_C =$$

$$= \sqrt{\frac{2}{\pi}}\left\| -\sum_{k=1}^{n}\sum_{j=1}^{k}\frac{\sin(jx)}{2j}\int_0^{\pi} f(t)\cdot\sin(jt)\int_0^{t} u^2(s)\sin(2js)\,ds\,dt - \right.$$

$$- \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{\sin(jx)}{\pi}\int_0^{\pi} f(t)t\cos(jt)\int_0^{\pi} u(s)\sin(2js)\,ds\,dt +$$

$$+ \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{\sin(jx)}{2\pi j}\int_0^{\pi} f(t)t\cos(jt)\int_0^{\pi} u^2(s)\cdot(\cos(2js)-1)\,ds\,dt -$$

$$- \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{2\sin(jx)}{\pi}\int_0^{\pi} f(t)t\cos(jt)\int_0^{\pi}\int_0^{s} u(s)u(\tau)\cos(2js)\sin(2j\tau)d\tau\,ds\,dt +$$

$$+ \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{\sin(jx)}{2j}\int_0^{\pi} f(t)\cos(jt)\int_0^{t} u^2(s)(1-\cos(2js))\,ds\,dt +$$

$$\left. + \sum_{k=1}^{n}\sum_{j=1}^{k}2\sin(jx)\int_0^{\pi} f(t)\cos(jt)\int_0^{t}\int_0^{s} u(s)u(\tau)\cos(2js)\sin(2j\tau)\,d\tau\,ds\,dt\right\|_C.$$

Оценим первой слагаемое правой части. При этом введем функцию $u_2(s)$ такую, что

$$\begin{cases} u_2(s) = u(s), \ s\in[0;t] \\ u_2(s) = 0, \ s\in(t;\pi] \end{cases}$$

и периодически продолжим ее за отрезок $[0;\pi]$.

$$\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}\frac{\sin(jx)}{2j}\int_0^{\pi} f(t)\cdot\sin(jt)\int_0^{t} u^2(s)\sin(2js)\,ds\,dt\right\|_C \leqslant$$

$$\leqslant \left\| f(t)\|_{L_1}\cdot\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}\frac{1}{2j}\cdot u_j\right\|_C = \|f(t)\|_{L_1}\cdot\sum_{k=1}^{n}\left(\sum_{j=1}^{k}\frac{1}{4j^2}\right)^{\frac{1}{2}}\cdot\left(\sum_{j=1}^{k}\|u_j^2\right\|_C^2\right)^{\frac{1}{2}} \leqslant \\ \leqslant \|f(t)\|_{L_1}\cdot C^2$$

где C^2 — константа, не зависящая от n; $u_j = \int_0^\pi u_2^2(s)\cdot\sin(2js)\,ds$. Неравенства верны, так как $\{\|u_j\|\}_{j=1}^\infty \in l_2$.

Второе слагаемое оценивается так же.

Переходим к пятому слагаемому правой части (11) (третье оценивается аналогично). Оно разбивается на два из-за сомножителя $(\cos(2js)-1)$. Первая часть рассматривается аналогично предыдущему выражению. Что касается второй части, введем функцию $F(t) = \int_0^t sf(s)\,ds \in AC$. Для этой функции выполнено следующее: $|F(\pi)| \leqslant \pi\|f(t)\|_{L_1}$, $\quad F(0) = 0$. Тогда:

$$\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^n\sum_{j=1}^k\frac{\sin(jx)}{2\pi j}\int\limits_0^\pi f(t)t\cos(jt)\int\limits_0^\pi u^2(s)\,ds\,dt\right\|_C \leqslant$$
$$\leqslant \left\|C_u\sum_{k=1}^n\sum_{j=1}^k\frac{\sin(jx)}{j}\int\limits_0^\pi \cos(jt)dF(t)\right\|_C \leqslant$$
$$\leqslant \left\|C_1\sum_{k=1}^n\sum_{j=1}^k\frac{\sin(jx)\cos(j\pi)}{j}\cdot F(\pi)\right\|_C + \left\|\frac{1}{2}C_u\sum_{k=1}^n\int\limits_0^\pi F(t)(D_k(x-t)-D_k(x+t))\,dt\right\|_C \leqslant$$
$$\leqslant C_2\cdot n\cdot\|f(x)\|_{L_1} + \left\|\sum_{k=1}^n S_k(x-t)S_k(x+t)\right\|_C \leqslant C\cdot n\cdot\|f(x)\|_{L_1}.$$

Последнее верно, так как частичные суммы ряда Фурье абсолютно-непрерывной функции $F(t)$ ограничены: $\|S_k(x-t)\| \leqslant \sup_{0\leqslant t\leqslant\pi}|F(t)| \leqslant \|f(x)\|_{L_1}$.

Осталось рассмотреть слагаемые с двойными интегралами перед u в правой части (12). Проведем анализ последнего, так как оно является более сложным (второе слагаемое с двойными интегралами рассматривается аналогично). Заметим, что

$$\cos(2js)\cdot\cos(jt)\cdot\sin(2j\tau)\cdot\sin(jx) =$$
$$= \frac{1}{8}(\cos(j(2s-t-2\tau+x)) + \cos(j(2s-t+2\tau-x)) + \cos(j(2s+t-2\tau+x)) + \cos(j(2s+t+2\tau-x)) -$$
$$-\cos(j(2s-t-2\tau-x)) - \cos(j(2s-t+2\tau+x)) - \cos(j(2s+t-2\tau-x)) - \cos(j(2s+t+2\tau+x)).$$

Перейдем от косинусов к ядрам Фейера, как это делалось ранее. Выпишем одно из получившихся слагаемых (остальные семь будут отличаться от него только аргументом ядра Фейера):

$$\sqrt{\frac{2}{\pi}}\cdot n\cdot\left\|\int\limits_0^\pi\int\limits_0^t\int\limits_0^s f(t)u(s)u(\tau)K_n(2s-t-2\tau+x)d\tau\,ds\,dt\right\|_C \leqslant$$
$$\leqslant C\cdot\|f(t)\|_{L_1}\cdot n\cdot\left\|\int\limits_0^\pi\int\limits_0^s u(s)u(\tau)K_n(2s-t-2\tau+x)d\tau\,ds\right\|_C \leqslant$$
$$\leqslant C\cdot\|f(t)\|_{L_1}\cdot n\cdot C_u\cdot\left\|\int\limits_0^\pi u(\tau)K_n(2s-t-2\tau+x)d\tau\right\|_C \leqslant$$
$$\leqslant C_2^1\cdot n\cdot\|f(t)\|_{L_1},$$

где константа C_2^1 не зависит от n. Здесь при необходимости функции $u(s)$, $u(\tau)$ продолжались нулем на $[0;\pi]$, как это делалось выше.

1.6 Последнее слагаемое правой части (4) оценивается легко:

$$\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_{j,0}(t))\sin(jx)\right\|_C \leqslant$$
$$\leqslant \sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}\|\varphi_{j,2}(t)\|_C\cdot\|f(t)\|_{L_1}\leqslant$$
$$\leqslant C_3^1\cdot n\cdot\|f(t)\|_{L_1},$$

в силу асимптотических формул вспомогательной леммы А.

В результате приходим к оценке всего первого слагаемого главного выражения:

$$\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_j(t))\sin(jx)\right\|_C \leqslant C^1\|f(x)\|_{L_1}(n+1)$$

, где C^1 не зависит от n.

2.Переходим ко второму слагаемому основного выражения. Оно имеет вид:

$$\begin{aligned}
&\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\sin(jt))\varphi_j(x)=\\
=\;&\sqrt{\frac{2}{\pi}}\left(\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^{\pi}\int\limits_0^{\pi}f(t)u(s)(1-\frac{s}{\pi})\cos(2js)\cdot\sin(jt)\cdot\sin(jx)\,ds\,dt+\right.\\
+\;&\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^{\pi}\int\limits_0^{\pi}(-\frac{x}{\pi})f(t)u(s)\sin(2js)\cdot\cos(jx)\cdot\sin(jt)\,ds\,dt+\\
+\;&\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^{\pi}\int\limits_0^{x}f(t)u(s)\sin(2js)\cdot\cos(jx)\cdot\sin(jt)\,ds\,dt-\\
-\;&\sum_{k=1}^{n}\sum_{j=1}^{k}\int\limits_0^{\pi}\int\limits_0^{x}f(t)u(s)\cos(2js)\cdot\sin(jx)\cdot\sin(jt)\,ds\,dt+\\
+\;&\left.\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\sin(jt))\varphi_{j,1}(x)+\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\sin(jt))\varphi_{j,2}(x)\right).
\end{aligned}$$

Проделывая те же действия, что в пункте 1, получим, что

$$\left\|\sqrt{\frac{2}{\pi}}\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\sin(jt))\varphi_j(x)\right\|_C \leqslant C^2\|f(x)\|_{L_1}(n+1),$$

где C^2 — постоянная, не зависящая от n.

3. Осталось рассмотреть третье слагаемое основного выражения:

$$\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}(f(t),\varphi_j(t))\varphi_j(x)\right\|_C = \left\|\sum_{k=1}^{n}\sum_{j=1}^{k}\int_0^{\pi} f(t)\cdot\varphi_j(t)\,dt\cdot\varphi_j(x)\right\|_C \leqslant$$
$$\leqslant \sum_{k=1}^{n}\sum_{j=1}^{k}\int_0^{\pi}|f(t)|\,dt\cdot\|\varphi_j(t)\|_C\cdot\|\varphi_j(x)\|_C \leqslant$$
$$\leqslant \sum_{k=1}^{n}\|f(t)\|_{L_1}\cdot\sum_{j=1}^{k}\|\varphi_j(x)\|_C^2 \leqslant$$
$$\leqslant C_u\cdot n\cdot\|f(t)\|_{L_1},$$

где C_u — константа, зависящая от первообразной потенциала u и не зависящая от n.

Вернемся к главному выражению

$$\|\sigma_n(x)-S_n(x)\| \leqslant$$
$$\frac{1}{n+1}(C^1\|f(x)\|_{L_1}(n+1)+C^2\|f(x)\|_{L_1}(n+1)+C_u\cdot n\cdot\|f(t)\|_{L_1}) \leqslant$$
$$\leqslant C\cdot\|f(x)\|_{L_1},$$

где постоянная C зависит от первообразной потенциала u и не зависит от n.

Таким образом, мы доказали, что оператор $B_n : L_1 \to C$, действующий по правилу $B_nf(x)=\sigma_n(x)S_n(x)$, ограничен: $\|B_n(x)f(x)\|_C \leqslant C\cdot\|f(x)\|_{L_1}$. Далее доказательство теоремы завершим стандартным приемом. Рассмотрим действие оператора B_n на собственные функции оператора L $(m<n)$:

$$\|B_n(x)y_m(x)\|_C = \frac{1}{n+1}\left\|\sum_{k=1}^{n}\sum_{j=1}^{k}(y_m(t),y_j(t))y_j(x)-\frac{1}{n+1}\sum_{k=1}^{n}\sum_{j=1}^{k}(y_m(t),\sin(jt))\sin(jx)\right\|_C \leqslant$$
$$\leqslant \left\|\frac{n-m}{n+1}y_m(x)-S_n(x)\right\|_C \to 0 \ \text{ при } \ n\to\infty,$$

здесь $S_n(x)$ — чезаровские частичные суммы ряда Фурье функции $y_m(x)$, причем из того, что $y_m(x)\in AC$, следует равномерная сходимость этих сумм к $y_m(x)$ (см. [16], с. 92).

В силу полноты системы $\{y_j\}_{n=1}^{\infty}$ в L_2 (лемма А) подпространство конечных линейных комбинаций собственных функций плотно в L_2, а следовательно, и в L_1. Значит, $\|B_n(x)f(x)\|_C \to 0$, при $n\to\infty$, $\forall f\in L_1$. Теорема доказана.

Список литературы

[1] А.А. Шкаликов А.М. Савчук. Операторы Штурма-Лиувилля с сингулярными потенциалами. *Математические заметки*, 66:897–912, 1999.

[2] Шкаликов А.А. Савчук А.М. Операторы Штурма-Лиувилля с потенциалами - распределениями. *Труды Московского Мат. Общества*, 64:159–219, 2003.

[3] Шкаликов А.А. Савчук А.М. О собственных значениях оператора Штурма-Лиувилля с потенциалами из пространств Соболева. *Математические заметки*, 80:864–884, 1999.

[4] Савчук А.М. О собственных функциях оператора Штурма-Лиувилля с потенциалами из пространств Соболева. *Математические заметки*, arXiv 1003.3172, 2010.

[5] Dini U. *Fonamenti per la teorica delle funzioni di variabli reali.* Piza, 1878.

[6] В.А. Стеклов. Sur les expressions asymptotics de certaines fonctions definies par des equations differentielles lineares du dexieme ordre, et leurs applications au probleme du developpement d'une fonction arbitraire en nseries procedant suivant les dites fonctions. *Сообщения матем. общества (2)*, 10 (2 - 6):97–199, 1907-1909.

[7] Haar A. Zur teorie der orthogonalen funktionesysteme. *Math. Ann.*, 69:331–371, 1910.

[8] Stone M.H. A comparison of the series of furier and birkhoff. *Trans. Amer. Math. Soc.*, 28:659–761, 1926.

[9] Ильин В.А. Равносходимость с тригонометрическим рядом разложений по корневым функциям одномерного оператора Шредингера с комплексным потенциалом класса l_1. *Дифф. уравнения*, 27 №4:577–597, 1991.

[10] Радзиевский Г.В. Гомилко А.М. Равносходимость рядов по собственным фунцкиям обыкновенных функционально-дифференциальных операторов. *Докл. АН*, 316 №2:265–269, 1991.

[11] Хромов А.П. О равносходимости разложений по собственным фунцкиям дифференциальных операторов второго порядка. *Дифф. уравнения и выч. математика, изд-во Саратовского ун-та*, 5, Ч.2:3–20, 1975.

[12] Minkin A.M. Equiconvergence theorems for differential operators. arXiv:math/0602406vl, 2006.

[13] Садовничий В.А. Винокуров В.А. Равномерная равносходимость ряда Фурье по собственным функциям первой краевой задачи и тригонометрического ряда Фурье. *Докл. АН*, 380 №6:731–735, 2001.

[14] Садовничая И.В. О равносходимости разложений в ряды по собственным функциям операторов Штурма-Лиувилля с потенциалами -распределениями. *Труды МИАН*, 261:249–257, 2008.

[15] Зигмунд А. *Тригонометрические ряды.* Издательство "МИР", Москва, 1965.

[16] Бари Н.К. *Тригонометрические ряды.* Государственное издательство физико-математической литературы, Москва, 1961.

УДК 004.43+372.8

ЯЗЫК СИ И НАЧАЛЬНОЕ ОБУЧЕНИЕ ПРОГРАММИРОВАНИЮ

© 2010 г. А. В. Столяров

avst@cs.msu.ru

Кафедра алгоритмических языков

1 Предисловие

Недавно коллеги задали мне вопрос, как я отношусь к идее замены языка Паскаль на первом курсе языком Си. Когда я категорично ответил «ни в коем случае», меня спросили «а почему?» Предлагаемый текст — мой ответ на этот вопрос.

Мои основные возражения сводятся к следующему. Во-первых, для языка Си высок барьер вхождения: для понимания даже самых простых программ требуются достаточно серьёзные знания, поэтому, если Си изучать как первый язык, начало курса будет состоять из сплошных ссылок вперёд. Во-вторых, язык Си стимулирует применение хаков и формирует соответствующее мышление, причём в отсутствие уже сформировавшейся культуры программирования студенты лишаются возможности отличать случаи оправданного применения хаков от бессмысленного лихачества. В-третьих, в языке Си попросту отсутствует ряд механизмов, необходимых как иллюстрация общей программистской теории. Так, в Си возможен лишь один вид передачи параметров, именно — передача по значению; в Си отсутствуют массивы как полноправный тип данных; модульности в её истинном смысле в Си также нет. Кроме того, в начальном обучении программированию полезно применять строгую типизацию, тогда как в Си символ и его код суть одно и то же, а логические и перечислимые значения являются не более чем константами целого типа.

Сказанное не означает, что язык Си не следует изучать; утверждается лишь, что с Си нельзя *начинать* обучение программированию.

Язык Си, созданный в первой половине 1970-х годов, стал, как известно, подлинным прорывом в области программирования. Внезапно программисты обнаружили, что в некоторых областях, считавшихся до той поры сугубо «ассемблерными», таких как ядра операционных систем, прошивки ЭВМ специального назначения и т. п., автокоды и языки ассемблеров вдруг утратили монопольное положение, а сами операционные системы стало возможно создавать переносимыми с одной аппаратной архитектуры на другую. В настоящее время язык Си, история которого перевалила за 35-летний рубеж, используется для создания широчайшего спектра программ, от прошивок микроконтроллеров до офисных пакетов; его компиляторы доступны для практически всех существующих архитектур процессоров; объём программного кода, написанного на Си за долгую историю его существования, составляет заметную долю всех текстов программ, написанных человечеством. Язык прочно удерживает первое место по популярности, и в особенности это заметно в мире свободно распространяемых программ. Знание языка Си неявно, но от этого не менее категорично считается обязательным для всех профессиональных программистов вне зависимости от их специализации; представляется сомнительным, чтобы программист, не знающий Си, смог вообще найти работу по специальности. Пожалуй, в этом Си уникален: ни про какой другой язык программирования такого сказать нельзя.

Обучение студентов программистских специальностей языку Си, вне всякого сомнения, представляется мне обязательным; не обсуждая этот вопрос (в силу очевидности ответа на него), я хотел бы в предлагаемом эссе уделить внимание вопросу о возможности **начинать** обучение студентов именно с этого языка.

2 На кого ориентироваться

Одной из основных особенностей аудитории, если говорить о первокурсниках, является практически полное отсутствие базовой подготовки по программированию. Информатика в школах чаще всего не преподаётся вообще (по рассказам многих родителей нынешних школьников, чаще всего «уроки» информатики в школе представляют собой 45 минут компьютерных игр), а если преподаётся, то на столь низком уровне, что чудес из этой области можно не ждать. Введение ЕГЭ по информатике проблемы, скорее всего, не решит. Во-первых, преподавать информатику в школах просто некому: люди, одновременно умеющие и программировать, и учить, вообще встречаются чрезвычайно редко, а среди российских школьных учителей не встречаются никогда. Во-вторых, для школьной информатики начисто отсутствует методическая база. Мне пришлось за последние несколько лет рецензировать рукописи, предлагавшиеся к изданию в качестве школьных учебников информатики; судя по этим рукописям, ситуация в этой области плачевная, если не сказать катастрофическая. Относительно **всех** без исключения текстов, попавших ко мне на рецензию, я могу заявить совершенно определённо, что их авторы о предмете программирования не имеют ни малейшего представления. Единственный учебник, который я смог одобрить, отличался от остальных тем, что его автор даже не пытался излагать программирование; текст был посвящён общим вопросам информатики, истории информации, письменности, арифметики, уделялось внимание началам теории кодирования, логики и т. п. но не программированию. Наконец, даже если удастся заставить (например, с помощью ЕГЭ по информатике) школьных учителей исполнять свои обязанности на сколь бы то ни было должном уровне, не следует забывать, например, что динамические структуры данных (и вообще указатели) в школьный курс информатики не входят, считаясь неким высшим пилотажем.

Несомненно, среди студентов, поступающих на первый курс, имеется небольшой, но достаточно стабильный процент тех, кто, невзирая на всё вышеизложенное, программировать уже умеет. Я сам, поступив на факультет в 1992 году, имел к тому времени опыт получения денег за написанные мной программы, и в те времена это не было редкостью: только в одной со мной группе училось не менее четырёх человек, ни в чём не уступавших мне в области программирования. Сейчас экономическая ситуация не столь экстремальна, как в начале девяностых, так что количество состоявшихся профессионалов среди поступающих на факультет стало меньше, но они, тем не менее, есть. Однако следует ли ориентироваться на них при составлении программ программистских курсов? Прежде чем дать ответ на этот вопрос, отмечу одно обстоятельство, неизменно ускользающее от внимания публики. До поступления на факультет этих людей учить программированию было некому, и, следовательно, *если вчерашний абитуриент умеет программировать — это прежде всего означает, что программированию он может учиться самостоятельно, без всякой помощи со стороны преподавателей.* Иначе говоря, такой студент не нуждается в том, чтобы его «учили программировать» — рассказывали синтаксис языков программирования, показывали всяческие алгоритмы, приёмы и техники. Как ни странно, чаще такие студенты нуждаются в совершенно ином, а именно — чтобы кто-нибудь убедил их, что их без пяти минут профессиональные навыки — это ещё не всё, что есть в мире, и что умение писать программы как таковое ещё не делает человека программистом. Однако мир, к сожалению, не идеален, и у некоторых преподавателей попросту не хватает квалификации (не преподавательской, а профессионально-программистской), чтобы поддерживать на должном уровне свой авторитет среди студентов рассматриваемой категории. Лучшее, что можно сделать в таком случае — это не мешать студенту учиться самостоятельно и не прививать отвращение к предмету.

Итак, среди первокурсников есть прослойка тех, кто умеет программировать (в той или иной степени), но эта прослойка, во-первых, малочисленна, и, во-вторых, как раз эти студенты в услугах преподавателей (по программистским предметам) нуждаются меньше всего. Если в своей работе мы будем ориентироваться на них, то от нашей деятельности окажется на удивление мало пользы: *эти* студенты без нашей заботы вполне обошлись бы, а остальные на занятиях ничего не поймут (то есть вообще) и нам придётся смириться с тем, что подавляющему большинству студентов по программистским предметам ставятся безобразно натянутые «тройки», просто чтобы не выгонять почти весь курс. В этот производственный брак попадут и те студенты, кто потенциально мог бы прекрасно программировать, просто им не повезло со школой и окружением.

Приходится сделать неизбежный, хотя, возможно, и не очень приятный вывод: расчитывать следует на *средний* уровень студентов, а это, в частности, означает, что программировать наша целевая аудитория не умеет вообще ни в каком виде и ни до какой степени, так что начинать обучение программированию приходится **действительно с нуля**, и эта ситуация в обозримом будущем никак не изменится.

3 «Hello, world», или барьер, который возьмут не все

Как дорога начинается с первого шага, так и обучение начинается с первого занятия. Если начинать рассказ о языке программирования с долгих перечислений операторов, операций, типов, переменных и прочих составляющих языка, слушатели эту информацию не воспримут, поскольку не будут понимать, о чём вообще идёт речь. Чтобы это стало понятно, систематическое изложение необходимо предварить неким введением в предмет, в ходе которого показать, как вообще выглядят программы на том языке программирования, который мы начинаем рассматривать.

По традиции, введённой авторами Си Брайаном Керниганом и Денисом Ритчи, обучение этому языку начинают с программы «Hello, world». Вот её текст[1]:

```
#include <stdio.h>
int main() {
    printf("Hello, world\n");
    return 0;
}
```

Написать этот текст на доске не сложно, сложности начинаются в тот момент, когда нарисована последняя фигурная скобка и необходимо переходить к пояснениям. Что это там в первой строке? Правильный ответ звучит примерно так: это директива макропроцессора, которая включает в нашу трансляцию специальный «заголовочный» файл, содержащий объявления библиотечных функций. А теперь давайте вспомним, что мы проводим *первое* занятие для абсолютно неподготовленной аудитории. О смысле слова «директива» наши слушатели ещё могут догадаться из общих соображений и окажутся недалеки от истины, но остаток фразы — макропроцессор, функции (к тому же ещё «библиотечные»), объявления... все эти слова не оставляют слушателям ни единого шанса на понимание. Если в этот момент пуститься в пространные объяснения, мы можем где-то через полчаса с удивлением обнаружить, что увлечённо рассказываем, например, о конвенциях вызовов функций, в то время как наши слушатели читают книжки, играют в тетрис на мобильных телефонах, рисуют в тетрадках цветочки и занимаются другими делами, имеющими столь же прямое отношение к программированию. Реальность такова, что объяснить неподготовленному слушателю смысл директивы `include` **невозможно**, нравится это нам или нет. Поэтому приходится произнести сакраментальное «так надо, а зачем — мы узнаем позже». Заметим, мы ещё ничего не объяснили, а ссылка вперёд нам уже потребовалась.

Дальше — больше. Что такое `int`? Что такое `main`? Правильный ответ таков: `main` — это имя главной функции нашей программы, а `int` — тип значения, которое она возвращает. Конечно же, это очень просто, но только не для человека, который вообще не умеет программировать. Слово «функция» он раньше встречал только в математическом смысле, ну и ещё в общезначимом (функция утюга — гладить); к нашему случаю не подходит ни то, ни другое. Термин «возврат значения» слушателю и вовсе представляется странно звучащей бессмыслицей, а «тип значения» окончательно добьёт тех, кому до сей поры ещё казалось, что они что-то понимают. Всё, 90% аудитории потеряно, и с нами остались только те, кто уже программировал (хотя бы на каком-нибудь языке, где были типы и функции).

[1]На всякий случай отмечу, что, согласно современным представлениям, функция `main` **не может** иметь никакой тип, кроме `int`, в том числе не может она и быть `void`'овой, ну а в функции, возвращающей значение, присутствие оператора `return` обязательно. Программа, в которой опущено слово `int` перед `main` и оператор `return`, откомпилируется с двумя предупреждениями, что, конечно же, недопустимо в учебных примерах.

Частично спасти положение можно ещё одной сакраментальной фразой: «если это непонятно, ничего страшного, скоро всё станет понятно». Внимание некоторой части аудитории это нам вернёт. Правда, ненадолго, но пояснить следующую строку (к счастью, это единственная более-менее понятная строка в программе) мы успеем. Итак, следующая строка печатает фразу «Hello, world» на экране. У этого утверждения есть одно достоинство: оно понятно. Вместе с тем, оно под завязку наполнено фактическими ошибками и недоговорками, если же попытаться их устранить, у нас опять получится целый ряд ссылок вперёд. Чего стоит одна только комбинация «`\n`», обозначающая символ перевода строки. Большинство нынешних студентов никогда не видели пишущей машинки, во всяком случае, в действии, да и матричных принтеров тоже уже не застали, а с компьютером работали только в графическом режиме, так что вот это вот понятие «символ перевода строки» им придётся долго разъяснять. Ну то есть что такое «перевод строки», наверное, поймут все, но как ЭТО может быть символом?! Конечно, всё это станет понятно после первых же двух-трёх часов работы с эмулятором алфавитно-цифрового терминала, но не забывайте, что сейчас мы проводим **первое** занятие, и знакомство с терминалом у студентов ещё впереди.

Разъяснив с грехом пополам строчку, в которой вызывается `printf`, мы будем вынуждены перейти к следующей строке, в которой опять возникнет пресловутый «возврат значения из функции». Вне всякого сомнения, это очень простое и очень важное для программирования понятие, но объяснять его на **этом** примере — идея вряд ли удачная. Даже если мы как-то объясним, что же такое «функция» в терминах языка Си, слушатель вряд ли поймёт, какое это имеет отношение к программе, печатающей строку; ну а загадочный ноль в качестве параметра потребует экскурса в управление процессами, чтобы объяснить, что это такое за загадочный зверь по имени «код завершения задачи».

Посмотрим теперь, как всё то же самое выглядит на Паскале. Текст программы будет таким:

```
program hello;
begin
    WriteLn('Hello, world');
end.
```

При её объяснении не потребуется никаких особо сложных концепций, никаких ссылок вперёд. Слова «начало» и «конец» достаточно перевести с английского и сказать слушателям, что таковы правила языка Паскаль; мы нимало не погрешим здесь против истины, поскольку и заголовок программы, и то, что все её операторы следует поставить между словами `begin` и `end` — это именно что синтаксические соглашения, ничем особенным не обусловленные и не являющиеся частным случаем каких-либо более общих (и, значит, более сложных) правил. Можно даже не уточнять, что `begin` и `end` — это пресловутые «операторные скобки», потому что в данном случае они выступают не в этой роли. Более того, процедура `WriteLn` переходит на печати на следующую строку после распечатки своих аргументов (это слушатели поймут), и страшноватая для неподготовленного уха концепция *символа* перевода строки оказывается за кадром — до наступления удобного момента.

Но как же быть с Си, ведь все его «сложности» никуда не денутся, а рассказывать его студентам рано или поздно придётся? Так-то оно так, но если ту же самую программу «Hello, world» показать и объяснить студенту, уже занимавшемуся программированием (на том же Паскале или на чём-то ещё), мы сможем избежать практически всех трудностей, перечисленных выше. С понятием «макропроцессор» студент уже знаком, а если не знаком — то можно ему сказать (не упоминая макропроцессор), что директива `#include` подставляет вместо себя содержимое файла, в данном случае это файл `stdio.h`, прилагающийся к стандартной библиотеке, он тоже написан на Си и содержит информацию для компилятора о том, какие библиотечные функции есть в этой библиотеке. Понятие библиотеки студенту уже известно, понятие функции в программистском смысле — тоже (попутно уточняем, что в Си, в отличие от Паскаля, есть только функции, а процедур нет — и нас понимают). Слово `int` обозначает целочисленный тип выражений — и нас снова понимают, потому что со всеми этими терминами уже не раз встречались. Слово `main` — это имя функции, попутно уточняем, что главной программы, в отличие от Паскаля, здесь нет, программа вся состоит из функций, просто есть соглашение, что одна из них — а именно вот эта вот `main` — вызывается при старте программы. Никаких проблем не вызывает и `return`, и замечание, что `printf` представляет собой библиотечную функцию (попутно уточняем, что в Паскале `WriteLn` есть часть языка, тогда как здесь мы можем `printf` написать сами). Да и с символом перевода строки наш подготовленный слушатель, вне всякого сомнения, уже сталкивался. Совсем, как говорится, другой коленкор — а всё потому, что на этот раз Си для наших слушателей не первый в жизни язык программирования.

4 Второй пример программы: барьер, который не возьмёт никто

Обычно программы не ограничиваются выводом информации: в общем случае, да и в большинстве частных, алгоритм (и программа, как запись алгоритма) представляет собой преобразователь исходных данных в данные, составляющие результат. Вполне естественно поэтому, что уже вторая программа, предъявляемая слушателям в качестве примера, будет содержать ввод данных извне.

Совершенно не сговариваясь между собой, многие авторы пособий в качестве такого второго примера используют решение квадратного уравнения. Пример действительно удачный, он позволяет продемонстрировать переменные разных типов, арифметику, ветвление; вычисление дискриминанта можно вынести в отдельную функцию, заодно показав, как на Си пишутся функции и как из них возвращать значение (после этого слушателям становится проще понять, что собой в действительности представляет `main`). Наконец, необходимость использования функции `sqrt` заставляет нас сначала подключить второй заголовочный файл (в этот раз `math.h`), пояснив, что прототип функции `sqrt` находится именно в нём; затем приходится отметить, что функция сама по себе не входит в стандартную библиотеку, так что при сборке программы необходимо эту библиотеку подключить, используя параметры командной строки компилятора. Впрочем, многие из этих пояснений можно и опустить без особого ущерба для изложения; речь же сейчас пойдёт о другом.

Чтобы *получить исходные данные* (в рассматриваемом случае — коэффициенты уравнения) извне, нам в языке Си придётся воспользоваться функцией `scanf`. Квадратное уравнение тут ничем особенным не примечательно, ввод неких числовых данных «с клавиатуры» (то есть из стандартного потока ввода) предполагается большинством учебных заданий для начинающих. И здесь мы сталкиваемся с отсутствием в Си передачи параметров иначе как по значению. Описываем переменные (тут всё в порядке), выводим приглашение к вводу (тут тоже всё понятно, с `printf`'ом студенты уже знакомы), начинаем писать следующую строку... и в процессе её написания понимаем, что теперь нам предстоит объяснить аудитории нечто, для понимания чего слушателям не просто не хватает знаний — им не хватает примерно целого семестра этих знаний. А строка выглядит так:

```
scanf("%lf %lf %lf", &a, &b, &c);
```

Объяснить концепцию «форматной строки» ещё возможно, хотя нельзя не отметить, что даже на втором курсе далеко не все студенты с первого раза понимают, о чём идёт речь. Но вот зачем перед именами переменных стоит знак «`&`», неподготовленный слушатель не поймёт, что бы вы ни делали и к каким хитростям ни прибегали. Концепция «указателей» и адресных выражений кажется простой лишь тем, кто с ней давно и прочно знаком. Каждый, кто хотя бы раз объяснял работу с динамической памятью людям, имеющим нулевой уровень подготовки, знает, что это на самом деле сложнейшая педагогическая задача — как всегда бывает в случаях, когда необходимо сформировать в мышлении ученика новый уровень абстракции, которого там раньше не было. Можно, конечно, произнести несколько раз что-то вроде «у каждой переменной есть то место в памяти, где она располагается, и это место обозначается адресом; чтобы функция `scanf` знала, куда занести значения, которые она прочитает, мы передаём ей адреса наших переменных, и вот знак амперсанда как раз и есть операция взятия адреса» и тешить себя надеждой, что мы всё понятно объяснили, да только это будет не более чем самообман. Понять нас в данном случае есть шанс лишь у тех, кто уже сталкивался с указателями, причём даже среди этих (весьма немногочисленных) первокурсников нас поймут не все.

Остаётся лишь поднять белый флаг и сказать уже порядком поднадоевшее и нам, и студентам «так надо, а почему — поймёте потом». Студенты, разумеется, просто *запомнят* (читай — зазубрят), что перед переменными в `scanf` надо ставить знак «`&`», и будут очень удивлены, обнаружив, что это правило почему-то не работает при вводе текстовых строк с использованием `%s`. О массивах и адресной арифметике разговор ещё впереди, пока же отметим, что Паскаль позволяет не упоминать о явной работе с адресами до тех пор, пока мы не решим (сами, в силу поставленной перед собой учебной задачи) перейти к изложению динамических структур данных, что, в свою очередь, делается ближе к концу первого семестра, когда студенты в основной своей массе к этому уже более-менее готовы. С другой стороны, при изложении языка Си студентам второго курса можно опираться на уже имеющуюся у них понятийную базу, припомнить `var`-параметры Паскаля, пояснить, что здесь их нет и приходится поэтому прибегать к передаче адресов. Поскольку и с указателями, и с возвратом значений через параметры ко второму курсу студенты уже знакомы, проблемы с пониманием если и возникают, то не столь фатальные.

5 Хак как стиль мышления

Возможно, кто-то из моих коллег, которым адресовано это эссе, сочтёт, что сложности первого занятия так или иначе можно преодолеть; я соглашусь, что это действительно так, ведь существуют даже школы с углублённым изучением информатики, где каким-то образом ухитряются вбить Си в головы ученикам старших классов. Конечно, барьер минимального понимания тут оказывается очень высок, потому что для изучающего программирование на примере Си не понять указатели означает не понять ничего (то есть вообще ничего), а указатели понимает, увы, даже не всякий среди поступивших на ВМК; такие студенты часто становятся жертвами зачётных комиссий третьего и четвёртого семестра, но некоторые всё же выживают и продолжают «обучение», чтобы потом никогда даже не помышлять о работе по специальности. В этом плане основы программирования, излагаемые в рамках школьной информатики, играют не столь критическую роль: далеко не каждый школьник собирается становиться программистом, и не понять ничего в отдельно взятом школьном предмете — это не так уж страшно; в профильном ВУЗе ситуация совершенно иная.

Так или иначе, допустим, что трудности, описанные в двух предыдущих параграфах, благополучно преодолены. Это никак не отменяет того факта, что уровень подготовки бóльшей части аудитории остался нулевым и, следовательно, именно материал первого семестра закладывает у студентов основы стиля программирования и программистского мышления. И в этом плане язык Си предлагает широчайший ассортимент средств, прекрасно пригодных, чтобы мышление ученика раз и навсегда искалечить.

В первой половине девяностых появился шуточный текст, написанный якобы от лица Кена Томпсона, в котором говорится, что Си и Unix — это просто первоапрельская шутка, и что их создатели (Брайан Керниган, Деннис Ритчи и Кен Томпсон) никак не ожидали, что им удастся столь качественно разыграть всё мировое программистское сообщество. В частности, в этом тексте есть такие строки:

> Когда мы обнаружили, что другие действительно пытаются писать программы на A, мы быстро добавили еще парочку хитрых примочек, создав B, BCPL, и, наконец, Си. Мы остановились, добившись успешной компиляции следующего:
>
> ```
> for(;P("\n"),R--;P("\""))for(e=C;e--;P("_"+(*u++/8)%2))P("\" "+(*u/4)%2);
> ```
>
> Мы не могли даже представить, что современные программисты будут пытаться использовать язык, допускающий подобный оператор!

Конечно, это всего лишь шутка; известно, что на самом деле язык B был наследником языка BCPL, а не наоборот, и именно из него получился Си, а язык A в этой истории вообще никак не замешан. Однако, как говорится, в каждой шутке есть доля шутки, а всё остальное — правда. Так, например, многие приверженцы языка Си убеждены, что конструкция

```
while(*p++=*q++);
```

представляет собой образец подлинного изящества. В условное выражение цикла `while` здесь вставлено три побочных эффекта (сдвиг одного указателя, сдвиг второго указателя, присваивание), тело цикла в результате оказалось пустым, поскольку вся функциональность помещена в условное выражение, ну а сама по себе эта конструкция предназначена для копирования строки.

В качестве другого образчика программистского мастерства приведу следующее:

```
while(~wait(NULL));
```

Поясню, что системный вызов `wait`, предназначенный, чтобы дождаться завершения порождённого процесса, возвращает (как и практически все системные вызовы ОС Unix) значение `-1` как индикатор ошибки. Двоичное представление числа `-1` представляет собой цепочку из двоичных единиц во всех разрядах регистра или переменной; вспомнив, что в Си числом `0` обозначается логическая ложь, а любое другое число считается обозначением логической истины, мы сможем догадаться, что условное выражение, использующее побитовую инверсию (символ «~»), будет ложным в случае, если вызов вернёт ошибку, и истинным во всех остальных случаях. Приведённая конструкция, таким образом, означает «выполнять `wait`, пока он не вернёт ошибку», и предназначена, чтобы дождаться завершения всех имеющихся в настоящий момент порождённых процессов. Аналогичным образом часто поступают и с другими вызовами, используя побитовую инверсию, чтобы `-1` превратить в логическую ложь; конструкция вроде

```
while(wait(NULL) != -1) {}
```

делающая абсолютно то же самое, видимо, кажется истинным почитателям Си слишком банальной.

Конечно, никто не заставляет учить студентов таким выкрутасам; более того, студентов можно и нужно предостеречь от использования подобных трюков. Давайте, однако, рассмотрим другой случай. Существует целый класс программ, называемых «фильтрами»; эти программы принимают некий текст из стандартного потока ввода, производят над ним те или иные преобразования и полученный результат выдают в стандартный поток вывода; как правило, они продолжают работать, пока в стандартном потоке ввода не возникнет ситуация «конец файла». Поскольку многие задачи, связанные с разбором и преобразованием текста, можно решить с помощью регулярных автоматов или вариаций на тему оных, чаще всего такая программа на верхнем уровне управления представляет собой цикл посимвольного чтения, выполняемого до возникновения ситуации «конец файла». Записывается такой цикл обычно следующим образом:

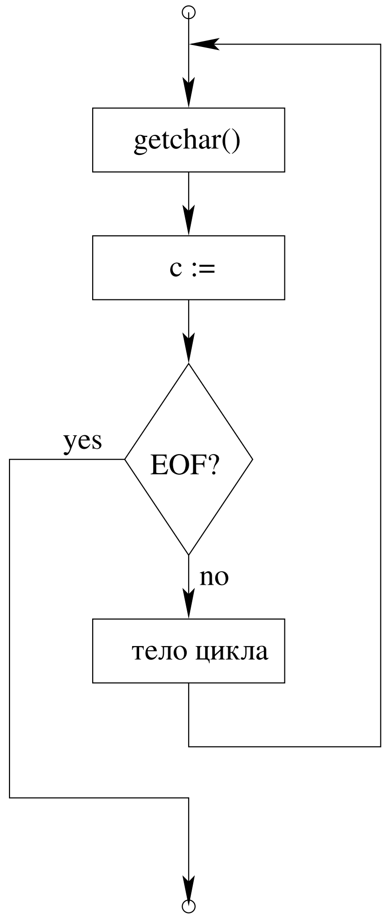

Рис. 1: Блоксхема традиционного цикла с `getchar`

```
while((c = getchar()) != EOF) {
    /* тело цикла */
}
```

Заголовок цикла `while`, построенный таким способом, можно встретить практически в любом учебнике по Си. Профессиональным программистам эта конструкция знакома и привычна, и никто даже не замечает, что она тоже представляет собою ни что иное, как хак. Чтобы убедиться в этом, рассмотрим блок-схему получившегося цикла (см. рис. 1). Что это, простите? Цикл с предусловием или с постусловием? Или структурное программирование уже отменили?

Итак, мы имеем ещё один хак. Однако, в отличие от двух предыдущих примеров, представляющих собой не более чем бессмысленное лихачество, в этом случае не всё так просто. Чтобы привести этот цикл к каноническому виду, нам придётся применить в тексте два совершенно одинаковых оператора:

```
c = getchar();
while(c != EOF) {
    /* тело цикла */
    c = getchar();
}
```

С концептуальной точки зрения этот текст, конечно, чище, но любого профессионального программиста он заставит недовольно скривиться, и есть от чего: ведь в Си присваивание сделали операцией, а не оператором, специально для таких случаев, а практическое программирование, как это часто бывает, диктует свои условия и не желает приносить лаконичность кода в жертву пусть и красивым, но всё-таки теоретическим канонам. Надо отметить, что я и сам использую в своих программах побочные эффекты в заголовках операторов, и в примерах, приводимых мной на лекциях и семинарах, встречаются такие приёмы. Более того, ещё несколько лет назад я вообще не обращал на это никакого внимания, воспринимая побочные эффекты в заголовках как нечто само собой разумеющееся, пока не заметил, что некоторые студенты из числа слабых не могут понять объяснения, например, управления процессами в ОС Unix именно потому, что в примерах используется что-то вроде

```
if(fork() == 0) {
    /* ... */
```

Применение вместо этого более подробного фрагмента, а именно

```
int pid;
pid = fork();
if(pid == 0) {
    /* ... */
```

к моему немалому удивлению **уменьшает в группе долю студентов, не понявших, что такое `fork()`**. А натолкнуло меня на мысль о трудностях понимания `if(fork()==0)` то обстоятельство, что на зачётных комиссиях и экзаменах в третьем семестре я несколько раз обнаруживал, например, такое:

```
fork();
if(fork() == 0) {
    /* ... */
```

Очевидно, что ошибка такого рода свидетельствует о том, что студент просто не владеет концепцией побочного эффекта.

Сделав для себя определённые методологические выводы, я теперь при объяснениях нового материала никогда не применяю побочные эффекты в заголовках `if`, `while` и других операторов; тем не менее, в дальнейших примерах я возвращаюсь к использованию побочных эффектов, делая на первых порах замечание, что это просто сокращённая и более элегантная запись того же самого. Полностью отказаться от использования этой техники мне представляется неправильным, поскольку это противоречит сложившейся культуре программирования на языке Си и может сформировать неправильное отношение к возможностям этого языка. Более того, я убеждён, что пытаться оградить студентов от «вредных» знаний — занятие, заведомо обречённое на неудачу, ведь соответствующие хаки встретятся им в литературе, в Интернете, в примерах, приводимых другими преподавателями, а иногда и другими студентами. Применение трюков в некоторых случаях есть часть, как я уже сказал, культуры, сложившейся вокруг языка Си, и при изучении этого языка отречься от всей этой культуры всё равно не получится.

Однако не следует забывать, что это делается на втором курсе, когда за плечами у аудитории уже имеется Паскаль и язык ассемблера. Это означает, во-первых, что большинство студентов к этому времени всё-таки чувствует на интуитивном уровне, что такое «побочный эффект» и как с этим работать. Во-вторых, и это даже важнее, **благодаря имеющемуся опыту программирования на Паскале студенты имеют возможность отличить трюк грязный и ненужный**, вроде вышеописанной побитовой инверсии, **от трюка, применение которого оправдано**. Чтобы уметь различать, что хорошо, а что плохо, нужно иметь хотя бы минимальный опыт программирования.

6 Инструменты, которых нет

В контексте фундаментального университетского образования овладение конкретным языком программирования нельзя рассматривать как основную цель обучения; языки и другие инструменты меняются со временем, устаревают, к тому же далеко не все студенты, обучающиеся на «программистских» специальностях, в будущем становятся именно программистами, большинство же тех, кто становится, способны (как это уже отмечалось выше) освоить практически любой язык программирования самостятельно. Целью университетских курсов должно быть овладение общими принципами, лежащими в основе дисциплины, и в этом плане конкретный язык программирования, конкретная операционная среда и т. п. — не более чем иллюстративный материал.

Основной лекционный программистский курс третьего семестра ВМК посвящён операционным системам, а на семинарах в его поддержку рассматривается программирование на уровне системных вызовов; язык Си представляется вполне подходящим иллюстративным примером в этой ситуации. Однако подходит ли Си в качестве учебного пособия для *начального* обучения? Выше уже приведён ряд причин, почему на этот вопрос следует ответить отрицательно. Добавим к ним ещё одну: в Си попросту *отсутствуют* некоторые базовые концепции, которые явно заслуживают того, чтобы быть проиллюстрированными.

Начнём с того, что в Си присутствует один, и только один способ передачи параметров в подпрограммы — а именно, передача по значению. Передачу по ссылке приходится имитировать с помощью явной передачи адресов. Проблема тут не только в том, что это сложно для понимания неподготовленной аудиторией, но и в том, что у студента сложится впечатление (неправильное), что в языках программирования только и бывает передача по значению. Паскаль в этом плане подходит гораздо лучше, ведь в нём есть и `var`-параметры. Конечно, всего многообразия способов передачи параметров это не исчерпывает, ведь бывает ещё и передача по имени, и ленивые вычисления, предполагающие передачу невычисленного выражения. Но наличие хотя бы двух, а не одного способа передачи параметров демонстрирует обучаемому очень важный факт: **существует больше одного способа передачи параметров в подпрограммы**. Зная два таких способа, человек, как правило, на уровне подсознания оказывается лучше подготовлен к восприятию третьего и последующих способов.

Возражение, что-де `var`-параметры Паскаля всё равно реализуются именно через передачу адреса, тут несколько неуместны: речь ведь идёт об изобразительных средствах языков программирования, а не о том, как эти средства реализуются. Программисты, использующие языки очень высокого уровня, такие как Лисп, Пролог, Haskell и тому подобные, часто вовсе не задумываются о том, как же этот вычислитель реализован.

В качестве второго важного объекта, отсутствующего в Си, я назову обыкновенные массивы. Утверждение, что **в языке Си нет массивов**, часто вызывает удивление — но ведь при описании массива у нас не возникает никакого имени, которое представляло бы собою весь этот массив целиком, как единый объект; имя массива в действительности представляет собой константу адресного типа, указывающую на первый элемент массива, а вовсе не массив как таковой. Над массивами (как таковыми) в Си нет никаких операций — собственно говоря, потому, что такие операции было бы не над чем проводить. Пикантность ситуации ещё и в том, что в Си присутствуют не только средства описания массива, но даже тип «указатель на массив»[2], однако самих массивов как не было, так и нет. Более того, операция индексирования в Си представляет собой не более чем арифметическую операцию над адресом и числом, и для её существования сами массивы оказываются не обязательны.

[2] Выражения такого типа возникают при описании многомерных массивов.

Здесь можно возразить, что используемая в Си для работы с массивами адресная арифметика, вообще говоря, мощнее, чем массивы того же Паскаля; так, в Си можно создавать массивы динамически изменяемого размера, к тому же с любой частью любого массива можно работать как с массивом (например, если есть массив из 20 элементов, то можно рассматривать его элементы с 3-го по 12-й как массив из 10 элементов, причём для этого достаточно одной операции сложения). Всё это так, но я ведь и не утверждаю, что Си хуже Паскаля в качестве профессионального инструмента, я и сам часто пишу на Си, тогда как на Паскале последний раз всерьёз что-то делал лет пятнадцать назад. Речь идёт не о том, какой из языков лучше использовать для написания программ, и даже не о том, надо или не надо изучать Си в рамках основных курсов (я сразу отметил, что для меня необходимость этого очевидна), а о том, пригоден ли Си для начального обучения.

Массив не просто как область памяти, в которой располагаются подряд N переменных одного типа, но как отдельный полноценный *тип переменной* — это концепция, несомненно достойная того, чтобы донести её до студентов. Чтобы развеять сомнения в этом, я хотел бы подчеркнуть, что отсутствие массива как полноценного типа переменной нарушает в Си концептуальную целостность системы типов, ведь *указатель на массив* (именно на массив, а не на его элемент) в языке в качестве отдельной сущности присутствует:

```
int (*p)[20];
```

есть сущность совершенно иная, нежели

```
int *q;
```

Объяснить, что за странная штука «указатель на массив», проще человеку, знакомому с Паскалем, нежели тем, кто начал сразу с Си, потому что у последних оборот «p указывает на массив» прочно ассоциируется с указанием именно что на первый элемент массива; дело осложняется ещё и тем, что адрес-то в обоих случаях используется один и тот же, иначе говоря, численное значение указателя на массив (например, на строку двумерной матрицы) такое же, как и указателя на её же первый элемент. Таким образом, отсутствие в Си массивов как полноценного типа переменных осложняет объяснение даже самого языка Си, не говоря о других языках; у тех учащихся, для кого странные массивы Си стали первыми массивами в жизни, позже часто наблюдаются трудности с пониманием, что представляет собой объект-вектор, который можно ввести в Си++.

Мне как-то раз довелось общаться с профессиональным программистом (то есть человеком, работающим по этой специальности и получающим за это деньги), который всерьёз утверждал, что понятия «массив» и «адрес массива» тождественны. Его не могли убедить в обратном никакие доводы, включая и тривиальное соображение о том, что адрес массива занимает 4 байта, тогда как сам массив — в общем случае сколько угодно. Сущность «указатель на массив» он, как и следовало ожидать, не понимал, а на вопрос, какой же тип, по его мнению, будет иметь имя, описанное как двумерный массив, он ответить не смог, но в своей уверенности так и не поколебался.

Продолжая тему типизации, отмечу, что отсутствие строгой типизации в Си тоже не слишком способствует пониманию общезначимых программистских концепций. Всё-таки символ и его ASCII-код — *это не одно и то же*, хотя бы потому, что операция сложения двух однобайтовых чисел вполне осмысленна, тогда как операция сложения двух букв (если, конечно, это не конкатенация в строковом смысле, но в Си такого нет) — бессмысленна. Да и перечислимый тип изначально всё-таки не то же самое, что набор целочисленных констант, ну а отсутствие отдельного булевского типа оказалось, как мы знаем, настолько неудобно, что в Си++ тип `bool` ввели, хотя и не сразу. По правде говоря, Паскаль тоже не умеет запрещать складывать килограммы с километрами, но имеющиеся в нём зачатки строгой типизации подготавливают слушателя к встрече с действительно строгими типами в Аде и, например, доменами в базах данных.

Наконец, в Си нет полноценной модульности, а то, что есть вместо неё, требует использования нетривиальной техники, основанной на директивах условной компиляции[3]. Здесь можно возразить, что в «стандартном Паскале» модульности тоже нет, но ведь в реальной жизни

[3]Имеются в виду защитные макросы заголовочных файлов.

практически все реализации Паскаля поддерживают вполне полноценные модули. Отсутствие знакомства с правильной системой модульного программирования заставляет студентов поверить, что техника раздельной компиляции исходников на языке Си, включающая заголовочные файлы и буквально *состоящая* из хаков — это и есть модульность, но ведь на самом деле это не так.

7 Заключение

Итак, заменять Паскаль на Си в начальном обучении, с моей точки зрения, недопустимо; позволю себе в заключение высказать ряд соображений относительно того, какие изменения в программистских предметах на первом курсе не только допустимы, но и желательны.

Прежде всего следует, видимо, отказаться от использования мёртвых операционных сред и систем программирования. Язык Паскаль сам по себе к этой категории не относится: существует несколько поддерживаемых кроссплатформенных реализаций оного, имеющих профессиональное качество. Однако с популярной в наших школах и ВУЗах средой Turbo Pascal ситуация совершенно иная, этот инструмент мёртв давно, прочно и безнадёжно. Больше того, обучение обычно проводят на программах для MSDOS, которая сама по себе также мертва. Почему учеников заставляют писать именно под MSDOS, понятно — ведь писать полноценные программы под Windows достаточно сложно, и задача научиться этому — явно не для начального обучения. Однако отказаться от Windows преподавателям мешает иррациональный страх перед другими платформами. В итоге ученикам предлагается писать, в сущности, не программы под Windows, а такие программы, которые *худо-бедно запускаются под Windows*; абсурдность ситуации можно, например, проиллюстрировать возникающей как у обучаемых, так и у некоторых преподавателей привычкой любую программу обязательно заканчивать оператором `ReadLn`. Делается это, естественно, чтобы не закрывалось окошко и можно было прочитать результаты, но зачастую этот факт забывается, настолько привычным оказывается этот вот оператор в конце; какое, спрашивается, отношение это имеет к обучению программированию? Полагаю, что никакого.

Можно, конечно, сказать, что мы и не ставим перед собой цели прямо на первом курсе сделать из студентов профессиональных програмистов, так что, казалось бы, «неполноценность» консольных приложений под Windows, а равно и «мёртвость» MSDOS не должны быть препятствием учебному процессу. К сожалению, утверждая так, мы не учитываем, что студенты — живые люди и их личное отношение к предмету (на эмоциональном уровне) способно весьма и весьма существенно повлиять на усвоение материала и успеваемость. Использование устаревших или в том или ином смысле «ненастоящих» инструментов, нравится нам это или нет, расхолаживает студентов и формирует отношение к программистским предметам как к бесполезной трате времени.

Итак, первая конструктивная рекомендация — отказ от устаревших систем и платформ в пользу современной ОС семейства Unix и одной из имеющихся в мире Unix свободных реализаций Паскаля (Free Pascal или GNU Pascal). В качестве второй рекомендации я бы назвал постепенное вытеснение из учебного процесса «программирования на бумажке», которое, с одной стороны, раздражает студентов, умеющих программировать, а с другой стороны, даёт возможность «аутсайдерам», что-то списав, где-то выехав на неразборчивости почерка и т. п., получить зачёт, так и не написав за свою жизнь ни одной программы. В итоге мы на втором курсе обнаруживаем студентов, не понимающих, что такое (например) вложенные циклы.

За счёт отказа от проведения письменных работ по программированию можно, очевидно, больше внимания и учебного времени уделить непосредственно работе с машиной. К сожалению, тут тоже не всё так просто. Многие преподаватели тщательно тестируют программы, которые студенты сдают в качестве выполненного задания, и при этом полностью забывают, что программа сама по себе — не самоцель, целью являются полученные студентом навыки, которые как раз и не проверяются. Задания по некоторым дисциплинам вообще принимаются по электронной почте или иным способом, не предполагающим личного общения со студентом. Неудивительно, что студенты старших курсов создали настоящую индустрию написания заданий для младшекурсников за денежное вознаграждение.

Установить, является ли студент автором сдаваемой программы, достаточно просто; проблема, очевидно, в личном отношении преподавателей к практике сдачи заданий, выполненных кем-то посторонним. Важно понимать, что *любые разговоры о том, что-де студент якобы разобрался в чужой программе, суть не более чем самообман*, и самообман опасный. Всерьёз разобраться в чужом коде многократно трудней, чем написать его самому. Чтобы убедиться в нулевой ценности слов «я разобрался», достаточно предложить студенту прямо в присутствии преподавателя применить полученные навыки, написав (вот прямо здесь и сейчас) новый фрагмент кода или простую программу целиком.

Выход из создавшейся ситуации мне видится в применении двух основных принципов. Во-первых, необходимо решить для себя, что чужие или «не совсем свои» программы не только не могут приниматься и рассматриваться вместо выполненных студентом самостоятельно, но, напротив, попытки предъявления таких программ к сдаче должны рассматриваться как **преступление** (а как ещё можно относиться к злонамеренному обману преподавателя?), и последствия этого должны быть более серьёзными, чем предложение прийти в следующий раз.

Во-вторых, все сомнительные случаи должны в обязательном порядке заканчиваться контрольной работой, проводимой на компьютере, на котором заблокирован доступ к любым внешним источникам. Всегда можно найти небольшую задачу, которая может быть выполнена за 15-20 минут, дать на неё для верности, скажем, час и посмотреть на результаты. Чтобы отмести любые обвинения в необъективности, следует в таких случаях категорически отказываться рассматривать в качестве результата что-либо кроме работающей и полностью соответствующей заданию программы. Контрольную работу такого рода можно при желании провести и для всей группы; результаты будут в большинстве случаев более релевантными, чем результаты проверки обычных заданий.

Как показыват практика, научиться программировать в объёме обязательного практикума может `любой` студент, поступивший на ВМК, и проблема только в том, чтобы убедить студента в отсутствии альтернативы.

8 Благодарности

Автор считает своим приятным долгом выразить признательность А. Чернову и А. Сальникову за конструктивные дискуссии, а также Е. А. Бордаченковой за проявленную инициативу. Если бы не они, это эссе вряд ли когда-нибудь было бы написано.

УДК 519.6

МОДЕЛИРОВАНИЕ ЭКОЛОГИЧЕСКИХ РИСКОВ ЗАГРЯЗНЕНИЯ ПОВЕРХНОСТНЫХ ВОД ОКЕАНОВ И МОРЕЙ НА ОСНОВЕ ТЕОРИИ СОПРЯЖЕННЫХ УРАВНЕНИЙ

© 2010 г. Заячковский Антон Олегович

Кафедра вычислительных технологий и моделирования

Введение

Развитие новых технологий, увеличение объемов промышленного и сельскохозяйственного производства, расширение сети транспортных систем и систем передачи энергии и энергоносителей сопровождаются ростом техногенной нагрузки на биосферу. Следствием этого являются все чаще возникающие чрезвычайные ситуации, аварии и катастрофы, характеризующиеся значительными материальными, социальными и экологическими последствиями. При этом, как показали события последних десятилетий, реализуются считавшиеся ранее весьма маловероятными крупные аварии и катастрофы на различных объектах высокой технологии (примерами могут служить аварии на атомных электростанциях, химических комбинатах, нефте- и газопроводах и т. д.). Стала очевидной необходимость в разработке новых подходов к обеспечению безопасности природной среды. Именно поэтому в странах с развитой экономикой сформировалась новая отрасль знания — анализ экологических рисков и управление ими. Естественно, что актуальной стала задача исследования рисков. Основная цель таких исследований — вырабатывать для лиц, ответственных за принятие решений, рекомендации по эффективным мерам управления рисками.

Фундаментальное значение в экологических исследованиях имеют понятия рисков и уязвимости территорий к естественным и антропогенным воздействиям. С их помощью можно выявить потенциально опасные ситуации и объекты, количественно оценить степень тяжести возможных последствий катастрофических событий.

Анализ риска, связанный с загрязнением окружающей среды токсичными химическими веществами, опирается на использование различных математических моделей. Среди них определяющую роль играют модели распространения и трансформации примесей.

В настоящей работе предлагается алгоритм моделирования экологических рисков загрязнения поверхностных вод океанов и морей на основе теории сопряженных уравнений.

1 Понятие «риска» и вычисление величины экологического риска

Прежде чем оценивать риск, надо определить сам термин «риск». Различные определения, встречающиеся в литературе:

- Термин «риск» употребляется как тождественный термину «опасность» («риск — опасность будущего ущерба» или «риск — это опасность возникновения неблагоприятных последствий рассматриваемого события»).

- Под риском подразумевают возможность или вероятность неблагоприятного события или процесса («опасность, возможность убытка или ущерба», «возможность или вероятность факта или события, рассматриваемого как некое зло или некий ущерб» или «риск представляет собой шанс того, что может случиться нечто нежелательное»).

- Все большее распространение получает такой подход к определению риска неблагоприятного события, который учитывает не только вероятность этого события, но также все его возможные последствия. Вероятность события или процесса здесь выступает одним из компонентов риска, а мера последствий (ущерба) — другим. Такое *двумерное* определение риска используется при количественном оценивании риска и применяется в настоящей работе.

- Существует и иной подход к определению риска — *многомерный*. Он основан на многочисленных факторах, ответственных за восприятие риска и влияющих на принятие связанных с риском решений. Эти факторы, выявленные психологами, имеют качественный характер. Чтобы сравнивать степени проявления этих факторов, им приписывают условные единицы (например, по пятибалльной системе: если данный фактор считается очень сильным, то его «вес» принимают за 5, а если очень слабым, то за 1). После этого все «веса» суммируются, в этом заключается сущность так называемого психометрического подхода к риску, использующего его многомерное определение. Многомерное определение носит качественный характер, оно полезно при выявлении приоритетов людей в их отношении к совокупности опасных событий или процессов.

Мы принимаем, что *риск — количественная мера опасности с учетом ее последствий*. Последствия проявления опасности всегда приносят ущерб. Следовательно, оценка риска должна быть связана с оценкой ущерба. Чем больше ожидаемый ущерб, тем значительнее риск. Кроме того, риск будет тем больше, чем больше вероятность проявления соответствующей опасности. Поэтому риск R может быть определен как произведение вероятности опасности рассматриваемого события или процесса P на величину ожидаемых последствий (ущерба) Q:

$$R = P \cdot Q. \tag{1}$$

Чтобы придать параметрам качественный смысл, выразим опасность как вероятность нарушения санитарных норм (при заданной относительной частоте аварий). Тогда если показатель ущерба характеризует штрафы за такое нарушение санитарных норм (аварийный выброс поллютанта), то значение риска — это ожидаемые выплаты за нарушения природоохранного законодательства в среднем за расчетный период.

Для задачи оценки экологической опасности используются комбинированные методы решения прямых и сопряженных задач и методы теории чувствительности моделей распространения загрязняющих веществ.

2 Алгоритм расчета величины экологической опасности и оценка уязвимости

Моделирование экологической опасности (и экологического риска) будет вестись в следующих исходных положениях. Распространение концентрации загрязняющего вещества (*поллютанта*) C происходит по части поверхности сферы D, поэтому используются сферические координаты $(\varphi, \theta, R_{\text{зем}})$, где $R_{\text{зем}}$ — радиус сферы, концентрация рассматривается в момент времени t и $C \equiv C(\varphi, \theta, t)$ в $D_T \equiv (0, T) \times D$ — область изменения пространственно-временных координат.

Прежде всего определим набор оцениваемых характеристик в виде функционалов

$$\Phi_k(C) = \int\limits_{D_T} F_k(C)\chi_k(\varphi, \theta, t)dDdt, \quad k = \overline{1, K}, \quad K \geq 1, \tag{2}$$

где $F_k(C)$ — функции заданного вида, определенные и дифференцируемые на множестве функций $C \in Q(D_T) \subset L_2(D_T)$; $\chi_k(\varphi, \theta, t) \geqslant 0$ — весовые функции, принадлежащие сопряженному пространству $Q^*(D_T)$. При подходящем выборе функций $F_k(C)$ и χ_k с помощью функционалов этого вида можно описать обобщенные характеристики поведения системы, экологические ограничения на качество природной среды, результаты наблюдений различных типов, критерии управления, критерии качества моделей и т. д.

Важнейшей характеристикой при оценке экологического риска является функционал, формирующийся следующим образов: область водоема задается как носитель ненулевых значений весовой функции в (2) и $F_k(C) \equiv C$. Функция C описывает концентрацию пассивной примеси в регионе. Функционал представляет суммарное количество загрязнений, которое может поступить в водоем от действующих и потенциально возможных источников. Его цель — понять, до какой степени водоем подвержен воздействию антропогенных источников, и оценить степень экологической опасности для объекта получить загрязнения от действующих и потенциально возможных источников. В этом случае функционал примет вид:

$$\Phi(C) = \int\limits_{D_T} C \cdot \chi(\varphi, \theta) dDdt. \tag{3}$$

Вообще говоря рассматривается не вся область водоема, нас будет интересовать лишь определенная часть (например, прибрежная зона), и весовая функция задается следующим образом:

$$\chi(\varphi, \theta) = \begin{cases} 1, & (\varphi, \theta) \text{ из подобласти } D \\ 0, & \text{иначе.} \end{cases}$$

Для количественной оценки экологических рисков введем пороговые значения величин вариаций $\delta\Phi_k$ функционалов (2). Обозначим их через Δ_k, $k = \overline{1, K}$. Тогда условия, при которых выполняются неравенства

$$|\delta\Phi_k| < \Delta_k, \tag{4}$$

можно условно считать *экологически благополучными*, а условия, при которых они нарушаются, — *ситуациями экологической опасности*.

Основой для дальнейшего анализа будет служить диффузионная модель переноса. Согласно [2] она наиболее точно и оперативно позволяет получить оценочное представление о концентрации поллютанта в определенной точке водного тела в фиксированный момент времени.

Запишем уравнение в следующей дифференциально-операторной форме:

$$LC \equiv \frac{dC}{dt} - \nabla \cdot \mu \nabla C + \Lambda(C) = 0 \qquad \text{в } D_T, \tag{5}$$

где полная производная переноса имеет вид

$$\frac{dC}{dt} \equiv \frac{\partial C}{\partial t} + \frac{u}{R_{\text{зем}} \sin\theta} \frac{\partial C}{\partial \varphi} + \frac{v}{R_{\text{зем}}} \frac{\partial C}{\partial \theta},$$

а $\Lambda(C)$ – член, отражающий скорость трансформации поллютанта, линейный относительно концентрации C.

Рассмотрим подробнее, что из себя представляет слагаемое $\Lambda(C)$. Согласно [5], его можно задать линейным по концентрации, если считать, что при таком распространении поллютанта процесс деструкции протекает по закону мономолекулярной реакции первого порядка. Тогда кинетическая кривая распада загрязняющего вещества в точке описывается уравнением:

$$C_1(t) = C\lambda \exp(-\lambda t),$$

где t — время превращения загрязняющего вещества;

λ — скорость химической реакции, которая зависит от τ — периода полураспада загрязняющего вещества $\lambda = \ln(2/\tau)$

Данное выражение используется, например, при нефтяном загрязнении согласно работе [5].

В первом приближении процессы осаждения и испарения загрязняющего вещества могут быть описаны линейной функцией:

$$C_2(t) = C\beta t.$$

Таким образом, скорость удаления примеси из бассейна q параметризуется так:

$$q \equiv (C_1 + C_2)\rho_0 = \lambda C \rho_0 \exp(-\lambda t) + C\rho_0 \beta t,$$

где λ — параметр, определяющий темп биохимической и микробиологической деструкции (окисления) загрязняющего вещества, β — параметр, характеризующий испарение примеси в атмосферу и удаление ее за счет осаждения (биоседиментации), ρ_0 — средняя плотность воды.

Тогда

$$\Lambda(C) \equiv \frac{q(C)}{\rho_0} = (\lambda \exp(-\lambda t) + \beta t)C. \tag{6}$$

Слагаемое $\Lambda(C)$ описано и именно оно принимается в (5).

Начальное условия при $t = 0$ можно записать в виде

$$C(\varphi, \theta, 0) = C_0(\varphi, \theta). \tag{7}$$

В данной работе будем называть его *управлением* или *параметром*, так как ниже принимается, что только оно задает источник загрязняющего вещества в начальный момент времени, а других источников загрязнения нет.

Граничные условия для замыкания модели определяются физическим содержанием изучаемой проблемы. Считая, что выброс загрязняющего вещества происходит внутри рассматриваемой акватории вдалеке от границы, примем граничные условия вида

$$C|_\Gamma = 0. \tag{8}$$

Далее мы полагаем, что L действует из $Q(D_T)$ в $Q^*(D_T)$, где $Q(D_T) \equiv L_2\left(0, T; \overset{0}{W}{}_2^1(D)\right)$ и $Q^*(D_T) \equiv L_2\left(0, T; \overset{0}{W}{}_2^{-1}(D)\right)$.

Связь между вариацией концентрации загрязняющего вещества и вариацией функционала (3) имеет вид:

$$\delta\Phi(C) = \Phi(C + \delta C) - \Phi(C) = \int\limits_{D_T} \delta C \cdot \chi(\varphi, \theta) dD dt. \tag{9}$$

Следовательно, задача оценки вариации функционала сводится к задаче оценки вариации концентрации загрязняющего вещества, зависящей от вариации управления. Для получения явной зависимости $\delta\Phi$ от δC_0 проведем некоторые преобразования.

Считаем, что в исходной операторной задаче (5) каждому начальному условию C_0 соответствует единственное решение C и операторы линейны, поэтому в D_T вариация δC удовлетворяет следующей системе:

$$\begin{cases} L\delta C = 0 \\ \delta C|_\Gamma = 0 \\ \delta C|_{t=0} = \delta C_0. \end{cases} \tag{10}$$

Для (5) справедливо следующее утверждение:

$$LC = 0 \Leftrightarrow (LC, C^*)_{Q(D_T)} = 0 \qquad \forall C^* \in Q^*(D_T),$$

при условии $L_2 \subseteq Q^*$ и т. к. $\delta C \in Q(D_T) \Rightarrow L\delta C \in Q^*(D_T)$, значит для (10)

$$L\delta C = 0 \Leftrightarrow (L\delta C, C^*)_{Q(D_T)} = 0 \qquad \forall C^* \in Q^*(D_T).$$

Но тогда получаем (переходом к L^*, интегрированием по частям и т. д.):

$$\int\limits_{D_T} L\delta C \cdot C^* dD dt = \int\limits_{D_T} L^* C^* \cdot \delta C dD dt - \int\limits_{D} C^* \delta C dD \Big|_{t=0}^{t=T} = 0.$$

Здесь C^* есть решение сопряженной задачи в D_T:

$$\left\{\begin{array}{c} L^*C^* = \chi(\varphi,\theta) \\ C^*|_\Gamma = 0, \\ C^*(\varphi,\theta,T) = 0. \end{array}\right. \tag{11}$$

Функция чувствительности — это функция $C^*\Big|_{t=0}$(см. [3, 4]). Действительно, выражаем вариацию функционала в виде:

$$\delta\Phi(C) \equiv \int\limits_{D_T} \delta C \cdot \chi(\varphi,\theta) dDdt = \int\limits_{D_T} \delta C \cdot L^*C^* dDdt =$$
$$= \int\limits_{D} C^* \cdot \delta C dD\Big|_{t=0} = \int\limits_{D} C^*\Big|_{t=0} \cdot \delta C_0 dD,$$

откуда следует, что $C^*\Big|_{t=0}$ есть функция чувствительности вариации функционала к вариации управления.

Итак, выведено основное соотношение теории чувствительности, определяющее связь между вариацией $\Phi(C)$ и вариацией параметра модели (см. [2]):

$$\delta\Phi \equiv (\Gamma, \delta Y), \tag{12}$$

где $\delta Y \equiv \delta C_0$ — вариация управления модели; $\Gamma \equiv C^*\Big|_{t=0}$ — функция чувствительности функционала $\Phi(C)$ к этим вариациям.

Информационный смысл функции опасности для функционалов, определяющих качество воды в рассматриваемой области, можно описать следующим образом. Ее значение в точке $(\varphi,\theta,t) \in D_T$ — величина относительного вклада выброса загрязнений от источника, расположенного в этой точке (в интервале времени его действия), в представленное значением функционала суммарное количество загрязнений, поступающих в наблюдаемую область за период наблюдений. Поэтому соотношения чувствительности и функции чувствительности содержат количественную информацию для измерения степени экологических рисков для зоны наблюдения и показывают характер ее уязвимости к возмущениям от потенциально опасных источников.

В конечном итоге нас интересует вероятность невыполнения неравенства (4), выражающего условия попадания изучаемой ситуации в категорию экологической опасности для рассмотренного случая:

$$P \equiv P(|\delta\Phi| \geqslant \Delta). \tag{13}$$

Предположим, что поверхность акватории разбита на подобласти $\bigcup\limits_i D_i = D$ (мера $mes(D_i)$) таким образом, что можно было задать вероятность аварии (выброса загрязняющего вещества массой M от точечного источника, представленного вариацией управления) на конкретной части акватории как $P_i((\varphi,\theta) \in D_i)$ (разбиение по торгово-промышленным путям, по сейсмически опасным зонам). Для каждой поверхности D_i определим зоны $\hat{D}_i \in D_i$ (возможно нулевой меры) — *зоны экологической опасности*, на основе «сопряженных функций» — решений сопряженных уравнений следующим образом (см. (10),(11),(12),(13)):

$$\hat{D}_i : \left\{ (\hat{\varphi},\hat{\theta}) \in D_i : \left(M \int\limits_D C^*\Big|_{t=0} \cdot \frac{1}{2\pi} e^{-\frac{(\hat{\varphi}-\varphi)^2+(\hat{\theta}-\theta)^2}{2}} dD \right) \geqslant \Delta \right\}. \tag{14}$$

Тогда согласно результатам теории вероятностей имеем:

$$P(|\delta\Phi| \geqslant \Delta) = \sum_i \frac{mes(\hat{D}_i)}{mes(D_i)} P_i, \tag{15}$$

что завершает решение задачи оценки экологической опасности. Таким образом, способ нахождения вероятности экологического загрязнения определен.

Далее рассмотрим возможные значения ущерба при распространении примеси в ситуациях экологической опасности.

3 Расчет экологического ущерба

Под экономическим ущербом, наносимым окружающей среде, понимаются выраженные в стоимостной форме фактические и будущие (возможные) негативные изменения окружающей среды в результате ее загрязнения, повлекшие за собой деградацию естественных экологических систем и истощение природных ресурсов. Другими словами это изменение полезности окружающей среды вследствие ее загрязнения. В связи с этим вред окружающей среде оценивается как затраты общества, связанные с изменением окружающей среды, и складывается из следующих затрат:

- затраты общества на возврат окружающей среды в прежнее состояние;
- затраты общества в связи с безвозвратным изъятием части дефицитных природных ресурсов;
- дополнительные затраты общества в связи с изменением в окружающей среде.

Рассмотрим методы оценки затрат. При оценке экономического ущерба от загрязнения используется два основных методологических подхода: *прямой счет* и *косвенная оценка.* Оценка ущерба прямым счетом, требует сбора и обработки огромного объема информации, вследствие большой трудоемкости неудобны для широкого использования в экономических расчетах, и, как правило, служат лишь инструментом для создания информационной базы при разработке косвенных методов определения ущерба.

При прямом счете различают три метода выявления составляющих ущерба: контрольных районов (базирующийся на сравнении показателей загрязненного и условно чистого районов), аналитических зависимостей, основанных на получении математических зависимостей (например, при помощи многофакторного анализа) между показателями состояния соответствующей экономической системы и уровнем загрязнения окружающей среды, и комбинированный.

Для оценки ущерба (т. е. для оценки Q), связанного с аварийным выбросом загрязняющего вещества в поверхностные воды, используется косвенный подход к данной задаче. Косвенный метод оценки экономического ущерба основан на принципе перенесения на конкретный исследуемый объект общих закономерностей и предполагает использование системы нормативных показателей, фиксирующих зависимость негативных последствий от основных ущербообразующих факторов.

Основным этапом расчета экономического ущерба является определение показателя условной нагрузки на исследуемый объект, создаваемой источником загрязнения. С помощью поправочных коэффициентов (эколого-экономической) опасности все загрязняющие вещества приводятся к сопоставимым массам и суммируются в агрегированном (объединенном) показателе нагрузки. Коэффициенты опасности рассчитываются в зависимости от предельно допустимых концентраций (ПДК) и предельно допустимых сбросов (ПДС) загрязняющих веществ.

Ущерб от загрязнения водного объекта рассчитывается как плата за сверхлимитный сброс путем умножения приведенной массы M загрязняющих веществ, поступивших в водный объект (масса загрязняющего вещества за вычетом допустимых лимитов), на базовые нормативы платы $H_{\text{уд}}$ за сброс загрязняющих веществ в поверхностные с применением коэффициентов индексации $K_{\text{и}}$ и коэффициентов опасности (экологической ситуации) $K_{\text{эк}}$ [1]:

$$Q = K_{\text{эк}} \cdot K_{\text{и}} \cdot H_{\text{уд}} \cdot M. \tag{16}$$

Таким образом, мы рассмотрели факторы, влияющие на значение величины экологического риска, исходя из его *двумерного* определения, и связанные с его количественным оцениванием.

Покажем теперь, как рассчитывать сами показатели экологического риска.

4 Численные эксперименты по расчету экологического риска

Рассмотренные факторы, влияющие на значение величины экологического риска, позволяют рассчитать сами показатели экологического риска. В данной работе значение экологического риска R представляет собой ожидаемый среднемесячный уровень платежей государственным органам за нарушение природоохранного законодательства при крупных авариях.

На рисунках 1-2 представлены функции чувствительности $\{C^*\big|_{t=0}\}^h$ для двух различных прибрежных зон наблюдения, являющиеся решением системы (11), и поле скоростей на верхней границе водного слоя. Систему для C^* решаем в обратном времени с применением неявной схемы по времени и метода конечных разностей по пространству. В таблицах приведены тестовые расчеты величин экологических рисков для выбросов загрязняющего вещества массы M.

На примере Мадагаскара

Интенсивность источника массы M (тыс. тонн)	Величина экологического ущерба R (тыс. долларов)
0.1	0.630267
0.5	4.147
1	8.84178
5	50.2667
10	105.238

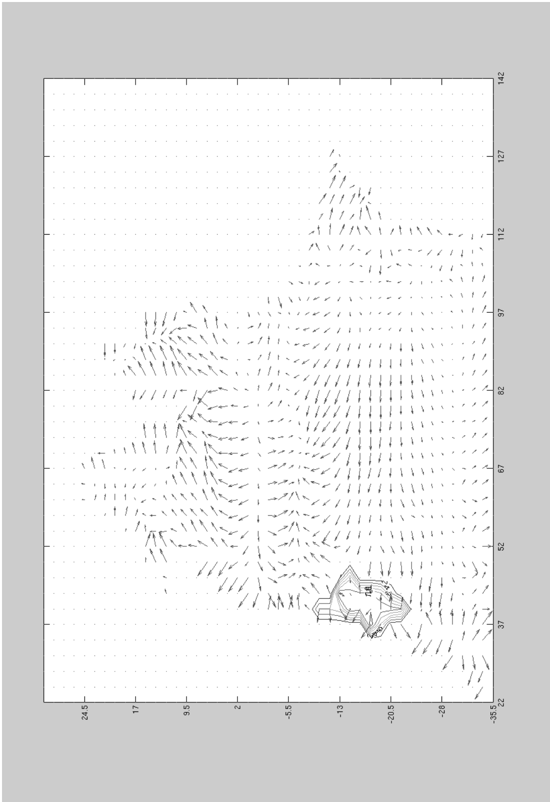

Рис. 1: Функция чувствительности для прибрежной зоны Мадагаскара

На примере Никобарских островов

Результатом настоящей работы является разработка алгоритма моделирования распространения загрязняющего вещества и создание методики расчета экологических рисков с применением теории сопряженных уравнений, а также приложение разработанной методики для тестовых расчетов значения величины риска в акватории Индийского океана.

Автор выражает благодарность научному руководителю д. ф.-м. н., профессору Агошкову В. И. за поддержку настоящей работы и ряд важных замечаний.

Работа выполнена при финансовой поддержке РФФИ (коды проектов 09-01-12029-офи-м, 09-05-00421, 10-01-00806), программы ФЦП «Кадры»(НК-408Р-42, НК-421Р-67).

Интенсивность источника массы M (тыс. тонн)	Величина экологического ущерба R (тыс. долларов)
0.1	0.0832802
0.5	0.49567
1	1.04355
5	5.73665
10	11.919

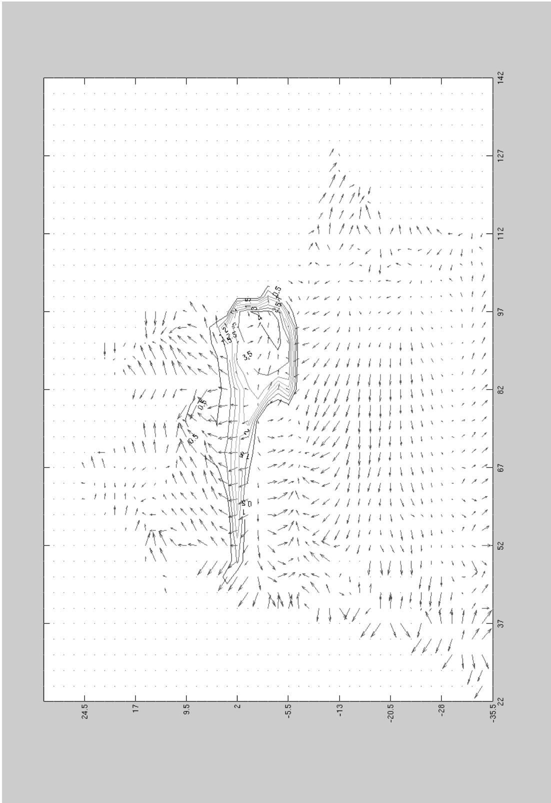

Рис. 2: Функция чувствительности для прибрежной зоны Никобарских островов

Список литературы

[1] *Ваганов П. А., Им М.-С.* Экологические риски — СПБ., 2001.

[2] *Пененко В. В., Цветова Е. А.* Математические модели для изучения рисков загрязнения природной среды // ПМТФ. 2004 Т. 45, № 2. С. 136-146.

[3] *Марчук Г. И.* Математическое моделирование в проблеме окружающей среды — М.: Наука, 1982.

[4] *Агошков В. И.* Методы оптимального управления и сопряженных уравнений в задачах математической физики — М.: ИВМ РАН, 2003.

[5] *Лебедев С. А.* Модельные расчеты фоновых значений антропогенного загрязнения нефтепродуктами и ассимиляционной емкости Черного моря (с исользованием данных дистанционного зондирования) — М.: Инженерная экология, 2008.

УДК 519.224.24

ОБ АСИМПТОТИЧЕСКОМ АНАЛИЗЕ U-СТАТИСТИК: ОЦЕНКА ТОЧНОСТИ АППРОКСИМАЦИИ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

© 2010 г. Т. А. Зубайраев

Кафедра Математической статистики

1 Введение

Пусть $X, \bar{X}, X_1, ..., X_N$ - независимые одинаково распределенные случайные величины, принимающие значения в произвольном измеримом пространстве $(\mathfrak{X}, \mathfrak{B})$. Пусть ϕ_1: $\mathfrak{X} \to \mathbb{R}$ и ϕ: $\mathfrak{X}^2 \to \mathbb{R}$ измеримые функции, принимающие вещественные значения. Предположим, что ϕ симметрична, т.е. $\phi(x,\ y) = \phi(y,\ x)$, для любых $x,\ y \in \mathfrak{X}$. Рассмотрим U-статистику

$$T = \frac{1}{N} \sum_{1 \le i < j \le N} \phi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \phi_1(X_i),$$

предполагая, что

$$\mathbb{E}\phi_1(X) = 0, \mathbb{E}\phi(x, X) = 0, \text{для всех } x \in \mathfrak{X},$$

$$\mathbb{E}\phi^2(x, X) < \infty, \mathbb{E}\phi_1^2(X) < \infty.$$

Одной из задач математической статистики является аппроксимация функций распределения статистик, имеющих сложный вид, некоторой более простой функцией распределения. Эта задача решается для функции распределения U-статистики, которую обозначим через $F(x)$, при этом дается приближение так называемым коротким разложением, включающим как предельную функцию $F_0(x)$, так и первый член асимптотического разложения $F_1(x)$ порядка $O(1/\sqrt{N})$. В работе получена оценка сверху для величины

$$\Delta_N = \sup_x |F(x) - F_0(x) - F_1(x)|, \text{ где } F_1(x) - \text{ поправка Эджворта,}$$

оценка имеет порядок $O(1/N)$. Ранее оценка для Δ_N была получена в [2] с оценкой константы C_1 экспоненциально зависящей от собственных значений оператора. Зависимость оценки от собственных значений в данной работе имеет степенной вид. Данный результат существенным образом опирается на условия невырожденности аналогичные работе [1], которые отличаются от использованных в работе [2]. Результат сформулирован в теореме 2.

Введем ряд обозначений:

$$\beta_s = \mathbb{E}|\phi_1(X)|^s, \gamma_s = \mathbb{E}|\phi(X, \bar{X})|^s,$$

$$\sigma^2 = \gamma_2, \gamma_{s,r} = \mathbb{E}(\mathbb{E}\{|\phi(X, \bar{X})|^s|X\})^r,$$

также преположим, что

$$\beta_2 < \infty, 0 < \sigma^2 < \infty.$$

Тогда дисперсию T можно переписать в виде

$$\mathbb{E}T^2 = \beta_2 + \frac{N-1}{2N}\sigma^2.$$

Статистика T является вырожденной по той причине, что квадратическая ее часть не является асимптотически несмещенной из-за условия $\sigma^2 > 0$ и, следовательно, статистика T

не является асимптотически нормальной. Более точно, асимптотическое распределение T не является гауссовским и задается распределением случайной величины

$$T_0 = \frac{1}{2}\sum_{j\geq 1} q_j(\eta_j^2 - 1) + \sum_{j\geq 0} a_j\eta_j,$$

где η_j последовательность независимых одинаково распределенных стандартных нормальных случайных величин, $a_0, a_1, \ldots$ последовательность суммируемых в квадрате весов и $q_1, q_2, \ldots$ собственные значения оператора Гильберта-Шмидта, например $\mathbb{Q}$, ассоциированного с ядром ϕ.

Рассмотрим функции распределения

$$F(x) = \mathbb{P}\{T \leq x\}, F_0(x) = \mathbb{P}\{T_0 \leq x\},$$

$$\Delta_N = \sup_x |\Delta_N(x)|, \Delta_N(x) = F(x) - F_0(x) - F_1(x),$$

$F_1(x)$ - поправка Эджворта, определенная в (4.6). Заметим, что F_1 исчезает если $\phi_1 = 0$ или равенства

$$\mathbb{E}\phi_1^3(X) = \mathbb{E}\phi_1^2(X)\phi(X,x) = \mathbb{E}\phi_1(X)\phi(x,x) = \mathbb{E}\phi^3(X) = 0, \tag{1}$$

справедливы для всех $x \in \mathfrak{X}$. Обозначим $c, c_1, \ldots$ константы, в случае, если константа зависит, например, от s мы будем писать c(s)=c_s.

Рассмотрим функцию концентрации статистики T_*

$$Q(T_*;\lambda) = \sup_x \mathbb{P}\{x \leq T_* \leq x+\lambda\}, \lambda \geq 0, \tag{2}$$

$$T_* = \sum_{1\leq i<k\leq N} \phi(X_j, X_k) + f_1(X_1, \ldots, X_M) + f_2(X_{M+1}, \ldots, X_N), 1 \leq M \leq N/2,$$

где $f_1 = f_1(X_1, \ldots, X_M)$ произвольная статистика, зависящая только от $X_1, \ldots, X_M$, $f_2 = f_2(X_{M+1}, \ldots, X_N)$ также произвольная статистика, но не являющаяся зависимой от $X_1, \ldots, X_M$. Заметим, что класс статистик T_* является более общим, чем класс статистик T.

2 Основные понятия и обозначения

Пусть $\mathbb{Q}$ оператор Гильберта-Шмидта, ассоциированный с ядром ϕ , а $q_1, q_2...$ - его собственные значения. Не ограничивая общности рассуждений, будем считать, что $|q_1| \geq |q_2| \geq ...$ Оператор $\mathbb{Q}$: $L^2 \to L^2$ определяется следующим образом

$$\mathbb{Q}f(x) = \int_{\mathfrak{X}} \phi(x,y)f(y)\mu(dy) = \mathbb{E}\phi(x,X)f(x).$$

Пусть $\{e_j : j \geq 1\}$ полная ортонормированная система составленная из собственных функций оператора $\mathbb{Q}$ соответствующих собственным значениям $q_1, q_2....$ Тогда

$$\sigma^2 = \mathbb{E}\phi(\bar{X}, X) = \sum_{j\geq 1} q_j^2 \ , \ \phi(x,y) = \sum_{j\geq 1} q_j e_j(x) e_j(y), \tag{2.3}$$

т.к. $\mathbb{Q}$ оператор Гильберта-Шмидта и ядро ϕ вырождено. Ряды в (2.3) сходятся в $L^2(\mathfrak{X}^2, \mathfrak{B}^2, \mu\times\mu)$. Рассмотрим подпространство $L^2(\phi, \phi_1) \subset L^2$ образованное ϕ_1 и собственными функциями e_j, которые соответствуют ненулевым собственным значениям q_j. Добавляя , если необходимо,собственную функцию e_0, $\mathbb{Q}e_0 = 0$,мы можем предположить, что функции $e_0, e_1, \ldots$ образуют ортонормированный базис в L^2 (ϕ, ϕ_1). Таким образом, справедливо следующее разложение

$$\phi_1(x) = \sum_{j\geq 0} a_j e_j(x) \text{ в } L_2 \ , \ \beta_2 = \mathbb{E}\phi_1^2(X) = \sum_{j\geq 0} a_j^2, \tag{2.4}$$

где $a_j = \mathbb{E}\phi_1(X)e_j(X)$ и $\mathbb{E}e_j(X) = 0$, для любого j. Следовательно, система $(e_j(X))_{j\geq 0}$ является ортонормированной системой случайных чисел с нулевым средним.

Гильбертово пространство $\ell_2 \subset \mathbb{R}^\infty$ состоит из элементов
$x = (x_1, x_2 \ldots) \in \mathbb{R}^\infty$ таких, что

$$| x |^2 =_{def} \langle x, x \rangle \ , \ \ | x | < \infty \ , \ \langle x, x \rangle = \sum_{j\geq 0} x_j y_j.$$

Рассмотрим случайный вектор

$$\mathbf{X} =_{def} (e_0(X), e_1(X), e_2(X), \ldots), \tag{2.5}$$

принимающий значения в $\mathbb{R}^\infty$. Поскольку $(e_j(X))_{j\geq 0}$ система некоррелированных случайных величин с единичными дисперсиями, случайный вектор $\mathbf{X}$ имеет единичную матрицу ковариаций и нулевое среднее. В силу (2.3) и (2.4), справедливы равенства

$$\phi(X, \bar{X}) = \langle \mathbb{Q}\mathbf{X}, \bar{\mathbf{X}} \rangle, \ \phi_1(X) = \langle a, \mathbf{X} \rangle, \tag{2.6}$$

где определяем $\mathbb{Q}x = (0, q_1 x_1, q_2 x_2, \ldots)$ для $x \in \mathbb{R}^\infty$ и $a = (a_j)_{j\geq 0} \in \mathbb{R}^\infty$. Равенства в (2.6) позволяют взять нам в качестве измеримого пространства $\mathfrak{X}$ простанство $\mathbb{R}^\infty$. Случайная переменная X - случайный вектор, принимающий значения в $\mathbb{R}^\infty$ с нулевым средним и единичной матрицей ковариаций и такой, что

$$\phi(X, \bar{X}) = \langle \mathbb{Q}X, \bar{X} \rangle, \ \phi_1(X) = \langle a, X \rangle. \tag{2.7}$$

Не ограничивая общности рассуждений, мы будем предполагать, что ядра $\phi(x, y)$ и $\phi_1(x)$ - линейные функции по каждому из своих аргументов (см. Goetze, Bentkus).

3 Основные результаты

Лемма 1 *Пусть $Y_1, \ldots, Y_d, Y'_1, \ldots, Y'_d$ независимые, одинаково распределенные гауссовские случайные величины с параметрами $(0, \mathbb{I})$ в гильбертовом пространстве H. Пусть $|q_1| \geq |q_2| \geq \ldots$ - собственные значения оператора Гильберта-Шмидта $\mathbb{Q}$. $W = (detA)^2$, где $A(Y) = \{a_{ij}(Y)\}_{i,j=1}^d$, $a_{ij}(Y) = \phi(Y_i, Y'_j) = \langle \mathbb{Q}Y_i, Y_j \rangle$.*

Тогда

$$\mathbb{E}W = (d!)^2 \sum_{1\leq i_1 < \ldots < i_d < \infty} (q_{i_1} \ldots q_{i_d})^2,$$
$$(\mathbb{E}W^2)^{1/2} \leq c(d)\mathbb{E}W.$$

Условия невырожденности

Будем считать, что для некоторого случайного вектора Z, ядра ϕ, параметров c, c_1, s и p выполнены условия невырожденности, если

$$\mathbb{P}\{W(\bar{Z}) > q_1^2 \ldots q_9^2\} \geq p, \tag{3.1}$$
$$\mathbb{P}\{|\phi(Z_i, \bar{Z}_j)| \leq c\} \geq c_1, \ 1 \leq i, j \leq s,$$
$$\text{где } W(\bar{Z}) = (detA)^2, A = \{a_{ij}\}_{i,j=1}^s, a_{ij} = \phi(Z_i, \bar{Z}_j),$$
$$Z_i, \bar{Z}_j - \text{ независимые копии вектора } Z.$$

При этом параметр p мал, а параметр c_1 близок к единице.

Рассмотрим вектор $G = (\eta_0, \eta_1 \ldots)$ со значениями в $\mathbb{R}^\infty$. Пусть ковариации и средние векторов G и X совпадают, тогда

$$\mathbb{E}\phi_1(G) = \mathbb{E}\phi(G, x) = 0, \mathbb{E}\phi_1^2(G) = \mathbb{E}\phi_1^2(X),$$

$$\mathbb{E}\phi_1(G)\phi(G,x) = \mathbb{E}\phi_1(X)\phi(X,x),$$

$$\mathbb{E}\phi(G,x)\phi(G,y) = \mathbb{E}\phi(X,x)\phi(X,y).$$

Лемма 2 *Пусть для случайного гауссовского вектора G, рассмотренного выше, выполнены условия невырожденности и $\mathbb{P}\{W(\bar{G}) > q_1^2 \dots q_9^2\} \geq 2p$. Тогда при $m \geq c_s|q_1 \dots q_9|^{-3}p^{-1}(|q_1 \dots q_9|^{-3}p^{-1}\gamma_{2,3/2}+\gamma_3)$ случайная сумма $S_m = m^{-1/2}(X_1+\ldots+X_m)$ также удовлетворяет условиям невырожденности , при этом*

$$\mathbb{P}\{W(\bar{S_m}) > 4q_1^2 \dots q_9^2\} \geq p.$$

Далее необходимо оценить характеристическую функцию статистики T_*.

Сформулируем лемму, которая позволит ограничить характеристическую функцию статистики T_* характеристической функцией специального вида, зависящей от дискретных случайных величин. Пусть $\varepsilon_1, \varepsilon_2, ...$ случайные, одинаково распределенные величины Радемахера, т.е $\mathbb{P}\{\varepsilon_j = -1\} = \mathbb{P}\{\varepsilon_j = 1\} = 1/2$. Мы предполагаем, что все случайные величины и случайные вектора независимы, если обратное не вытекает из контекста. Пусть $m \in \mathbb{N}$,определим случайную величину $\vartheta_1 = \varepsilon_1, ..., \vartheta_m = \varepsilon_m$. Аналогично, величины $\vartheta_j, j > m$ определены следующим образом:

$$\vartheta_j = \varepsilon_l \text{ для } j \in I(l) =_{def} (ml-m, ml],$$

Для натуральных чисел m и s мы введем неотрицательные числа

$$L = [M/(2ms)], \quad K = sL \text{ и } K_0 = mK = smL.$$

Лемма 3 *Пусть $m, s \in \mathbb{N}$. Тогда статистика T_* удовлетворяет*

$$|\mathbb{E}exp\{itT_*\}| \leq \mathbb{E}|\mathbb{E}_\vartheta exp\{itT^\vartheta\}|,$$

где $\mathbb{E}_\vartheta = \mathbb{E}_{\vartheta_j, 1\leq j\leq 3K_0}$

$$T^\vartheta = \sum_{1\leq j\leq 3K_0} \vartheta_j\vartheta_k a_{jk} + F_1 + F_2, \; a_{jk} =_{def} 1/4\phi(\tilde{X}_j, \tilde{X}_k),$$

номер K_0 определен как $K_0 = mK = smL$ и $F_1(F_2)$ является функцией от $\vartheta_j, j \in [1, K_0]$, $(\vartheta_j, j \in (K_0, 3K_0])$.

Лемма 4 *Пусть $m \in \mathbb{N}$. Обозначим:*

$$Y = (2m)^{-1/2}\sum_{j=1}^{m} \tilde{X}_j, \quad \Lambda = \sum_{j=1}^{K} \tilde{\varepsilon}_j Y_j,$$

где $Y_1, Y_2, ...$ - независимые копии Y и K=sL. Тогда мы имеем

$$|\mathbb{E}\{tT_*\}|^2 \leq \mathbb{E}exp\{2^{-1}itm\phi(\Lambda, \bar{\Lambda}\}.$$

Определим $\tau, \tau_1, \tau_2 \dots$ как независимые копии симметричной случайной величины τ с неотрицательной характеристической функцией и такие, что выполнены условия:

$$1 \leq \mathbb{E}\tau^2 \leq 2 \text{ , } \mathbb{P}\{|\tau| \leq 2\} = 1. \tag{3.2}$$

Лемма 5 *Пусть $s \in \mathbb{N}$ и $L \in \mathbb{Z}_+$. Предположим, что вектор Y принимает значения в R^∞ и удовлетворяет условиям невырожденности. Пусть далее*

$$\Lambda = \sum_{j=1}^{sL} \tau_j Y_j,\ \bar{\Lambda} = \sum_{j=1}^{sL} \bar{\tau}_j \bar{Y}_j,\ q = [pL/4],$$

где Y_j и $\bar{Y}_j$ являются независимыми копиями Y. Тогда

$$\mathbb{E}e\{t\phi(\Lambda, \bar{\Lambda}\} \le c_d(s)(pL)^{-d} + \sup_{\mathbb{A}} \mathbb{E}e\{t\langle \mathbb{A}U, V\rangle\},\ t \in \mathbb{R},\ d \ge 0,$$

где $\sup_{\mathbb{A}}$ обозначает супремум по всем неслучайным матрицам $\mathbb{A}$ размера $s \times s$ таких, что выполнено неравенство $(detA)^2 > q_1^2 \dots q_9^2$.
U и V - независимые вектора в $\mathbb{R}^s$, являющиеся суммами n независимых копий $W = (\tau_1, \dots, \tau_s)$.

Лемма 6 *Пусть A невырожденная матрица размера $s \times s$. Пусть $X \in \mathbb{R}^s$- случайный вектор с ковариационной матрицей . Предположим, что существует константа c_s такая , что*

$$\mathbb{P}\{|X| \le c_s\} = 1,\ |A| \le c_s,\ |C^{-1}| \le c_s. \tag{3.3}$$

Пусть U и V независимые случайные векторы, являющиеся суммами n независимых копий X. Тогда

$$|\mathbb{E}e\{t\langle AU, V\rangle\}| \le c(s)|detA|^{-1}\mathcal{M}^{2s}(t; N) \text{ для } |t| > 0,$$
$$\text{где } \mathcal{M}(t; N) = 1/\sqrt{|t|N} + \sqrt{|t|} \text{ для } |t| > 0.$$

Лемма 7 *Пусть U и V - независимые вектора в $\mathbb{R}^s$, являющиеся суммами n независимых копий $W = (\tau_1, \dots, \tau_s)$. Предположим, что выполнено неравенство $(detA)^2 > q_1^2 \dots q_9^2$. Тогда справедливо неравенство:*

$$\mathbb{E}e\{t\langle AU, V\rangle\} \le c(s)\frac{1}{|q_9|^9}\mathcal{M}^{2s}(t; q).$$

Доказательство. Оценим детерминант матрицы A в правой части неравенства для характеристической функции из Леммы 6:

$$|detA|^2 > q_1^2 \dots q_9^2 \sim |detA| > |q_1| \dots |q_9| \sim |detA|^{-1} < \frac{1}{|q_1| \dots |q_9|} < \frac{1}{|q_9|^9}.$$

Результат следует из предыдущей леммы, с заменой N на q и X на W. Матрица A удовлетворяет неравенству $|A| \le c_s$.

■

Теорема 1 *Пусть $m \in \mathbb{N}$, предположим, что сумма $Y = (2m)^{-1/2}(\tilde{X_1} + \dots + \tilde{X_m})$ удовлетворяет условию невырожденности. Тогда для любой статистики T_* справедливо неравенство:*

$$|\mathbb{E}e\{tT_*\}| \ll_s \frac{1}{|q_9|^9}\mathcal{M}^{2s}(tm; pM/m).$$

Доказательство. Предположим, что $pM \ge c_s m$ с достаточно большой константой c_s. Иначе $pM \le c_s m$ и результат будет следовать из простого неравенства $|\mathbb{E}e\{tT_*\}| \le 1$. Вспомним ряд обозначений, введенных ранее:

$$L = [M/(2ms)], \quad K = sL \text{ и } K_0 = mK = smL.$$

Далее,

$$L \asymp_s K \asymp_s K_0/m \asymp_s M/m \asymp q/p,\ q = [pL/4]. \tag{3.4}$$

Лемма 6 дает нам оценку

$$|\mathbb{E}\{tT_*\}|^2 \le \mathbb{E}e\{\beta\phi(\Lambda,\bar{\Lambda})\},\ \Lambda,\bar{\Lambda}\text{ - определены в Лемме 5.} \tag{3.5}$$

С помощью леммы 7 получаем оценку

$$\mathbb{E}e\{\beta\phi(\Lambda,\bar{\Lambda}\} \le c_d(s)(pL)^{-d} + \sup_{\mathbb{A}} \mathbb{E}e\{\beta\langle\mathbb{A}U,V\rangle\}, \tag{3.6}$$

$$\mathbb{E}e\{\beta\langle AU,V\rangle\} \le c(s)\frac{1}{|q_9|^9}\mathcal{M}^{2s}(\beta;q), \tag{3.7}$$

Собирая оценки (3.5)-(3.7) и подставляя $\beta = tm/2$, используем (3.4) и получаем оценку

$$\mathbb{E}e\{tT_*\} \ll_{d,s} (pM/m)^{-d/2} + \frac{1}{|q_9|^9}\mathcal{M}^{2s}(tm;pM/m). \tag{3.8}$$

Известно, что $\inf_x \mathcal{M}^s(x;pM/m) = (pM/m)^{-s/4}$. Таким образом (3.8) доказывает желаемый результат при выборе $d = s/2$.

■

Введем обозначение (см. Лемму 3):

$$\psi(t) = |\mathbb{E}_\vartheta e\{tT^\vartheta\}|. \tag{3.9}$$

Заметим, что $\psi(t)$ является случайной величиной в силу того, что зависит от $X_1,\dots,X_N$ и $\tilde{X_1},\dots,\tilde{X_N}$.

Лемма 8 *Пусть $d \ge 0$ и $s \in \mathbb{N}$. Предположим, что $Y = (2m)^{-1}\sum_{k=1}^{k=m}\tilde{X}_k$ удовлетворяет условию невырожденности. Тогда существуют константы $c_1(s,d)$ и $c_2(s,d)$ такие, что событие*

$$D = \{\psi(t-\gamma)\psi(t+\gamma) \le c_1(s,d)\frac{1}{|q_9|^9 v}\mathcal{M}^s(\gamma m;pM/m)\}, \tag{3.10}$$

удовлетворяет условию

$$\mathbb{P}\{D\} \ge 1 - c_2(s,d)(pM/m)^{-d}. \tag{3.11}$$

Для $A \ge t_0, t_1 \ge 0$ определим интегралы

$$I_0 = \int_{-t_1}^{t_1} |\hat{\Psi}(t)|dt, I_1 = \int_{t_0\le|t|\le A} |\hat{\Psi}(t)|\frac{dt}{|t|},$$

где $\hat{\Psi} = \int_{\mathbb{R}} e\{tx\}d\Psi(x)$ преобразование Фурье-Стилтьеса функции распределения $\Psi(x) = \mathbb{P}\{T_* \le x\}$.

Лемма 9 *Пусть $m \in \mathbb{N}$. Предположим, что случайный вектор $Y = (2m)^{-1/2}(\tilde{X_1}+\ldots+\tilde{X_m})$ удовлетворяет условиям невырожденности и $s \ge 9$. Введем ряд обозначений и условий*

$$k = \frac{pM}{m}, t_0 = \frac{c_0(s)}{m}k^{-1+2/s}, t_1 = \frac{c_1(s)}{m}k^{-1/2},$$

$$\frac{c_2(s)}{m} \le A \le \frac{c_3(s)}{m},$$

где $c_j(s), 0 \le j \le 3$ - некоторые положительные константы.

Тогда

$$I_0 \ll_s |q_9|^{-9}(pM)^{-1}, I_1 \ll_s max\{1,|q_9|^{-18}\}m(pM)^{-1}. \tag{3.12}$$

4 Оценка точности аппроксимации

Для $r \in \mathbb{Z}_+$ и функций f_i рассмотрим статистику следующего вида:

$$T^{(r)} = \frac{1}{N} \sum_{1 \le i < j \le N} \phi(Z_i, Z_j) + \sum_{1 \le i \le N} f_i(Z_i), \tag{4.1}$$

где

$$Z_j = X_j \text{ для } 1 \le j \le r, Z_j = G_j \text{ для } r < j \le N.$$

Обозначим $l = [(N-2)/20]$ и рассмотрим следующую оценку характеристических функций

$$\kappa(t) = \kappa(t, N, \phi, \mathcal{L}(X)) = \kappa_1(t) + \kappa_2(t), \tag{4.2}$$

где

$$\kappa_1(t) = \sup_L |\mathbb{E}e\{tN^{-1} \sum_{1 \le j < k \le l} \phi(X_j, X_k) + L(X_1, \dots, X_l)\}|, \tag{4.3}$$

$$\kappa_2(t) = \sup_L |\mathbb{E}e\{tN^{-1} \sum_{1 \le j < k \le l} \phi(G_j, G_k) + L(G_1, \dots, G_l)\}|, \tag{4.4}$$

супремум берется по всем линейным статистикам L которые могут быть представлены как $L(x_1, \dots, x_l) = \sum_{j=1}^{l} f_j(x_j)$ с некоторыми функциями $f_1, \dots, f_l$.

Лемма 10 *Пусть $m \in \mathbb{N}, s \ge 9$ и $t_0 = m^{-1}(pN/m)^{-1+2/s}$. Предположим, что случайный вектор $Y = (2m)^{-1/2}(\tilde{X_1} + \ldots + \tilde{X_m})$ удовлетворяет условиям невырожденноти. Тогда, для $pN > m$ и $m^{-1} \ge t_* \ge t_0$ функция распределения $F^{(r)}$ статистики $T^{(r)}$ удовлетворяет равенству*

$$F^{(r)}(x) = \frac{1}{2} + \frac{i}{2\pi} V.P. \int_{-Nt_*}^{Nt_*} e\{-xt\} \widehat{F^{(r)}}(t) \frac{dt}{t} + R, \tag{4.5}$$

$$\text{где } |R| \ll_s (|q_9|^{-9} + max\{1, |q_9|^{-18}\})m/(pN).$$

Поправка Эджворта $F_1(x) = F_1(x; \mathcal{L}(X), \phi_1, \phi)$ определена как функция с ограниченной вариацией удовлетворяющая условию $F_1(-\infty) = 0$ и с преобразованием Фурье-Стилтьеса следующего вида:

$$\hat{F}_1(t) = \frac{(it)^3}{6\sqrt{N}} \mathbb{E}(\phi_1(X) + \phi(X, G))^3 e\{tT_0\}. \tag{4.6}$$

Лемма 11 *Справедливо неравенство:*

$$\hat{\Delta}_N \ll \kappa N^{-1}(t^4\beta_4 + t^6\beta_3^2 + t^2\gamma_2 + |t|^3\gamma_3 + |t|^5\gamma_2\gamma_3 + t^2\gamma_{2,2} + t^6\gamma_2\gamma_{2,2}), \tag{4.7}$$

Если выполнено условие (1), тогда $\hat{\Delta}_N = |\hat{F}(t) - \hat{F}_0(t)|$ и выполнено неравенство

$$\hat{\Delta}_N \ll \kappa N^{-1}(t^4\beta_4 + t^2\gamma_2 + |t|^3\gamma_3 + t^4\gamma_2\gamma_{2,2}). \tag{4.8}$$

Лемма 12 *Поправка Эджворта F_1 удовлетворяет неравенству:*

$$|\hat{F}_1(t)| \ll N^{-1/2}|t|^3(\beta_3^2 + \sigma^2\gamma_{2,2})^{1/2} \prod_{j \ge 1}(1 + 2t^2q_j^2/25)^{-1/4}.$$

Далее, при $s \ge 7$ справедливы неравенства:

$$\int_{|t| \ge \lambda} |\hat{F}_1(t)| \frac{dt}{|t|} \ll N^{-1/2}(\beta_3^2 + \sigma^2\gamma_{2,2})^{1/2}|q_s|^{-s/2}\lambda^{3-s/2}, \text{ для } \lambda > 0, \tag{4.9}$$

$$\int_{\mathbb{R}} |\hat{F}_1(t)| \frac{dt}{|t|} \ll N^{-1/2}(\beta_3^2 + \sigma^2\gamma_{2,2})^{1/2}|q_s|^{-3}. \tag{4.10}$$

Лемма 13 *Предположим, что выполнены условия невырожденности.*
(i) Пусть $s \geq 13$ *и* $m_0 \asymp |q_1 \dots q_9|^{-3} p^{-1} (|q_1 \dots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3)$. *Тогда*

$$\Delta_N \ll_s \frac{m_0(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{pN} + N^{-1}(\beta_3^2 + \sigma^2 \gamma_{2,2})(\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6}) + \tag{4.11}$$

$$+ \frac{|q_9|^{-9}}{p^6 N} \cdot (\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2 \gamma_3 + \gamma_{2,2} + \sigma^2 \gamma_{2,2}),$$

(ii) Пусть выполнено условие (1) и $s \geq 9$. *Тогда*

$$\Delta_N \ll_s \frac{m_0(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{pN} + N^{-1}(\beta_3^2 + \sigma^2 \gamma_{2,2})(\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6}) + \tag{4.12}$$

$$+ \frac{|q_9|^{-9}}{p^4 N} \cdot (\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}).$$

Доказательство. Представим первый член разложения Эджворта в виде следующего обратного преобразования Фурье-Стилтьеса. Для любой функции $F : \mathbb{R} \to \mathbb{R}$ с ограниченной вариацией такой, что $F(-\infty) = 0$ и $2F(x) = F(x+) + F(x-)$, для всех $x \in \mathbb{R}$ имеем

$$F(x) = \frac{1}{2} F(\infty) + \frac{i}{2\pi} \lim_{M \to \infty} V.P. \int_{|t| \leq M} e\{-xt\} \hat{F}(t) \frac{dt}{t}. \tag{4.13}$$

Пусть $m \in \mathbb{N}$. Предполагая, что случайный вектор $Y = (1/2m)^{-1/2} \tilde{X}_1 + \ldots + \tilde{X_m}$ удовлетворяет условиям невырожденности мы докажем, что выполнено неравенство

$$\Delta_N \ll_s \frac{m(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{pN} + \alpha + \frac{|q_9|^{-9}}{p^6 N} \cdot \tag{4.14}$$

$$\cdot (\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2 + \gamma_3 + \sigma^2 \gamma_3 + \gamma_{2,2} + \sigma^2 \gamma_{2,2}),$$

при условии, что $s \geq 13$, где

$$\alpha = \frac{\beta_3^2 + \sigma^2 \gamma_{2,2}}{N} (\frac{1}{|q_s|^s} + \frac{1}{|q_s|^6}).$$

В случае, если условие (1) выполнено и $s \geq 9$, мы докажем, что

$$\Delta_N \ll_s \frac{m(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{pN} + \alpha + \frac{|q_9|^{-9}}{p^4 N} \cdot (\beta_4 + \sigma^4 + \gamma_3 + \gamma_{2,2}), \tag{4.15}$$

Оценки (4.14) и (4.15) дают нам (4.21) и (4.22). Используя минимально возможное $m = m_0$ мы получим оценки (4.21) и (4.22).

В доказательстве (4.14) и (4.15) мы предполагаем, что $pN > m$.

Выберем $t_* = m^{-1}(pN/m)^{-1/2}$. Применяя лемму 9 к функциям распределения F и F_0, лемму 13 при $\lambda = N t_*$ и (4.13) к поправке Эджворта F_1, мы получим

$$\Delta_N \ll_s J + R, J =_{def} \int_{-Nt*}^{Nt_*} |\hat{F}(t) - \hat{F_0}(t) + \hat{F_1}(t)| \frac{dt}{|t|}, \tag{4.16}$$

где

$$|R| \ll_s (|q_9|^{-9} + max\{1, |q_9|^{-18}\}) m/(pN) + N^{-1/2} (\beta_3^2 + \sigma^2 \gamma_{2,2})^{1/2} |q_s|^{-s/2} (N t_*)^{3-s/2}.$$

Имеем $Nt_* = (N/m)(pN/m)^{-1/2}$. Следовательно $ab \ll a^2 + b^2$ вместе с $pN \geq m, 0 < p \leq 1$ и $s \geq 9$ дают нам оценку

$$|R| \ll_s (|q_9|^{-9} + max\{1, |q_9|^{-18}\})\frac{m}{pN} + \frac{\beta_3^2 + \sigma^2\gamma_{2,2}}{|q_s|^s}. \tag{4.17}$$

Далее мы докажем, что

$$I_r =_{def} \int_{-Nt*}^{Nt_*} |t|^r \kappa(t)\frac{dt}{|t|} \ll_s |q_9|^{-9}p^{-r}. \tag{4.18}$$

Используя (4.18) и $p \leq 1$, интегрируя оценку для $\hat{\Delta}_N = |\hat{F}(t) - \hat{F}_0(t) + \hat{F}_1(t)|$ из леммы 15 мы получим

$$J \leq_s |q_9|^{-9}\frac{1}{p^6 N}(\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2 + \gamma_3 + \sigma^2\gamma_3 + \gamma_{2,2} + \sigma^2\gamma_{2,2}), \tag{4.19}$$

при условии, что $s \geq 13$, и

$$J \leq_s |q_9|^{-9}\frac{1}{p^4 N}(\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}), \tag{4.20}$$

если (1) выполнено и $s \geq 9$. Собирая (4.16) - (4.20), мы получим (4.14) и (4.15).

Остается доказать (4.18). Чтобы оценить $\kappa(t)$ воспользуемся теоремой 1. Статистика в определении $\kappa(t)$ является статистикой типа (2), где N заменено на $l = [(N-2)/20]$. M может быть выбрано произвольным и удовлетворять неравенству $1 \leq M \leq l$; например, положим $M = [l/2]$. Имеем $\kappa_1(t) \ll_s |q_9|^{-9} min\{1, \mathcal{M}^s(tm/N; pN/m\}$. Аналогичная оценка справедлива и для $\kappa_2(t)$ Далее имеем

$$I_R \ll_s \int_{-Nt_*}^{Nt_*} |t|^r |q_9|^{-9} min\{1, \mathcal{M}^s(tm/N; pN/m\}\frac{dt}{|t|} =$$

$$2N^r|q_9|^{-9}\int_0^{t_*} t^r min\{1, \mathcal{M}^s(tm; pN/m\}\frac{dt}{t}$$

$$\ll N^r|q_9|^{-9}\int_0^{1/pN} t^r\frac{dt}{t} + N^r|q_9|^{-9}\int_{1/pN}^{t_*} t^r\frac{1}{(tpN)^{s/2}}\frac{dt}{t} \ll p^{-r}|q_9|^{-9}.$$

■

Используя Леммы 9-12 мы можем сформулировать следующую теорему:

Теорема 2 *(i) Пусть* $s \geq 13$ *и* $m_0 \asymp |q_1 \dots q_9|^{-3}p^{-1}(|q_1 \dots q_9|^{-3}p^{-1}\gamma_{2,3/2} + \gamma_3)$,$p_0 \asymp c(s)$. *Тогда*

$$\Delta_N \ll \frac{m_0(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{cN} + N^{-1}(\beta_3^2 + \sigma^2\gamma_{2,2})(\frac{1}{|q_{13}|^{13}} + \frac{1}{|q_{13}|^6})+ \tag{4.21}$$

$$+\frac{|q_9|^{-9}}{cN}\cdot(\beta_4 + \beta_3^2 + \sigma^2 + \gamma_3 + \sigma^2\gamma_3 + \gamma_{2,2} + \sigma^2\gamma_{2,2}),$$

(ii) Пусть выполнено условие (1) и $s \geq 9$. *Тогда*

$$\Delta_N \ll \frac{m_0(|q_9|^{-9} + max\{1, |q_9|^{-18}\})}{cN} + N^{-1}(\beta_3^2 + \sigma^2\gamma_{2,2})(\frac{1}{|q_9|^9} + \frac{1}{|q_9|^6})+ \tag{4.22}$$

$$+\frac{|q_9|^{-9}}{cN}\cdot(\beta_4 + \sigma^2 + \gamma_3 + \gamma_{2,2}).$$

Список литературы

[1] Ulyanov, V., Götze, F. Uniform approximations in the CLT for balls in euclidian spaces, 00-034, SFB 343. University of Bielefeld, 2000, 26 p. <http://www.math.uni-bielefeld.de/sfb343/preprints/pr00034.pdf.gz>.

[2] Bentkus, V., Götze, F. Optimal bounds in non-Gaussian limit theorems for U-statistics. Ann.Prob. 27, no.1, 454–521 (1999)

[3] S.A. Bogatyrev, Götze F., V.V. Ulyanov. Non-uniform bounds for short asymptotic expansions in the CLT for balls in a Hilbert space, Journal of Multivariate Analysis 97 (2006) 2041 – 2056,<http://www.elsevier.com/locate/jmva>

РЕФЕРАТЫ

Т. В. Андреева, С. С. Лебедев. О ЗАМКНУТЫХ ЛЕКСИКОГРАФИЧЕСКИХ ОТРЕЗКАХ В k–ЗНАЧНОЙ n–МЕРНОЙ РЕШЕТКЕ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 5–10. Получено необходимое и достаточное условие замкнутости лексикографических отрезков в слоях n–мерной k–значной решетки. Подсчитано их количество.

Библиография 8 раб.

А. И. Аристов. ИССЛЕДОВАНИЕ КАЧЕСТВЕННЫХ СВОЙСТВ РЕШЕНИЙ ОДНОГО НЕЛИНЕЙНОГО СОБОЛЕВСКОГО УРАВНЕНИЯ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 11–22. Работа посвящена изучению свойств решений одного уравнения соболевского типа со степенными нелинейностями специального вида. Рассмотрены задача Коши и начально-краевая задача. Для задачи Коши доказана глобальная по времени однозначная разрешимость, исследовано асимптотическое поведение решений при больших временах. Для начально-краевой задачи найдены достаточные условия локальной и глобальной по времени разрешимости, в случае локальной (но не глобальной) разрешимости найдены двусторонние оценки времени разрушения решения для двух подклассов нелинейностей, входящих в рассматриваемое уравнение.

Библиография 6 раб.

В. С. Федорова. SFE-ЗАМКНУТЫЕ КЛАССЫ ТРЕХЗНАЧНОЙ ЛОГИКИ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 23–33. В статье на множестве функций многозначной логики вводится оператор SFE-замыкания, основывающийся на системах функциональных уравнений. Строится полная решетка SFE-замкнутых классов на множестве функций трехзначной логики.

Библиография 17 раб.

М. А. Федюков. РЕКОНСТРУКЦИЯ ФОРМЫ И ТЕКСТУРЫ ГОЛОВЫ ЧЕЛОВЕКА ПО НАБОРУ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ СИНТЕТИЧЕСКИХ ГИБКИХ МОДЕЛЕЙ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 34–41. В статье представлен краткий обзор методов реконструкции моделей головы человека, описывается общая схема и разработанный алгоритм построения модели головы с использованием синтетических гибких моделей

Библиография 20 раб.

В. Б. Ларионов. О НАДСТРУКТУРЕ КЛАССОВ МОНОТОННЫХ ФУНКЦИЙ В МНОГОЗНАЧНЫХ ЛОГИКАХ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 42–59. В работе изучены основные вопросы о строении надструктуры замкнутых классов монотонных k-значных функций.

Библиография 7 раб.

Н. А. Найденов. ИССЛЕДОВАНИЕ МЕТРИЧЕСКОГО АНАЛОГА МЕТОДА ГЛАВНЫХ КОМПОНЕНТ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 60–69. В данной статье предложен и исследован аналог метода главных компонент для метрических описаний. На получаемое новое описание малой размерности накладывается дополнительное ограничение в виде условия, что каждая новая компонента описания должна удовлетворять аксиомам полуметрики. Рассматриваемый метод назван метрическим методом главных компонент.

Библиография 8 раб.

О. А. Швейкина. РАВНОСХОДИМОСТЬ В СМЫСЛЕ ЧЕЗАРОВСКИХ СРЕДНИХ РАЗЛОЖЕНИЙ В РЯДЫ ПО СОБСТВЕННЫМ ФУНКЦИЯМ ОПЕРАТОРОВ ШТУРМА-ЛИУВИЛЛЯ С ПОТЕНЦИАЛАМИ - РАСПРЕДЕЛЕНИЯМИ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 70–78. Статья посвящена изучению равносходимости разложений в ряды по собственным функциям операторов Штурма-Лиувилля с потенциалами-распределениями

Библиография 16 раб.

А. В. Столяров. ЯЗЫК СИ И НАЧАЛЬНОЕ ОБУЧЕНИЕ ПРОГРАММИРОВАНИЮ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 79–90. В статье рассматриваются причины недопустимости использования языка Си в качестве первого языка при обучении начинающих программистов

Библиография 0 раб.

Заячковский Антон Олегович. МОДЕЛИРОВАНИЕ ЭКОЛОГИЧЕСКИХ РИСКОВ ЗАГРЯЗНЕНИЯ ПОВЕРХНОСТНЫХ ВОД ОКЕАНОВ И МОРЕЙ НА ОСНОВЕ ТЕОРИИ СОПРЯЖЕННЫХ УРАВНЕНИЙ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 91–98. В настоящей работе предлагается алгоритм для количественного подсчета экологического риска на основе математических моделей и методов, предназначенных для решения задач природоохранного направления. Используется методика построения экологических оценок для областей-рецепторов, базирующаяся на прямом и обратном моделировании, теории чувствительности моделей и математической теории риска.

Библиография 5 раб.

Т. А. Зубайраев. ОБ АСИМПТОТИЧЕСКОМ АНАЛИЗЕ U-СТАТИСТИК: ОЦЕНКА ТОЧНОСТИ АППРОКСИМАЦИИ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ // СБОРНИК СТАТЕЙ МОЛОДЫХ УЧЁНЫХ факультета ВМК МГУ выпуск №7 (2010 г.), стр. 99–108. В статье приводится оценка аппроксимации степенного порядка и краткое доказательство результата

Библиография 3 раб.
