

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ В СИСТЕМАХ ИНФОРМИРОВАНИЯ И ИНФОРМАЦИОННОГО ОБМЕНА В ОБЩЕСТВЕ¹

В.Э. Джашитов, В.М. Панкратов, А.Ф. Резчиков, А.Э. Джашитов

Построены и исследованы математические модели, позволяющие решать задачи управления информационным обменом в обществе. Для построения и анализа математических моделей предложен метод информационных элементарных балансов и применены методы нелинейной динамики систем. Приведены результаты численных расчетов, получены качественные и количественные оценки, подтверждающие работоспособность построенных моделей. Даны рекомендации по оптимизации информационного обмена в общественных структурах.

Ключевые слова: математическое моделирование, информационный обмен, управление в социальных системах, нелинейные динамические системы.

ВВЕДЕНИЕ И ИСХОДНЫЕ ПРЕДПОЛОЖЕНИЯ

В современном мире существует множество систем и структур информирования и информационного обмена в социальных системах. Их можно условно разделить на следующие основные группы (рис. 1):

— средства массовой информации, охватывающие практически все цивилизованное человечество (телевидение, регулярно выходящие массовые печатные издания и др.) — они играют роль *глобальной системы*;

— Интернет, охватывающий существенно меньшие, но все же значительные слои общества — он играет роль *условно локальной системы*;

— средства индивидуальной информации (обмен информацией происходит на индивидуальном уровне по телефонной или мобильной связи, вербально или письменно и т. п.) — назовем эти средства *сугубо локальной системой*.

Все эти системы и структуры могут работать как с обратной связью, так и без неё и представляют собой достаточно сложные динамические системы с распределенными параметрами. Например, с од-

ной стороны, Интернет можно считать в определенной мере промежуточным и одновременно связующим звеном между *глобальными* и *сугубо локальными* системами. С другой стороны, Интернет — это чрезвычайно быстро развивающаяся система, обладающая возможностями, близкими к возможностям средств массовой информации, и важным свойством реализации обратной связи при информационном обмене. В дальнейшем будем рассматривать Интернет как важное средство информирования и информационного обмена в обществе.

В настоящее время широко проводятся важные и актуальные фундаментальные исследования по математическим моделям в социально-экономических системах, например, [1, 2]. Вместе с тем попытки применения математических моделей и ме-

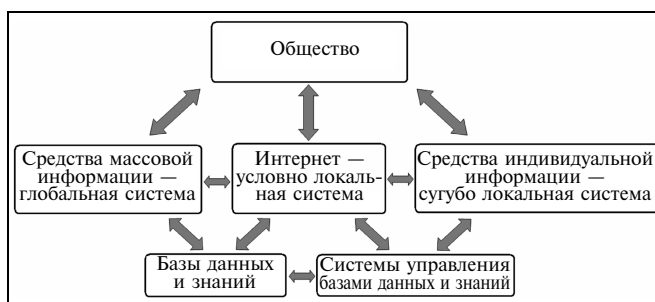


Рис. 1. Структура информирования и информационного обмена

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 08-08-00084а).



тодов (например, метода элементарных балансов [3] и методов нелинейной динамики [4, 5]) к решению задач управления в социальных науках, связанных с информационным обменом, приводят чаще к новому осмыслению общих проблем методологии этих наук, чем к имеющим реальный физический смысл решениям задач [6–8]. При составлении математических моделей в этих задачах принципиальный вопрос заключается в том, какие переменные действительно описывают сложную динамическую систему и как они связаны между собой. Кроме того, в задачах информационного обмена в сложных социальных системах многие из параметров системы неизвестны и (или) неуправляемы. Помочь здесь может лишь лучшее понимание влияния «лишних» параметров или анализ моделей, учитывающих большее их число или дополнительные (возможно, нелинейные) связи.

Авторы считают для себя важным стремиться к той ситуации, когда современный математический формализм не затруднял бы, а наоборот, помогал наиболее ясно отображать сущность процессов информационного обмена, постановки задач и методов их решения.

Сформулируем некоторые принципы, положения и допущения, касающиеся методов и математических моделей, которые предлагаются в настоящей работе для исследования конкретных задач информационного обмена в обществе с помощью имеющихся средств.

Для построения математических моделей в качестве одного из основных предлагается *метод информационных балансов*.

Введем *основные* предположения.

- Основной принцип метода информационных балансов: обмен информацией (в отличие от обмена энергией [3]) *не приводит* к уменьшению информации у дающего ее, а приводит к увеличению информации у принимающего ее, без потери (или почти без потери) информации у дающего.

Количество информации у участников информационного обмена все же уменьшается, но только путем ее частичной «устареваемости», «забываемости» и т. п.

- Базовая гипотеза для составления обыкновенных дифференциальных уравнений, определяющих основу метода информационных балансов: имеется сколь угодно большое, но ограниченное множество A людей (граждан), разделенных на группы X , Y , Z и S и имеющих возможность обмена информацией с помощью Интернета, индивидуальных (и) или массовых средств информации:

в группу X входят люди, заинтересованные (назовем их активными) в создании некоторого «общества» (например, структуры местного самоуправления, товарищества собственников жилья, сознательного электората при проведении выбо-

ров, болельщиков футбольного клуба и т. д.) для достижения поставленной цели (например, оптимального функционирования структуры, достижения определенного результата выборов, качественного ремонта дома, поездки в отдаленный район на матч клуба и т. д.), но неинформированные (как создать «общество», какие правовые документы нужны, что это дает, каковы его экономическая и социальная основа и структура, полномочия его исполнительных органов и т. п.);

в группу Y — заинтересованные (активные) и информированные;

в группу Z — незаинтересованные (неактивные), но информированные;

в группу S — незаинтересованные (неактивные), и неинформированные.

Будем обозначать число людей в группе той же буквой, что и группу.

Из основного принципа метода информационных балансов следует, что в такой информационной системе должны отсутствовать колебательные движения, а значения $Y(t)$ и $Z(t)$ могут только возрастать, а $X(t)$ и $S(t)$ — только убывать (информированных людей становится все больше, а неинформированных все меньше в процессе обмена информацией). Для начала процесса обмена информацией численности людей в группах Y и Z не должно быть нулевыми. Будем полагать справедливыми начальными условиями $0 < Y_0 \leq Z_0 \leq X_0 \leq S_0$.

Полагаем также, что выполняются условия $Y(t) > 0$, $Z(t) > 0$, $X(t) \geq 0$ и $S(t) \geq 0$ и информация может свободно передаваться от информированных групп к неинформированным.

Введем и некоторые *дополнительные* важные предположения.

- В реальной жизни позиция «активности» или «неактивности» граждан в группах X и S может измениться на противоположную по мере поступления к ним информации. Другими словами, часть граждан в этих группах не вполне определилась в своей позиции и определяется в ней по мере поступления информации.

Свойство «активности» в группах Y и Z информированных граждан полагаем неизменяемым, поскольку они информированы и определились в своей позиции.

Таким образом, уже в упрощенной (линейной) постановке учитывается то обстоятельство, что число граждан группы Z (неактивных, но информированных) может увеличиваться не только за счет численности группы S (неактивных и неинформированных), но и за счет численности группы X (активных, но неинформированных), в которой по мере поступления информации меняются сразу два свойства на противоположные.

Аналогично, число граждан группы Y (активных и информированных) может увеличиваться не

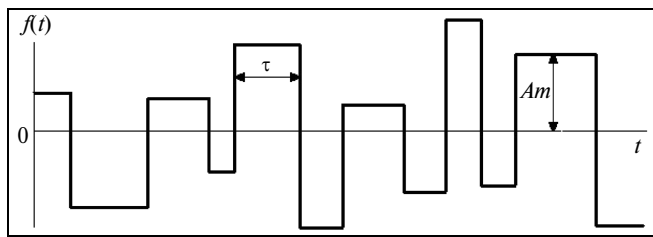


Рис. 2. Математическая модель стационарного импульсного случайного процесса Пуассона в информационном обмене

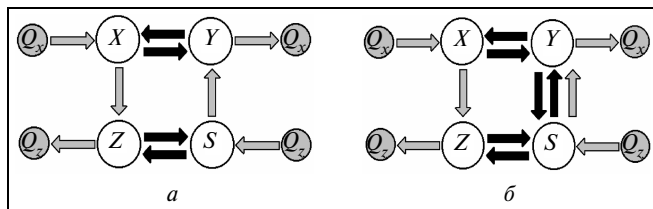


Рис. 3. Схемы изменения численности граждан в группах в процессе информационного обмена:
а — линейная постановка; б — нелинейная постановка

только за счет численности группы X при поступлении к ним необходимой информации, но и за счет численности группы S , в которой, по мере поступления информации, меняются на противоположные сразу два свойства, и пополняется, тем самым, группа Y .

- Информация в ту или иную группу поступает как от соседних групп, так и от посторонних (не входящих ни в одну из групп) источников информации.
- Характер поступающей информации от посторонних источников носит, вообще говоря, случайный характер, что обуславливает случайные изменения численности групп.
- Суммарное число A граждан изменяется (как и в реальной жизни) случайным образом около некоторого среднего значения с некоторым среднеквадратическим отклонением.
- Математическую модель случайных процессов, используемую в рассматриваемом информационном обмене, примем в виде [5, 7] стационарного импульсного случайного процесса Пуассона (рис. 2).

Амплитуда Am случайной функции $f(t)$ и ширина импульсов τ изменяются случайно с заданными математическими ожиданиями $M(A)$ и $M(\tau)$ и среднеквадратическими отклонениями $\sigma(A)$ и $\sigma(\tau)$. Между моментами изменений функция $f(t)$ полагается постоянной. Этот процесс стационарный и эргодический.

- Рассматриваются две постановки задачи — линейная (когда не учитываются нелинейные эффекты взаимодействий между группами в процессе информационного обмена) и нелинейная (когда

эти эффекты учитываются). В процессе информационного обмена численности групп изменяются в соответствии с показанными на рис. 3 связями. Связи между группами отражены черными стрелками. Дополнительные и нелинейные связи и элементы Q_x и Q_z , учитывающие поступление информации со стороны, выделены серым.

Ставится задача: путем управляемого оптимизированного информационного обмена между группами *максимально увеличить* численность группы Y (людей активных и информированных) $Y \rightarrow Y_{\max}$.

1. ЛИНЕЙНАЯ ПОСТАНОВКА ЗАДАЧИ

При введенных предположениях система линейных обыкновенных дифференциальных уравнений информационных балансов может быть представлена в виде:

$$\dot{X} = -q_{yx}X + Q_x + Q_{xc}(t), \quad (1)$$

$$\dot{Y} = -q_{yx}Y + q_{yx}(Y_0 + X_0) - (Q_x + Q_{xc}(t)) + q_{ys}(Y - S), \quad (2)$$

$$\dot{Z} = -q_{zs}Z + q_{zs}(Z_0 + S_0) - (Q_z + Q_{zc}(t)) + q_{zx}(Z - X), \quad (3)$$

$$S = A + A_c(t) - X - Y - Z, \quad (4)$$

где q_{yx} — коэффициент, характеризующий интенсивность убывания числа граждан в группе X при поступлении в нее информации или интенсивность возрастания числа граждан в группе Y и зависящий от многих факторов; q_{zs} — коэффициент, характеризующий интенсивность убывания числа граждан в группе S при поступлении в него информации или интенсивность возрастания числа граждан в группе Z и также зависящий от многих факторов; q_{ys} — коэффициент, характеризующий интенсивность возрастания числа граждан в группе Y благодаря изменению свойств неактивности и неинформированности граждан в группе S на противоположные; q_{zx} — коэффициент, характеризующий интенсивность возрастания числа граждан в группе Z из-за изменения свойств активности и неинформированности граждан в группе X на противоположные; Q_x и Q_z — математическое ожидание скорости изменения числа граждан в группах X и Y и группах S и Z , соответственно, благодаря поступлению посторонней информации; A — суммарное число людей в группах в детерминированной постановке или математическое ожидание этой величины, когда учитывается случайный характер процессов; $Q_{xc}(t)$, $Q_{zc}(t)$ и $A_c(t)$ —



случайные составляющие соответствующих параметров.

Знак минус перед членом, характеризующим источник посторонней информации в уравнении (2) означает, что с увеличением скорости изменения числа граждан в группе X одновременно уменьшается скорость изменения числа граждан в группе Y . Аналогичные рассуждения справедливы и для уравнения (3).

Полагаем, что $q_{ys} < q_{yx}$ и $q_{zx} < q_{xs}$.

Управляющие параметры могут быть выбраны из коэффициентов q_{ys} , q_{yx} , q_{zx} , q_{zs} , ... по мере выяснения значимости их влияния на достижение заявленной цели $Y \rightarrow Y_{\max}$ и возможности их практической реализации.

Размерность величин Q_x , Q_z , $Q_{xc}(t)$ и $Q_{zc}(t)$ [число граждан/ед. времени], размерность коэффициентов q_{ys} , q_{yx} , q_{zx} и q_{zs} [1/ед. времени].

Для получения качественных оценок аналитически получим стационарное решение линейной системы (1)–(4). Для этого положим $\dot{X} = \dot{Y} = \dot{Z} = 0$. Пренебрегая случайными составляющими, получаем систему алгебраических уравнений:

$$-q_{yx}X + Q_x = 0,$$

$$-q_{yx}Y + q_{yx}(Y_0 + X_0) - Q_x + q_{ys}(Y - S) = 0,$$

$$-q_{zs}Z + q_{zs}(Z_0 + S_0) - Q_z + q_{zx}(Z - X) = 0,$$

$$S = A - X - Y - Z.$$

Отсюда находим стационарное решение системы (1)–(4):

$$X = Q_x/q_{yx}, \quad (5)$$

$$Z = \frac{q_{zs}(Z_0 + S_0) - Q_z - q_{zx}X}{q_{zs} - q_{zx}}, \quad (6)$$

$$Y = \frac{q_{yx}(Y_0 + X_0) - q_{ys}A + q_{ys}X + q_{ys}Z - Q_x}{q_{yx} - 2q_{ys}}, \quad (7)$$

$$S = A - X - Y - Z. \quad (8)$$

Положим в частном случае, что информация со стороны в группы не поступает, т. е. $Q_x = Q_z = 0$, и свойства активности и неактивности в группах X и S не изменяются, т. е. $q_{ys} = q_{zx} = 0$.

Тогда поставленная задача достижения максимума функции Y в линейной постановке имеет в установившемся режиме тривиальное решение $X = 0$, $Z = Z_0 + S_0$, $Y = Y_0 + X_0$, $S = 0$.

Анализ полученных соотношений (5)–(8) показывает, что учет, даже в линейной постановке, поступления в группы информации со стороны (Q_x , $Q_z \neq 0$), а также учет возможного изменения

свойств активности и неактивности на противоположные в группах X и S (q_{ys} , $q_{zx} \neq 0$) «ухудшает» решение поставленной задачи достижения максимума функции Y .

Численность групп X и S уменьшается уже не до нуля, а до некоторых постоянных значений, определяемых интенсивностью поступления посторонней информации и изменением у части граждан в этих группах свойств активности и неактивности на противоположные. Соответственно уменьшается число людей в группе Z и, что особенно важно, в группе Y .

Например, полагая в выражениях (5)–(8) $q_{ys} \ll q_{yx}$, $q_{zx} = 0$, $Q_x = Q_z = 0$, получим $X = 0$, $Z = Z_0 + S_0$, $Y \approx Y_0 + X_0 - \frac{q_{ys}}{q_{yx}}(A - Z_0 - S_0)$. Как видим, $Y < Y_0 + X_0$, т. е. действительно учет в линейной постановке возможного изменения свойств неактивности (или активности) на противоположные в группах X и S ($q_{ys} \neq 0$) «ухудшает» решение задачи достижения максимума Y .

2. НЕЛИНЕЙНАЯ ПОСТАНОВКА ЗАДАЧИ

Учтем взаимное влияние изменения числа граждан в группах на скорость изменения числа граждан в каждой группе. Будем также учитывать влияние взаимодействия между двумя различными группами на скорость изменения числа граждан в данной группе.

Методика постановки и решения такого рода нелинейных задач, составления и исследования нелинейных обыкновенных дифференциальных уравнений подробно рассмотрена в работах [3, 4, 7, 8].

В конечном итоге система нелинейных обыкновенных дифференциальных уравнений информационных балансов примет следующий вид:

$$\dot{X} = -q_{yx}X + Q_x + Q_{xc}(t) + b_{yz}YZ + b_{ys}YS + b_{zs}ZS, \quad (9)$$

$$\dot{Y} = -q_{yx}Y + q_{yx}(Y_0 + X_0) - (Q_x + Q_{xc}(t)) + q_{ys}(Y - S) + b_{zx}ZX + b_{xs}XS + b_{zs}ZS, \quad (10)$$

$$\dot{Z} = -q_{zs}Z + q_{zs}(Z_0 + S_0) - (Q_z + Q_{zc}(t)) + q_{zx}(Z - X) + b_{yx}YX + b_{ys}YS + b_{xs}XS, \quad (11)$$

$$S = A + A_c(t) - X - Y - Z. \quad (12)$$

Присутствие члена YZ в уравнении (9) означает, что состав группы X изменяется (уменьшается) как за счет изменения (увеличения) числа граждан в группе Y , так и в группе Z . Коэффициент b_{yz} характеризует интенсивность этого влияния. Аналогичен смысл и других нелинейных членов.

Проведенный качественный анализ системы (9)–(12) апробированными методами нелинейной динамики [3–5] (оценка значимости коэффициентов, нахождение стационарных точек, линеаризация системы около этих точек, построение матрицы устойчивости, получение и исследование характеристического уравнения, получение условий для устойчивых стационарных точек как функций параметров системы и др.) показал, что существуют значения численностей групп Y , X , Z и S , удовлетворяющие требованиям поставленной задачи $Y \rightarrow Y_{\max}$.

Для количественного решения этой задачи и получения конкретных практических результатов с учетом различных факторов использовались построенные линейная (1)–(4) и нелинейная (9)–(12) математические модели информационной системы.

Системы уравнений интегрировались численными методами с помощью стандартного и специально разработанного программного обеспечения.

3. КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ И АНАЛИЗ РЕЗУЛЬТАТОВ

Рассмотрим систему линейных уравнений (1)–(4) с детерминированными значениями параметров, т. е. $Q_{xc} = Q_{zc} = 0$, $A_c = 0$.

Все данные по числу граждан в группах будем рассматривать в процентах по отношению к суммарному числу A людей в группах, которое принято за 100 %. Время переходных процессов берется в условных единицах, отнесенных к некоторому номинальному значению, которое должно быть определено экспериментально.

Основные исходные данные: начальные условия $X_0 = 0,3 \cdot A$; $Y_0 = Z_0 = 0,02 \cdot A$; $S_0 = 0,66 \cdot A$; коэффициенты интенсивности обмена гражданами между группами $q_{yx} = q_{zs} = 0,02$.

На рис. 4 представлены результаты численного интегрирования системы линейных уравнений (1)–(4).

Как видим, учет источников посторонней информации и возможного изменения свойств активности и неактивности граждан группы X и S приводит к существенному ухудшению качества переходных процессов.

Значительно (в 1,5–2 раза) возрастает время переходных процессов. Не все граждане переходят из групп X и S в группы Y и Z и хуже выполняется достижение поставленной цели — обеспечение максимального числа граждан в группе Y . Без учета источников посторонней информации и возможного изменения свойств активности и неактивно-

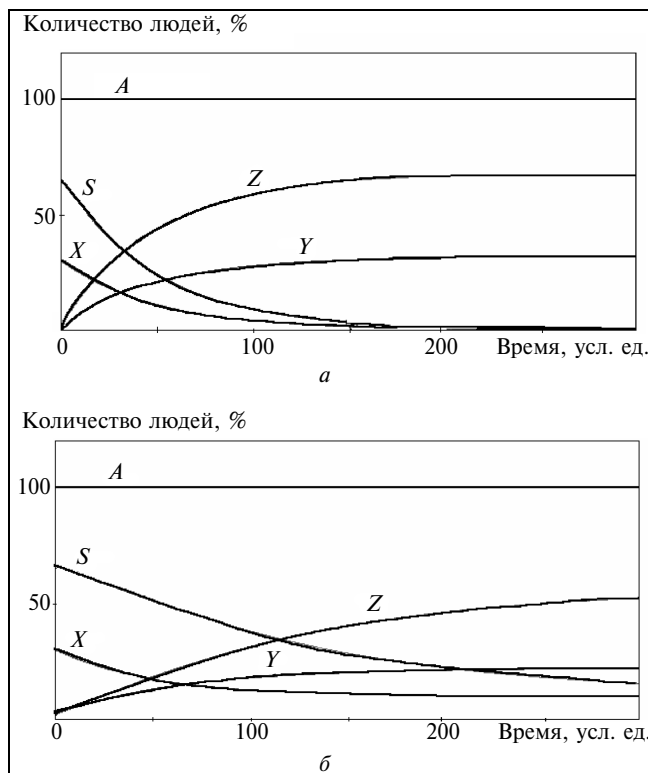


Рис. 4. Переходные процессы в линейной детерминированной системе (1)–(4) информационного обмена:

a — не учитывается поступление информации в группы со стороны ($Q_x = Q_z = 0$) и свойства активности и неактивности в группах X и S не изменяются ($q_{ys} = q_{zx} = 0$); b — учитываются поступление информации со стороны ($Q_x = 0,2$; $Q_z = 0,7$) и изменения свойств активности и неактивности на противоположные в группах X и S ($q_{ys} = 0,001$; $q_{zx} = 0,01$)

сти граждан групп X и S после окончания переходных процессов число граждан в группе Y составило $Y = 0,32 \cdot A$. С учетом названных факторов это число составило $Y = 0,22 \cdot A$. Полученные количественные оценки полностью подтверждают данные качественного анализа.

На рис. 5 представлены результаты численного интегрирования линейной модели (1)–(4) с учетом случайных составляющих параметров системы — среднеквадратических отклонений $\sigma(Q_{xc}) = 0,15$; $\sigma(Q_{zc}) = 0,6$; $\sigma(A_c) = 5$; $\tau = 10$; $\sigma(\tau) = 1$.

Как видим, переходные процессы еще более ухудшились. Появились случайные флуктуации около средних значений. «Амплитуды» и «частоты» этих флуктуаций определяются «амплитудами» и «частотами» случайных внешних воздействий — источников посторонней информации и флуктуациями суммарного числа граждан в группах. Среднеквадратические отклонения по группам граждан после окончания переходных процессов приняли следующие значения: $\sigma(X) = 0,02 \cdot A$, $\sigma(Y) = 0,02 \cdot A$, $\sigma(Z) = 0,04 \cdot A$, $\sigma(S) = 0,05 \cdot A$.

В целом математическое моделирование линейной системы с учетом рассмотренных детерминированных и случайных возмущений (наличие источников посторонней информации и случайные колебания суммарного числа людей в группах) позволило получить следующие осредненные данные по решению поставленной задачи.

Начальные численности групп: $X_0 = 0,3 \cdot A$; $Y_0 = Z_0 = 0,02 \cdot A$; $S_0 = 0,66 \cdot A$, конечные — $X = (0,1 \pm 0,02)A$; $Y = (0,22 \pm 0,02)A$; $Z = (0,52 \pm 0,04)A$; $S = (0,16 \pm 0,05)A$.

Рассмотрим систему нелинейных уравнений (9)–(12).

Исходные данные: $q_{yx} = q_{zs} = 0,02$; $b_{zs} = b_{xs} = 0$; $b_{yx} = 0$; $b_{yz} = 0,002$; $b_{ys} = 0,5$; $b_{zx} = 0,5$; $q_{ys} = 0,001$; $q_{zx} = 0,01$; $Q_x = 0,2$; $Q_z = 0,7$; $\sigma(Q_{xc}) = 0 \div 0,15$; $\sigma(Q_{zc}) = 0 \div 0,6$; $\sigma(A_c) = 0 \div 5$; $\tau = 10$; $\sigma(\tau) = 1$.

На рис. 6 представлены результаты численного интегрирования системы уравнений (9)–(12).

Математическое моделирование системы уравнений (9)–(12) позволило получить следующие осредненные данные по решению поставленной задачи: $X_0 = 0,3 \cdot A$; $Y_0 = Z_0 = 0,2 \cdot A$; $S_0 = 0,66 \cdot A$; $X = (0,3 \pm 0,03)A$; $Y = (0,97 \pm 0,02)A$; $Z = (0,3 \pm 0,03)A$; $S = (0 \pm 0)A$.

Учет нелинейных взаимодействий между группами граждан в процессе информационного обмена позволил существенно (примерно в 10 раз) сократить время переходных процессов и значительно точнее (в 4,4 раза) решить поставленную задачу по сравнению с линейной постановкой.

Вместе с тем в результате проведенных компьютерных экспериментов на нелинейной математической модели выявился следующий важный факт. Учет случайной составляющей в суммарном числе граждан A позволил установить, что, при определенных параметрах нелинейной системы и

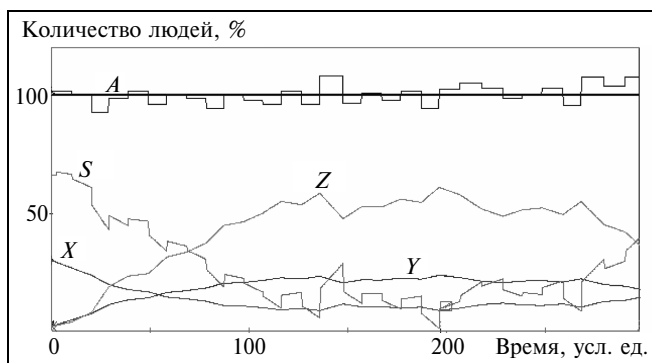


Рис. 5. Переходные процессы в детерминированной системе (1)–(4) информационного обмена с учетом случайных составляющих факторов посторонней информации и суммарного числа граждан

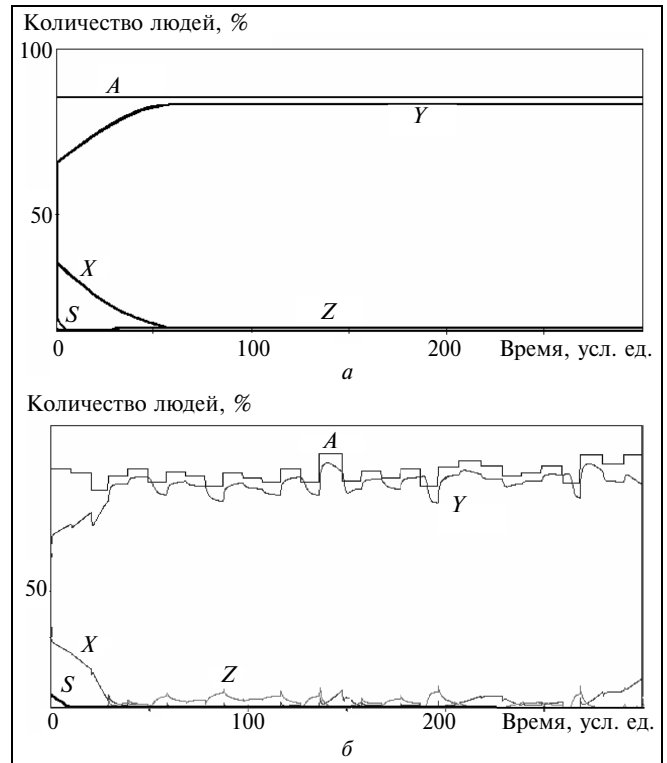


Рис. 6. Переходные процессы в детерминированной системе (9)–(12) информационного обмена:

a — с учетом только детерминированных факторов; b — с учетом детерминированных и случайных факторов ($\sigma(Q_{xc}) = 0,15$; $\sigma(Q_{zc}) = 0,6$; $\sigma(A_c) = 5$)

некотором критическом значении среднеквадратического отклонения $\sigma(A_c)$ от математического ожидания суммарного числа граждан, переходные процессы теряют устойчивость (система «разваливается») и поставленная цель $Y \rightarrow Y_{\max}$ не достигается. Это отклонение составило величину $\sigma(A_c) = 5$ для рассматриваемой системы с выбранными параметрами. Другими словами, данная информационная система с учтенными нелинейными взаимодействиями может быть неработоспособна при определенных параметрах системы и если флуктуации суммарного числа граждан в группах будут больше $3\sigma(A_c) \approx \pm 0,15 \cdot A$.

РЕКОМЕНДАЦИИ И ЗАКЛЮЧЕНИЕ

Сформулируем ряд рекомендаций по стратегии создания системы информационного обмена между группами населения в целях достижения максимального увеличения числа активных и информированных граждан $Y \rightarrow Y_{\max}$.

1. Прежде всего, необходимо организовать интенсивное регулярное информационное взаимо-

действие группы информированных и неактивных граждан с группой неинформированных и активных граждан $Z \rightarrow X$, а также группы информированных и активных с группой неинформированных и неактивных граждан $Y \rightarrow S$. Это можно сделать, например, через Интернет, методом кутовых собраний, путем создания информационно-консультативных пунктов, когда информация за единицу времени передается и становится доступной сразу нескольким гражданам и группам граждан.

2. Информационное взаимодействие между группой активных и информированных граждан и группой активных и неинформированных граждан $Y \rightarrow X$ может быть обычным прямым, а не через посредство граждан из других групп, его не надо форсировать (нет большой необходимости проводить для них кутовые собрания), но оно обязательно должно быть.

3. На информационное взаимодействие между группой неактивных и информированных граждан и группой неактивных и неинформированных граждан $Z \rightarrow S$ можно не обращать серьезного внимания, оно слабо влияет на достижение желаемого результата $Y \rightarrow Y_{\max}$.

4. Общее число A граждан, участвующих в информационном обмене должно быть достаточно велико (1000 чел. и более). Тогда более ощутимо достигается цель $Y \rightarrow Y_{\max}$ и, в зависимости от начальных условий, возможно, кроме того, сокращение времени достижения этого результата. Это может оказаться очень важным обстоятельством, когда необходимо организовать информационный обмен в достаточно сжатые сроки.

5. Следует принять возможные меры, направленные на предотвращение больших изменений (до 15 % по данным моделирования) суммарного числа A граждан, участвующих в информационном обмене; в противном случае возможно разрушение системы.

6. При выполнении рекомендаций 1–5 действие дополнительных возмущающих факторов, связанных с наличием источников посторонней информации, колебаний суммарного числа A и их случайного характера, и возможного изменения свойств активности и неактивности в группах X и S не приведет к существенному ухудшению картины информационного обмена и поставленная задача будет решена.

В заключение отметим, что построенные математические модели — это первые приближения в моделировании таких сложных социологических процессов, как взаимодействие информационных потоков между группами населения разной активности и степени информированности, в целях до-

стижения в течение реального срока желаемого результата.

В дальнейшем следует учесть еще многие факторы, имеющие место в реальной жизни. Это существование других видов групп населения, различия их по составу и плотности, сложные зависимости параметров системы от различных факторов и др. Необходимо учесть также дозировку информации, ее объем, последовательность, повторяемость, интенсивность распространения и многое другое.

Однако уже имеющаяся математическая модель не только достаточно адекватно объясняет сущность процессов информирования населения и его влияния на активность граждан, но и позволяет предпринимать реальные шаги для достижения желаемых результатов.

ЛИТЕРАТУРА

1. Новиков Д.А. Математические модели формирования и функционирования команд. — М.: Изд-во физ.-мат. лит., 2008. — 184 с.
2. Леонов Г.А. Динамические принципы прогнозирования и управления // Проблемы управления. — 2008. — № 5. — С. 31–35.
3. Джашинов В.Э., Панкратов В.М. Датчики, приборы и системы авиакосмического и морского приборостроения в условиях тепловых воздействий / Под ред. В.Г. Пешехонова. — СПб.: ГНЦ РФ ЦНИИ «Электроприбор», 2005. — 404 с.
4. Бутенин Н.В., Неймарк Ю.И., Фуфаев Н.А. Введение в теорию нелинейных колебаний. — М.: Наука, 1987. — 384 с.
5. Анищенко В.С., Вадивасова Т.Е., Астахов В.В. Нелинейная динамика хаотических и стохастических систем. — Саратов: Изд-во Саратовского ун-та, 1999. — 368 с.
6. Николис Г., Пригожин И. Познание сложного. — М.: Мир, 1990.
7. Вышинский С.А., Деветьярова А.А. О динамике некоторых математических моделей в социологии // Изв. вузов. Сер. Прикладная нелинейная динамика. — 1994. — № 2. — С. 17–26.
8. Шустер Г. Детерминированный хаос. — М.: Мир, 1988. — 240 с.

Статья представлена к публикации членом редколлегии В.В. Кульбой.

Джашинов Виктор Эмануилович — д-р техн. наук, зав. лабораторией, Институт проблем точной механики и управления РАН, г. Саратов, ☎(8452) 22-16-93,

Панкратов Владимир Михайлович — д-р техн. наук, зам. директора по науке, Институт проблем точной механики и управления РАН, г. Саратов, ☎(8452) 22-10-48,

Резчиков Александр Федорович — чл.-корр. РАН, д-р техн. наук, директор, Институт проблем точной механики и управления РАН, г. Саратов, ☎(8452) 22-23-76,

Джашинов Александр Эмануилович — канд. техн. наук, доцент, Саратовский государственный технический университет, ☎(8452) 23-37-55, ✉iptmuran@san.ru.



УПРАВЛЕНИЕ ОДНОСЕКТОРНОЙ ЭКОНОМИКОЙ ПРИ ОГРАНИЧЕНИЯХ НА НАКОПЛЕНИЕ И ПОТРЕБЛЕНИЕ

Н.С. Дёмин, Е.В. Кулешова

На классе линейно-однородных производственных функций исследована задача оптимального управления односекторной экономикой на конечном интервале времени при наличии ограничений на накопление и потребление с учетом производственных затрат и налоговых отчислений. Основной результат сформулирован в виде магистральной теоремы. Получено золотое правило накопления, определяющее распределение произведенного экономикой продукта на магистрали. Результаты конкретизируются для случая производственной функции Кобба — Дугласа.

Ключевые слова: управление, односекторная экономика, производственная функция, магистральная теорема, золотое правило накопления.

ВВЕДЕНИЕ

При анализе динамических моделей макроэкономики экономическое развитие рассматривается как научно-технический прогресс [1–6], когда находится производственная функция, соответствующая какому-либо типу научно-технического прогресса (нейтральность по Хиксу, Харроду, Солоу и более сложного вида Ф-нейтральности). Распределение произведенного экономикой продукта рассматривается как некоторая задача оптимизации [7–10] — на заданном классе производственных функций решается некоторая задача оптимального управления.

Основной задачей управления односекторной (агрегированной, однопродуктовой) экономикой на конечном интервале времени является задача максимизации потребления за весь плановый период при выполнении условия экономического горизонта в конечный момент времени. При этом решение ищется в предположении, что на отдельных временных интервалах планового периода на накопление может направляться весь произведенный экономикой продукт либо не направляться вообще, когда весь продукт направляется на потребление. Решение получается в форме «магистральной теоремы» [7, 8, 10], суть которой состоит в том, что выход экономики на магистраль и сход экономики с магистрали для удовлетворения

условия экономического горизонта происходит максимально быстро за счет направления всего продукта целиком либо только на потребление, либо только на накопление, а нахождению экономики на магистрали соответствует распределение валового продукта на накопление и потребление согласно принципу «золотого правила накопления» [4, 6].

В данной работе рассматривается задача максимизации интегрального (т. е. за весь плановый период) потребления с дисконтированием при предположениях, в большей мере соответствующих сути макроэкономического процесса, а именно:

— накопление и потребление ограничены сверху и снизу, т. е. даже на отдельных интервалах планового периода целиком весь произведенный продукт не может направляться только на накопление либо только на потребление;

— суммарно на накопление и потребление идет только часть произведенного экономикой продукта после изъятия из него некоторой части на налоги и производственные затраты.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть на интервале времени $t \in [0, T]$, составляющем плановый период, задано соотношение $Y(t) = F(K(t), L(t))$, где $Y(t)$ — произведенный экономикой продукт, $K(t)$ — основные фонды (капитал), $L(t)$ — трудовые ресурсы, $F(K, L)$ — линейно-

однородная производственная функция [5, 7, 10], причем $L(t) = L_0 \exp\{\lambda t\}$, $L_0 > 0$, $\lambda > 0$. Весь продукт делится на четыре части в виде

$$Y(t) = \Psi(t) + I(t) + C(t) + N(t), \quad (1.1)$$

где $I(t)$ — накопление, $C(t)$ — потребление, $N(t)$ — налоговые отчисления, $\Psi(t)$ — производственные затраты с нормой материалоемкости γ , т. е. $\Psi(t) = \gamma Y(t)$, $0 \leq \gamma \leq 1$. Пусть $D(t)$ — прибыль, облагаемая налогом со ставкой u , такой, что $0 \leq u < 1$. Тогда $D(t) = Y(t) - \Psi(t) = (1 - \gamma)Y(t)$, $N(t) = uD(t) = (1 - \gamma)uY(t)$, а $G(t) = D(t) - N(t) = (1 - \gamma)(1 - u)Y(t)$ является прибылью, которая осталась после выплаты налогов и которая направляется на накопление и потребление, т. е.

$$G(t) = I(t) + C(t) = (1 - \gamma)(1 - u)Y(t). \quad (1.2)$$

Пусть $s(t)$ — норма накопления, $0 \leq s_0 \leq s(t) \leq s_1 \leq 1$, а $\tilde{s}(t) = (1 - s(t))$ — норма потребления, т. е.

$$\begin{aligned} I(t) &= s(t)G(t) = (1 - \gamma)(1 - u)s(t)Y(t), \\ C(t) &= (1 - s(t))G(t) = \\ &= (1 - \gamma)(1 - u)(1 - s(t))Y(t). \end{aligned} \quad (1.3)$$

Если $\mu > 0$ есть коэффициент амортизации основных фондов, то для $K(t)$ справедливо дифференциальное уравнение $\dot{K}(t) = I(t) - \mu K(t)$ [5, 7, 8], которое с учетом выражений (1.3) принимает вид

$$\dot{K}(t) = (1 - \gamma)(1 - u)s(t)F(K(t), L(t)) - \mu K(t). \quad (1.4)$$

Перейдя к нормированным (удельным) относительно трудовых ресурсов величинам, получаем с учетом линейной однородности $F(K(t), L(t))$ [5, 7, 10]:

$$\begin{aligned} y(t) &= \frac{Y(t)}{L(t)} = f(k(t)), \quad k(t) = \frac{K(t)}{L(t)}, \quad i(t) = \frac{I(t)}{L(t)}, \\ c(t) &= \frac{C(t)}{L(t)}, \quad n(t) = \frac{N(t)}{L(t)}. \end{aligned} \quad (1.5)$$

Тогда балансовое соотношение (1.1) с учетом выражений (1.3) и (1.5) преобразуется к виду

$$y(t) = \gamma y(t) + i(t) + c(t) + n(t), \quad (1.6)$$

$$\begin{aligned} i(t) &= (1 - \gamma)(1 - u)s(t)y(t), \quad c(t) = \\ &= (1 - \gamma)(1 - u)(1 - s(t))y(t), \quad n(t) = \\ &= (1 - \gamma)uy(t). \end{aligned} \quad (1.7)$$

Приняв в качестве критерия оптимальности удельное потребление с дисконтированием за весь

плановый период, приходим к следующей задаче оптимального управления [11]:

$$\begin{aligned} \dot{k}(t) &= (1 - \gamma)(1 - u)s(t)f(k(t)) - vk(t); \quad t \in [0, T], \\ k(0) &= k_0, \quad k(T) \geq k_T > 0; \quad v = \mu + \lambda; \end{aligned} \quad (1.8)$$

$$\begin{aligned} J &= \int_0^T (1 - \gamma)(1 - u)(1 - s(t))f(k(t))\exp\{-\delta t\}dt \rightarrow \\ &\rightarrow \max_{s_0 \leq \{s(t)\} \leq s_1}. \end{aligned} \quad (1.9)$$

Уравнение (1.8) следует из уравнения (1.4) с учетом соотношений (1.5) и того, что $L(t) = L_0 \exp\{\lambda t\}$. Критерий (1.9), где $\delta > 0$ — коэффициент дисконтирования, следует из соотношений (1.5) и (1.7). Условие $k(T) \geq k_T > 0$ в выражении (1.8) представляет собой условие экономического горизонта, которое определяет, что к моменту окончания планового периода T уровень фондовооруженности $k(T)$ не может опуститься ниже величины k_T .

Замечание. Считаем, что для функции $f(k)$ выполняются неоклассические условия [5, 7, 8, 10]:

$$\begin{aligned} 1) & f(k) > 0, \quad k > 0; \quad f(k) = 0, \quad k = 0; \\ 2) & f'(k) > 0, \quad f''(k) < 0; \\ 3) & f'(0) = \infty, \quad f'(\infty) = 0. \end{aligned} \quad (1.10)$$

Кроме того, потребуем выполнения дополнительно условий

$$0 \leq s_0 < v(v + \delta)^{-1}\alpha(k) < v(v + \delta)^{-1} < s_1 \leq 1, \quad (1.11)$$

где $\alpha(k) = kf'(k)/f(k)$ является коэффициентом эластичности по основным фондам. Так как для линейно-однородных производственных функций $\alpha(k)$ обладает свойством $0 < \alpha(k) < 1$, то условия (1.11) являются корректными.

2. ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Лемма 1. Если $s(t)$ оптимальное управление, то:

$$\begin{aligned} s(t) &= s_1, \quad q(t) > 1; \quad s(t) = s_0, \quad q(t) < 1; \\ s(t) &\in [s_0; s_1], \quad q(t) = 1. \end{aligned} \quad (2.1)$$

Функция $q(t)$ определяется дифференциальным уравнением

$$\begin{aligned} \dot{q}(t) &= [(v + \delta) - (1 - \gamma)(1 - u)s(t)f'(k(t))]q(t) - \\ &- (1 - \gamma)(1 - u)(1 - s(t))f''(k(t)), \end{aligned} \quad (2.2)$$

где $f'(k(t)) = df(k)/dk|_{k=k(t)}$, решения которого удовлетворяют условиям

$$q(T) \geq 0, \quad q(T)[k(T) - k_T] = 0. \quad \blacklozenge \quad (2.3)$$

Доказательства леммы 1 и других утверждений приведены в Приложении.



Лемма 2. Функция $q(t)$ неотрицательная, т. е. $q(t) \geq 0$, $t \in [0, T]$. ♦

Лемма 3. Если $q(t) > 1$, чему соответствует значение $s(t) = s_1$, то области знакопостоянства производных $\dot{k}(t)$ и $\dot{q}(t)$ имеют вид:

$$\begin{aligned} \dot{k} &= 0, \quad k = k_1; \quad \dot{k} > 0, \quad k < k_1; \\ \dot{k} &< 0, \quad k > k_1; \end{aligned} \quad (2.4)$$

$$\begin{aligned} \dot{q} &= 0, \quad q = \Phi_1(k); \quad \dot{q} < 0, \quad q < \Phi_1(k), \quad k < k^*; \\ \dot{q} &> 0, \quad q > \Phi_1(k), \quad k < k^*; \quad \dot{q} > 0, \quad k > k^*. \end{aligned} \quad (2.5)$$

Значения $k = k_1$ и $k = k^*$ такие, что $k^* < k_1$, являются единственными корнями уравнений

$$\begin{aligned} k_1: f(k) &= v[(1 - \gamma)(1 - u)s_1]^{-1}k; \\ k^*: f'(k) &= (v + \delta)[(1 - \gamma)(1 - u)]^{-1}. \end{aligned} \quad (2.6)$$

Для $k \leq k^*$ функция

$$\begin{aligned} \Phi_1(k) &= [(1 - \gamma)(1 - u)(1 - s_1)f(k)] \times \\ &\times [(v + \delta) - (1 - \gamma)(1 - u)s_1f'(k)]^{-1} \end{aligned} \quad (2.7)$$

убывающая, и при этом $\Phi_1(k) > 1$ для $k \in [0; k^*)$, $\Phi_1(k^*) = 1$. ♦

Лемма 4. Если $q(t) < 1$, чему соответствует значение $s(t) = s_0$, то области знакопостоянства $\dot{k}(t)$ и $\dot{q}(t)$ имеют вид:

$$\dot{k} = 0, \quad k = k_2; \quad \dot{k} < 0, \quad k > k_2; \quad \dot{k} > 0, \quad k < k_2; \quad (2.8)$$

$$\begin{aligned} \dot{q} &= 0, \quad q = \Phi_0(k); \quad \dot{q} > 0, \quad q > \Phi_0(k), \quad k > k^*; \\ \dot{q} &< 0, \quad q < \Phi_0(k), \quad k > k^*; \quad \dot{q} < 0, \quad k < k^*. \end{aligned} \quad (2.9)$$

Значение $k = k_2$ такое, что $k^* > k_2$ и является единственным корнем уравнения

$$k_2: f(k) = [v/(1 - \gamma)(1 - u)s_0]k. \quad (2.10)$$

Для $k \geq k^*$ функция

$$\begin{aligned} \Phi_0(k) &= [(1 - \gamma)(1 - u)(1 - s_0)f(k)] \times \\ &\times [(v + \delta) - (1 - \gamma)(1 - u)s_0f'(k)]^{-1} \end{aligned} \quad (2.11)$$

убывающая и при этом $\Phi_0(k) < 1$ для $k > k^*$, $\Phi_0(k^*) = 1$. ♦

Лемма 5. На плоскости (k, q) особая точка $X^* = (k^*; 1)$ является стационарным решением системы (1.8), (2.2). При этом $s(t) \equiv s^* = \text{const}$ определяется двумя эквивалентными формулами

$$\begin{aligned} s^* &= vk^*[(1 - \gamma)(1 - u)f(k^*)]^{-1}, \\ s^* &= [kf'(k^*)/f(k^*)] - \\ &- \delta k^*[(1 - \gamma)(1 - u)f(k^*)]^{-1} \end{aligned} \quad (2.12)$$

и удовлетворяет условию $s_0 < s^* < s_1$, которое раскрывает неопределенность третьего из соотношений (2.1). ♦

Следствие 1. Пусть $\alpha(k)$ и $\beta(k)$ являются коэффициентами эластичности соответственно по основным фондам и трудовым ресурсам. Тогда s^* имеет следующие эквивалентные формулам (2.12) представления:

$$\begin{aligned} s^* &= \alpha(k^*) - \delta k^*[(1 - \gamma)(1 - u)f(k^*)]^{-1}, \\ s^* &= v(v + \delta)^{-1}\alpha(k^*). \end{aligned} \quad (2.13)$$

Так как для линейно-однородных производственных функций [5] $\alpha(k) = kf'(k)/f(k)$, $\beta(k) = 1 - \alpha(k)$, то формулы (2.13) следуют непосредственно из выражений (2.6) и (2.12). ♦

Следствие 2. Норма потребления $\tilde{s}(t) = 1 - s(t)$, для которой выполняются условия $0 \leq \tilde{s}_0 \leq \tilde{s}(t) \leq \tilde{s}_1 \leq 1$, $\tilde{s}_0 = 1 - s_1$, $\tilde{s}_1 = 1 - s_0$, в особой точке определяется следующими эквивалентными формулами

$$\begin{aligned} \tilde{s}^* &= 1 - vk^*[(1 - \gamma)(1 - u)f(k^*)]^{-1}, \\ \tilde{s}^* &= [1 - (k^*f'(k^*)/f(k^*))] + \\ &+ \delta k^*[(1 - \gamma)(1 - u)f(k^*)]^{-1}, \end{aligned} \quad (2.14)$$

$$\begin{aligned} \tilde{s}^* &= \beta(k^*) + \delta k^*[(1 - \gamma)(1 - u)f(k^*)]^{-1}, \\ \tilde{s}^* &= [\delta/(v + \delta)] + [v/(v + \delta)]\beta(k^*) \end{aligned} \quad (2.15)$$

и удовлетворяет условию $\tilde{s}_0 \leq \tilde{s}^* \leq \tilde{s}_1$.

Соотношения (2.14) и (2.15) очевидным образом следуют из соотношений (2.12) и (2.13). ♦

3. МАГИСТРАЛЬНАЯ ТЕОРЕМА

Исходя из доказанных результатов, сформулируем для рассматриваемой задачи магистральную теорему, предполагая достаточно общие условия: время управления T достаточно большое (см. § 5); интервал принадлежности фондовооруженности $k \in [0, k_1]$; величины k_0 , k^* и k_T находятся между собой в произвольных соотношениях с дополнительным условием $k_T > k_2$.

На рис. 1 в соответствии с проведенными исследованиями жирными линиями выделены области знакопостоянства $\dot{k}$ и $\dot{q}$ (области I–V), а также изображены типовые фазовые траектории (экстремали) $X(t) = \{k(t); q(t)\}$ под номерами I–7. Траектории, входящие в особую точку $X^* = (k^*; 1)$ и выходящие из нее, обозначены соответственно Γ_0 и Γ_T и при этом $\Gamma_0 = \Gamma_0^1 \cup \Gamma_0^3$, $\Gamma_T = \Gamma_T^2 \cup \Gamma_T^4$.

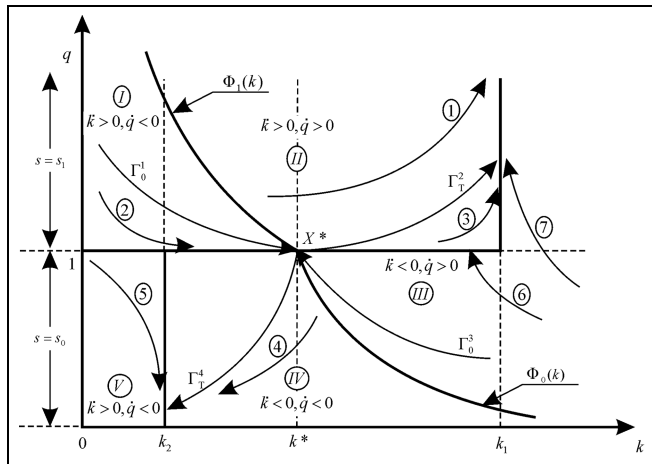


Рис. 1. Типовые фазовые траектории

Введем для $k < k^*$, $\tau_1 < \tau_2$ по обозначению величины, смысл которых следующий: $T_1(k', k'')$ — время возрастания фондовооруженности $k(t)$ от значения k' до значения k'' , когда $s(t) = s_1$; $T_0(k', k'')$ — время убывания фондовооруженности $k(t)$ от значения k'' до значения k' , когда $s(t) = s_0$; $J^1(\tau_1, \tau_2)$, $J^0(\tau_1, \tau_2)$ и $J^*(t_1, t_2)$ — значения критерия качества на интервале $t \in [\tau_1, \tau_2]$, когда соответственно $s(t) = s_1$, $s(t) = s_0$ и $s(t) = s^*$. Тогда непосредственно из выражений (1.8) и (1.9) с учетом лемм 3—5 следует:

$$T_1(k', k'') = \int_{k'}^{k''} \frac{dk}{(1-\gamma)(1-u)s_1 f(k) - vk},$$

$$T_0(k', k'') = \int_{k'}^{k''} \frac{dk}{vk - (1-\gamma)(1-u)s_0 f(k)}, \quad (3.1)$$

$$J^1(\tau_1, \tau_2) = (1-\gamma)(1-u)(1-s_1) \times$$

$$\times \int_{\tau_1}^{\tau_2} f(k(t); s_1) \exp\{-\delta t\} dt, \quad (3.2)$$

$$J^0(\tau_1, \tau_2) = (1-\gamma)(1-u)(1-s_0) \times$$

$$\times \int_{\tau_1}^{\tau_2} f(k(t); s_0) \exp\{-\delta t\} dt, \quad (3.3)$$

$$J^*(\tau_1, \tau_2) = [(1-\gamma)(1-u)(1-s^*)f(k^*)] \times$$

$$\times \delta^{-1} [\exp\{-\delta \tau_1\} - \exp\{-\delta \tau_2\}]. \quad (3.4)$$

Таким образом, основной результат в форме магистральной теоремы формулируется в виде следующего утверждения (рис. 1).

Теорема 1. При достаточно большом плановом периоде $[0, T]$ решение задачи имеет следующий вид:

1) интервал времени $[0, T]$ разбивается на три интервала $[0, T] = [0, T^*] \cup [T^*, T^{**}] \cup [T^{**}, T]$;

2) управление $s(t) \in \{s_1; s_0; s^*\}$, т. е. является кусочно-постоянным;

3) на магистральном интервале времени $t \in [T^*, T^{**}]$ $s(t) = s^*$, которое определяется формулами (2.12) и (2.13), а фондовооруженность $k(t)$ сохраняет постоянное значение k^* , являющееся единственным корнем уравнения (2.6);

4) на начальном интервале времени $t \in [0, T^*]$, когда происходит выход экономики на магистраль, $s(t) = s_1$, если $k_0 < k^*$, и $s(t) = s_0$, если $k_0 > k^*$, и происходит соответственно возрастание либо убывание $k(t)$ от k_0 до k^* ;

5) на конечном интервале времени $t \in [T^{**}, T]$, когда происходит сход экономики с магистрали для удовлетворения условия экономического горизонта $k(T) = k_T$, $s(t) = s_1$, если $k_T > k^*$, и $s(t) = s_0$, если $k_T < k^*$, и происходит соответственно возрастание либо убывание $k(t)$ от k^* до k_T ;

6) значения T^* , T^{**} и J определяются следующими формулами:

а) если $k_0 < k^*$, $k_T < k^*$, то

$$T^* = T_1(k_0, k^*), \quad T^{**} = T - T_0(k_T, k^*),$$

$$J = J^1(0, T^*) + J^*(T^*, T^{**}) + J^0(T^{**}, T); \quad (3.5)$$

б) если $k_0 < k^*$, $k_T > k^*$, то

$$T^* = T_1(k_0, k^*), \quad T^{**} = T - T_1(k^*, k_T),$$

$$J = J^1(0, T^*) + J^*(T^*, T^{**}) + J^1(T^{**}, T); \quad (3.6)$$

в) если $k_0 > k^*$, $k_T < k^*$, то

$$T^* = T_0(k^*, k_0), \quad T^{**} = T - T_0(k_T, k^*),$$

$$J = J^0(0, T^*) + J^*(T^*, T^{**}) + J^0(T^{**}, T); \quad (3.7)$$

г) если $k_0 > k^*$, $k_T > k^*$, то

$$T^* = T_0(k^*, k_0), \quad T^{**} = T - T_1(k^*, k_T),$$

$$J = J^0(0, T^*) + J^*(T^*, T^{**}) + J^1(T^{**}, T). \quad (3.8)$$

Следующее утверждение определяет золотое правило накопления [4, 5, 7, 8, 10] в рассматриваемой задаче, а именно, каким образом произведенный экономикой продукт $Y^*(t) = F(K^*(t), L(t))$, где $K^*(t) = k^*L(t)$, как сумма доходов с основных фондов $Y_K^*(t)$ и трудовых ресурсов $Y_L^*(t)$, распределяется на магистрали между накоплением $I^*(t)$, потреблением $C^*(t)$, налоговыми отчислениями $N^*(t)$ и материальными затратами $\Psi^*(t)$.



Теорема 2. На интервале времени $t \in [T^*, T^{**}]$, когда экономика находится на магистрали и $s(t) = s^*$, $k(t) = k^*$:

1) на накопление используется $(1 - \gamma)(1 - u)$ -я часть дохода с основных фондов $Y_K^*(t)$ минус величина $\delta K^*(t)$, т. е.

$$I^*(t) = (1 - \gamma)(1 - u)Y_K^*(t) - \delta K^*(t),$$

$$Y_K^*(t) = [\partial F(K^*(t), L(t))/\partial K^*(t)]K^*(t); \quad (3.9)$$

2) на потребление используется $(1 - \gamma)(1 - u)$ -я часть дохода с трудовых ресурсов $Y_L^*(t)$ плюс величина $\delta K^*(t)$, т. е.

$$C^*(t) = (1 - \gamma)(1 - u)Y_L^*(t) + \delta K^*(t),$$

$$Y_L^*(t) = [\partial F(K^*(t), L(t))/\partial L(t)]L(t); \quad (3.10)$$

3) на налоговые отчисления используется сумма $[(1 - \gamma)u]$ -х частей доходов с основных фондов и трудовых ресурсов, т. е.

$$N^*(t) = (1 - \gamma)u[Y_K^*(t) + Y_L^*(t)] =$$

$$= (1 - \gamma)uF(K^*(t), L(t)); \quad (3.11)$$

4) материальные затраты равны сумме оставшихся $[1 - (1 - \gamma)(1 - u) - (1 - \gamma)u]$ -х частей доходов с основных фондов и трудовых ресурсов, т. е.

$$\Psi^*(t) = [1 - (1 - \gamma)(1 - u) - (1 - \gamma)u] \times$$

$$\times [Y_K^*(t) + Y_L^*(t)] = \gamma F(K^*(t), L(t)). \quad \blacklozenge \quad (3.12)$$

4. СЛУЧАЙ ПРОИЗВОДСТВЕННОЙ ФУНКЦИИ КОББА—ДУГЛАСА

Широко используемой в исследованиях линейно-однородной производственной функцией служит функция Кобба — Дугласа [5, 7, 8]

$$F(K, L) = AK^\alpha L^\beta, \quad f(k) = Ak^\alpha,$$

$$A > 0, \quad \alpha > 0, \quad \beta > 0, \quad \alpha + \beta = 1, \quad (4.1)$$

где α — коэффициент эластичности по основным фондам, β — коэффициент эластичности по трудовым ресурсам. Конкретизируем для нее теорему 1.

Теорема 3. В случае производственной функции Кобба—Дугласа решение задачи в форме магистральной теоремы имеет следующий вид:

1) имеют место формулы:

$$s^* = v(v + \delta)^{-1}\alpha, \quad \tilde{s}^* = \delta(v + \delta)^{-1} + v(v + \delta)^{-1}\beta,$$

$$k^* = [(1 - \gamma)(1 - u)\alpha A/(v + \delta)]^{\frac{1}{1-\alpha}}, \quad (4.2)$$

$$k_1 = [v^{-1}(1 - \gamma)(1 - u)s_1 A]^{\frac{1}{1-\alpha}},$$

$$k_2 = [v^{-1}(1 - \gamma)(1 - u)s_0 A]^{\frac{1}{1-\alpha}}; \quad (4.3)$$

2) для T^* и T^{**} имеют место формулы (3.5)—(3.8), где

$$T_1(k', k'') = [v(1 - \alpha)]^{-1} \ln\{[(1 - \gamma)(1 - u)s_1 A -$$

$$- v(k')^{1-\alpha}][(1 - \gamma)(1 - u)s_1 A - v(k'')^{1-\alpha}]^{-1}\},$$

$$T_2(k', k'') = [v(1 - \alpha)]^{-1} \ln\{[v(k'')^{1-\alpha} - (1 - \gamma) \times$$

$$\times (1 - u)s_0 A][v(k')^{1-\alpha} - (1 - \gamma)(1 - u)s_0 A]^{-1}\}; \quad (4.4)$$

3) на интервале $t \in [0, T^*]$:

$$k(t) = [v^{-1}((1 - \gamma)(1 - u)s_1 A - [(1 - \gamma)(1 - u)s_1 A -$$

$$- v k_0^{1-\alpha}] \exp\{-v(1 - \alpha)t\})]^{\frac{1}{1-\alpha}}, \quad k_0 < k^*, \quad (4.5)$$

$$k(t) = [v^{-1}((1 - \gamma)(1 - u)s_0 A + [v k_0^{1-\alpha} - (1 - \gamma) \times$$

$$\times (1 - u)s_0 A] \exp\{-v(1 - \alpha)t\})]^{\frac{1}{1-\alpha}}, \quad k_0 > k^*; \quad (4.6)$$

4) на интервале $t \in (T^*, T]$:

$$k(t) = [v^{-1}((1 - \gamma)(1 - u)s_1 A - [(1 - \gamma)(1 - u)s_1 A -$$

$$- v k_T^{1-\alpha}] \exp\{v(1 - \alpha)(T - t)\})]^{\frac{1}{1-\alpha}}, \quad k_T > k^*, \quad (4.7)$$

$$k(t) = [v^{-1}((1 - \gamma)(1 - u)s_0 A + [v k_T^{1-\alpha} - (1 - \gamma) \times$$

$$\times (1 - u)s_0 A] \exp\{v(1 - \alpha)(T - t)\})]^{\frac{1}{1-\alpha}}, \quad k_T < k^*; \quad (4.8)$$

5) для J имеют место формулы (3.5)—(3.8), а для $J^*(T^*, T^{**})$, $J^1(0, T^*)$, $J^1(T^{**}, T)$, $J^0(0, T^*)$ и $J^0(T^{**}, T)$ — формулы (3.2)—(3.4), в которых $f(k) = Ak^\alpha$, а τ_1 и τ_2 присваиваются соответственно значения: $\tau_1 = T^*$, $\tau_2 = T^{**}$; $\tau_1 = 0$, $\tau_2 = T^*$; $\tau_1 = T^{**}$, $\tau_2 = T$; $\tau_1 = 0$, $\tau_2 = T^*$; $\tau_1 = T^{**}$, $\tau_2 = T$. $\blacklozenge$

5. АНАЛИЗ РЕЗУЛЬТАТОВ

1. Как и в классической задаче при отсутствии ограничений на накопление и потребление ($s_0 = 0$, $s_1 = 1$), а также налоговых отчислений ($u = 0$) и производственных затрат ($\gamma = 0$) [2, 4], в рассмотренной задаче при наличии перечисленных факторов суть решения в форме магистральной теоремы (теорема 1) заключается в наискорейшем выводе экономики на магистраль, где осуществляется сбалансированный рост и распределение про-

дукта происходит в соответствии с золотым правилом накопления (теорема 2), и наискорейшем сходе экономики с магистрали для удовлетворения условия экономического горизонта в момент T окончания планового периода. Тем самым обеспечивается максимально возможное время нахождения экономики на магистрали.

2. Условием существования решения в форме магистральной теоремы является условие $\Delta T^* = T^{**} - T^* > 0$. Так как $T^{**} = T - \tilde{T}$, где $\tilde{T}$ равно времени достижения траекторией $k(t)$ значения k_T на интервале времени $t \in (T^{**}, T]$, то данное условие принимает вид $T > T^* + \tilde{T}$, а именно (см. п. 6 теоремы 1): $T > T_1(k_0, k^*) + T_0(k_T, k^*)$, $k_0 < k^*$, $k_T < k^*$; $T > T_1(k_0, k^*) + T_1(k^*, k_T)$, $k_0 < k^*$, $k_T > k^*$; $T > T_0(k^*, k_0) + T_0(k_T, k^*)$, $k_0 > k^*$, $k_T < k^*$; $T > T_0(k^*, k_0) + T_1(k^*, k_T)$, $k_0 > k^*$, $k_T > k^*$.

3. Формулы (2.13) и (2.15) определяют вклад коэффициентов эластичности по основным фондам $\alpha(k^*)$ и трудовым ресурсам $\beta(k^*)$ в структуру накопления и потребления на магистрали, т. е. определяют, в какой степени накопление и потребление зависят от эффективности факторов производства. Из этих формул следует, что s^* и $\tilde{s}^*$ являются возрастающими функциями соответственно α и β . Поскольку с ростом коэффициента эластичности увеличивается эффективность соответствующего фактора производства, то это означает, что накопление возрастает с ростом эффективности основных фондов, а потребление возрастает с ростом эффективности трудовых ресурсов, что согласуется с золотым правилом накопления (теорема 2), согласно которому накопление $I^*(t)$ формируется на основе дохода с основных фондов, а потребление $C^*(t)$ — на основе дохода с трудовых ресурсов.

4. Устойчивый экономический рост, когда за весь плановый период времени $t \in [0, T]$ не происходит уменьшения фондовооруженности $k(t)$, осуществляется, согласно теореме 1, в ситуации, когда $k_0 < k^*$, $k_T > k^*$. В этом случае на интервалах $t \in [0, T^*)$ и $t \in (T^{**}, T)$ используется максимальная норма накопления $s(t) = s_1$, а условие существования решения в форме магистральной теоремы имеет вид $T > T_1(k_0, k^*) + T_1(k^*, k_T)$ в общем случае (см. п. 2) и вид

$$T > [v(1 - \alpha)]^{-1} \ln\{[(1 - \gamma)(1 - u)s_1 A - v k_0^{1-\alpha}] \times [(1 - \gamma)(1 - u)s_1 A - v k_T^{1-\alpha}]^{-1}\} \quad (5.1)$$

в случае производственной функции Кобба—Дугласа, что следует из соотношения (4.4). Условием экономического роста являются два эквивалент-

ных условия $dY(t)/dt > 0$ и $\hat{Y}(t) > 0$, где $\hat{Y}(t) = [1/Y(t)][dY(t)/dt]$ — относительная скорость роста валового продукта. Так как $Y(t) = F(K(t), L(t))$, $L(t) = L_0 e^{\lambda t}$ и $K(t) = k(t)L(t)$, на интервалах $t \in [0, T^*)$

и $t \in (T^{**}, T]$ в случае $k_0 < k^*$, $k_T > k^*$, когда $\dot{k}(t) > 0$, с учетом неоклассических условий $\partial F(\cdot)/\partial K > 0$ и $\partial F(\cdot)/\partial L > 0$ следует выполнение условий экономического роста. На магистрали, где $k^*(t) \equiv k^* = \text{const}$ и $K^*(t) = k^*L(t)$, условия экономического роста также выполняются. Так как $\alpha = [\partial F/\partial K][K/F]$, $\beta = [\partial F/\partial L][L/F]$, причем $\alpha + \beta = 1$ для линейно-однородных функций $F(K, L)$ [5], то непосред-

ственно получаем, что $\hat{Y}^*(t) \equiv \hat{Y}^* = \lambda$, т. е. на магистрали относительная скорость роста валового продукта равна коэффициенту экспоненциального роста трудовых ресурсов. Из проведенного анализа следует, что в случае $k_0 < k^*$, $k_T > k^*$ на всем интервале времени $t \in [0, T]$ обеспечивается экономический рост. При этом на магистрали — сбалансированный рост ($\dot{k}(t) = 0$, $k(t) \equiv k^*$, а на интервалах $t \in [0, T^*)$ и $t \in (T^{**}, T]$ — расширенный рост ($\dot{k}(t) > 0$).

5. При отсутствии ограничений на накопление ($s_0 = 0$, $s_1 = 1$), а также производственных затрат ($\gamma = 0$) и налоговых отчислений ($u = 0$) результаты магистральной теоремы (теорема 1) и золотого правила накопления (теорема 2) переходят в классические варианты этих теорем [5, 7, 8], когда выход экономики на магистраль и сход экономики с неё осуществляется направлением валового продукта целиком только на накопление либо только на потребление, а распределение продукта на магистрали осуществляется только между накоплением и потреблением с учетом дисконтирования. Если дополнительно отсутствует и дисконтирование ($\delta = 0$), когда текущему потреблению не отдается предпочтение перед будущим потреблением, то золотое правило накопления получается в его первоначальном виде [4], сформулированном лауреатом Нобелевской премии по экономике 2006 г. Э. Фелпсом.

ЗАКЛЮЧЕНИЕ

Проведено полное исследование задачи управления односекторной экономикой при наличии ограничений на нормы накопления и потребления и при учете ненулевой материалоемкости и налоговых отчислений. Доказано существование магистрального принципа управления экономикой при сделанных предположениях. Получено золотое правило накопления, определяющее единственный способ распределения валового продукта



на магистрали. Выделен случай, для которого на всем временном интервале планового периода осуществляется экономический рост, а именно на магистрали — сбалансированный рост, а на начальном и конечном интервалах времени — расширенный рост. Все результаты конкретизируются для производственной функции Кобба—Дугласа. Совместно с работами [12–14] проведенное исследование утверждает универсальность принципа магистрали независимо от критерия оптимальности и дополнительных условий функционирования экономики и, вместе с тем, зависимость золотого правила накопления, как способа распределения произведенного продукта на магистрали, от указанных факторов.

ПРИЛОЖЕНИЕ

Доказательство леммы 1. Согласно принципу максимума Понтрягина из выражений (1.8) и (1.9) следует [11]

$$H(k, s, \psi) = \psi[(1 - \gamma)(1 - u)sf(k) - vk] + (1 - \gamma)(1 - u)(1 - s)f(k)e^{-\delta t}, \quad (\text{П.1})$$

$$s(t) = \arg \max_{s_0 \leq s \leq s_1} H(k(t), s, \psi(t)), \quad (\text{П.2})$$

$$\dot{\psi}(t) = -\partial H(k(t), s(t), \psi(t)) / \partial k(t), \\ \psi(T) \geq 0, \psi(T)[k(T) - k_T] = 0. \quad (\text{П.3})$$

Так как $(1 - \gamma) > 0$, $(1 - u) > 0$, $f(k) > 0$, то из формул (П.1) и (П.2) следует:

$$s(t) = s_1, \quad \psi(t) - e^{-\delta t} > 0; \quad s(t) = s_0, \quad \psi(t) - e^{-\delta t} < 0; \\ s(t) \in [s_0; s_1], \quad \psi(t) - e^{-\delta t} = 0. \quad (\text{П.4})$$

Из выражений (П.1) и (П.3) для сопряженной переменной $\psi(t)$ следует уравнение

$$\dot{\psi}(t) = -[(1 - \gamma)(1 - u)s(t)f'(k(t)) - v]\psi(t) - (1 - \gamma)(1 - u)(1 - s(t))f(k(t))e^{-\delta t}. \quad (\text{П.5})$$

Введем новую сопряженную переменную $q(t) = e^{\delta t}\psi(t)$. Тогда формулы (2.1)–(2.3) следуют соответственно из выражений (П.3)–(П.5).

Доказательство леммы 2. Согласно формулам (2.1) свойство $q(t) < 0$ может реализоваться только при $s(t) = s_0$, когда уравнение (2.2) принимает вид

$$\dot{q}(t) = a(t)q(t) - b(t), \quad (\text{П.6})$$

где $a(t) = (v + \delta) - (1 - \gamma)(1 - u)s_0f'(k(t))$, $b(t) = (1 - \gamma) \times (1 - u)(1 - s_0)f(k(t))$. Решение этого уравнения при условии $q(T) = q_T$ имеет вид [15]

$$q(t) = \exp\left\{\int_0^t a(\tau) d\tau\right\} \times \\ \times \left[\int_t^T b(\xi) \exp\left\{-\int_0^\xi a(\tau) d\tau\right\} d\xi + q_T \exp\left\{-\int_0^T a(\tau) d\tau\right\}\right]. \quad (\text{П.7})$$

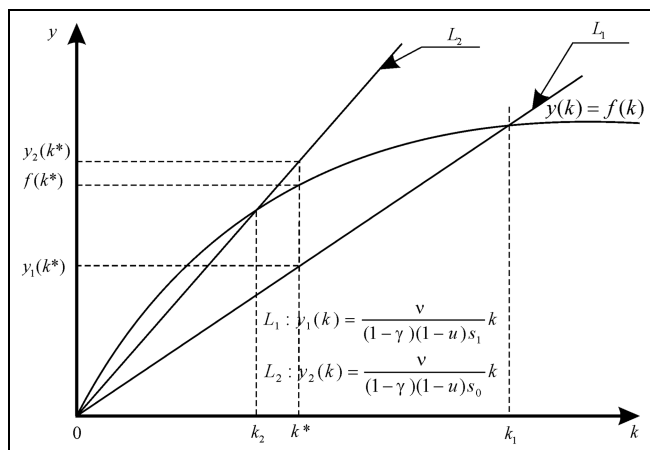


Рис. 2. К единственности корней k_1 и k_2

Так как, учитывая условия (1.10) и (2.3), $b(t) > 0$, $q_T \geq 0$, то из решения (П.7) следует, что $q(t) \geq 0$.

Доказательство леммы 3. При $s(t) = s_1$ уравнения (1.8) и (2.2) принимают вид

$$\dot{k}(t) = (1 - \gamma)(1 - u)s_1f(k(t)) - vk(t), \quad (\text{П.8})$$

$$\dot{q}(t) = [(v + \delta) - (1 - \gamma)(1 - u)s_1f'(k(t))]q(t) - (1 - \gamma)(1 - u)(1 - s_1)f'(k(t)). \quad (\text{П.9})$$

Свойства (2.4) для $\dot{k}$ и единственность решения k_1 уравнения (2.6) следуют с учетом условий (1.10) из уравнения (П.8) (рис. 2). Свойство $\dot{q} = 0$ для $q = \Phi_1(k)$ следует из выражений (2.7) и (П.9). Условие $\Phi_1(k) > 1$ соответствует, согласно формуле (2.7), условию $(1 - \gamma) \times (1 - u)(1 - s_1)f'(k) > (v + \delta) - (1 - \gamma)(1 - u)s_1f'(k)$, из которого следует

$$f'(k) > (v + \delta)[(1 - \gamma)(1 - u)]^{-1}. \quad (\text{П.10})$$

Единственность решения k^* уравнения (2.6) следует с учетом условий (1.10). Так как, согласно второму из уравнений (2.6), $f(k) > (v + \delta)/[(1 - \gamma)(1 - u)]$ для $k < k^*$, то из выражения (П.10) следует, что $\Phi_1(k) > 1$ для $k \in (0; k^*)$. Свойство $\Phi_1(k^*) = 1$ следует в результате использования подстановки второго из выражений (2.6) в формулу (2.7). Для доказательства свойства $k^* < k_1$ нужно доказать, что (см. рис. 2)

$$f(k^*) > v[(1 - \gamma)(1 - u)s_1]^{-1}k^* = y_1(k^*). \quad (\text{П.11})$$

Так как, согласно условиям (1.10), $f'(k)$ — монотонно убывающая функция, то учитывая формулу (П.10), имеем

$$f(k^*) = \int_0^{k^*} f'(k) dk > (v + \delta)[(1 - \gamma)(1 - u)]^{-1}k^*. \quad (\text{П.12})$$

Из правой части неравенств (1.11) следует $(v + \delta) > v/s_1$. Учет этого неравенства в выражении (П.12)

приводит к формуле (П.11), что доказывает требуемое свойство. Согласно выражению (2.7)

$$\begin{aligned}\Phi_1'(k) &= (v + \delta)(1 - \gamma)(1 - u)(1 - s_1)f''(k)\Psi_1^{-2}(k), \\ \Psi_1(k) &= (v + \delta) - (1 - \gamma)(1 - u)s_1f'(k).\end{aligned}\quad (\text{П.13})$$

Отсюда с учетом условий (1.10) следует $\Phi_1'(k) < 0$, т.е. функция $\Phi_1(k)$ убывающая. Свойства $\dot{q} < 0$ и $\dot{q} > 0$ в выражении (2.5) для $k < k^*$ следуют из доказанных свойств функции $\Phi_1(k)$. Пусть теперь $k > k^*$ (см. рис. 1). Так как $q > 1$, то из уравнения (П.9) следует

$$\begin{aligned}\dot{q} &> (v + \delta) - (1 - \gamma)(1 - u)s_1f'(k) - (1 - \gamma)(1 - u) \times \\ &\times (1 - s_1)f'(k) = (v + \delta) - (1 - \gamma)(1 - u)f'(k).\end{aligned}\quad (\text{П.14})$$

Так как, согласно уравнению (2.6) и условиям (1.10), $0 < f'(k) < [(v + \delta)/(1 - \gamma)(1 - u)]$ для $k > k^*$, то из неравенства (П.14) следует требуемое свойство $\dot{q} > 0$ для $k > k^*$ в выражении (2.5).

Доказательство леммы 4. При $s(t) = s_0$ уравнения (1.8) и (2.2) принимают вид

$$\dot{k}(t) = (1 - \gamma)(1 - u)s_0f(k(t)) - vk(t), \quad (\text{П.15})$$

$$\begin{aligned}\dot{q}(t) &= [(v + \delta) - (1 - \gamma)(1 - u)s_0f'(k(t))]q(t) - \\ &- (1 - \gamma)(1 - u)(1 - s_0)f''(k(t)).\end{aligned}\quad (\text{П.16})$$

Свойства (2.8) для $\dot{k}$ и единственность решения k_2 уравнения (2.10) следуют с учетом условий (1.10) из уравнения (П.15) (см. рис. 2). Свойство $\dot{q} = 0$ для $q = \Phi_0(k)$ следует из выражений (2.11) и (П.16). Условие $\Phi_0(k) < 1$ соответствует, согласно формуле (2.11), условию $(1 - \gamma)(1 - u)(1 - s_0)f'(k) < (v + \delta) - (1 - \gamma) \times (1 - u)s_0f'(k)$, из которого следует

$$f'(k) < (v + \delta)[(1 - \gamma)(1 - u)]^{-1} \quad (\text{П.17})$$

Так как, согласно второму из уравнений (2.6), $f'(k) < (v + \delta)/[(1 - \gamma)(1 - u)]$ для $k > k^*$, то из неравенства (П.17) следует, что $\Phi_0(k) < 1$ для $k > k^*$. Свойство $\Phi_0(k^*) = 1$ следует в результате учета второго из уравнений (2.6) в формуле (2.11). Для доказательства свойства $k^* > k_2$ нужно доказать, что (см. рис. 2)

$$f(k^*) < v[(1 - \gamma)(1 - u)s_0]^{-1}k^* = y_2(k^*). \quad (\text{П.18})$$

Из левой части неравенств (1.11) следует $s_0 < vk^*f'(k^*)/(v + \delta)f(k^*)$. Тогда

$$\begin{aligned}v[(1 - \gamma)(1 - u)s_0]^{-1}k^* &> (v + \delta) \times \\ &\times [(1 - \gamma)(1 - u)f'(k^*)]^{-1}f(k^*).\end{aligned}\quad (\text{П.19})$$

Так как, согласно выражению (2.6), $(v + \delta) = (1 - \gamma)(1 - u)f'(k^*)$, то неравенство (П.18) следует из неравенства (П.19). Аналогично формуле (П.13)

$$\begin{aligned}\Phi_0'(k) &= (v + \delta)(1 - \gamma)(1 - u)(1 - s_0)f''(k)\Psi_0^{-2}(k), \\ \Psi_0(k) &= (v + \delta) - (1 - \gamma)(1 - u)s_0f'(k).\end{aligned}\quad (\text{П.20})$$

Тогда из формулы (П.20) и условий (1.10) следует $\Phi_0'(k) < 0$, т.е. функция $\Phi_0(k)$ убывающая. Свойства

$\dot{q} > 0$ и $\dot{q} < 0$ в выражении (2.9) для $k > k^*$ следуют из доказанных свойств функции $\Phi_0(k)$. Пусть теперь $k < k^*$ (см. рис. 1). Так как $q < 1$, то из уравнения (П.16) следует

$$\begin{aligned}\dot{q} &< (v + \delta) - (1 - \gamma)(1 - u)s_0f'(k) - (1 - \gamma)(1 - u) \times \\ &\times (1 - s_0)f'(k) = (v + \delta) - (1 - \gamma)(1 - u)f'(k).\end{aligned}\quad (\text{П.21})$$

Так как согласно уравнениям (2.6) и условиям (1.10) $f'(k) > [(v + \delta)/(1 - \gamma)(1 - u)]$ для $k < k^*$, то из неравенства (П.21) следует требуемое свойство $\dot{q} < 0$ для $k < k^*$ в выражении (2.9).

Доказательство леммы 5. Условиями существования стационарных решений уравнений (1.8) и (2.2) являются условия $\dot{k} = 0$, $\dot{q} = 0$. Пусть эти условия реализуются при управлении $s(t) \equiv s^* = \text{const}$, которому соответствует стационарное решение $k(t) \equiv k^* = \text{const}$, при которых согласно уравнениям (1.8) и (2.2)

$$(1 - \gamma)(1 - u)s^*f(k^*) - vk^* = 0, \quad (\text{П.22})$$

$$\begin{aligned}[(v + \delta) - (1 - \gamma)(1 - u)s^*f'(k^*)]q - \\ - (1 - \gamma)(1 - u)(1 - s^*)f''(k^*) = 0.\end{aligned}\quad (\text{П.23})$$

Из выражений (2.6) и (П.23) следует, что $q = 1$, а формулы (2.12) следуют из выражений (2.6) и (П.22).

Пусть $g(k) = f(k)/k$. С учетом условий (1.10): $\lim_{k \downarrow 0} g(k) = \lim_{k \downarrow 0} f'(k) = \infty$; $\lim_{k \uparrow \infty} g(k) = \lim_{k \uparrow \infty} f'(k) = 0$. Так как

$$\begin{aligned}g'(k) &= [kf'(k) - f(k)]/k^2, \text{ то доказать свойство } g'(k) < 0 \\ \text{означает доказать, что } kf'(k) - f(k) < 0. \text{ С учетом усло-} \\ \text{вий (1.10) получаем } f(k) &= \int_0^k f'(y)dy > \int_0^k f'(k)dy = kf'(k),\end{aligned}$$

что доказывает требуемое свойство. Таким образом $g(k) \downarrow_0^\infty$ при $k \uparrow_0^\infty$. Так как $k^* < k_1$, то $g(k^*) > g(k_1)$, т.е. $[k^*/f(k^*)] < [k_1/f(k_1)]$. Тогда из формул (2.12) с учетом последнего неравенства и выражения (2.6) следует $s^* < vk_1[(1 - \gamma)(1 - u)f(k_1)]^{-1} = s_1$, что доказывает свойство $s^* < s_1$. Так как $k^* > k_2$, то аналогично предыдущему получаем, что $[k^*/f(k^*)] > [k_2/f(k_2)]$. Тогда из формул (2.12) с учетом последнего неравенства и уравнения (2.10) следует $s^* > vk_2[(1 - \gamma)(1 - u)f(k_2)]^{-1} = s_0$, что доказывает свойство $s^* > s_0$.

Доказательство теоремы 2. Формула Эйлера для линейно-однородных производственных функций дает представление для валового продукта $Y(t) = F(K(t), L(t))$, как суммы доходов с основных фондов $Y_K(t)$ и трудовых ресурсов $Y_L(t)$, в виде [5]

$$\begin{aligned}F(K(t), L(t)) &= [\partial F(K(t), L(t))/\partial K(t)]K(t) + \\ &+ [\partial F(K(t), L(t))/\partial L(t)]L(t) = Y_K(t) + Y_L(t).\end{aligned}\quad (\text{П.24})$$

Из формул (2.12) следует $(1 - \gamma)(1 - u)s^*f(k^*) = (1 - \gamma)(1 - u)k^*f'(k^*) - \delta k^*$. Тогда, учитывая формулы (1.5) и (П.24), получаем, что

$$\begin{aligned}(1 - \gamma)(1 - u)s^*F(K^*(t), L(t)) &= \\ = (1 - \gamma)(1 - u)Y_K^*(t) - \delta K^*(t).\end{aligned}\quad (\text{П.25})$$



Согласно формуле (1.3) $I^*(t) = (1 - \gamma)(1 - u)s^* \times F(K^*(t), L(t))$. Тогда выражение (3.9) следует из равенства (П.25). Согласно формуле (П.24) $Y_K^*(t) = F(K^*(t), L(t)) - Y_L^*(t)$. Тогда из равенства (П.25) следует

$$(1 - \gamma)(1 - u)s^*F(K^*(t), L(t)) = (1 - \gamma)(1 - u) \times F(K^*(t), L(t)) - (1 - \gamma)(1 - u)Y_L^*(t) - \delta K^*(t). \quad (\text{П.26})$$

Согласно формуле (1.3) $C^*(t) = (1 - \gamma)(1 - u)(1 - s^*) \times F(K^*(t), L(t))$. Тогда выражение (3.10) следует из равенства (П. 26). Так как $N^*(t) = (1 - \gamma)uF(K^*(t), L(t))$, то выражение (3.11) следует из формулы (П. 24). Из выражений (3.9)–(3.11) следует, что остаются неиспользованными на $I^*(t)$, $C^*(t)$ и $N^*(t)$ суммарно $[1 - ((1 - \gamma) \times (1 - u) + (1 - \gamma)u)]$ -е части $Y_K^*(t)$ и $Y_L^*(t)$, которые остаются на возмещение материальных затрат Ψ^* . Таким образом $[1 - ((1 - \gamma)(1 - u) + (1 - \gamma)u)][Y_K^*(t) + Y_L^*(t)] = \Psi^*(t)$. Тогда выражение (3.12) следует из последнего выражения с учетом записи (П.24) очевидным образом.

Доказательство теоремы 3. Формулы (4.2) и (4.3) для s^* и $\tilde{s}^*$ следуют из выражений (2.13), (2.15) и (4.1), а для k^* , k_1 и k_2 получаются в результате учета формулы (4.1) в уравнениях (2.6) и (2.10). Учет формулы (4.1) в выражениях (3.1) дает, что

$$T_1(k', k'') = \int_{k'}^{k''} \frac{dk}{(1 - \gamma)(1 - u)s_1 A k^\alpha - \nu k},$$

$$T_2(k', k'') = \int_{k'}^{k''} \frac{dk}{\nu k - (1 - \gamma)(1 - u)s_0 A k^\alpha}.$$

Интегрирование приводит к формулам (4.4). При $s(t) \equiv s = \text{const}$ общее решение уравнения (1.8) в случае функции $f(k)$ вида (4.1) имеет вид

$$k(t) = (\nu^{-1}[(1 - \gamma)(1 - u)sA - C \exp\{-\nu(1 - \alpha)t\}])^{\frac{1}{1-\alpha}},$$

где C — константа интегрирования. Выражения (4.5) и (4.6) следуют отсюда соответственно для $s = s_1$ и $s = s_0$ при условии $k(0) = k_0$. Выражения (4.7) и (4.8) следуют соответственно для $s = s_1$ и $s = s_0$ при условии $k(T) = k_T$. Последнее утверждение теоремы очевидно.

ЛИТЕРАТУРА

1. Ramsey F.P. Mathematical theory of savings // Econ. J. — 1928. — Vol. 38. — P. 543–559.
2. Solow R.A. Contribution to the theory of economic growth // Quart. J. Economics. — 1956. — Vol. 70. — P. 65–94.
3. Солоу Р.А. Перспективы теории роста // Мировая экономика и международные отношения. — 1966. — № 8. — С. 69–77.
4. Phelps E.S. Golden rules of economic growth. — New-York: Norton, 1966. — 189 p.
5. Ашманов С.А. Введение в математическую экономику. — М.: Наука, 1984. — 293 с.
6. Mankiw N.G. Economics. — New-York: Worth Publ., 2003. — 256 p.
7. Митягин Б.С. Заметки по математической экономике // Успехи математических наук. — 1972. — Т. 27, № 3. — С. 3–19.
8. Эрроу К. Применение теории управления к экономическому росту / В кн.: Математическая экономика / Под ред. Б.С. Митягина. — М.: Мир, 1974. — С. 7–45.
9. Москаленко А.И. Оптимальное управление моделями экономической динамики. — Новосибирск: Наука, 1999. — 185 с.
10. Интрилигатор М. Математические методы оптимизации и экономическая теория. — М.: Айрисс пресс, 2002. — 576 с.
11. Математическая теория оптимальных процессов / Л.С. Понтрягин, В.Г. Болтянский, Р.В. Гамкрелидзе, Е.Ф. Мищенко. — М.: Наука, 1978. — 392 с.
12. Демин Н.С., Кулешова Е.В. Управление односекторной экономикой на конечном интервале времени с учетом потребления работодателей // Автоматика и телемеханика. — 2008. — № 9. — С. 140–155.
13. Демин Н.С., Кулешова Е.В. Управление односекторной экономикой на конечном интервале времени с учетом налоговых отчислений // Изв. РАН. Теория и системы управления. — 2008. — № 6. — С. 87–98.
14. Демин Н.С., Кулешова Е.В. Управление односекторной экономикой на конечном интервале времени по критерию максимизации налоговых отчислений // Сибирский журнал промышленной математики. — 2009. — Т. 12, № 1(37). — С. 74–88.
15. Матвеев А.М. Методы интегрирования обыкновенных дифференциальных уравнений. — М.: Высшая школа, 1967. — 564 с.

Статья представлена к публикации членом редколлегии Р.М. Нижегородцевым.

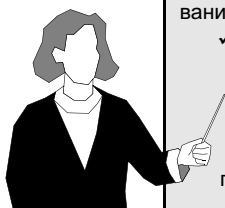
Дёмин Николай Серапионович — д-р физ.-мат. наук, профессор, ☎(3822) 52-92-99,

Кулешова Елена Викторовна — аспирантка, ☎(3822) 55-45-11, ✉ kuleshova.e@mail.ru,

Томский государственный университет.

Читайте в следующем номере

- ✓ Ашимов А.А., Боровский Ю.В., Новиков Д.А. и др. Структурная устойчивость и параметрическое регулирование на примере моделей циклической динамики макросистем
- ✓ Анулова С.В. Максимизация времени выхода управляемого случайного блуждания на границу квадранта
 - ✓ Белоцерковский Д.Л. Об одной задаче перечисления образующих графов с ограничением на диаметр
 - ✓ Угольников Г.А., Усов А.Б. Трехуровневые системы управления эколого-экономическими объектами веерной структуры
 - ✓ Чеботарев П.Ю., Логинов А.К., Цодикова Я.Ю. и др. Голосование в стохастической среде: случай двух групп



АНАЛИЗ РЕАЛИЗУЕМОСТИ РЕЗУЛЬТАТОВ МНОГОКРИТЕРИАЛЬНОЙ ЭКСПЕРТИЗЫ – ПРИМЕНЕНИЕ «СВОЙСТВА ПЕРЕСЕЧЕНИЯ»

Н.А. Коргин

Исследованы свойства неманипулируемых механизмов многокритериальной экспертизы, представимых в виде обобщенных медианных схем, позволяющих уменьшить трудоемкость проверки «свойства пересечения» — условия, необходимого и достаточного для того, чтобы результаты экспертизы принадлежали множеству допустимых результатов. Предложен алгоритм, позволяющий для произвольной обобщенной медианной схемы определить — для каких множеств допустимых результатов экспертизы эта схема удовлетворяет свойству пересечения.

Ключевые слова: активная экспертиза, коллективный выбор, неманипулируемый механизм, обобщенная медианная схема.

ВВЕДЕНИЕ

При проведении экспертизы, в которой участвуют эксперты, заинтересованные в ее результате, актуальна проблема манипулирования сообщаемой информацией с их стороны. В этом проявляется так называемое «оппортунистическое» поведение экспертов — действия в собственных интересах вопреки интересам всей системы в целом. Механизмы экспертизы с участием заинтересованных экспертов получили название механизмов активной экспертизы [1, 2], которые также можно рассматривать как механизмы коллективного выбора [3] с совпадающими непрерывными множествами допустимых альтернатив и результатов выбора (поэтому, в рамках данной статьи, понятия «результат экспертизы» и «результат выбора» считаются эквивалентными). Механизмы активной экспертизы (правила агрегирования сообщений экспертов о своих функциях предпочтения), в которых все эксперты предпочитают сообщать истину, называются неманипулируемыми механизмами активной экспертизы [1, 2]. В работе [4] был предложен конструктивный механизм построения многокритериальных механизмов на основе так называемых *обобщенных медианных схем* [3, 5, 6] для случая, когда множество допустимых результатов экспертизы (или множество допустимых альтернатив в терминологии теории коллективного выбора [3, 5, 6]) является многомерным прямоугольником — декартовым произведением мно-

жеств допустимых значений каждого из критериев. Было доказано, что для множества допустимых альтернатив, представимого в виде многомерного компакта, любой неманипулируемый механизм коллективного выбора также представим в виде обобщенной медианной схемы, при условии, что данная схема удовлетворяет особому условию на данном множестве — *свойству пересечения* [6]. Данное свойство гарантирует, что, если все сообщения участников выбора (экспертов в нашем понимании) принадлежат множеству допустимых альтернатив, то результат выбора в соответствии с применяемым неманипулируемым механизмом также будет принадлежать этому множеству. Однако, как отмечалось [4], проверка данного свойства достаточно трудоемка и осуществляется для комбинации «проверяемый неманипулируемый механизм выбора — исследуемое множество допустимых альтернатив». В рамках данной статьи исследуются свойства обобщенных медианных схем, позволяющие уменьшить трудоемкость проверки свойства пересечения. Предлагается алгоритм, позволяющий для произвольной обобщенной медианной схемы определить — для каких множеств допустимых альтернатив эта схема обладает свойством пересечения, а для каких нет. Данный подход актуален в том смысле, что позволяет организатору экспертизы зафиксировать устраивающий его неманипулируемый механизм многокритериальной активной экспертизы и проверять, при каких условиях на допустимые сообщения экспертов этот механизм можно будет применять.



1. ОСНОВНЫЕ ПОНЯТИЯ И ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Определим необходимые основные понятия. В развернутой форме все эти понятия даны в работе [4], поэтому здесь они будут изложены кратко, без подробных комментариев.

Пусть заданы множество критериев, по которым производится экспертиза $M = \{1, \dots, m\}$, множество допустимых альтернатив $A \in \mathcal{R}^m$, являющееся компактным и множество агентов (избирателей или экспертов) $N = \{1, \dots, n\}$. На основании сообщений агентов о наиболее предпочтительных с их точек зрения альтернативах должна быть выбрана одна — как результат коллективного выбора, получаемый с помощью некоторой функции коллективного выбора $f(\cdot): A^n \rightarrow A$. В рамках данной статьи будет предполагаться, что множество A выпукло, хотя часть определений и результатов будет обсуждаться и для случая невыпуклых множеств допустимых альтернатив. Предпочтения агентов над множеством альтернатив описывается многомерно однопиковыми функциями предпочтения [6] $u^i(\cdot): \mathcal{R}^m \rightarrow \mathcal{R}$, $i \in N$:

— существует единственная альтернатива $\tau^i = \arg \max_{y \in A} u^i(y)$ — точка «пика» агента i , $\forall i \in N$ $\tau^i \in A$;

— $\forall z, z' \in \mathcal{R}^m$,

$$[z' \in \widehat{B}(\{z, \tau^i\}), z' \neq z] \Rightarrow [u^i(z') > u^i(z)],$$

где $\widehat{B}(\{z, \tau^i\})$ — минимальный m -мерный прямоугольник, содержащий альтернативы z и τ^i , определяемый следующим образом. Для произвольного множества $A \subseteq \mathcal{R}^m$ обозначим через A_k проекцию данного множества на координатную ось $k \in M$, т. е. $A_k = \text{Proj}_k(A)$. Нижнюю и верхнюю границы множества A_k обозначим $\min A_k$ и $\max A_k$ соответственно. Тогда минимальный m -мерный прямоугольник, содержащий множество $A \subseteq \mathcal{R}^m$, определяется как

$$\widehat{B}(A) = \prod_{k \in M} [\min A_k, \max A_k].$$

Будем называть допустимыми альтернативы $z \in A$, а недопустимыми — $z \in \widehat{B}(A) \setminus A$.

Далее речь будет вестись о задаче построения такой неманипулируемой функции коллективного выбора¹ $f(\cdot)$, при которой агенты заинтересованы

¹ В рамках данной статьи понятия «механизм активной экспертизы», «механизм коллективного выбора» «функция коллективного выбора» считаются эквивалентными.

в сообщении истинного значения своих точек пиков. В работах [3, 5, 6] доказано, что такие функции неманипулируемы тогда и только тогда, когда они представимы в виде обобщенных медианных схем². Далее приводятся основные понятия, необходимые для данного представления.

Для описания медианных схем используется понятие систем правых и левых коалиций, изначально определяемое на одномерном пространстве альтернатив.

Определение 1 [6]. Система правых (левых)³ коалиций W определяет для каждой альтернативы $z \in [d, D] \subset \mathcal{R}^1$ набор коалиций агентов $W(z)$ в соответствии со следующими требованиями:

— принцип суверенитета: $\forall z \in (d, D] ([d, D))$, $W(z) \neq \emptyset$, $\emptyset \notin W(z)$ и $W(d) = 2^N \setminus \emptyset$ ($W(D) = 2^N \setminus \emptyset$);

— монотонность коалиций: если $S \in W(z)$ и $S \subset S'$, то $S' \in W(z)$;

— монотонность результата: если $z' < (>) z$ и $S \in W(z)$, то $S \in W(z')$;

— полунепрерывность сверху: $\forall S \subseteq N$, $\forall z \in [d, D]$ и для любой последовательности $\{z^t\} \subseteq [d, D]$ такой, что $\lim_{t \rightarrow \infty} z^t = z$, верно $[\forall t, S \in W(z^t)] \Rightarrow [S \in W(z)]$. ♦

Иными словами, вводится функция $W(\cdot): [d, D] \rightarrow 2^N \setminus \emptyset$, ставящая в соответствие каждой альтернативе $z \in [d, D]$ некоторое подмножество множества всех непустых коалиций агентов $2^N \setminus \emptyset$. В случае многокритериального выбора, для минимального многомерного прямоугольника, содержащего множество допустимых альтернатив A ,

$\widehat{B}(A) = \prod_{k=1}^m A_k$ можно определить семейство систем

правых коалиций R — набор $\{R_k\}_{k=1}^m$, где R_k является системой правых коалиций. Аналогичным образом можно определить семейство систем левых коалиций L — набор $\{L_k\}_{k=1}^m$.

Используя данные понятия, вводится следующее формальное определение обобщенной медианной схемы.

Определение 2 [6]. Пусть задано множество допустимых результатов выбора A и семейство систем правых R (левых L) коалиций на множестве $\widehat{B}(A)$. Тогда обобщенная медианная схема $x = h(\tau)$ ⁴,

² Поэтому, в рамках данной статьи, понятия «прямой неманипулируемый механизм активной экспертизы» и «обобщенная медианная схема» считаются эквивалентными.

³ В данном определении в скобках идет запись соответствующего условия для левой системы коалиций.

⁴ Подразумевается, что агенты сообщают свои наилучшие альтернативы (точки пика).

порожденная совокупностью (A, R) (соответственно (A, L)), определяется следующим образом:

$$h_k(\tau_k) = \max\{z_k \in A_k | \{i \in N | \tau_k^i \geq z_k\} \in R_k(z_k)\},$$

$$(h_k(\tau_k) = \min\{z_k \in A_k | \{i \in N | \tau_k^i \leq z_k\} \in L_k(z_k)\})$$

для $\forall \tau \in A^n$ и $\forall k \in M$. ♦

Иными словами, в качестве результата экспертизы по каждому критерию выбирается максимальная (минимальная) проекция альтернативы $z_k \in A_k$, для которой множество экспертов, чьи точки пика лежат справа (слева) от данной проекции, удовлетворяет требованиям выбранной системы правых (левых) коалиций по данному критерию для данной проекции альтернативы.

Связь между системами левых и правых коалиций, порождающими одну и ту же медианную схему (для одного критерия), определялась следующим образом [6]:

$$L^*(z) = \{S \in 2^N | \forall z' > z, \forall S' \in R(z'), S \cap S' \neq \emptyset\}.$$

Проиллюстрируем введенные определения.

Пример 1. Рассмотрим механизм активной экспертизы, усредняющий сообщения агентов по каждому критерию:

$$x = f(s): \forall k \in M, x_j = \frac{1}{n} \sum_{i=1}^n s_k^i. \quad (1)$$

Данный механизм манипулируем. Если агенты сообщают честно значения своих точек пика, то результатом выбора будет

$$x(\tau): \forall j \in M, x_j(\tau_j) = \frac{1}{n} \sum_{i=1}^n \tau_j^i,$$

где τ_k^i — наилучшее значение критерия k для агента i (или наилучшая альтернатива по данному критерию), а выигрыш каждого агента будет $u^i(x(\tau))$, $i \in N$.

Любому агенту $i \in N$, для которого $x(\tau^i, s^{-i}) \neq \tau^i$ при любом наборе сообщений остальных агентов s^{-i} , будет выгодно исказить свое сообщение. Если результат выбора по какому-либо критерию $k \in M$ будет меньше, чем значение наилучшей альтернативы агента по данному критерию, т. е. $x_k(\tau_k^i, s_k^{-i}) < \tau_k^i$, то сообщение искаженного значения своей наилучшей альтернативы $s_k^i > \tau_k^i$ увеличит значение x_k , приблизив его к τ_k^i , тем самым улучшив выигрыш агента $u^i(x(s^i, s^{-i})) \geq u^i(x(\tau^i, s^{-i}))$. Аналогично, для каждого критерия $k \in M$, такого что $x_k(\tau_k^i, s_k^{-i}) > \tau_k^i$, сообщение $s_k^i < \tau_k^i$, уменьшит значение $x_k(\tau_k)$, тем самым увеличив выигрыш агента.

Однако, проанализировав подобным образом поведение агента, можно определить равновесные по Нэшу сообщения $s^*(\tau)$ при заданном профиле (наборе) τ точек пика агентов. Механизм $h(\tau) = f(s^*(\tau))$, $\forall \tau \in A^n$ называется соответствующим $f(s)$ прямым механизмом. Если механизм $h(\tau)$ неманипулируемый, то он называется эквивалентным $f(s)$ [1, 2].

В работе [4] было показано, что семейства систем правых и левых коалиций, порождающих прямой механизм, эквивалентный механизму $f(s)$ и определяемый выражением (1), имеют следующий вид:

$$R_k(z_k) = \left\{ S \subset N | \#S = r \geq n \frac{z_k - \min A_k}{\max A_k - \min A_k} \right\},$$

$$\forall z_k \in (\min A_k, \max A_k],$$

$$R_k(\min A_k) = 2^N \setminus \emptyset, \forall k \in M,$$

$$L_k^*(z_k) = \left\{ S \subset N | \#S = l \leq n \frac{\min A_k - z_k}{\max A_k - \min A_k} \right\},$$

$$\forall z_k \in [\min A_k, \max A_k),$$

$$L_k^*(\max A_k) = 2^N \setminus \emptyset, \forall k \in M.$$

Символ $\#$ обозначает мощность множества.

На рис. 1 представлен данный прямой неманипулируемый механизм для случая двух критериев и трех экспертов. В качестве множества допустимых результатов выбора рассматривается «бюджетное ограничение» $A = \{x \in \mathbb{R}^2 | x_1 + x_2 \leq 1, x_1, x_2 \geq 0\}$. Для данного

множества допустимых альтернатив $\hat{B}(A) = [0, 1]^2$. Символами r_k и l_k , $k = \{1, 2\}$ обозначено число участников коалиции, необходимое для удовлетворения требований системы правых и левых коалиций по критерию $k \in \{1, 2\}$. Точки пиков агентов отмечены белыми кружками. Пунктирными линиями обозначены проекции точек пиков трех агентов на каждый из критериев. Штриховыми — изменение требований к коалициям по каждому из критериев. В первоначальном варианте формулировки обобщенных медианных схем, предложенной Э. Муленом [3] — точки пика «фантомных» агентов для определения результатов выбора по каждому из критериев. В качестве результата выбора по каждому критерию берется медиана проекций точек пика реальных и фантомных агентов. Результат выбора отмечен черным кружком.

Для прямого механизма, эквивалентного механизму (1), результат выбора на множестве A определяется следующим образом [4]:

$$h_k(\tau_k) = \max\{z_k \in [0, 1] | \#\{i \in N | \tau_k^i \geq z_k\} \geq n z_k\},$$

$$k \in \{1, 2\}. \quad (2)$$

Из рис. 1 видно, что результат выбора по критерию 1 h_1 равен 1/3, что пропорционально доле аген-

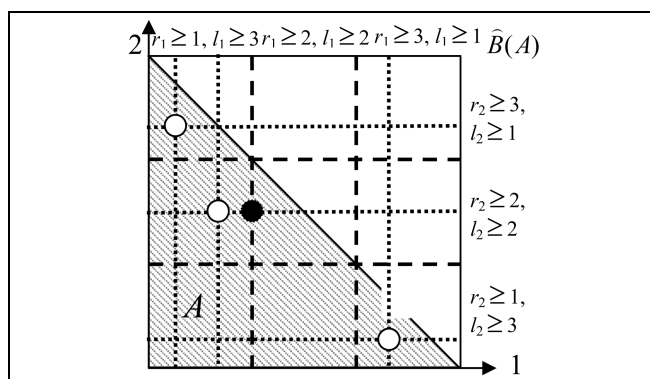


Рис. 1. Обобщенная медианная схема



тов (один из трех), значение проекций точек пика которых не меньше h_1 . Аналогично по второму критерию $h_2 = 2/3$. ♦

В ходе дальнейшего изложения термин «правые» или «левые коалиции» заменим на *выбирающие коалиции*, уточняя, где необходимо, о каких именно коалициях будет идти речь.

Обобщенные медианные схемы определяются на множестве $\widehat{B}(A)$, т. е. на m -мерном прямоугольнике. Поэтому любой прямой неманипулируемый механизм коллективного выбора, определенный в терминах *семейства систем выбирающих коалиций* (ССВК) для произвольного m -мерного прямоугольника B , останется таковым для любого множества допустимых результатов выбора A , такого что $\widehat{B}(A) = B$. Для проверки реализуемости прямого неманипулируемого механизма используется понятие *свойства пересечения* для порождающего его ССВК.

Для любой пары альтернатив $y, z \in \widehat{B}(A)$ обозначим множества $M^+(y, z) = \{k \in M | z_k > y_k\}$ и $M^-(y, z) = \{k \in M | z_k < y_k\}$, которые определяют набор критериев, для которых значение соответствующей компоненты альтернативы z строго больше y — альтернатива z расположена правее по данным критериям y , и наоборот — альтернатива z расположена левее по данным критериям y . Если $k \notin M^+(y, z) \cup M^-(y, z)$, то альтернативы неразличимы по данному критерию. Таким образом, положение альтернативы y относительно альтернативы z описывается парой множеств $\{M^-(y, z), M^+(y, z)\}$.

Определение 3 [6]. Семейство систем правых коалиций $R = \{R_k\}_{k=1}^m$, определенное на множестве $\widehat{B}(A)$, обладает *свойством пересечения* для множества A , если для $\forall y \in \widehat{B}(A) \setminus A$ и любого конечного подмножества $\{z^1, \dots, z^T\} \subset A$ верно:

$$\bigcap_{t=1}^T \left\{ \left[\bigcup_{k \in M^+(y, z^t)} l_k(y_k) \right] \cup \left[\bigcup_{k \in M^-(y, z^t)} r_k(y_k) \right] \right\} \neq \emptyset \quad (3)$$

для любой коалиции $r_k(y_k) \in R_k(y_k)$, где $k \in \bigcup_{t=1}^T M^-(y, z^t)$

и любой коалиции $l_k(y_k) \in L_k^*(y_k)$, где $k \in \bigcup_{t=1}^T M^+(y, z^t)$. ♦

Качественно, наличие свойства пересечения у ССВК означает, что любая недопустимая альтернатива $y \in \widehat{B}(A) \setminus A$ может быть выбрана в результате применения прямого неманипулируемого механизма только в том случае, если найдется хотя

бы один агент, чья точка пика также не принадлежит множеству A . Если ССВК обладает свойством пересечения на множестве A , то порождаемая им обобщенная схема удовлетворяет свойству пересечения на этом же множестве. Если ССВК удовлетворяет условию (3) для какой-либо альтернативы $y \in \widehat{B}(A) \setminus A$, то будем считать, что данное семейство обладает *свойством пересечения для A в y* .

Введя все необходимые определения, приведем основной результат, полученный группой С. Барберы:

«Функция коллективного выбора на профиле многомерно однопиковых предпочтений над множеством допустимых альтернатив $A \subseteq \mathbb{R}^m$ неманипулируема тогда и только тогда, когда она представима в виде обобщенной медианной схемы, удовлетворяющей свойству пересечения» [6].

В работах [6, 7] было показано, что условие (3) достаточно проверять лишь для одного *решающего* множества $\widehat{S}(y) \subset A$ для каждой недопустимой альтернативы $y \in \widehat{B}(A) \setminus A$. Для выпуклых компактных множеств $\forall y \in \widehat{B}(A) \setminus A$ решающее подмножество $\widehat{S}(y) = \{z^1, \dots, z^m\} \subset A$ определяется следующим образом — из каждого сечения A гиперплоскостью, проходящей через альтернативу y , выбирается по одной альтернативе; $\forall t = 1, \dots, m \ z_k^t = y_k$, если $k = t$, а z_k^t для $k \neq t$ любые такие, что $z^t \in A$. Из условия [7] следует, что для решающих множеств условие (3) может быть записано следующим образом:

$$\left[\bigcap_{k \in M^+(y, A)} l_k(y_k) \right] \cap \left[\bigcap_{k \in M^-(y, A)} r_k(y_k) \right] \neq \emptyset, \quad (4)$$

для любой коалиции $r_k(y_k) \in R_k(y_k)$, где $k \in M^-(y, A)$, и любой коалиции $l_k(y_k) \in L_k^*(y_k)$, где $k \in M^+(y, A)$.

Однако проверка, реализуем или нет прямой неманипулируемый механизм выбора на заданном множестве, остается отдельной сложной задачей, так как требуется проверять все недопустимые альтернативы $y \in \widehat{B}(A) \setminus A$ и не существует конструктивного алгоритма, позволяющего уменьшить сложность проверки. Разработке подобного алгоритма посвящено дальнейшее изложение.

2. МОНОТОННОСТЬ СВОЙСТВА ПЕРЕСЕЧЕНИЯ

Покажем, что проверку свойства пересечения можно существенно облегчить, так как достаточно проверять недопустимые альтернативы $y \in \widehat{B}(A) \setminus A$, лежащие как можно «ближе» к замыканию множества $A - clA$.



Тогда можно сформулировать

Определение 4 по своей сути является аналогом определения положения двух альтернатив относительно друг друга по отдельному критерию. Иллюстрацией этого определения служит рис. 2 для $\mathcal{R}^2$ (множество допустимых альтернатив A взято из примера 1). Альтернатива $y \in \widehat{B}(A) \setminus A$ расположена правее относительно A по обоим критериям, так как для $\forall z \in [z^1, z^2]$ $M^+(y, z) = \emptyset$, и $\forall z \in (z^1, z^2)$ $M^-(y, z) = \{1, 2\}$.

Для произвольной недопустимой альтернативы $y \in \widehat{B}(A) \setminus A$ определим *противоположный многомерный прямоугольник относительно множества A* :

$$\begin{aligned} & \check{B}(\{y, A\}) = \\ & = \left(\prod_{k \in M^+(y, A)} [\min A_k, y_k] \right) \times \left(\prod_{k \in M^-(y, A)} [y_k, \max A_k] \right). \end{aligned}$$

Рис. 3. Противоположный многомерный прямоугольник



альтернативы $y' \in \tilde{B}(\{y, A\})$ расположены дальше от A , чем y . Очевидно, что для всех $y' \in \tilde{B}(\{y, A\})$ $d(y', A) = d(y, A)$. Аналогично, все недопустимые альтернативы $y'' \in \hat{B}(A) \setminus A$, такие, что $\exists z \in cIA(y)$ и $y'' \in \hat{B}(y, z)$ расположены ближе к A , чем y . Для этих альтернатив также верно, что $d(y'', A) = d(y, A)$.

Можно доказать, что если требования, необходимые для обладания ССВК свойством пересечения выполнены для какой-либо альтернативы $y \in \hat{B}(A) \setminus A$, то они также выполнены для всех недопустимых альтернатив, расположенных дальше от множества A , чем альтернатива y :

Лемма 1. Пусть задано ССВК $W = \{W_k\}_{k=1}^m$, определенное на множестве $\hat{B}(A)$ и обладающее свойством пересечения для множества A в альтернативе $y \in \hat{B}(A) \setminus A$. Тогда данное ССВК обладает свойством пересечения для A в любой недопустимой альтернативе $y' \in \tilde{B}(\{y, A\})$. ♦

Доказательства данной леммы и всех дальнейших утверждений вынесены в приложение.

Следствие леммы 1. Пусть $\exists y \in \hat{B}(A) \setminus A$ такая, что ССВК $W = \{W_k\}_{k=1}^m$, определенное на множестве $\hat{B}(A)$, не обладает свойством пересечения для A в y . Тогда это ССВК не обладает свойством пересечения для множества A в любой недопустимой альтернативе y' , расположенной ближе к A , чем y . ♦

Лемма 1 и ее следствие показывают «монотонность» свойства пересечения и позволяет существенно облегчить проверку наличия свойства пересечения, так как достаточно проверять недопустимые альтернативы $y \in \hat{B}(A) \setminus A$, лежащие как можно «ближе» к cIA . Но, в силу открытости множества $\hat{B}(A) \setminus A$, решить эту задачу достаточно трудно. Покажем, что возможно дальнейшее упрощение алгоритма проверки наличия свойства пересечения.

3. ПУСТЫЕ НАПРАВЛЕНИЯ ОБОБЩЕННЫХ МЕДИАННЫХ СХЕМ

Следующим этапом в построении конструктивного алгоритма проверки наличия свойства пересечения для множества A у ССВК $W = \{W_k\}_{k=1}^m$, определенном на множестве $\hat{B}(A)$, будет формулировка свойств, которым оно наделяет все альтернативы из $\hat{B}(A)$. Эти свойства определяются на $\hat{B}(A)$ и являются критичными для проверки наличия свойства пересечения у $W = \{W_k\}_{k=1}^m$ для любого A' , такого, что $\hat{B}(A') = \hat{B}(A)$.

Определение 5. Пусть задано ССВК $W = \{W_k\}_{k=1}^m$, определенное на $\hat{B}(A)$. Подмножества критериев $M_r, M_l \subseteq M$, $M_r \cap M_l = \emptyset$ определяют для альтернативы $y \in \hat{B}(A)$ пустое направление $\{M_r, M_l\}_y$, если найдется такой набор выбирающих коалиций по всем критериям, что их пересечение пусто — т. е. $\forall k \in M_r \exists r_k(y) \in R_k(y)$ и $\forall k \in M_l \exists l_k(y) \in L_k^*(y)$ такие, что

$$\left[\bigcap_{k \in M_l} l_k(y) \right] \cap \left[\bigcap_{k \in M_r} r_k(y) \right] = \emptyset. \quad \blacklozenge$$

Пустое направление находится «правее» по всем критериям из множества M_r и «левее» по критериям из M_l . Альтернативным описанием пустых направлений будет запись $d(y) = \{d_1(y), \dots, d_m(y)\}$, $d(y) \in 3^M \setminus \emptyset$, где $d_k(y) = r$, если $k \in M_r$, $d_k(y) = l$, если $k \in M_l$, $d_k(y) = 0$, если $k \notin M_l \cup M_r$.

Легко заметить, что пустое направление означает, что для альтернативы $y \in \hat{B}(A)$ не выполняется условие (4), если будет проверяться именно эта комбинация критериев и выбирающих коалиций по ним. Качественно, смысл пустого направления очень простой и важный — если для множества A существует недопустимая альтернатива $y \in \hat{B}(A) \setminus A$, лежащая в своем пустом направлении относительно множества A : $d(y) = d(y, A)$, то ССВК $W = \{W_k\}_{k=1}^m$ не обладает свойством пересечения для A в y .

Для любой альтернативы может существовать не одно пустое направление, поэтому определим множество пустых направлений $D^0(y) = \{d(y)\}$. Справедлива следующая теорема, позволяющая проверять наличие свойства пересечения через пустые направления:

Теорема 1. Семейство систем выбирающих коалиций $W = \{W_k\}_{k=1}^m$, определенное на множестве $\hat{B}(A)$, обладает свойством пересечения для A , тогда и только тогда, когда $\forall y \in \hat{B}(A) \setminus A$ верно, что $d(y, A) \notin D^0(y)$. ♦

Рис. 4 иллюстрирует проверку свойства пересечения в терминах пустых направлений. Если направление «правее» по обоим критериям $d = (r, r)$ является пустым для альтернативы $y \in \hat{B}(A) \setminus A$, то ССВК, которым было определено данное пустое направление, не обладает свойством пересечения для множества A . Для ССВК из примера 1 (случай трех экспертов) множество альтернатив, для которого направление (r, r) является пустым, множество $\hat{B}_{rr}(A)$ изображено на рисунке серым цветом.

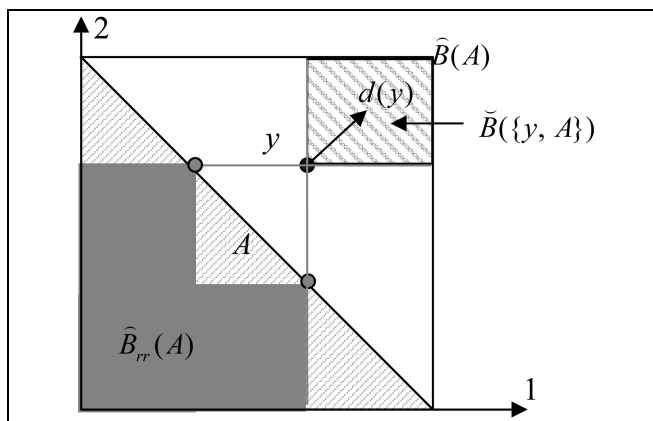


Рис. 4. К проверке свойства пересечения пустых направлений в терминах

Пустые направления для альтернатив обладают свойством, аналогичным монотонности свойства пересечения — если направление d пустое для какой-либо альтернативы $y \in \hat{B}(A)$, то оно также пустое для всех других альтернатив $\forall z \in \hat{B}(A)$ таких, что альтернатива y лежит в этом направлении относительно них: $d(y, z) = d$.

Лемма 2. Пусть вектор $d \in 3^M \setminus \emptyset$ задает пустое направление для альтернативы $y \in \hat{B}(A) - d \in D^0(y)$. Тогда для $\forall z \in \hat{B}(A)$, таких что $d(y, z) = d$ данный вектор также задает пустое направление: $d \in D^0(z)$. ♦

Из леммы 2 следует, что $D^0(y) \subseteq D^0(z)$.

Если направление (r, r) — (см. рис. 4) — пустое для альтернативы y , то для любой альтернативы, расположенной левее по обоим критериям от y (ближе к A), оно также пустое.

Опираясь на этот результат, можно определить понятие *критичных* альтернатив для каждого из пустых направлений.

Определение 6. Альтернатива $y \in \hat{B}(A)$ является критичной для пустого направления $d \in 3^M \setminus \emptyset$, если $\forall z \in \hat{B}(A)$, таких что $d(z, y) = d$, верно $d \notin D^0(z)$. ♦

Другими словами, альтернатива y является критичной для своего пустого направления, если для любой альтернативы, лежащей в этом направлении от y , оно не пустое. Если альтернатива y (см. рис. 4) является критичной для пустого направления (r, r) , то для всех альтернатив $z \in \hat{B}(\{y, A\})$, расположенных дальше от A , чем y , данное направление не пустое. Определим множество критичных альтернатив для произвольного пустого направления d :

$$Emp_d \hat{B}(A) = \{y \in \hat{B}(A) : d \in D^0(y), \forall z \in \hat{B}(A), d = d(z, y) \Rightarrow d \notin D^0(z)\}.$$

По сути, множество критичных альтернатив $Emp_d \hat{B}(A)$ является граничной поверхностью для всего подмножества альтернатив из множества $\hat{B}(A)$, для которых направление d пустое: $Emp_d \hat{B}(A) = cl\{y \in \hat{B}(A) : d \in D^0(y)\} \setminus cl \hat{B}(A)$. На рис. 5 жирной линией выделено множество $Emp_{(r,r)} \hat{B}(A)$ критичных альтернатив для направления (r, r) для ССВК из примера 1.

Отметим следующее достаточно очевидное, но важное свойство пустых направлений — их «расширяемость» по критериям.

Лемма 3. Пусть существует пустое направление $d(y)$ для альтернативы $y \in \hat{B}(A)$, такое что $\exists M_0 \subset M, \forall k \in M_0 d_k(y) = 0$. Тогда, пустыми являются все направления $d'(y) \neq d(y)$ такие что $\forall k \in M \setminus M_0 d'_k(y) = d_k(y)$. ♦

Пустые направления $d(y)$, для которых $\forall k \in M d_k(y) \neq 0$, будем называть *полноразмерными пустыми направлениями* и обозначим множество полноразмерных пустых направлений как $Df^0(y) \subseteq D^0(y)$. Результат леммы 3 значителен по следующей причине. В терминах определения 5 существует $3^M \setminus \emptyset$ возможных пустых направлений относительно практически любой альтернативы $y \in \hat{B}(A)$. В то время как полноразмерных пустых направлений относительно той же альтернативы может существовать $2^M \setminus \emptyset$, и для проверки свойства пересечения достаточно исследовать только их.

Леммы 2 и 3, вкуче с понятием пустого направления, позволяют сформулировать основной результат данной статьи.

Теорема 2. Семейство систем выбирающих коалиций $W = \{W_k\}_{k=1}^m$, определенное на множестве $\hat{B}(A)$, обладает свойством пересечения для множес-

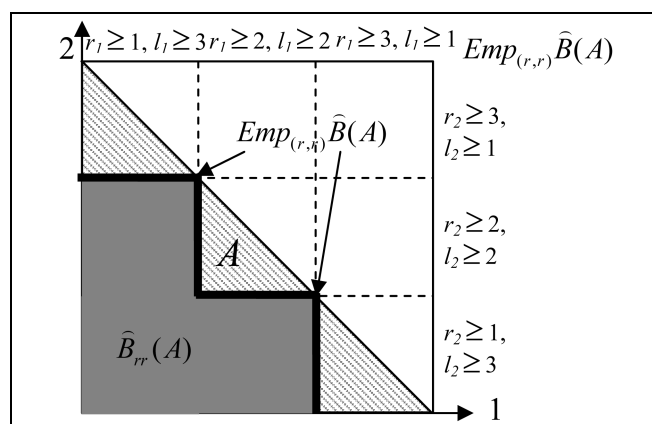


Рис. 5. Множество критичных альтернатив

тва A тогда и только тогда, когда для любого вектора направлений $d \in 2^M \setminus \emptyset$, такого что $\exists z \in \widehat{B}(A) \setminus A$, $d(z, A) = d$, верно, что для $\forall y \in \text{Emp}_d \widehat{B}(A)$ $d(y, A) \notin D^0(y)$.

Теорема 2 дает конструктивный способ проверки наличия свойства пересечения у определенной на многомерном прямоугольнике ССВК для произвольных множеств допустимых результатов выбора. Отметим, что перспективным представляется исследование множеств критичных альтернатив для дальнейшего упрощения алгоритма проверки свойства пересечения.

Алгоритм, позволяющий проверить ССВК на обладание свойством пересечения для некоторого множества допустимых альтернатив, выглядит следующим образом.

1. Для множества допустимых альтернатив A определяется минимальный многомерный прямоугольник $\widehat{B}(A)$, содержащий это множество.

2. Определяется множество направлений проверки $D(A)$: $d \in 3^M \setminus \emptyset \mid \exists y \in \widehat{B}(A) \setminus A: d(y, A) = d$.

3. Для каждого направления $d \in D(A)$:

а) определяется множество критичных альтернатив $\text{Emp}_d \widehat{B}(A)$;

б) проверяется, существуют ли недопустимые альтернативы, в которых ССВК не обладает свойством пересечения для A , т. е. $y \in \widehat{B}(A) \setminus A \cap \cap \text{Emp}_d \widehat{B}(A)$: $d(y, A) = d$.

Если, в соответствии с данным алгоритмом, найдена хотя бы одна такая недопустимая альтернатива, то исследуемая ССВК не обладает свойством пересечения для множества A . Если таких альтернатив не найдено, то ССВК обладает свойством пересечения для A . Следовательно, прямой неманипулируемый механизм, порождаемый исследуемой ССВК реализуем на данном множестве допустимых альтернатив A .

4. ПРИМЕНЕНИЕ ПУСТЫХ НАПРАВЛЕНИЙ ДЛЯ ПРОВЕРКИ СВОЙСТВА ПЕРЕСЕЧЕНИЯ

Проиллюстрируем проверку наличия свойства пересечения у ССВК предложенным алгоритмом.

Пример 2. Рассмотрим ССВК, порождающую прямой неманипулируемый механизм (2), эквивалентный механизму усреднения по каждому критерию (см. пример 1) для случая двух критериев и трех агентов:

$$h_k(\tau_k) = \max\{z_k \in [0, 1] \mid \# \{i \in \{1, 2, 3\} \mid \tau_k^i \geq z_k\} \geq nz_k\}, \\ k \in \{1, 2\}.$$

Проверим наличие свойства пересечения у этой ССВК для множества допустимых альтернатив A , моделирующего бюджетное ограничение $A = \{x \in \mathbb{R}^2 \mid x_1 +$

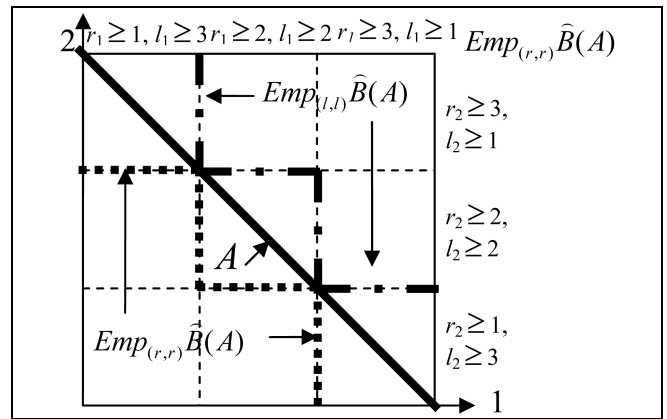


Рис. 6. Множество критичных альтернатив для балансового ограничения

$+ x_2 \leq 1, x_1, x_2 \geq 0\}$. На рис. 5 изображены ССВК и множество A . Все недопустимые альтернативы из множества $\widehat{B}(A) \setminus A$ в данном примере лежат в направлении (r, r) относительно множества A . Поэтому достаточно исследовать только множество критичных альтернатив для данной ССВК для направления (r, r) — $\text{Emp}_{(r,r)} \widehat{B}(A)$, которое также изображено на рис. 5. Видно, что $\text{Emp}_{(r,r)} \widehat{B}(A) \subseteq A$, следовательно, любая альтернатива из множества $\text{Emp}_{(r,r)} \widehat{B}(A)$ не лежит в направлении (r, r) от множества допустимых альтернатив A . Поэтому по теореме 2 ССВК, порождающая прямой неманипулируемый механизм, эквивалентный механизму усреднения по двум критериям для трех экспертов:

$$x = f(s): \forall j \in \{1, 2\} \ x_j = \frac{1}{3} \sum_{i=1}^3 s_j^i,$$

обладает свойством пересечения для множества $A = \{x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 1, x_1, x_2 \geq 0\}$. ♦

Пример 3. Проверим, обладает ли ССВК из примера 2 свойством пересечения для множества допустимых альтернатив A , моделирующего «балансовое» ограничение — $A = \{x \in \mathbb{R}^2 \mid x_1 + x_2 = 1, x_1, x_2 \geq 0\}$. На рис. 6 изображены ССВК и множество A (жирная линия).

Все недопустимые альтернативы из множества $\widehat{B}(A) \setminus A$ в данном примере лежат в направлениях (r, r) и (l, l) относительно множества A . Поэтому достаточно исследовать только множества критичных альтернатив для данной ССВК для данных направлений — $\text{Emp}_{(r,r)} \widehat{B}(A)$, которое изображено жирной пунктирной линией (см. рис. 6) и $\text{Emp}_{(l,l)} \widehat{B}(A)$ — жирная штрихпунктирная линия. В данном примере $\text{Emp}_{(r,r)} \widehat{B}(A) \not\subseteq A$ и $\text{Emp}_{(l,l)} \widehat{B}(A) \not\subseteq A$, но любая альтернатива из множества $\text{Emp}_{(r,r)} \widehat{B}(A)$ не лежит в направлении (r, r) допустимых альтернатив A и любая альтернатива из $\text{Emp}_{(l,l)} \widehat{B}(A)$ не лежит в направлении (l, l) от множества допустимых

альтернатив A . Следовательно, по теореме 2, ССВК, прямой неманипулируемый механизм, эквивалентный механизму усреднения по двум критериям для трех экспертов:

$$x = f(s): \forall j \in \{1, 2\} x_j = \frac{1}{3} \sum_{i=1}^3 s_j^i,$$

обладает свойством пересечения для множества $A = \{x \in \mathbb{R}^2 | x_1 + x_2 = 1, x_1, x_2 \geq 0\}$. ♦

Пример 4. Покажем, как предложенный алгоритм показывает, что ССВК не обладает свойством пересечения для некоторого множества допустимых альтернатив. Для этого рассмотрим ССВК, порождающую прямой неманипулируемый механизм, эквивалентный механизму усреднения по каждому критерию (см. пример 1) для случая трех критериев и трех агентов:

$$h_k(\tau_k) = \max\{z_k \in [0, 1] | \#\{i \in \{1, 2, 3\} | \tau_k^i \geq z_k\} \geq n z_k\},$$

$$k \in \{1, 2, 3\}. \quad (5)$$

Проверим наличие свойства пересечения у этой ССВК для множества допустимых альтернатив A , моделирующего бюджетное ограничение

$$A = \{x \in \mathbb{R}^3 | x_1 + x_2 + x_3 \leq 1, x_1, x_2, x_3 \geq 0\}.$$

По аналогии с примером 2, все недопустимые альтернативы из множества $\hat{B}(A) \setminus A$ лежат в направлении (r, r, r) относительно множества A . На рис. 7 изображены ССВК, множество A (заштрихованная область) и $Emp_{(r,r,r)} \hat{B}(A)$ (затемненная область).

Недопустимая альтернатива $y = (2/3, 1/3, 1/3)$, изображенная белым кружком на рис. 7 принадлежит множеству $Emp_{(r,r,r)} \hat{B}(A)$. Следовательно, по теореме 2, ССВК, порождающая прямой неманипулируемый механизм, эквивалентный механизму усреднения по трем критериям для трех экспертов:

$$x = f(s): \forall j \in \{1, 2, 3\} x_j = \frac{1}{3} \sum_{i=1}^3 s_j^i,$$

не обладает свойством пересечения для $A = \{x \in \mathbb{R}^3 | x_1 + x_2 + x_3 = 1, x_1, x_2, x_3 \geq 0\}$.

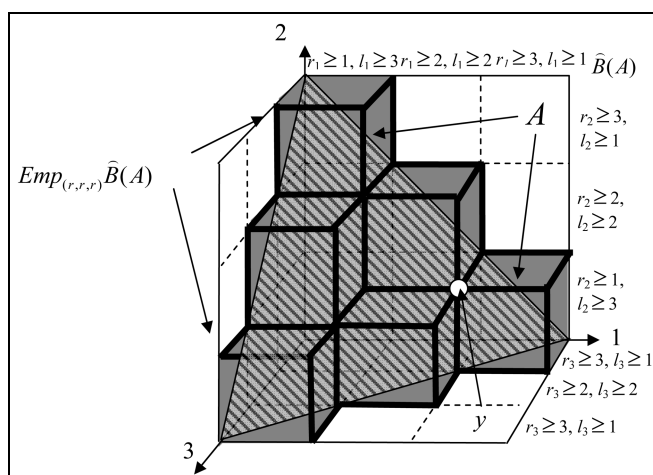


Рис. 7. Пример для случая трех критериев

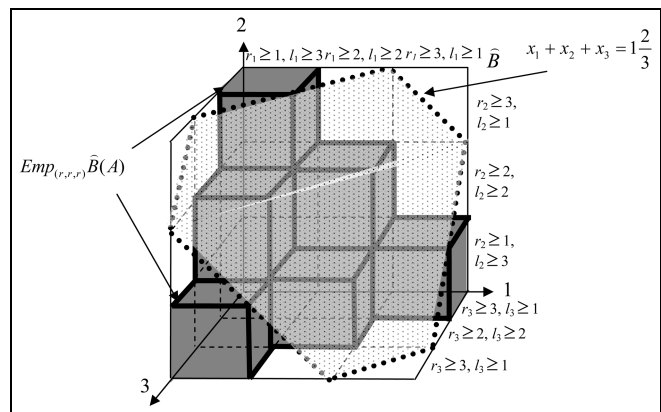


Рис. 8. Определение допустимых бюджетных ограничений

Действительно, возьмем трех агентов со следующими точками пиков: $\tau^1 = (2/3, 0, 1/3)$, $\tau^2 = (1, 0, 0)$, $\tau^3 = (0, 1, 0)$. Очевидно, что $\forall i \in \{1, 2, 3\} \tau^i \in A$. Из выражения (5) получаем, что результат выбора по критерию 1 $h_1(2/3, 1, 0) = 2/3$, по критерию 2 $h_2(0, 0, 1) = 1/3$, по критерию 3 $h_3(1/3, 0, 0) = 1/3$, т. е. результатом выбора будет недопустимая альтернатива $y = (2/3, 1/3, 1/3)$. ♦

На этом же примере покажем, как полученные результаты могут быть применены для (r, r) множества допустимых альтернатив, для которого заданная ССВК обладает свойством пересечения. Определим, для каких «бюджетных» ограничений $A = \{x \in \mathbb{R}^3 | x_1 + x_2 + x_3 \leq C, x_1, x_2, x_3 \in [0, 1]\}$ прямой неманипулируемый механизм (5) реализуем.

Легко видеть, что при $C \geq 1 \frac{2}{3}$ любая альтернатива

из множества $Emp_{(r,r,r)} \hat{B}(A)$ не лежит в направлении (r, r) от множества допустимых альтернатив A — (рис. 8); т. е. ССВК, порождающая прямой механизм (5), обладает свойством пересечения для множества

$$A = \{x \in \mathbb{R}^3 | x_1 + x_2 + x_3 \leq 1 \frac{2}{3}, x_1, x_2, x_3 \in [0, 1]\}.$$

Если рассмотреть данный пример как задачу распределения ресурсов между тремя проектами на основании сообщений трех экспертов, то это будет означать, что при применении прямого неманипулируемого механизма (5) из суммарного бюджета C на каждый проект в отдельности можно выделять не более $\frac{3}{5} C$ средств.

ЗАКЛЮЧЕНИЕ

Показана возможность упрощения проверки свойства пересечения для обобщенных медианных схем. Предложен алгоритм, позволяющий для



произвольной обобщенной медианной схемы определить — для каких множеств допустимых результатов выбора эта схема удовлетворяет свойству пересечения, а для каких нет. К недостатку алгоритма следует отнести трудоемкость построения множеств критичных альтернатив. В дальнейшем будет предложено развитие данного алгоритма, основанное на том факте, что пустые направления для альтернатив из минимального многомерного прямоугольника, в который вписывается множество допустимых результатов выбора, изменяются дискретно — семейство систем выбирающих коалиций порождает разбиение данного прямоугольника на «блоки», также представляющие собой многомерные прямоугольники. Для проверки свойства пересечения у медианной схемы достаточно проверять по одной недопустимой альтернативе из каждого такого блока, через которые проходит граница множества допустимых результатов выбора.

ПРИЛОЖЕНИЕ

Доказательство леммы 1. Из определения систем выбирающих коалиций следует их монотонность: для любого вектора $y', y'' \in \hat{B}(A)$ и произвольного $k \in M$ если $y'_k \geq y''_k$, то $R_k(y'_k) \subseteq R_k(y''_k)$, если $y'_k \leq y''_k$, то $L_k^*(y'_k) \subseteq L_k^*(y''_k)$. Также достаточно очевиден тот факт, что $\forall y' \in \tilde{B}(\{y, A\})$ верно $M^+(\tilde{y}, A) = M^+(y, A)$ и $M^-(\tilde{y}, A) = M^-(y, A)$.

Из «монотонности» семейств систем выбирающих коалиций следует, что (4) верно также для $\forall \tilde{y} \in \tilde{B}(y, A)$, так как множества M^+ и M^- остаются неизменными. ♦

Доказательство теоремы 1. Пусть ССВК $W = \{W_k\}_{k=1}^m$ обладает свойством пересечения для A . Тогда $\forall y \in \hat{B}(A) \setminus A$ она обладает свойством пересечения для A в y . Следовательно, из определения пустого направления, $\forall y \in \hat{B}(A) \setminus A$ $d(y, A) \notin D^0(y)$, т. е. любая недопустимая альтернатива не лежит в своем пустом направлении относительно множества A .

Пусть $\forall y \in \hat{B}(A) \setminus A$ верно, что $d(y, A) \notin D^0(y)$, но ССВК $W = \{W_k\}_{k=1}^m$ не обладает свойством пересечения. Это значит, что $\exists y \in \hat{B}(A) \setminus A$ такая, что для нее не выполняется условие (4); т. е. $d(y, A) \in D^0(y)$. Получили противоречие. ♦

Доказательство леммы 2. Из определения систем выбирающих коалиций следует их монотонность: для любых альтернатив $y', y'' \in \hat{B}(A)$ и произвольного $k \in M$ если $y'_k \geq y''_k$, то $R_k(y'_k) \subseteq R_k(y''_k)$, если $y'_k \leq y''_k$, то $L_k^*(y'_k) \subseteq L_k^*(y''_k)$. Следовательно, для $\forall k \in M^-(y, z)$ $R_k(y_k) \subseteq R_k(z_k)$ и $\forall k \in M^+(y, z)$ $L_k^*(y_k) \subseteq L_k^*(z_k)$. Следовательно, для выбирающих коалиций для

альтернативы z таких, что $\forall k \in M_r$ $r_k(z) = r_k(y) \in R_k(z)$ и $\forall k \in M_l$ $l_k(z) = l_k(y) \in L_k^*(z)$ верно

$$\left[\bigcap_{k \in M_l} l_k(z) \right] \cap \left[\bigcap_{k \in M_r} r_k(z) \right] \cdot \blacklozenge$$

Доказательство леммы 3. Пусть направление $d(y)$ описывается множествами $M_r, M_l \subseteq M$, а направление $d'(y) = M'_r, M'_l \subseteq M$. Тогда $M_r \subseteq M'_r$ и $M_l \subseteq M'_l$. Следовательно, для направления $d'(y)$ верно:

$$\begin{aligned} & \left[\bigcap_{k \in M'_l} l_k(y) \right] \cap \left[\bigcap_{k \in M'_r} r_k(y) \right] = \\ & = \left[\bigcap_{k \in M'_l \setminus M_l} l_k(y) \right] \cap \left[\bigcap_{k \in M'_r \setminus M_r} r_k(y) \right] \cdot \blacklozenge \end{aligned}$$

Доказательство теоремы 2. Необходимость следует очевидным образом из теоремы 1.

Достаточность. Пусть для $\forall d \in 2^M/\emptyset$ такая, что $\exists y \in \hat{B}(A) \setminus A$, $d(y, A) = d$ верно, что $\forall y \in \text{Emp}_d \hat{B}(A)$ $d(y, A) \neq d$. Предположим, что $\exists \tilde{d} \in 2^M/\emptyset$ и $\exists \tilde{y} \in \hat{B}(A) \setminus A$ такая, что $d(\tilde{y}, A) = \tilde{d}$ и $\tilde{d} \in D^0(\tilde{y})$; т. е. ССВК не обладает свойством пересечения для A в $\tilde{y}$. Из определения 6 следует, что $\forall y' \in \text{Emp}_{\tilde{d}} \hat{B}(A)$ такая, что $d(\tilde{y}, y') = \tilde{d}$; т. е. альтернатива $\tilde{y}$ лежит ближе к множеству A , чем альтернатива y' ; т. е. $d(y', A) = \tilde{d}$. Получили противоречие. ♦

ЛИТЕРАТУРА

1. *Большие системы: моделирование организационных механизмов* / В.Н. Бурков, Б. Данев, А.К. Еналеев и др. — М.: Наука, 1989. — 245 с.
2. *Новиков Д.А., Петраков С.Н.* Курс теории активных систем. — М.: СИНТЕГ, 1999. — 108 с.
3. *Moulin H.* On strategy-proofness and single-peakedness // Public Choice. — 1980. — Vol. 35. — P. 437–455.
4. *Бурков В.Н., Исаков М.Б., Коргин Н.А.* Применение обобщенных медианных схем для построения неманипулируемого механизма многокритериальной активной экспертизы // Проблемы управления. — 2008. — № 4. — С. 38–47.
5. *Border K., Jordan J.* Straightforward elections, unanimity and phantom voters // Review of Economic Studies. — 1983. — Vol. 50. — P. 153–170.
6. *Barbera S., Masso J., Serizawa S.* Strategy-proof voting on compact ranges // Games and Behavior. — 1998. — Vol. 25. — P. 272–291.
7. *Barbera S., Masso J., Neme A.* Voting under Constraints // J. Econ. Theory. — 1997. — Vol. 76. — P. 298–321.

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

Коргин Николай Андреевич — канд. техн. наук, ст. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН,
☎(495) 335-60-37, ✉nkorgin@ipu.ru.

К ПРОБЛЕМЕ ОПТИМАЛЬНОГО РЕГУЛИРОВАНИЯ ПРОЦЕССА ИЗВЛЕЧЕНИЯ НЕФТИ С УЧЕТОМ ВЕРОЯТНОСТНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ НЕКОТОРОГО ПАРАМЕТРА СЛОИСТОГО ПЛАСТА¹

Р.А. Султанов, В.Д. Слабнов, И.А. Фукин, В.В. Скворцов

Предложена математическая модель процесса двухфазной фильтрации в слоистых пластах, параметры которых могут изменяться по толщине и простираию. Вероятностный закон распределения абсолютной проницаемости по толщине пласта описан дискретным аналогом распределения случайной величины. Сформулированы четыре гипотезы о вытеснении нефти водой из элементарного объема слоистого пласта. Выполнено сравнение основных технологических показателей извлечения нефти, полученных без учета и с учетом оптимального регулирования.

Ключевые слова: математическая модель, двухфазная фильтрация, вероятностный закон распределения, случайная величина, слоистый пласт, оптимальное регулирование.

ВВЕДЕНИЕ

Необходимость повышения объемов нефти, извлекаемой из пластов сложного геологического строения, характеристики которых изменяются по толщине и простираию пласта, делает актуальной проблему оптимального регулирования процесса её извлечения. Для её решения используются усредненные математические модели. Преимущество двухмерных моделей по сравнению с трехмерными связано с недостатком геологической информации о строении пласта и значительным сокращением времени численных расчетов благодаря уменьшению числа ячеек конечно-разностной сетки [1].

В данной работе используется модель Баклея — Леверетта процесса вытеснения нефти водой из слоистых пластов, не учитывающей гравитационные и капиллярные силы [2]. Принимаются во внимание четыре гипотезы о вытеснении нефти

водой из элементарного объема слоистого пласта. Предполагается, что абсолютная проницаемость слоистого пласта может изменяться по толщине и простираию. Вероятностный закон ее распределения в слоистом пласте определяется на основе заданных параметров распределения: математического ожидания, дисперсии и некоторого дополнительного условия. Полученный вероятностный закон используется для построения «фиктивных» (модифицированных) относительных фазовых проницаемостей в слоистом пласте [2]. Усреднение трехмерной модели по толщине пласта позволяет уменьшить размерность исходной модели на единицу.

Численное решение задачи управления для построенной усредненной математической модели позволяет определить симплекс-методом оптимальные дебиты и забойные давления источников и стоков при задании для них двусторонних ограничений и некоторых вариантов граничных условий в элементе заводнения. Критерием оптимизации служит максимизация целевой функции — суммарного отбора нефти [3, 4, 6, 7].

Цель данной работы — построение математической модели двухфазной фильтрации в слоистом

¹ Основное содержание статьи доложено на IV Международной конференции по проблемам управления, Москва, Институт проблем управления им. В.А. Трапезникова РАН, 26—30 января 2009 г.



пласте с учетом четырех гипотез о вытеснении нефти водой из элементарного объема; постановка задачи оптимального управления; разработка численного алгоритма ее решения; анализ результатов решения задачи оптимального регулирования извлечения нефти из слоистого пласта при варьировании параметров распределения для дискретных аналогов равномерного и логарифмически-нормального законов распределения абсолютной проницаемости; сравнение результатов численного решения задачи без учета и с учетом регулирования.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРОЦЕССА 2D ДВУХФАЗНОЙ ФИЛЬТРАЦИИ В СЛОИСТЫХ ПЛАСТАХ, НЕОДНОРОДНЫХ ПО ТОЛЩИНЕ И ПРОСТИРАНИЮ

Рассмотрим нефтяной пласт, в котором происходит процесс вытеснения нефти водой. Кровлю и подошву пласта будем считать непроницаемыми. Введем прямоугольную декартову систему координат x, y, z . Выделим элементарный объем в виде столбца, ограниченного сверху кровлей, а снизу — подошвой. Его основанием будет служить элементарная площадка dx, dy , а высотой — толщина пласта H .

Будем считать, что пласт неоднороден по абсолютной проницаемости, причем проницаемость изменяется по толщине (вертикали) и простиранию (горизонтали) пласта. Это изменение хаотично, и можно считать, что пласт представляет собой n_k сред с различными значениями абсолютной проницаемости. Поэтому j -й средой на произвольно выбранной вертикали пласта будем называть среду, содержащую все те пропластки этой вертикали, которые имеют абсолютную проницаемость $k_{a(j)}$. Будем полагать, что производные

$\frac{\partial p}{\partial x}$ и $\frac{\partial p}{\partial y}$, где p — пластовое давление, постоянны по вертикали и, в частности, не зависят от j [2].

Распределение абсолютной проницаемости в слоистом пласте с учетом дискретного аналога вероятностного закона распределения случайной величины можно задавать с помощью гистограммы или ряда распределения, приведенного в табл. 1, где β_j — вероятность того, что абсолютная проницаемость примет значение $k_{a(j)}$.

Таблица 1

Распределение абсолютной проницаемости

$k_{a(j)}$	$k_{a(1)}$	$k_{a(2)}$	$k_{a(3)}$	...	$k_{a(n_k)}$
β_j	β_1	β_2	β_3	...	β_{n_k}

Сделаем еще ряд предположений относительно механизма вытеснения нефти водой. Пусть каждая j -я среда с абсолютной проницаемостью $k_{a(j)}$ имеет свою остаточную нефтенасыщенность $s_{*n(j)}$ и долю связанной воды $s_{*в(j)}$, которые считаются известными. В этом случае максимальные значения объемов воды и нефти в каждом элементарном столбце имеют следующий вид

$$s_{*в(j)}^* = 1 - s_{*н(j)}, \quad s_{*н(j)}^* = 1 - s_{*в(j)},$$

где буквенный индекс «в» относится к водной фазе, «н» — к нефтяной фазе.

В общем случае j -я среда может состоять из отдельных частей, которые могут быть расположены в разных местах элементарного столбца.

Пусть $s_{в(j)}$ и $s_{н(j)}$ — водо- и нефтенасыщенность j -й среды соответственно.

Определим активные нефте- и водонасыщенность j -й среды как

$$\theta_j = \frac{s_{н(j)} - s_{*н(j)}}{s_{*н(j)}^* - s_{*н(j)}}, \quad 1 - \theta_j = \frac{s_{в(j)} - s_{*в(j)}}{s_{*в(j)}^* - s_{*в(j)}}.$$

Отсюда нефте- и водонасыщенность j -й среды можно представить в виде

$$s_{н(j)} = s_{*н(j)}(1 - \theta_j) + s_{*н(j)}^*, \\ s_{в(j)} = s_{*в(j)}\theta_j + s_{*в(j)}^*(1 - \theta_j).$$

Тогда нефте- и водонасыщенность в элементарном объеме имеют следующий вид:

$$s_{н} = \sum_{j=1}^{n_k} s_{*н(j)}(1 - \theta_j)\beta_j + \sum_{j=1}^{n_k} s_{*н(j)}^*\theta_j\beta_j, \\ s_{в} = \sum_{j=1}^{n_k} s_{*в(j)}\theta_j\beta_j + \sum_{j=1}^{n_k} s_{*в(j)}^*(1 - \theta_j)\beta_j. \quad (1)$$

Очевидно, что

$$\sum_{j=1}^{n_k} \theta_j\beta_j + \sum_{j=1}^{n_k} (1 - \theta_j)\beta_j = 1.$$

Обозначим через $k_{в(j)}(s_{в})$ и $k_{н(j)}(s_{н})$ относительные фазовые проницаемости воды и нефти в j -й среде соответственно. Как видно из формулы (1), они будут зависеть от величины θ_j .

Составим уравнение материального баланса для i -й жидкости:

$$\frac{\partial}{\partial x} \left[\frac{H \tilde{k}_a k_i^\phi}{\mu_i} \frac{\partial p}{\partial x} \right] + \frac{\partial}{\partial y} \left[\frac{H \tilde{k}_a k_i^\phi}{\mu_i} \frac{\partial p}{\partial y} \right] = -mH \frac{\partial s_i}{\partial t}, \quad (2)$$

где $i = \text{в, н}$, $\tilde{k}_a$ — осредненная абсолютная проницаемость, m — пористость пласта; $\mu_{\text{в}}$ и $\mu_{\text{н}}$ — вязкости воды и нефти.

В данной математической модели «фиктивные» относительные фазовые проницаемости $k_i^\Phi(s_i)$, $i = \text{в, н}$, построены на основе известных максимальных значений относительных фазовых проницаемостей, активных насыщенностей сред и вероятностного закона распределения абсолютной проницаемости по толщине пласта.

Таким образом, математическую модель вытеснения нефти из пласта толщиной H , в котором абсолютная проницаемость меняется по вертикали и горизонтали, при некоторых допущениях удастся свести к математической модели процесса фильтрации в некотором пласте с усредненной абсолютной проницаемостью по толщине пласта с учетом «фиктивных» (модифицированных) относительных фазовых проницаемостей.

Систему уравнений (2) после несложных преобразований и учета закона Дарси можно привести к системе уравнений относительно усредненной скорости фильтрации $\tilde{V}$ и нефтенасыщенности s :

$$\operatorname{div}(\tilde{V} H) = \tilde{q}, \operatorname{div}(\tilde{V} H \tilde{F}(s)) = m H \frac{\partial s}{\partial t}, \quad (3)$$

где $\tilde{V} = -\tilde{\sigma}(s) \operatorname{grad} p$ — суммарная усредненная скорость фильтрации нефти и воды, $\tilde{q}$ — усредненный дебит скважины, $\tilde{\sigma}(s) = \tilde{k}_a \left(\frac{k_{\text{н}}^\Phi(s)}{\mu_{\text{н}}} + \frac{k_{\text{в}}^\Phi(s)}{\mu_{\text{в}}} \right)$ — суммарная усредненная проводимость в слоистом пласте, $\tilde{F}(s) = \tilde{k}_a \frac{k_{\text{н}}^\Phi(s)}{\mu_{\text{н}}} / \tilde{\sigma}(s)$ — усредненная доля нефти в потоке жидкости;

$$\begin{aligned} k_{\text{н}}^\Phi(s_{\text{н}}) &= \frac{1}{k_a} \sum_{j=1}^{n_k} \beta_j k_{a(j)} k_{\text{н}(j)}(s_{\text{н}}), \\ k_{\text{в}}^\Phi(s_{\text{в}}) &= \frac{1}{k_a} \sum_{j=1}^{n_k} \beta_j k_{a(j)} k_{\text{в}(j)}(s_{\text{в}}). \end{aligned} \quad (4)$$

Для решения системы уравнений (3) требуется знать вид «фиктивных» относительных фазовых проницаемостей (4) по некоторому вероятностному закону распределения случайной величины или вид ряда распределения абсолютной проницаемости в каждой точке усредненного слоистого пласта.

2. ПОСТАНОВКА ЗАДАЧИ ОПТИМАЛЬНОГО РЕГУЛИРОВАНИЯ ПРОЦЕССА ИЗВЛЕЧЕНИЯ НЕФТИ

Постановка задачи регулирования при двухфазной фильтрации с учетом ввода скважин в эксплуатацию на основе математической модели Баклея — Леверетта (3) с начальными и граничными условиями и алгоритм ее решения была дана в работе [5]. Поставим в соответствие размещению k -й скважины в области фильтрации D ее номер. Номера добывающих скважин отнесем к множеству J^* , нагнетательных — к множеству $J \setminus J^*$, где J — множество номеров всех скважин. Предполагается, что срок T разработки залежи определяется временем отключения последней действующей добывающей скважины. Отключение k -й добывающей скважины происходит при $T_k \leq t \leq T$, где T_k — время достижения k -й скважиной заданного предельно-допустимого значения текущей обводненности ее продукции $1 - F(T_k) = F_{\text{вк}}^*$. Будем рассматривать регулирование процесса нефтедобычи из неоднородной залежи в условиях принятой модели по критерию максимизации текущей нефтеотдачи на заданные моменты времени $t_r, r = 0, 1, 2, \dots$

Математическая задача формулируется следующим образом: найти максимальный суммарный отбор нефти

$$\sum_{k \in J^*(t_r)} \int_{t_n}^{t_{n+1}} q_k(t_r) F_{\text{нк}}(t_r) dt_r \rightarrow \max, \quad t_r \in [t_n, t_{n+1}],$$

для заданного планового периода $[t_n, t_{n+1}]$ и ограничений на режимы действующего фонда скважин

$$p_* \leq p_{3k} \leq p^*, \quad q_* \leq |q_k| \leq q^*, \quad \forall k \in J^*(t_r),$$

с учетом момента времени t_r начала задачи прогнозирования, который определяется из условия

$$t_r = t(F_{\text{вк}}(t_r)) \geq t(F_{\text{вк}}^*(t_r)), \quad \forall k \in J^*(t_r),$$

где номера добывающих и нагнетательных скважин соответствуют оптимальному подмножеству $J^*(t_r) \cup J(t_r) \setminus J^*(t_r)$.

Здесь $q_k(t_r)$ — дебит k -й скважины на t_r -ю дату регулирования, $F_{\text{вк}}(t_r)$ — доля воды в потоке жидкости для k -й скважины на t_r -ю дату регулирования, $p_{3k}(t_r)$ — забойное давление k -й скважины на t_r -ю дату регулирования.

При численном решении задачи управления нефтеизвлечением на каждую контрольную дату t_r восстанавливается оптимальное поле давления в области фильтрации D в результате численного



решения задачи линейного программирования (5)–(7) [3, 5, 7, 8]:

$$\sum_{k \in J^*(t_r)} \alpha \sigma^k(s(t_r))(p_{ij}^k(t_r) - p_{3k}(t_r)) F_{nk}(t_r) dt_r \rightarrow \max, \\ t_r \in [t_n, t_{n+1}], \quad (5)$$

при ограничениях

$$\sigma_{i-1j} p_{i-1j} + \sigma_{i+1j} p_{i+1j} + \sigma_{ij-1} p_{ij-1} + \sigma_{ij+1} p_{ij+1} + \\ + (\sigma_{ij} + \alpha \delta_{ij}) p_{ij} + \alpha \delta_{ij} p_{3k} = 0 \quad (6)$$

$$p_* \leq p_{3k} \leq p^*, \quad q_* \leq |\alpha \sigma^k(s(t_r))(p_{ij}^k(t_r) - p_{3k}(t_r))| \leq q^*, \\ \forall k \in J^*(t_r), \quad (7)$$

где $i = \overline{1, M_x}, j = \overline{1, M_y}$, M_x и M_y — число узлов по оси OX и OY соответственно, p_{3k} — забойное давление скважин с координатами (x_k, y_k) , p_{ij} — сеточное давление по области фильтрации D , δ_{ij} — сеточный аналог функции Дирака, p_* и p^* — нижнее и верхнее ограничения на забойные давления скважин соответственно, q_* и q^* — нижнее и верхнее ограничения на дебиты скважин соответственно.

Найденное поле давления вместе со значениями оптимальных забойных давлений и дебитов действующих скважин обеспечивают при решении задачи управления максимальный суммарный отбор нефти из неоднородного пласта на t_r -ю контрольную дату. Здесь линейное программирование используется одновременно в качестве метода решения системы алгебраических уравнений, полученных из разностного аналога (6) для первого уравнения системы (3).

3. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Рассмотрена модель неоднородного пласта прямоугольной конфигурации $D = \{0 \leq x \leq a, 0 \leq y \leq b\}$, вскрытого девятиточечным элементом заводнения (рис. 1), для $a = b = 850$ м, $H = 5$ м, $m = 0,2$, $\mu_n = 3$ мПа·с, $\mu_v = 1$ мПа·с, $s_* = 0,26$, $s^* = 0,8$, $p_k = 22$ МПа, $p_* = 12,5$ МПа, $p^* = 22$ МПа, $q_* = 1$ м³/сут, $q^* = 500$ м³/сут, $h_x = h_y = 25$ м. Здесь h_x, h_y — шаги разностной сетки по осям координат. Относительные фазовые проницаемости $k_n(s)$ и $k_v(s)$ — линейные полиномы [5]. Проведена серия вычислительных экспериментов для элемента заводнения в соответствии со следующими гипотезами о характере вытеснения нефти водой.

1. Вода вытесняет нефть из участков (зон) пласта в первую очередь с наилучшей абсолютной проницаемостью, т. е. каждый элементарный объем разбивается на две зоны: водную и нефтяную

(в водной зоне нефтенасыщенность предполагается остаточной, в нефтяной зоне — максимальной). Разделение на две зоны определяется найденным значением абсолютной проницаемости, которое будем называть пороговым значением.

2. Вода вытесняет нефть из участков (зон) пласта в первую очередь с наихудшей абсолютной проницаемостью, т. е. каждый элементарный объем разбивается пороговым значением на две зоны: водную и нефтяную (в водной зоне нефтенасыщенность предполагается остаточной, в нефтяной зоне — максимальной).

3. Вода вытесняет нефть пропорционально найденному переменному коэффициенту.

4. Обобщение гипотез 1 и 3. Каждый элементарный объем разбивается пороговым значением на две зоны: водную и нефтяную. Причем в зоне с наихудшей абсолютной проницаемостью нефтенасыщенность максимальна, а в зоне с наилучшей абсолютной проницаемостью — происходит неполное вытеснение нефти водой пропорционально найденному переменному коэффициенту.

Оптимальные показатели извлечения нефти из слоистого пласта для этих гипотез определялись на основе численного решения задачи управления, где пластовое и забойные давления действующих скважин определялись по предложенному алгоритму. Они сравнивались с неоптимальными показателями, полученными на основе численного решения задачи прогнозирования, где пластовое давление определялось по методу переменных направлений, а забойные давления скважин были заданы.

На рис. 1 показано размещение 9 скважин в элементе заводнения для следующего варианта задания граничных условий: на одной стороне задано граничное условие I рода, контур питания КП, на трех остальных сторонах — граничное условие II рода, непроницаемая граница НГ. Математическое ожидание абсолютной проницаемости $k_a(x, y)$ в области двухфазной фильтрации D элемента за-

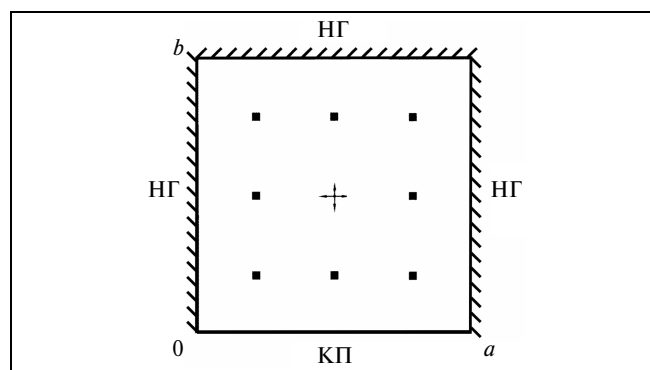


Рис. 1. Элемент заводнения слоистого пласта

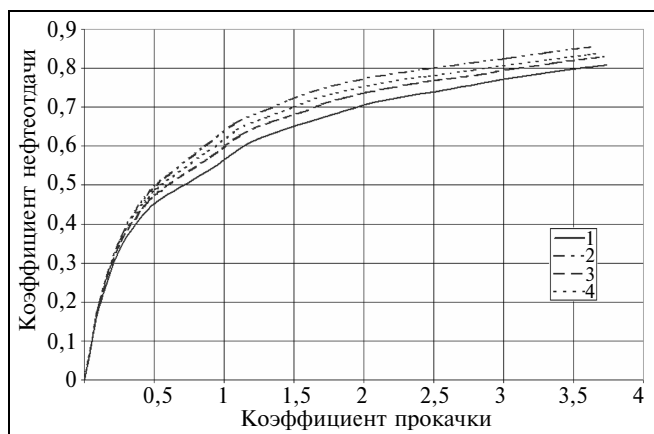


Рис. 2. Графики 1—4 коэффициента нефтеотдачи $\eta(\tau)$ для гипотез 1—4 соответственно. Дискретный аналог логарифмически-нормального закона, $\tilde{k}_a = 0,1 \text{ мкм}^2$; $\sigma^2 = 0,001$

воднения принимает следующие значения: $\tilde{k}_a = 0,1$; $0,3 \text{ мкм}^2$, дисперсия — $\sigma^2 = 0,001$; $0,01$.

Вычислительные эксперименты проводились для дискретных аналогов двух законов распределения:

— ряд распределения с одинаковыми вероятностями того, что случайная величина k_a по толщине пласта примет данное значение (дискретный аналог равномерного закона распределения случайной величины, табл. 2);

— ряд распределения с различными вероятностями того, что случайная величина k_a по толщине пласта примет данное значение (дискретный аналог логарифмически-нормального закона распределения случайной величины, табл. 3).

В табл. 4 приведены значения коэффициента прокачки в поровых объемах и соответствующие им значения коэффициента конечной нефтеотдачи для двух законов распределения случайной величины с учетом заданий математического ожидания, дисперсии и гипотез 1, 2 и 4 о вытеснении нефти водой из элементарного объема слоистого пласта. Анализируя данные табл. 4 и графики, представленные на рис. 2, можно сделать вывод, что гипотезы 1 и 2 соответствуют нижнему и верхнему графикам значений коэффициента нефтеотдачи η от коэффициента прокачки поровых объемов τ .

Таблица 2

Равномерный закон

$\tilde{k}_a = 0,1 \text{ мкм}^2$ $\sigma^2 = 0,001$	$k_{a(j)}$	0,051	0,063	0,075	0,087	0,100	0,112	0,124	0,136	0,149
	β_j	0,111	0,111	0,111	0,111	0,111	0,111	0,111	0,111	0,111

Таблица 3

Логарифмически-нормальный закон

$\tilde{k}_a = 0,1 \text{ мкм}^2$ $\sigma^2 = 0,001$	$k_{a(j)}$	0,058	0,078	0,098	0,118	0,137	0,157	0,177	0,196	0,216
	β_j	0,050	0,400	0,300	0,100	0,050	0,040	0,030	0,020	0,010

Таблица 4

Значения коэффициентов прокачки и нефтеотдачи

Вид закона распределения		Дискретный аналог равномерного закона			Дискретный аналог логарифмически-нормального закона		
Номер гипотезы		1	2	4	1	2	4
$\tilde{k}_a = 0,1 \text{ мкм}^2$ $\sigma^2 = 0,001$	τ	3,744	3,624	3,697	3,737	3,635	3,675
	η	0,791	0,847	0,831	0,807	0,853	0,837
$\tilde{k}_a = 0,3 \text{ мкм}^2$ $\sigma^2 = 0,001$	τ	3,719	3,702	3,719	3,725	3,72	3,719
	η	0,819	0,836	0,829	0,823	0,839	0,829
$\tilde{k}_a = 0,3 \text{ мкм}^2$ $\sigma^2 = 0,01$	τ	3,747	3,617	3,719	3,738	3,625	3,677
	η	0,788	0,848	0,829	0,806	0,854	0,836

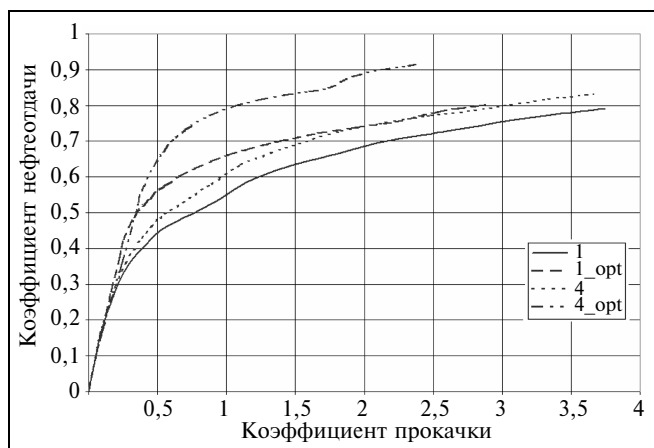


Рис. 3. Графики 1 , 4 и 1_{opt} , 4_{opt} коэффициента нефтеотдачи $\eta(\tau)$ для гипотез 1 и 4 без учета и с учетом оптимального регулирования соответственно. Дискретный аналог равномерного закона, $\tilde{k}_a = 0,1 \text{ мкм}^2$; $\sigma^2 = 0,001$

Характер расположения графиков сохраняется как для одинаковых $\tilde{k}_a$, но разных σ^2 , так и для одинаковых σ^2 , но разных $\tilde{k}_a$. Графики коэффициента нефтеотдачи η от коэффициента прокачки поровых объемов τ для гипотез 3 и 4 расположены между графиками для гипотез 1 и 2. Показатели извлечения нефти из слоистого пласта для гипотезы 4 лучше, чем для гипотезы 3 в случае дискретного аналога равномерного и логарифмически-нормального законов.

На рис. 3 приведены кривые зависимости коэффициента нефтеотдачи от коэффициента прокачки поровых объемов без учета и с учетом оптимального регулирования для дискретного аналога логарифмически-нормального закона в случае принятия гипотез 1 и 4. Анализ графиков показывает значительное превышение коэффициента текущей нефтеотдачи с учетом оптимального регулирования над аналогичным показателем без его учета.

Таблица 5

Дискретный аналог равномерного закона с учетом оптимального регулирования

Номер гипотезы		1	2	3	4
$\tilde{k}_a = 0,1 \text{ мкм}^2$ $\sigma^2 = 0,001$	τ	2,907	2,48	2,659	2,405
	$\Delta\eta$	0,01	0,015	0,011	0,084
	ΔBHF	2,046	2,593	2,449	3,367

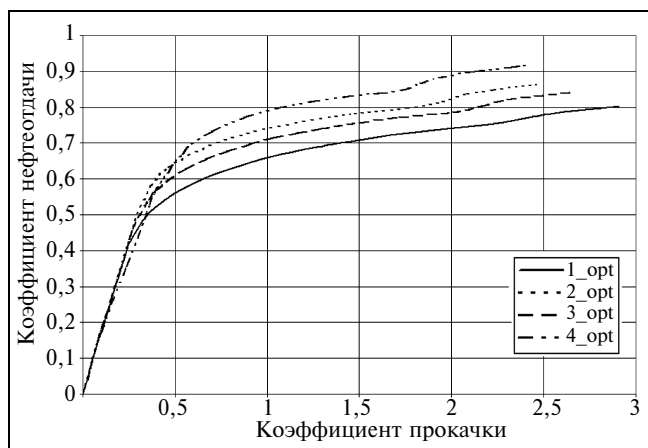


Рис. 4. Графики 1_{opt} — 4_{opt} коэффициента нефтеотдачи $\eta(\tau)$ с учетом оптимального регулирования для гипотез 1—4 соответственно. Дискретный аналог равномерного закона, $\tilde{k}_a = 0,1 \text{ мкм}^2$; $\sigma^2 = 0,001$

В табл. 5 для гипотез 1—4 и дискретного аналога равномерного закона приведены значения τ коэффициента прокачки поровых объемов и соответствующие им значения абсолютных разностей для коэффициентов конечной нефтеотдачи $\Delta\eta$ и водо-нефтяного фактора $\Delta BHF = \Sigma\Delta Q_v / \Sigma\Delta Q_n$ ($\Sigma\Delta Q_v$, $\Sigma\Delta Q_n$ — суммарная добыча воды и нефти) с учетом оптимального регулирования и без его учета, соответственно.

На рис. 4 приведены графики зависимости коэффициента нефтеотдачи η от коэффициента прокачки поровых объемов τ с учетом дискретного аналога равномерного закона и гипотез 1—4 для оптимального регулирования. Анализируя кривые, можно сделать вывод о том, что наилучшие показатели получены для гипотезы 4, наихудшие — для гипотезы 1. Для гипотез 2 и 3 данные показатели расположены между двумя графиками, соответствующими гипотезам 1 и 4. Это объясняется видом относительных фазовых проницаемостей для гипотез 1—4 и динамикой изменения поля скоростей нефти в области двухфазной фильтрации элемента заводнения слоистого пласта.

ЗАКЛЮЧЕНИЕ

Предложена математическая модель процесса двухфазной фильтрации в нефтяных пластах, параметры которых могут изменяться по толщине и простираю. Вероятностные законы — равномерный или логарифмически нормальный. Сформулированы четыре гипотезы о вытеснении нефти водой из элементарного объема слоистого пласта.

Разработан вычислительный алгоритм для определения модифицированных относительных фазовых проницаемостей на основе этих гипотез для произвольного дискретного аналога вероятностного закона распределения абсолютной проницаемости.

Предложена постановка двумерной нестационарной задачи оптимального регулирования извлечения нефти из слоистого пласта. Приведен алгоритм ее решения. Критерий качества решения — максимальный суммарный отбор нефти. Использование оптимального регулирования улучшило показатели извлечения нефти и водонефтяного фактора по сравнению с показателями прогнозирования.

Описанный подход позволяет решать задачи оптимального управления в элементе заводнения с нерегулярной сеткой скважин и заданием граничных условий I-го или II-го рода и их комбинации.

ЛИТЕРАТУРА

1. Северов Я.А. Сравнение гидродинамических показателей двухмерного с трехмерным моделированием пласта // Науч. тр. Оренб. нефтян. акцион. компании «ОНАКО». — 2001. — № 3. — С. 165–169.
2. Булыгин Д.В., Булыгин В.Я. Геология и имитация разработки залежей нефти. — М.: Недра, 1996. — 382 с.
3. Мееров М.В. Исследование и оптимизация многосвязных систем управления. — М.: Наука, 1986. — 236 с.
4. Закиров Э.С. Трехмерные многофазные задачи прогнозирования, анализа и регулирования разработки месторождений нефти и газа. — М.: Грааль, 2000. — 303 с.
5. Слабнов В.Д., Волков Ю.А., Скворцов В.В. Влияние некоторых факторов регулирования на основные показатели нефтеизвлечения из неоднородного пласта // Математическое моделирование. — 2002. — Т. 14, № 1. — С. 3–15.
6. Dolle N., Brouwer D.R., Jansen J.D. Dynamic Optimization of Water Flooding with Multiple Injectors and Producers using Optimal Control Theory // Proc. XIV Intern. Conf. on Computational Methods in Water Resources. — Delft, 2002. — P. 23–28.
7. Slabnov V.D., Volkov Yu.A. To a problem of optimum oil production process control from seams, development with application of horizontal wells // Вторая междунар. конф. по проблемам управления (Москва, 17 — 19 июня 2003 г. / ИПУ РАН) // Избранные труды в двух томах. — М., 2003. — Т. 2. — С. 57–67.
8. Азиз Х, Сеттари Э. Математическое моделирование пластовых систем. — М.: Недра, 1982. — 407 с.

Статья представлена к публикации членом редколлегии Л.П. Боровских.

Султанов Рустем Анварович — канд. физ.-мат. наук, доцент, Казанский государственный финансово-экономический институт, ☎(843) 291-13-71, ✉sultanov@mail.ru,

Слабнов Виктор Дмитриевич — канд. физ.-мат. наук, доцент, Институт экономики, управления и права, г. Казань, ☎(843) 231-92-90, ✉slabnov@ieml.ru,

Фукин Игорь Анатольевич — канд. физ.-мат. наук, доцент, зав. кафедрой, Институт экономики, управления и права, г. Казань, ☎(843) 231-92-90, ✉igor@ieml.ru,

Скворцов Владимир Викторович — д-р техн. наук, профессор, Казанский государственный технологический университет, ✉vskvortsov@mail.ru.

Новая книга

Новиков Д.А. Теория управления образовательными системами. — М.: Народное образование, 2009. — 452 с.

Современной теории управления давно известны законы, принципы и механизмы, позволяющие эффективно управлять любой системой. Автор демонстрирует, что эти общие процедуры возможно и целесообразно применять и в образовательных системах.

Управление образовательными системами рассмотрено с общекибернетических позиций, что позволяет изложить компоненты и структуру теории систематически. Значительное внимание уделено программно-целевому (проектному) подходу к управлению образовательными системами. Детально рассмотрены функции, задачи и методы управления образовательными системами. Их изложение сопровождается анализом практических ситуаций, с которыми сталкиваются руководители образовательных систем и на которых иллюстрируется применение теории.

Для сотрудников органов управления образованием, студентов, аспирантов и преподавателей педагогических и управленческих специальностей, руководителей и сотрудников инновационных образовательных учреждений, работников институтов повышения квалификации.

По вопросам приобретения обращайтесь в ООО «НИИ школьных технологий»

(109202, Москва, шоссе Фрезер, д. 17, 3-й этаж, офис 1.

☎/☎ (499) 270-20-65; ✉ market@narodnoe.org).

МОДЕЛИРОВАНИЕ ГОРМОНАЛЬНО УПРАВЛЯЕМОГО БИПОЛЯРНОГО РОСТА КЛЕТОЧНЫХ СТРУКТУР РАСТИТЕЛЬНОГО ТИПА

В.С. Суховеров, Г.А. Романов

Предложена модель пролиферативного роста клеточной структуры растительного типа, в которой управление делением клеток осуществляется с помощью двух гормонов-активаторов. Разработан программный комплекс для интерактивных расчетов на модели. Приведены результаты моделирования.

Ключевые слова модель, управление, программный комплекс, клетка, гормон, пролиферативный рост.

ВВЕДЕНИЕ

Биполярный рост — неотъемлемый признак жизнедеятельности большинства растений. Его сущность состоит в том, что стебель растения удлиняется вверх, а корень — вниз, углубляясь в почву. Биполярный рост обусловлен, в первую очередь, увеличением числа клеток благодаря размножению (пролиферации) в специальных клеточных структурах — апикальных меристемах, расположенных в самой верхушке стебля и в кончике корня. В природе растения представлены огромным многообразием жизненных форм, сильно различающихся по своим ростовым характеристикам [1]: числу клеток в клеточных цепочках от верхушки побега до кончика корня, соотношению размеров побега и корня, скорости роста.

Установлено, что для деления растительных клеток необходимо наличие двух гормонов — ауксина и цитокинина [2]. Существенно, что эти гормоны синтезируются в основном на полюсах растения: ауксин — в клеточных структурах верхушки стебля, а цитокинин — в клеточных структурах кончика корня. По транспортным каналам растения ауксины перемещаются вниз, тогда как цитокинины — им навстречу [3, 4]. До сих пор до конца не выяснена роль гормонального противотока в организации ростовых процессов. Экспериментально выделить ауксино-цитокениновый противоточный контур и исследовать его функции для роста растения не представляется возможным. Для

решения подобных задач часто эффективным оказывается применение моделирования, т. е. построение разного рода моделей, в том числе и математических, и расчеты на них [5—8]. Целью данной работы было создание модели пролиферативного роста клеточной структуры растительного типа. В результате предложена модель, в которой бигормональный регуляторный контур управляет пролиферативным ростом клеточной структуры растительного типа.

1. ИСХОДНАЯ МОДЕЛЬ РОСТА КЛЕТОЧНОЙ СТРУКТУРЫ

Предпосылки из области физиологии растений и прообраз настоящей модели были опубликованы ранее [9, 10], а затем представлены и первые результаты моделирования [11]. Предложенная в этих работах модель клеточной структуры растительного типа представляет собой одномерную цепочку идентичных клеток (рис. 1), образующих систему транспорта двух гормонов-активаторов

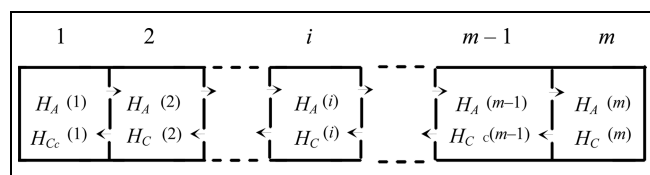


Рис. 1. Исходная модель клеточной структуры

клеточных делений. Фактически каналами транспорта гормонов в модели служат сами клетки. Каждая из m клеток структуры, как указано на рисунке, характеризуется концентрациями гормонов-активаторов: $H_A(i)$, $H_C(i)$, ($i = 1, \dots, m$), где i — номер клетки, а m — число клеток в клеточной структуре. Стрелками указано направление транспорта гормонов. Прототипами гормонов-активаторов модели у живых растений являются ауксины и цитокинины. Гормоны в модели производятся в концевых клетках (1-я и m -я клетки). Поскольку в природе направленность растений вертикальная, условно будем считать, что 1-я клетка соответствует верхней концевой клетке структуры, m -я клетка — нижней концевой клетке структуры. По аналогии с растениями верхняя концевая клетка в модели синтезирует гормон A (от лат. *auxin*), а нижняя концевая клетка синтезирует гормон C (от лат. *cytokinin*). Гормон, синтезируемый верхней концевой клеткой, перемещается по структуре клеток вниз, а гормон, синтезируемый нижней концевой клеткой, — в обратном направлении, по структуре клеток вверх. Каждая клетка структуры способна делиться. Деление происходит при условии, что концентрации каждого из гормонов в ней превышают заданные величины, называемые порогами чувствительности: SL_A ($0 \leq SL_A \leq 1$) и SL_C ($0 \leq SL_C \leq 1$):

$$\begin{aligned} H_A(i) &\geq SL_A, \\ H_C(i) &\geq SL_C. \end{aligned} \quad (1)$$

В процессе транспорта количество доставляемого из клетки в клетку гормона уменьшается в результате так называемых транспортных потерь. В реальности причина потерь состоит в том, что во время транспорта происходят процессы метаболизации гормона, иммобилизации его на рецепторах, разрушения под влиянием внешних факторов и др. [9]. Транспортные потери в модели отражают постоянные коэффициенты K передачи гормонов из клетки в клетку: K_A ($0 \leq K_A \leq 1$) — для «верхнего» и K_C ($0 \leq K_C \leq 1$) — для «нижнего» гормона. Очевидно, что из-за потерь концентрация транспортируемого гормона в клетках будет уменьшаться по мере удаления от концевой клетки, синтезирующей гормон. В концевых клетках доставленные с противоположного конца клеточной структуры гормоны аккумулируются благодаря одностороннему активному переносу и невозможности дальнейшего перемещения [9–11]. Для концевых клеток через какое-то конечное время оба неравенства сигнального условия (1) будут выполнены благодаря синтезу одного гормона и накоплению второго. Произойдет деление клетки на две. Это приведет к увеличению общего числа клеток в ка-

нале транспорта, и, соответственно, увеличению потерь гормонов на пути из синтезирующих клеток в противоположные концевые.

Количество гормона Q_m , дошедшего от синтезирующей до противоположной концевой m -й клетки, можно представить в виде m -го члена убывающей геометрической прогрессии: $Q_m = Q_1 K^{m-1}$, где Q_1 — количество гормона, передаваемого из первой клетки, синтезирующей гормон, во вторую; $K = K_A$ или $K = K_C$. В концевых клетках происходит не только аккумуляция гормона, но и его метаболическая инактивация. Эти потери в концевых клетках модели характеризуются постоянным коэффициентом M ($0 \leq M \leq 1$), который определяется как отношение доли метаболических потерь гормона к концентрации гормона. Если H_m — это количество «противоположного» гормона в концевой клетке, то с учетом распада гормона изменение концентрации гормона в концевой m -й клетке запишется следующим образом: $\Delta H_m = Q_1 K^{m-1} - M H_m$. Из последнего выражения видно, что в процессе роста клеточной структуры, т. е. при возрастании числа клеток m , наступит момент, когда $\Delta H_m \leq 0$. В этом случае концентрация доставляемого гормона в m -й клетке перестает расти и условие (1) становится невыполнимым. Это означает прекращение деления концевых клеток и завершение роста клеточной структуры.

Представленную модель гормональной регуляции деления клеток назовем ГР-моделью [11]. Таким образом, ГР-модель воспроизводит гормонально управляемый рост клеточной структуры путем деления клеток, главным образом концевых. Этот рост определяется параметрами гормональной системы регуляции. Очевидное достоинство модели состоит в том, что при внешней простоте и минимуме эффекторов (гормонов) она воспроизводит основные характеристики процесса биполярного роста, свойственного растениям.

2. МОДЕЛИРОВАНИЕ РОСТА КЛЕТОЧНОЙ СТРУКТУРЫ С УЧЕТОМ ДИНАМИКИ ТРАНСПОРТА ГОРМОНОВ

Рост растений путем деления клеток представляет собой процесс, происходящий и изменяющийся во времени. Поэтому представляется естественным с помощью исходной модели оценить динамику этого процесса. Для этого требуется, хотя бы в первом приближении, в рамках исходной модели математически описать изменение во времени процессов, происходящих в клетках, — синтеза, транспорта, утилизации гормонов и деления клеток надвое. Отметим, что в данной работе не рассматривается динамика процесса деления клеток. Модель только определяет условия, когда деление



клетки возможно (1), а затем отражает количественные изменения в структуре, произошедшие в результате деления клетки (появление новой клетки и изменение концентрации гормонов в поделенной клетке).

Для построения модели был применены методы теории управления [12—14]. Для исходной модели процессы синтеза, транспорта и инактивации гормонов для внутренней и концевой клеток были описаны с помощью системы линейных дифференциальных уравнений, в которых независимой переменной является время (см. приложение). Затем полученные системы уравнений были представлены в виде некоторого числа стандартных графических элементов (динамических звеньев), связанных между собой. Таким образом, для внутренней и концевой клеток модели были созданы схемы моделирования, отражающие динамику изменения концентрации гормонов. Затем эти схемы были дополнены блоками механизма деления. Далее из полученных схем была составлена схема моделирования клеточной структуры в целом, где клетки связаны между собой каналами транспорта гормонов A и C . Схема моделирования клеточной структуры отражает динамику изменения концентрации гормонов и процесс роста клеточной структуры путем деления клеток. В соответствии со схемой моделирования была составлена система уравнений в конечных разностях. На ее основе получены вычислительные формулы модифицированного метода Эйлера для расчетов на модели.

2.1. Динамика изменения концентрации гормонов в клетке

Для описания синтеза гормона в концевых клетках ограничимся единственным условием, как и в исходной модели: в любой момент времени в концевой синтезирующей клетке концентрация гормона является постоянной величиной, которая условно принимается равной единице (эта концентрация гормона заведомо превышает порог чувствительности SL). В аспекте изменения во времени рассмотрим процессы транспорта и роста клеточной структуры. В естественных условиях существует несколько путей транспорта гормонов из синтезирующих клеток. Поскольку в данном случае рассматривается клеточная структура, в которой гормоны перемещаются из клетки в клетку, будем считать, что в модели действует механизм полярного транспорта, аналогичный механизму перемещения ауксина [7, 8, 15]. В этом случае для проводящих гормон внутренних клеток можно определить временной интервал T , в течение которого поступивший внутрь клетки гормон достигнет ее противоположной стенки и начнет выводиться наружу, перемещаясь в следующую клетку. В соответствии с этим динамику транспорта гормона

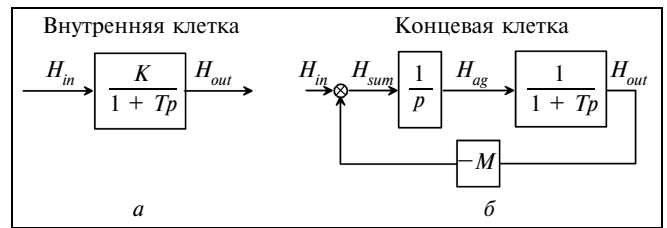


Рис. 2. Схемы моделирования:

a — транспорта гормона через внутреннюю клетку; b — накопления и инактивации гормона в концевой клетке

через внутреннюю клетку модели можно представить линейным дифференциальным уравнением в операторной форме:

$$H_{out} = KH_{in}/(1 + Tp). \quad (2)$$

Уравнению (2) соответствует схема моделирования, представленная на рис. 2, a , где H_{in} и H_{out} — количество доставляемого в клетку и количество выводимого из нее гормона соответственно, T — постоянная времени прохождения гормона через клетку, $K < 1$ — коэффициент транспортных потерь. Инерционное звено отражает временные процессы, происходящие при перемещении гормона через внутреннюю клетку структуры. *In planta* аналогичные процессы включают в себя перенос ауксина через клеточную мембрану внутрь клетки, ионизацию его молекул, их перемещение к нижнему участку клеточной мембраны и вывод гормона из цитоплазмы с помощью белков-переносчиков [7, 8, 15].

По аналогии с внутренними клетками можно представить динамику накопления и распада гормонов в концевых клетках:

$$\begin{aligned} H_{sum} &= H_{in} - MH_{out}, \\ H_{ag} &= H_{sum}/p, \\ H_{out} &= H_{ag}/(1 + Tp). \end{aligned} \quad (3)$$

Этой системе уравнений соответствует схема моделирования, представленная на рис. 2, b . Интегрирующее звено отражает аккумуляцию гормона в клетке H_{ag} , а отрицательная обратная связь с постоянным коэффициентом передачи M — потери гормона из-за его инактивации.

2.2. Механизм запуска клеточного деления

Запуск деления клетки в ГР-модели происходит при выполнении сигнального условия (1). Схематически его можно представить с помощью двух пороговых элементов ПЭ и элемента логического И (рис. 3). На входы пороговых элементов подаются сигналы, пропорциональные концентрациям гормонов A и C в клетке. Если концентрация гормона

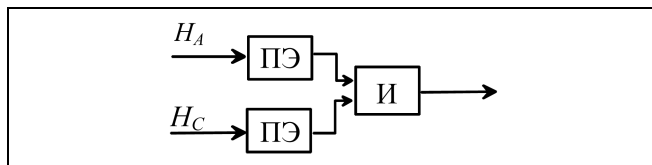
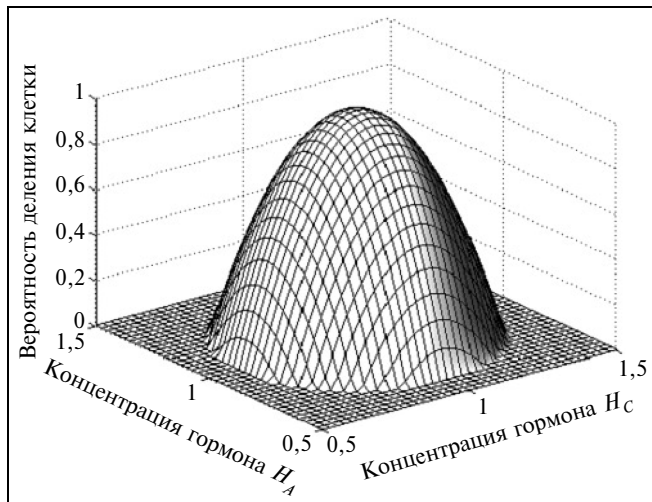


Рис. 3. Схема сигнального условия (1)


Рис. 4. Зависимость вероятности деления клеток от концентрации гормонов H_A и H_C

в клетке удовлетворяет сигнальному условию, то пороговый элемент выдает сигнал логической 1, если — нет, то сигнал логического 0.

Как показали эксперименты с клетками в культуре [16, 17], деления клеток при заданных концентрациях гормонов представляют собой статистически устойчивые события и характеризуются отношением $\nu = \mu/n$, где μ — число поделившихся клеток, а n — общее число клеток с таким же содержанием гормонов. Для вычислительных экспериментов на модели будем считать, что вероятность P_F деления клетки пропорциональна ν , более того, примем $P_F = \mu/n$. Тогда экспериментально полученные зависимости отношения ν от концентрации гормонов (при фиксированной концентрации одного из них) можно трактовать как зависимости вида:

$$\begin{aligned} P_F(H_A) &= P_F(H_A, H_C)|_{H_C = \text{const}}, \\ P_F(H_C) &= P_F(H_A, H_C)|_{H_A = \text{const}}. \end{aligned} \quad (4)$$

Зависимости (4) имеют форму кривых, близких к «перевернутой» параболы, причем для значений $H \leq H_1$ и $H \geq H_2$, где H_1 и H_2 — корни уравнения параболы, $P_F(H_A, H_C) = 0$. Строим для модели идеализированную пространственную характеристику

$P_F(H_A, H_C)$ в виде поверхности в координатах $X = H_A$, $Y = H_C$ и $Z = P_F$. Используем соотношения (4) в качестве сечений этой поверхности плоскостями, параллельными плоскостям XOZ и YOZ , совместив вершины парабол. Для сечений:

$$1 - Z = \frac{(X - H_{A0})^2}{2p}; \quad 1 - Z = \frac{(Y - H_{C0})^2}{2q}, \quad (5)$$

где H_{A0} и H_{C0} — координаты X и Y вершин парабол, p и q — параметры парабол. Тогда зависимость вероятности деления клеток от концентрации гормонов можно представить с помощью эллиптического параболоида (рис. 4), для которого параболические зависимости (5) являются сечениями, параллельными плоскостям XOZ и YOZ и проходящими через вершину параболоида $\{H_{A0}, H_{C0}, 1\}$. Уравнение поверхности параболоида, учитывая, что координата вершины параболоида $P_F = 1$, можно представить в виде:

$$1 - P_F = \frac{(H_A - H_{A0})^2}{2p} + \frac{(H_C - H_{C0})^2}{2q}. \quad (6)$$

Для произвольной i -й клетки модели ($i = 1, \dots, m$) вероятность запуска деления в данный дискретный момент модельного времени $P_F(i)$ определится как точка на поверхности параболоида, соответствующая концентрациям гормонов в клетке: $P_F(i) = P_F(H_A(i), H_C(i))$. Сигнальное условие (1) можно интерпретировать как условие ненулевой вероятности деления клетки, и выражение (6) позволяет вычислить эту вероятность по концентрациям гормонов в клетке. В итоге, схему запуска деления клетки с учетом вероятностного характера клеточных делений можно представить в виде, показанном на рис. 5. Из нее видно, что по значениям входных сигналов H_A и H_C вычисляется вероятность деления клетки, которая затем умножается на значение выходного сигнала элемента логического И. Поэтому в ситуации, когда сигнальное условие (1) не выполняется, вероятность деления клетки будет равна нулю.

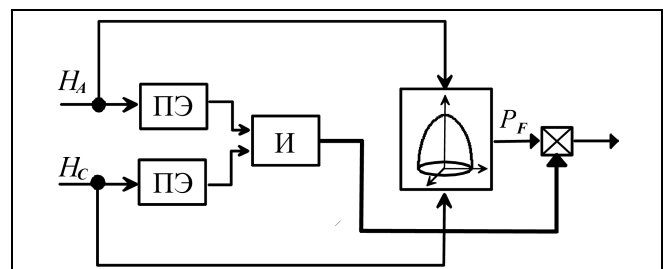


Рис. 5. Схема запуска деления клетки

2.3. Схема моделирования деления клетки с учетом динамики транспорта гормонов

Для получения схемы моделирования гормонально управляемого деления клетки с учетом динамики транспорта необходимо объединить схемы транспорта гормонов (см. рис. 2) со схемой механизма запуска деления клетки (см. рис. 5). Полученные путем такого объединения схемы для внутренней и концевой клеток представлены на рис. 6. Видно, что через каждую клетку структуры в противоположных направлениях транспортируются

два гормона: A и C , причем концентрации гормонов служат входными сигналами для схемы механизма запуска деления клетки.

2.4. Схема моделирования роста клеточной структуры для вычислительных экспериментов

Схема моделирования роста всей клеточной структуры с учетом динамики транспорта гормонов-активаторов клеточных делений представлена на рис. 7. В соответствии с этой схемой была составлена система уравнений, которая использова-

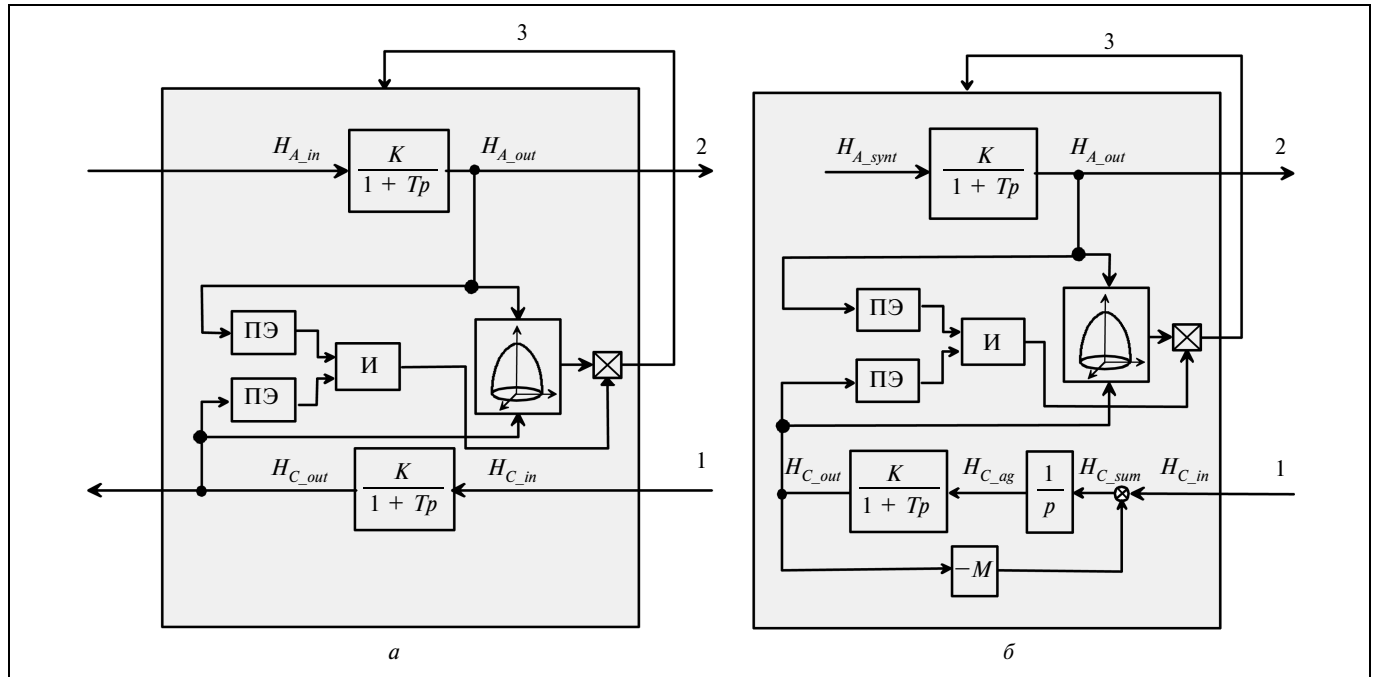


Рис. 6. Структурные схемы гормонально управляемого деления клеток:

a — внутренняя клетка; b — концевая клетка; 1 — канал транспорта гормона C ; 2 — канал транспорта гормона A ; 3 — сигнал запуска деления клетки

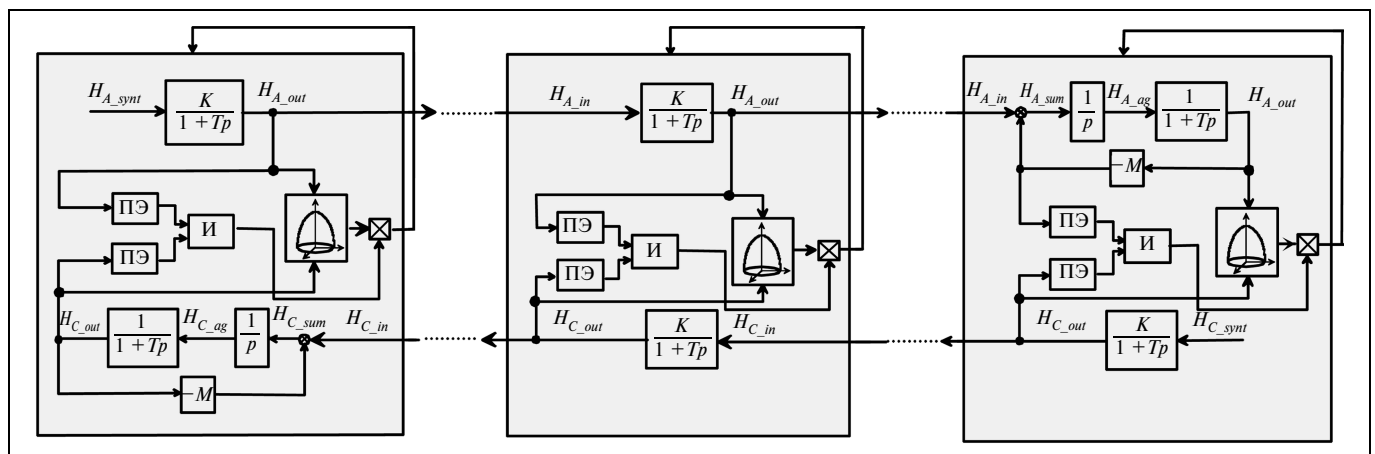


Рис. 7. Схема моделирования регуляции клеточных делений системой синтеза и транспорта двух гормонов

Начальное состояние клеточной структуры

Концентрация гормона	Верхняя клетка	Внутренняя клетка	Внутренняя клетка	Нижняя клетка
H_A	1	0	0	0
H_C	0	0	0	1

лась в программном комплексе для вычислений на модели (см. Приложение).

3. ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ РАБОТЫ С МОДЕЛЬЮ

Для проведения расчетов на ГР-модели был разработан программный комплекс на языке программирования C⁺⁺. В его состав входят:

- программный модуль собственно математической модели роста клеточной структуры с возможностью изменения параметров транспорта гормонов и механизма деления клеток;

- программный модуль экранного интерфейса, обеспечивающий интерактивную работу при вычислениях на модели и функции по сохранению результатов работы.

Экранный интерфейс позволяет в интерактивном режиме изменять параметры модели и графически отображать зависимости показателей роста клеточной структуры от времени (изменение во времени числа клеток, гистограммы концентрации гормонов в клетках и локализации зон клеточных делений, т. е. расположение делящихся клеток в структуре).

Уравнения (2) и (3), отражающие транспорт гормонов в модели, были преобразованы в уравнения в конечных разностях, позволившие получить вычислительные формулы модифицирован-

ного метода Эйлера, которые применялись в расчетах на ГР-модели. При расчетах на модели клеточная структура в начальном состоянии (см. таблицу) содержит четыре клетки с заданной концентрацией гормонов. Как указывалось ранее, концентрации гормонов, синтезируемых в концевых клетках, имеют постоянное значение, равное единице, независимо от времени и оттока в соседние клетки.

4. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Программный комплекс позволил исследовать характер роста структуры растительных клеток на ГР-модели в зависимости от параметров гормонов-активаторов: SL (порог чувствительности клеток к гормону), K (коэффициент передачи гормона в соседнюю клетку), M (коэффициент инактивации гормона в концевой клетке). Графические зависимости (рис. 8—10) иллюстрируют эти исследования. Чтобы отчетливее выявить зависимости характера роста клеточной структуры от перечисленных параметров, при моделировании вероятностный элемент в механизме запуска деления клетки был отключен, и использовалась только схема сигнального условия (см. рис. 3). На рис. 8—10 кривые 1 отражают рост «нижней» части клеточной структуры, кривые 2 — «верхней» ее части, а кривые 3 — общего числа клеток. В исходном варианте модели параметры для гормонов A и C одинаковы: $K_A = K_C = 0,9$; $SL_A = SL_C = 0,8$, $M_A = M_C = 0$. Для ГР-модели с такими параметрами регуляторный контур производит симметричный «организм», для которого вначале характерен биполярный рост с квазипостоянной скоростью, а после достижения «верхней» и «нижней» частями длин, равных примерно 50 клеткам следует резкое уменьшение скорости роста (см. рис. 8, А). Для

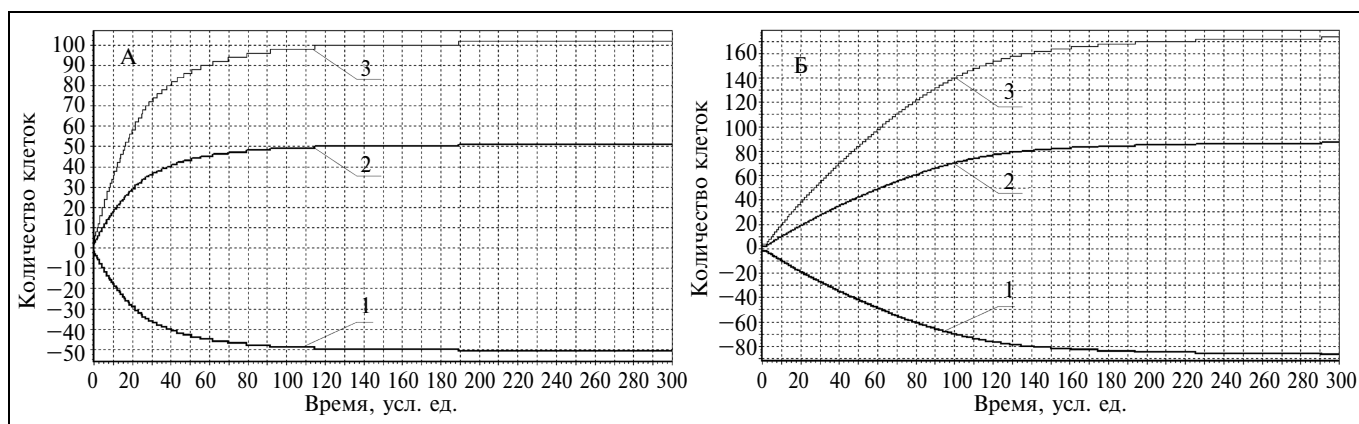


Рис. 8. Влияние коэффициента (K) транспортных потерь гормонов на рост клеточной структуры

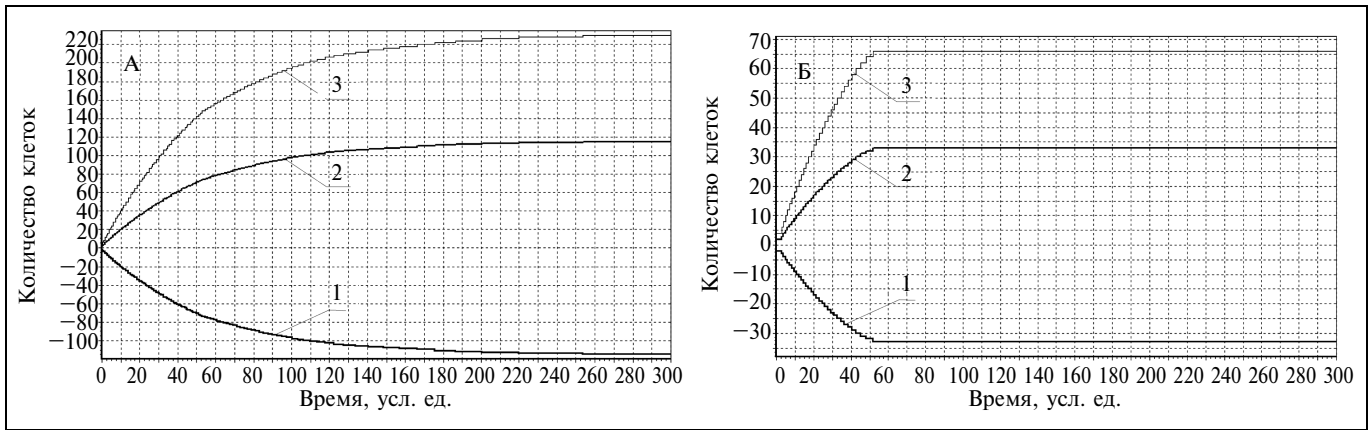


Рис. 9. Влияние порога (SL) чувствительности и коэффициента (M) инактивации гормонов в концевых клетках на рост клеточной структуры

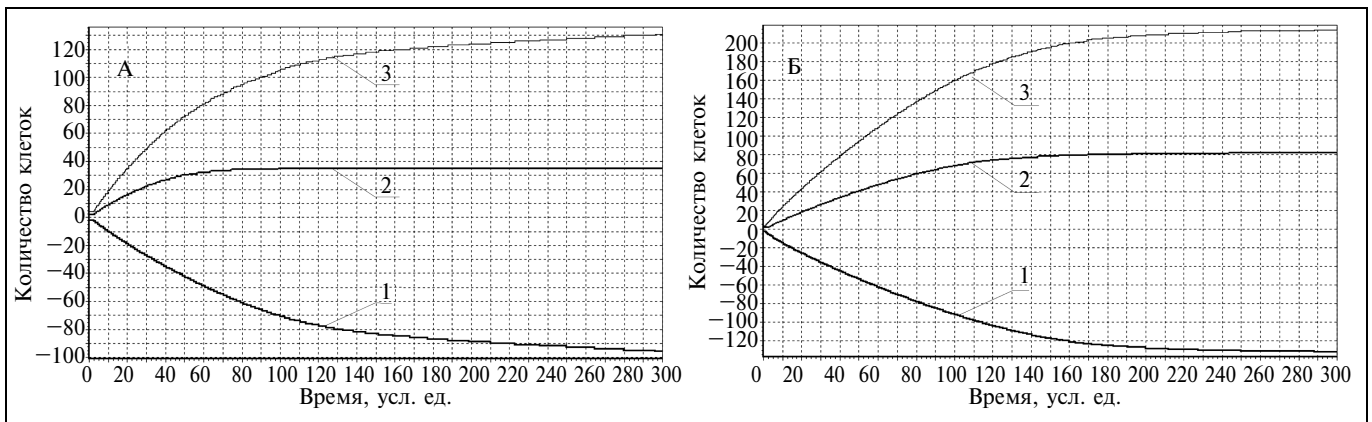


Рис. 10. Асимметричный рост при различающихся параметрах для гормонов A и C

сравнения, если первые 49 клеток «верхней» и «нижней» частей структуры «вырастают» за 100 условных временных единиц, то к моменту времени 300 усл. ед. структура удлинится на 2 клетки с каждого конца. Влияние коэффициента K передачи гормона из клетки в клетку иллюстрирует рис. 8, Б. При увеличении значения K до 0,95 характер роста «растения» качественно не изменится, но его размеры за время 300 усл. ед. вырастут до 90 клеток с каждой стороны, т. е. почти вдвое. Уменьшение порогов чувствительности SL в 10 раз по сравнению с первоначальными значениями (т. е. десятикратное повышение чувствительности к гормонам) приводит к увеличению конечных размеров «организма» (до 118-клеточной цепочки с каждой стороны) (см. рис. 9, А).

В рассмотренных случаях параметр $M = 0$, т. е. инактивация гормонов в концевых клетках не учитывалась. Если же параметр M получает небольшое положительное значение, то «рост» виртуального растения резко замедляется. Уже при $M = 0,001$

(остальные параметры: $SL = 0,8$, $K = 0,95$ и $T = 1$) рост сильно замедляется, а затем резко останавливается (см. рис. 9, Б). Размер такого «растения» по сравнению с его аналогом без инактивации гормонов уменьшается в три с лишним раза.

При неодинаковых значениях параметров для гормонов A и C образуются асимметричные растения с различными ростовыми характеристиками «верхней» и «нижней» частей. Например, большее значение K для «верхнего» гормона A (0,95) по сравнению с «нижним» гормоном C (0,9) приводит к быстрой и окончательной остановке роста верхней части («стебля») всего по достижении длины в 35 клеток, тогда как нижняя часть («корень») продолжает расти с постепенным замедлением и достигает к моменту времени 300 усл. ед. длины в 95 клеток (рис. 10, А). При таких параметрах «гормональной системы» размер «корня» превышает размер «побега» почти в 3 раза. Если сохранить равенство коэффициентов K , но снизить порог чувствительности для гормона A в 10 раз (при ос-

тальных исходных параметрах), то это снова приведет к преобладанию роста «нижней» части структуры по отношению к «верхней», хотя и не в такой степени: соотношение составит 135 : 82, т. е. в этом случае нижняя часть более чем в 1,5 раза длиннее верхней (см. рис. 10, Б).

ЗАКЛЮЧЕНИЕ

Рассмотренная ГР-модель наглядно демонстрирует возможности роста клеточной структуры путем внутренней регуляции, обусловленной действием двух гормонов-активаторов клеточных делений. В модели управление делением клеток осуществляется сигнальными молекулами гормонов всего двух типов. При этом сигналами, определяющими запуск деления клетки, служат концентрации гормонов-активаторов. В свою очередь, концентрации гормонов-активаторов в клетке зависят от параметров синтеза и транспорта гормонов в целостной структуре клеток. Принципиально важная характеристика полученной модели роста клеточной структуры состоит в локализации делящихся клеток в концевых областях. Такая локализация в модели не задается *a priori*, она является следствием особенностей гормональной системы регуляции размножения клеток. Это имеет прямую аналогию с наземными растениями, естественными прототипами модели, у которых основные зоны деления клеток расположены в верхушке стебля и кончике корня [2–4, 9, 10], т. е. в зонах синтеза ауксинов и цитокининов.

Модель показывает, что небольшие различия параметров для гормонов *A* и *C* могут приводить к существенной разнице в динамике роста и соотношении верхней и нижней частей клеточной структуры. Аналогии этому во множестве существуют в растительном мире, где растения меняют соотношение надземной и подземной частей в зависимости от условий произрастания. Модель адекватно воспроизводит все основные особенности роста целостного растения. Это позволяет рассматривать ГР-модель в качестве первой, хотя и предельно упрощенной, модели вегетативного роста многоклеточного наземного растения, управляемого градиентами эндогенных фитогормонов. Уже сейчас на качественном уровне компьютерная модель способна делать оценки влияния того или иного параметра транспорта и (или) восприятия гормонов (что эквивалентно генной мутации) на фенотип растения. Такие прогнозы важны для селекционеров и агротехников, так как предсказывают вероятные фенотипические эффекты при изменении конкретных характеристик систем гормональной регуляции растений.

В дальнейшем представляется целесообразным оценить возможности применения программного комплекса для прогнозирования темпов роста отдельных видов полезных растений на основе реальных данных, в частности, хвойных пород деревьев [18].

ПРИЛОЖЕНИЕ

Представленная в данной работе модель рассматривается как линейная стационарная динамическая система, уравнение состояния которой имеет вид:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad (\text{П1})$$

$$\mathbf{x}(t) \in \mathbb{R}^n; \quad \mathbf{u}(t) \in \mathbb{R}^p; \quad \dot{\mathbf{x}}(t) = \frac{d\mathbf{x}(t)}{dt};$$

где $\mathbf{x}(t)$ — вектор состояния; $\mathbf{u}(t)$ — вектор управления; $\mathbf{A} = \|a_{ij}\|$ ($i = 1, \dots, n; j = 1, \dots, n$) — матрица системы; $\mathbf{B} = \|b_{ij}\|$ ($i = 1, \dots, p; j = 1, \dots, n$) — матрица управления. Внешнее управление $\mathbf{u}(t)$ на модели не рассматривалось: $\mathbf{u}(t) = 0$, потому что управлять пролиферативным ростом растения можно либо путем изменения наследственности, либо путем воздействия гормональными препаратами. И то, и другое — предмет отдельных исследований. Рассматриваемая модель описывает лишь внутренние процессы, происходящие в клетках структуры и обуславливающие деление клеток. В вычислениях на модели исследовалось влияние параметров на пролиферативный рост клеточной структуры. В результате уравнение состояния системы записывается в виде: $\dot{\mathbf{x}}(t) = \mathbf{A}(\lambda)\mathbf{x}(t)$, где вектор параметров $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\}$, $\lambda_1 = K_A$, $\lambda_2 = K_C$, $\lambda_3 = SL_A$, $\lambda_4 = SL_C$, $\lambda_5 = M_A$, $\lambda_6 = M_C$.

В продолжение развития исходной модели (см. рис. 1) клеточную структуру можно описать с помощью компартментальной модели [14] (рис. П1), в которой каждая из m клеток состоит из компартментов. Каждая из внутренних клеток содержит два компартмента, а каждая концевая — три (см. рис. П1). Компартменты характеризуются концентрациями гормонов-активаторов: $H_A(i)$, $i = 1, \dots, m + 1$; и $H_C(i)$, $i = m + 2, \dots, 2m + 2$; где i — номер компартмента, а m — число клеток в клеточной структуре. При этом вектор состояния системы (П1) можно рассматривать как сумму двух векторов: $\mathbf{x}(t) = \mathbf{x}^A(t) + \mathbf{x}^C(t)$, где $\mathbf{x}^A(t)$ и $\mathbf{x}^C(t)$ — векторы концентрации гормонов *A* и *C* в компартментах $\mathbf{x}^A(t) = (x_1(t), \dots, x_{m+1}(t), 0, \dots, 0)$ и $\mathbf{x}^C(t) = (0, \dots, 0, x_{m+2}(t), \dots, x_{2m+2}(t))$.

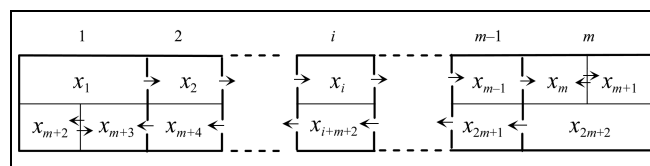


Рис. П1. Компартментальная модель клеточной структуры



Размерность вектора состояния $n = 2(m + 1)$, где n — число компартментов модели, m — число клеток структуры.

Покоординатно модель описывается с помощью системы линейных дифференциальных уравнений с постоянными коэффициентами. В модели выделены два типа компартментов: компартменты внутренних клеток и компартменты концевых клеток. Через компартменты внутренних клеток происходит транспорт гормонов. Для гормона A :

$$\dot{x}_i = a_{ii-1}x_{i-1} + a_{ii}x_i, \text{ где } i = 2, \dots, m-1. \quad (\text{П2})$$

Для гормона C :

$$\dot{x}_i = a_{ii}x_i + a_{ii+1}x_{i+1}, \text{ где } i = m+4, \dots, n-1. \quad (\text{П3})$$

В компартментах концевых клеток происходит синтез одного гормона, а также накопление и распад другого. Будем считать, что гормоны синтезируются непрерывно и скорость синтеза такова, что транспорт гормона из синтезирующего компартмента клетки не вызывает изменения концентрации гормона в синтезирующем компартменте. Это означает, что концентрация синтезируемых гормонов в синтезирующих компартментах клеток постоянна и условно равна единице: $x_1(t) = x_1(0) = \text{const} = 1$ (для гормона A) и $x_n(t) = x_n(0) = \text{const} = 1$ (для гормона C). В нижней концевой клетке происходит накопление и инактивация гормона A :

$$\begin{aligned} \dot{x}_m &= a_{mm-1}x_{m-1} + a_{mm}x_m + a_{mm+1}x_{m+1}, \\ \dot{x}_{m+1} &= a_{m+1m}x_m. \end{aligned} \quad (\text{П4})$$

В верхней концевой клетке происходит накопление и инактивация гормона C :

$$\begin{aligned} \dot{x}_{m+2} &= a_{m+2\ m+3}x_{m+3}, \\ \dot{x}_{m+3} &= a_{m+3\ m+2}x_{m+2} + a_{m+3\ m+3}x_{m+3} + \\ &\quad + a_{m+3\ m+4}x_{m+4}. \end{aligned} \quad (\text{П5})$$

При вычислениях на модели в начальном состоянии структура содержит четыре клетки — верхнюю концевую, две внутренних и нижнюю концевую. Для этого случая на основании выражений (П2)—(П5) система уравнений динамической системы в форме Коши имеет следующий вид:

$$\begin{aligned} \dot{x}_1 &= 0, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2, \\ \dot{x}_3 &= a_{32}x_2 + a_{33}x_3, \\ \dot{x}_4 &= a_{43}x_3 + a_{44}x_4 + a_{45}x_5, \\ \dot{x}_5 &= a_{54}x_4, \\ \dot{x}_6 &= a_{67}x_7, \\ \dot{x}_7 &= a_{76}x_6 + a_{77}x_7 + a_{78}x_8, \end{aligned}$$

$$\dot{x}_8 = a_{88}x_8 + a_{89}x_9,$$

$$\dot{x}_9 = a_{99}x_9 + a_{910}x_{10},$$

$$\dot{x}_{10} = 0.$$

Начальные условия: $x_1 = 1$; $x_2 = x_3 = x_4 = x_5 = x_6 = x_7 = x_8 = x_9 = 0$; $x_{10} = 1$.

ЛИТЕРАТУРА

1. Флиндт Р. Биология в цифрах. — М., Мир, 1992. — 303 с.
2. Stals H, Inze D. When Plant Cells Decide to Divide // Trends Plant Sci. — 2001. — Vol. 6, N 8. — P. 359–364.
3. Полевой В.В. Фитогормоны. — Л.: Изд-во ЛГУ, 1982. — 249с.
4. Полевой В.В., Саламатова Т.С. Физиология роста и развития растений. — Л.: Изд-во ЛГУ, 1991. — 239с.
5. Prusinkiewicz P. Modelling Plant Growth and Development // Curr. Opin Plant Biol. — 2004. — Vol. 7, N 1. — P. 79–83.
6. Lewis J. From Signals to Patterns: Space, Time, and Mathematics in Development Biology // Science — 2008, Vol. 322. — N.5900. — P. 399–403.
7. Berleth T, Scarpella E, Prusinkiewicz P. Towards the System Biology of Auxin-Transport-Mediated Patterning // Trends Plant Sci. — 2007. — Vol. 12, N 4. — P. 151–159.
8. Kramer E.M. Computer Models of Auxin Transport: a Review and Commentary // J. Exp Bot. — 2008. — Vol. 59, N 1. — P. 45–53.
9. Романов Г.А. Модель гормонально-организуемого пролиферативного роста: аналогии с ростом растений // Онтогенез. — 1992. — Т. 23, № 3. — С. 228–236.
10. Romanov G.A. A Model for Bipolar Plant-type Growth: Role of Auxin-cytokinin Countercurrent / in «Progress in Plant Growth Regulation». — Dordrecht, Boston, London: Kluwer, 1992. — P. 459–463.
11. Романов Г.А., Суховеров В.С. Исследование кинетики роста и динамики гормональных градиентов на модельных многоклеточных структурах растительного типа — компьютерных растениях // Доклады Академии наук. — 1997. — Т. 352, № 6. — С. 845–848.
12. Новосельцев В.Н. Теория управления и биосистемы. — М.: Наука, 1978
13. Теория автоматического регулирования / Л.С. Гольдфарб и др. — М.: Высшая школа, 1976.
14. Теория автоматического регулирования: в 3 кн. М.А. Айзерман и др. — М.: Машиностроение, 1967.
15. Heisler M.G., Jonsson H. Modeling Auxin Transport and Plant Development // J. Plant Growth Regul. — 2006. — Vol. 25, N 4. — P. 302–312.
16. Бутенко Р.Г. Культура изолированных тканей и физиология морфогенеза растений. — М.: Наука, 1964. — 272с.
17. Гамбург К.З. Фитогормоны и клетки. — М.: Наука, 1970. — 104 с.
18. Кузьмичев В.В. Рост деревьев как функция времени // Закономерности роста древостоев. — Новосибирск: Наука, 1977.

Статья представлена к публикации членом редколлегии В.Н. Новосельцевым.

Суховеров Виктор Степанович — канд. техн. наук, ст. науч. сотрудник, Институт проблем управления РАН, г. Москва, ☎(495) 334-76-90, ✉suhoverv@ipu.rssi.ru,

Романов Георгий Александрович — д-р биол. наук, зав. лабораторией, Институт физиологии растений РАН, г. Москва, ☎(495) 977-94-09, ✉gar@ippras.ru.

ПОИСК ПОДВИЖНОГО ОБЪЕКТА ПО ИНФОРМАЦИОННОМУ ПРИЗНАКУ «СЛЕД».

Ч. 2. Оптимизация поисковых траекторий¹

Т.Г. Абрамянц, Ю.А. Беланов, Е.П. Маслов, В.П. Яхно

В задаче об оптимизации поисковой траектории преследователя, осуществляющего обнаружение подвижного объекта и сближение с ним по информационному признаку «след», в явном виде найдены оптимальные траектории для различных начальных условий.

Ключевые слова: поиск подвижного объекта, поисковая ситуация, «след» объекта, оптимальная траектория поиска.

ВВЕДЕНИЕ

Постановка задачи об оптимизации программной поисковой траектории преследователя, осуществляющего обнаружение подвижного объекта (цели) и сближение с ним по информационному признаку «след», изложена в первой части статьи [1].

Задача решается при следующих предположениях:

- в начальный момент времени цель выбирает направление прямолинейной траектории, по которой движется с постоянной скоростью v_E ; за целью тянется след постоянной длины R ;

- обнаружение следа поисковой системой (преследователем) происходит в момент t_n пересечения следа траекторией преследователя; в случае обнаружения следа преследователь движется далее вдоль следа до момента встречи с целью.

Критерием служит максимальная длительность интервала времени до момента точечной встречи объектов.

В первой части статьи [1] была установлена общая структура оптимального управления преследователя — показано, что по траектории преследователь движется с постоянной скоростью v_p , а оптимальная траектория поиска включает в себя

прямолинейные участки и может включать участок логарифмической спирали. Конкретный вид поисковой траектории зависит от соотношений между входными параметрами задачи, к которым относятся: начальное расстояние D_0 ; длина следа R ; скорости объектов v_E и v_p ; расположение начальной ψ_n и конечной ψ_k границ диапазона возможных направлений движения цели $\Delta\psi = [\psi_n, \psi_k]$.

В настоящей, второй, части статьи определяется конкретный вид оптимальных поисковых траекторий для произвольных сочетаний указанных величин.

Вводится неподвижная прямоугольная система координат XOY , начало которой совмещено с начальным положением преследователя P_0 , а ось OY проходит через начальное положение цели E^0 (рис. 1); здесь и далее верхний индекс при обозначении позиций объектов указывает момент времени. Углы отсчитываются от оси OY ; положительным считается отсчет по часовой стрелке.

В момент $t = 0$ цель выбирает направление движения ψ (угол между траекторией и осью OY) из диапазона углов $\Delta\psi = [\psi_n, \psi_k]$. Далее символом (ψ) обозначается луч, исходящий из точки E^0 под углом ψ к оси OY ; символом $E(\psi)$ — цель, движущаяся вдоль луча (ψ) ; символом $E^t(\psi)$ — положение цели, движущейся вдоль луча (ψ) , в момент t ; символом $P^t = E^t(\psi)$ — событие, состоящее в совпадении в момент t позиций преследователя и цели, движущейся вдоль луча (ψ) (точечная встреча).

¹ Работа выполнена при финансовой поддержке Программы № 29 Президиума РАН «Математическая теория управления».

Преследователь управляет направлением своего движения.

Решение задачи находится для случая, когда цель движется в полуплоскости $x \geq 0$; $0 \leq \psi \leq \pi$.

При движении по программной поисковой траектории для любого $\psi \in \Delta\Psi = [\psi_n, \psi_k]$ должны выполняться следующие условия:

$$\langle e_z, e_v \rangle_{t_n} = 1, \quad (1)$$

$$\|z(t_n)\| \leq R, \quad (2)$$

где $z(t)$ — вектор, направленный от P к E в момент времени t ; e_z — единичный вектор, коллинеарный вектору $z(t_n)$, соединяющему позиции преследователя и цели в момент t_n пересечения луча траекторией преследователя; e_v — единичный вектор, коллинеарный вектору скорости цели в этот момент; $\langle \cdot, \cdot \rangle$ — скалярное произведение.

1. РЕШЕНИЕ ЗАДАЧИ

В первой части статьи [1] было установлено, что первым участком оптимальной поисковой траектории всегда является прямолинейный отрезок. В работе [2] было показано, что при определенных соотношениях между входными параметрами (D_0 ; R ; v_E , v_P ; ψ_n ; ψ_k) существует прямолинейная оптимальная поисковая траектория, обеспечивающая минимально возможное время движения преследователя до момента реализации терминального условия $P^T = E^T(\psi_n)$. Соответствующее программное управление преследователя имеет вид:

$$\varphi_n = \arcsin(\beta \sin \psi_n); \quad \beta = v_E/v_P, \quad (3)$$

Оптимальная прямолинейная поисковая траектория существует, если справедлива система неравенств:

$$\psi_k \leq \pi - \psi_n, \quad (4)$$

$$\max_{\psi_n \leq \psi \leq \psi_k} D(t_n, \varphi_n, \psi) = D_0 \cos \psi^*(\varphi_n) \leq R, \quad (5)$$

где (φ, ψ) — направления движения преследователя и цели, соответственно; символом $D(t_n, \varphi, \psi)$ обозначено расстояние между преследователем и целью в момент пересечения траекторией преследователя луча (ψ) ;

$$\psi^*(\varphi_n) = \varphi_n + \gamma \quad (6)$$

— то направление движения цели, на котором достигается максимальное расстояние между объектами в момент пересечения луча прямолинейной траекторией преследователя, отвечающей формуле (3); $\gamma = \arccos \beta$.

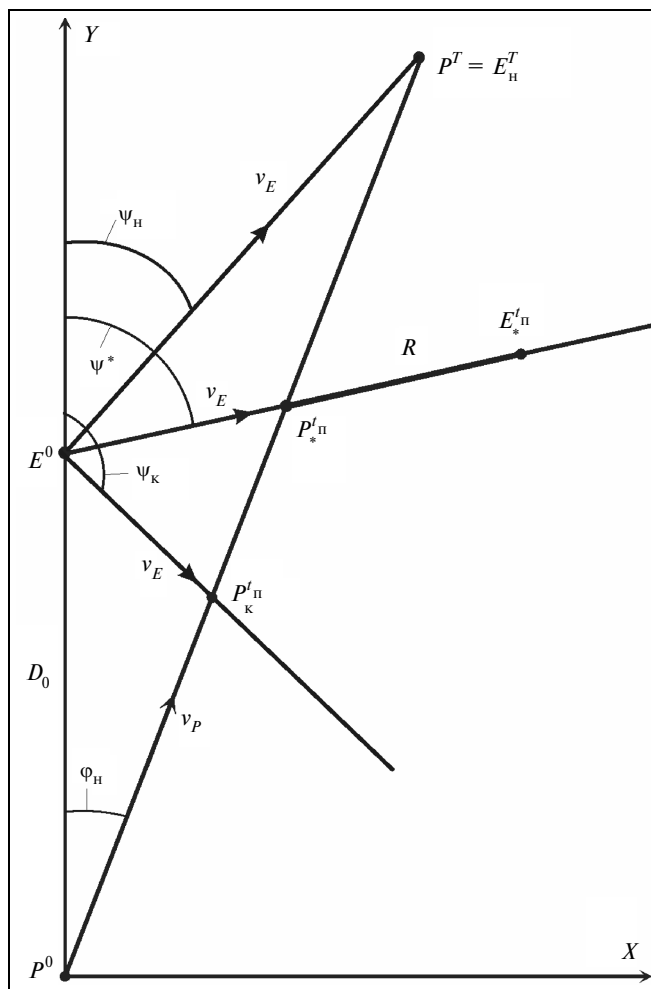


Рис. 1. Оптимальная прямолинейная траектория

В терминах параметров задачи неравенство (5) переписывается в виде:

$$\arccos(R/D_0) - \arccos \beta \leq \arcsin(\beta \sin \psi_n).$$

Решение оптимизационной задачи для указанного случая иллюстрируется рис. 1, построенном для предельного случая $\psi_k = \pi - \psi_n$. Здесь P'^n и E'^n — положения преследователя и цели в момент R -встречи с точкой $E(\psi^*(\varphi_n))$; P'^n — положение преследователя в момент точечной встречи $P'^n = E'^n(\varphi_k)$. Все остальные обозначения на рис. 1 совпадают с приведенными в тексте. Длительность интервала времени движения на прямолинейной поисковой оптимальной траектории

$$T = \frac{D_0}{v_P(\sqrt{1 - \beta^2 \sin^2 \psi_n} - \beta \cos \psi_n)}.$$

Пусть

$$\max_{\psi_H \leq \psi \leq \psi_K} D(t_n, \varphi_n, \psi) = D_0 \cos \psi^*(\varphi_n) > R.$$

В этом случае оптимальной прямолинейной поисковой траектории не существует. В фазовом пространстве первый прямолинейный участок поисковой траектории заканчивается выходом текущего положения цели на ограничения (1) и (2). Решение задачи об оптимизации сводится к построению траектории, которая удовлетворяла бы ограничениям (1) и (2), и длина которой наименее отличалась бы от длины прямолинейной поисковой траектории.

Структура оптимальных траекторий зависит от соотношений между углами (направлениями движения) $\psi_n, \psi^*(\varphi_n), \psi_K$.

Возможны три случая:

- $\psi_K \leq \pi - \psi_n, \psi^*(\varphi_n) \leq \psi_K$; сектор направлений движения цели $\psi \in \Delta\Psi = [\psi_n, \psi_K]$, удовлетворяющий этой системе, далее называется сектором типа 1;
- $\psi_K \leq \pi - \psi_n, \psi_K \leq \psi^*(\varphi_n)$; соответствующий сектор называется сектором типа 2;
- $\psi_K > \pi - \psi_n$; соответствующий сектор называется сектором типа 3.

1.1. Поиск в секторе типа 1

Справедлива система неравенств

$$\psi^*(\varphi_n) \leq \psi_K \leq \pi - \psi_n$$

При фиксированных управлениях (φ, ψ) момент пересечения луча (ψ), удовлетворяющий условию (1), определяется по формуле [1]

$$t_n = \frac{D_0 \sin \psi}{v_p \sin(\psi - \varphi)}, \quad (7)$$

а расстояние, разделяющее объекты в этот момент — по формуле

$$D(t_n, \varphi, \psi) = D_0 \frac{|\beta \sin \psi - \sin \varphi|}{\sin(\psi - \varphi)}. \quad (8)$$

Отмечалось [1], что решаемая в настоящей статье задача представляет собой задачу геодезического преследования на плоскости с выпуклым подвижным препятствием, где в качестве препятствия выступает логарифмическая спираль — огибающая семейства расширяющихся окружностей — геометрических мест концов следов, а терминальная точка принадлежит огибающей геометрических мест позиций цели. Из результатов работ [3—5] следует, что в том случае, когда экстремаль, лежащая внутри области возможных позиций цели, включает в себя участок спирали — границы

подвижного препятствия, прямолинейные отрезки экстремали являются касательными к этой границе.

В точке касания одновременно выполняются равенства

$$D_0 \frac{|\beta \sin \psi - \sin \varphi|}{\sin(\psi - \varphi)} = R, \quad \psi - \varphi = \gamma,$$

где φ — направление движения преследователя, $\psi \triangleq \psi(\varphi)$ — направление луча, на котором происходит касание спирали траекторией преследователя. Отсюда находим:

$$D_0 \cos(\varphi + \gamma) = D_0 \cos \psi(\varphi) = R.$$

Так как с ростом φ величина $D_0 \cos(\varphi + \gamma)$ уменьшается, а, по предположению,

$$\max_{\psi_H \leq \psi \leq \psi_K} D(t_n, \varphi_n, \psi) = D_0 \cos \psi^*(\varphi_n) > R,$$

то найдется управление $\varphi_1^R > \varphi_n$, для которого справедливо равенство

$$D_0 \cos(\varphi_1^R + \gamma) = D_0 \cos \psi^*(\varphi_1^R) = R. \quad (9)$$

В силу равенства (9) это управление имеет вид

$$\varphi_1^R = \arccos(R/D_0) - \arccos(v_E/v_P) \quad (10)$$

и зависит только от безразмерных величин; нижний индекс указывает номер участка программной поисковой траектории.

Из равенств (6) и (9) следует важный вывод: при фиксированном управлении φ в точке касания спирали траекторией преследователя достигается максимум расстояния между объектами.

Из формулы (8) следует, что с ростом φ увеличивается угол $\psi_m(\varphi)$, соответствующий тому направлению движения цели, при котором реализуется точечная встреча ее с преследователем, движущимся по направлению φ . В частности, точечной встрече преследователя с целью, движущейся вдоль граничного луча (ψ_K) диапазона $\Delta\Psi = [\psi_n, \psi_K]$, соответствует управление $\varphi_K = \arcsin(\beta \sin \psi_K)$.

Возможны два случая:

- $\varphi_1^R \leq \varphi_K$ — траектория преследователя, соответствующая управлению φ_1^R , пересекает след, расположенный на луче (φ_K);
- $\varphi_1^R > \varphi_K$ — траектория преследователя, соответствующая управлению φ_1^R , пересекает луч (ψ_K) в области, где след существовать не может; в момент пересечения луча (ψ_K) указанной траекторией $\langle e_z, e_v \rangle_{t_n} = -1$.



Пусть $\varphi_1^R \leq \varphi_k$.

В этом случае первый прямолинейный участок заканчивается R -встречей с некоторой точкой $E^{t_n}(\varphi)$, т. е. реализацией равенства $\|z(t_n, \psi)\| - R = 0$. Направление движения преследователя на первом прямолинейном отрезке определяется формулой (10). Далее текущее положение цели перемещается в фазовом пространстве по окружности (что соответствует движению преследователя по спирали в неподвижной системе координат) до момента схода на заключительный прямолинейный участок, касающийся спирали в точке схода и проходящий через терминальную точку $P^T = E^T(\psi_n)$. Следовательно, в рассматриваемом случае поисковая траектория содержит три участка, отличающихся характером движения — первый прямолинейный участок, спиральный участок, заключительный прямолинейный участок. Первый и заключительный прямолинейные участки являются касательными к логарифмической спирали.

Длительность интервала времени движения на первом прямолинейном участке поисковой траектории, вычисленная по формуле (7) с учетом выражений (9) и (10),

$$T = \sqrt{\frac{D_0^2 - R^2}{v_P^2 - v_E^2}}. \quad (11)$$

Длительности интервалов времени движения на двух других интервалах определяются формулами (17) и (19) первой части [1] статьи.

Пусть теперь $\varphi_1^R > \varphi_k$.

В этом случае направление движения преследователя на первом прямолинейном участке выбирается из условия точечной встречи $P^t = E^t(\varphi_k)$ преследователя с целью, движущейся вдоль луча (φ_k):

$$\varphi_1 = \varphi_k = \arcsin(\beta \sin \varphi_k).$$

Обозначим символом θ момент точечной встречи $P^\theta = E^\theta(\varphi_k)$. Длительность первого прямолинейного участка

$$\theta = \frac{D_0}{v_P(\cos \varphi_k - \beta \cos \varphi_k)}. \quad (12)$$

В момент θ преследователь изменяет направление движения и далее движется по направлению φ_2 до момента t_n пересечения некоторого луча (φ): $\varphi < \varphi_k$. Соответствующий фрагмент поисковой траектории изображен на рис. 2. Здесь: $P^\theta = E^\theta(\varphi_k)$ — позиция объектов в момент излома поисковой тра-

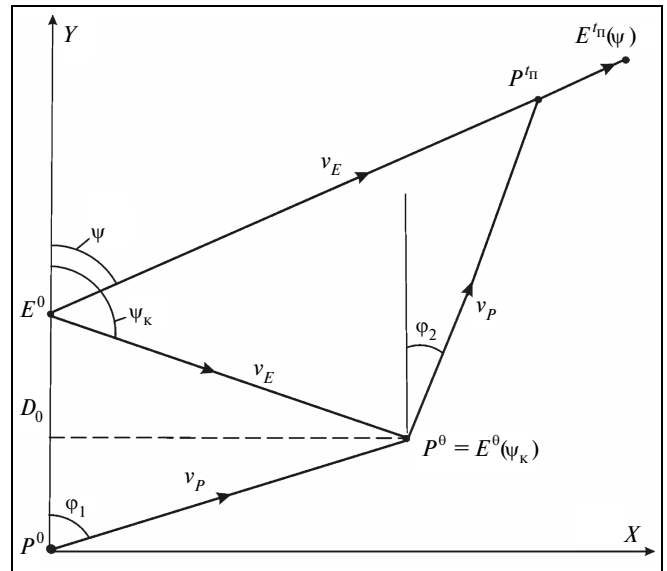


Рис. 2. Фрагмент кусочно-прямолинейной поисковой траектории

ектории; P^{t_n} , $E^{t_n}(\psi)$ — позиции объектов в момент пересечения луча (ψ). Остальные обозначения на рис. 2 совпадают с обозначениями, приведенными в тексте.

Так как лучи (ψ_k) и (ψ) исходят из одной точки E^0 , то условие пересечения произвольного луча (ψ) в момент $t_n > \theta$ записывается в виде:

$$\operatorname{tg} \psi = \left(\frac{\Delta x}{\Delta y} \right)_{t_n} = \frac{v_E t_n \sin \psi - v_E \theta \sin \psi_k - v_P \omega \sin \varphi_2}{v_E t_n \cos \psi - v_E \theta \cos \psi_k - v_P \omega \cos \varphi_2},$$

где ω — длительность интервала времени от момента θ излома поисковой траектории до момента t_n . Отсюда находится длительность второго прямолинейного участка:

$$\omega = \beta \theta \frac{\sin(\psi_k - \psi)}{\sin(\psi - \varphi_2)}. \quad (13)$$

Расстояние между точками P^{t_n} и $E^{t_n}(\psi)$ в момент $t_n = \theta + \omega$

$$\begin{aligned} \Delta(t_n, \psi, \varphi_2) &= \sqrt{(\Delta x)^2 + (\Delta y)^2} = \\ &= v_E \theta \left| 1 + \frac{\beta \sin(\psi_k - \psi) - \sin(\psi_k - \varphi_2)}{\sin(\psi - \varphi_2)} \right|. \end{aligned} \quad (14)$$

При этом

$$\max_{\psi} D(t_n, \psi, \varphi_2) = v_E \theta [1 + \cos(\varphi_k - \psi^*)],$$

где единственный максимум достигается в точке

$$\psi^* = \psi^*(\varphi_2) = \varphi_2 + \gamma, \quad \cos \gamma = \beta. \quad (15)$$

Управление φ_2 , реализующее R -встречу с лучом $\psi^*(\varphi_2)$, находится как корень уравнения

$$v_E \theta [1 + \cos(\psi_K - \psi^*)] = R.$$

Третьим участком оптимальной поисковой траектории является участок логарифмической спирали, начинающийся на луче (ψ^*), определяемом формулой (15). Траектория на втором прямолинейном участке является касательной к спирали. Таким образом, в рассматриваемом случае оптимальная поисковая траектория содержит четыре участка — двухзвенную ломаную (первый прямолинейный участок — до момента точечной встречи $P^I = E^I(\psi_K)$, второй прямолинейный участок — до момента пересечения луча ($\psi^*(\varphi_2^R)$), спиральный участок и заключительный прямолинейный участок.

1.2. Поиск в секторе типа 2

Справедливы неравенства

$$\psi_K \leq \pi - \psi_H, \quad \psi_K \leq \psi^*(\varphi_H).$$

Так как, по определению, в интервале $[\psi; \psi > \psi_K]$ траектории цели отсутствуют, то при $\psi_K < \psi^*(\varphi_H)$ максимальное расстояние между объектами на прямолинейной поисковой траектории, соответствующей равенству $\varphi_H = \arcsin(\beta \sin \psi_H)$, достигается в момент пересечения луча ψ_K . Функция $D(t_H, \psi, \varphi_H)$ является убывающей функцией аргумента ψ на интервале $[\psi_H, \psi_K]$. Поэтому

$$\begin{aligned} \max_{\psi_H \leq \psi \leq \psi_K} D(t_H, \psi, \varphi_H) &= D(t_H, \psi_K, \varphi_H) = \\ &= D_0 \frac{|\beta \sin \psi_K - \sin \varphi_H|}{\sin(\psi_K - \varphi_H)}. \end{aligned}$$

Для существования оптимальной прямолинейной поисковой траектории достаточно выполнение неравенства

$$D(t_H, \psi_K, \varphi_H) \leq R. \quad (16)$$

Длительность интервала времени движения по оптимальной прямолинейной траектории определяется по формуле (7).

Пусть неравенство (16) не выполняется. Тогда управление φ_1 на первом прямолинейном участке находится из условия R -встречи преследователя с лучом (ψ_K):

$$D(t_1, \varphi_1, \psi_K) = D_0 \frac{|\beta \sin \psi_K - \sin \varphi_1|}{\sin(\psi_K - \varphi_1)} = R, \quad (17)$$

где t_1 — длительность первого участка. Решение уравнения (17) имеет вид:

$$\begin{aligned} \sin \varphi_1 &= \\ &= \sin \psi_K \frac{-\beta(\delta \cos \psi_K - 1) - \delta \sqrt{1 + \delta^2 - 2\delta \cos \psi_K - \beta^2 \sin^2 \psi_K}}{1 + \delta^2 - 2\delta \cos \psi_K}, \end{aligned}$$

где $\delta = R/D_0$. Длительность первого участка находится по формуле

$$t_1 = \frac{D_0 \sin \psi_K}{v_P \sin(\psi_K - \varphi_1)}. \quad (18)$$

В момент t_1 происходит переход к следующему участку поисковой траектории.

Возможны два случая:

— существует прямолинейная траектория, про-

ходящая через точки $P^{t_1}(\psi_K)$ и $P^T = E^T(\psi_H)$; тогда оптимальная поисковая траектория состоит из двух прямолинейных отрезков;

— не существует прямолинейной траектории,

проходящей через точки $P^{t_1}(\psi_K)$ и $P^T = E^T(\psi_H)$; тогда оптимальная поисковая траектория содержит участок логарифмической спирали.

Рассмотрим первый случай. Пусть φ_2 — направление движения преследователя на втором прямолинейном участке.

Условие пересечения траекторией преследователя луча (ψ_H) в некоторый момент $T > t_1$ определяется по формуле

$$\operatorname{tg} \psi_H = \frac{v_E T \sin \psi_H - (v_E t_1 - R) \sin \psi_K - v_P \omega \sin \varphi_2}{v_E T \cos \psi_K - (v_E t_1 - R) \cos \psi_K - v_P \omega \cos \varphi_2}, \quad (19)$$

где символом ω обозначена длительность второго прямолинейного участка. Из формулы (19) находим:

$$\omega = \frac{(v_E t_1 - R) \sin(\psi_K - \psi_H)}{v_P \sin(\psi_H - \varphi_2)}. \quad (20)$$

Направление движения преследователя φ_2 , реализующее точечную встречу $P^T = E^T(\psi_H)$, находится из уравнения

$$\begin{aligned} D(T, \varphi_2, \psi_H) &= \left[v_E t_1 + \right. \\ &\left. + (v_E t_1 - R) \frac{\beta \sin(\psi_K - \psi_H) - \sin(\psi_K - \varphi_2)}{\sin(\psi_H - \varphi_2)} \right] = 0, \end{aligned}$$

где t_1 определяется формулой (18). Если на интервале $[\psi_H, \psi_K]$ справедливо неравенство

$$\max_{\psi_H \leq \psi \leq \psi_K} D(t_H, \varphi_2, \psi) \leq R, \quad (21)$$

то кусочно-прямолинейная траектория, второй участок которой соответствует управлению φ_2 , является оптимальной поисковой траекторией.

Полное время движения по оптимальной двухзвенной поисковой траектории $T = t_1 + \omega$, где слабые определяются формулами (18) и (20).

Пусть неравенство (21) несправедливо. Так как в момент t_1 реализуется R -встреча преследователя с лучом (ψ_k), т. е. текущее положение цели уже находится на границе допустимого множества, то вторым участком оптимальной поисковой траектории является спиральный участок, начинающийся на луче (ψ_k); третьим участком — прямолинейный отрезок, касающийся спирали и проходящий через точку $P^T = E^T(\psi_H)$.

1.3. Поиск в секторе типа 3

Справедливо неравенство

$$\psi_k > \pi - \psi_H.$$

Этот случай анализируется с помощью ряда формул, полученных ранее в п. 1.2.

Прямолинейной поисковой траектории не существует. Направление движения преследователя на первом прямолинейном участке выбирается из условия точечной встречи $P^t = E^t(\psi_k)$ преследователя с целью, движущейся вдоль луча (ψ_k):

$$\varphi_1 = \varphi_k = \arcsin(\beta \sin \psi_k)$$

Обозначим символом θ момент точечной встречи $P^0 = E^0(\psi_k)$. Длительность первого прямолинейного участка определяется формулой (12).

В момент θ преследователь изменяет направление движения и далее движется по направлению φ_2 до момента t_n пересечения траекторией преследователя некоторого луча (ψ): $\psi < \psi_k$.

В зависимости от соотношений между параметрами задачи возможны два случая:

- на управлении φ_2 реализуется точечная встреча $P^T = E^T(\psi_H)$;
- на управлении φ_2 реализуется R -встреча с лучом $\psi^*(\varphi_2)$.

Обозначим символом $\varphi_2 \triangleq \varphi_2^0$ управление, реализующее точечную встречу с лучом (ψ_H). Длительность второго прямолинейного участка определяется по формуле (13), а расстояние между точками P^{t_n} и $E^{t_n}(\psi)$ в момент $t_n = \theta + \omega$ — по формуле (14).

Управление φ_2^0 удовлетворяет равенству $D(t_n, \psi_H, \varphi_2^0) = 0$. Отсюда получается уравнение для нахождения φ_2^0 :

$$\cos\left(\frac{\psi_k + \psi_H}{2} - \varphi_2^0\right) = \beta \cos\left(\frac{\psi_k + \psi_H}{2}\right). \quad (22)$$

Управление φ_2^0 реализует прямолинейную поисковую траекторию, проходящую через точки $P^0 = E^0(\varphi_k)$ и $P^T = E^T(\psi_H)$.

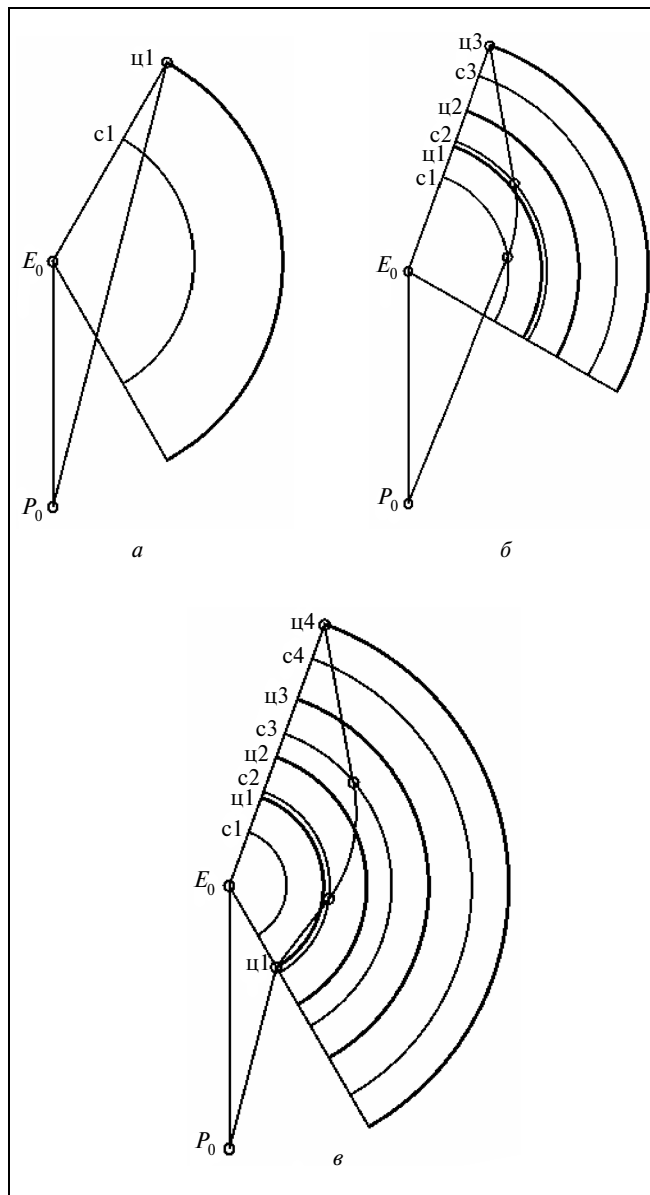


Рис. 3. Оптимальные траектории поиска в секторе типа 1:
 а — прямолинейная поисковая траектория, $\delta = 0,36$, $\psi_H = 30^\circ$, $\psi_k = 150^\circ$; б — трехзвенная поисковая траектория, $\delta = 0,14$, $\psi_H = 20^\circ$, $\psi_k = 120^\circ$; в — четырехзвенная поисковая траектория, $\delta = 0,14$, $\psi_H = 20^\circ$, $\psi_k = 150^\circ$

Обозначим символом $\varphi_2 \triangleq \varphi_2^R$ управление, на котором реализуется R -встреча с лучом $\psi^*(\varphi_2)$. Направление этого луча определяется уравнением

$$\max_{\psi} (t_{\Pi}, \psi, \varphi_2) = v_E \theta [1 + \cos(\psi_K - \psi^*)] = R.$$

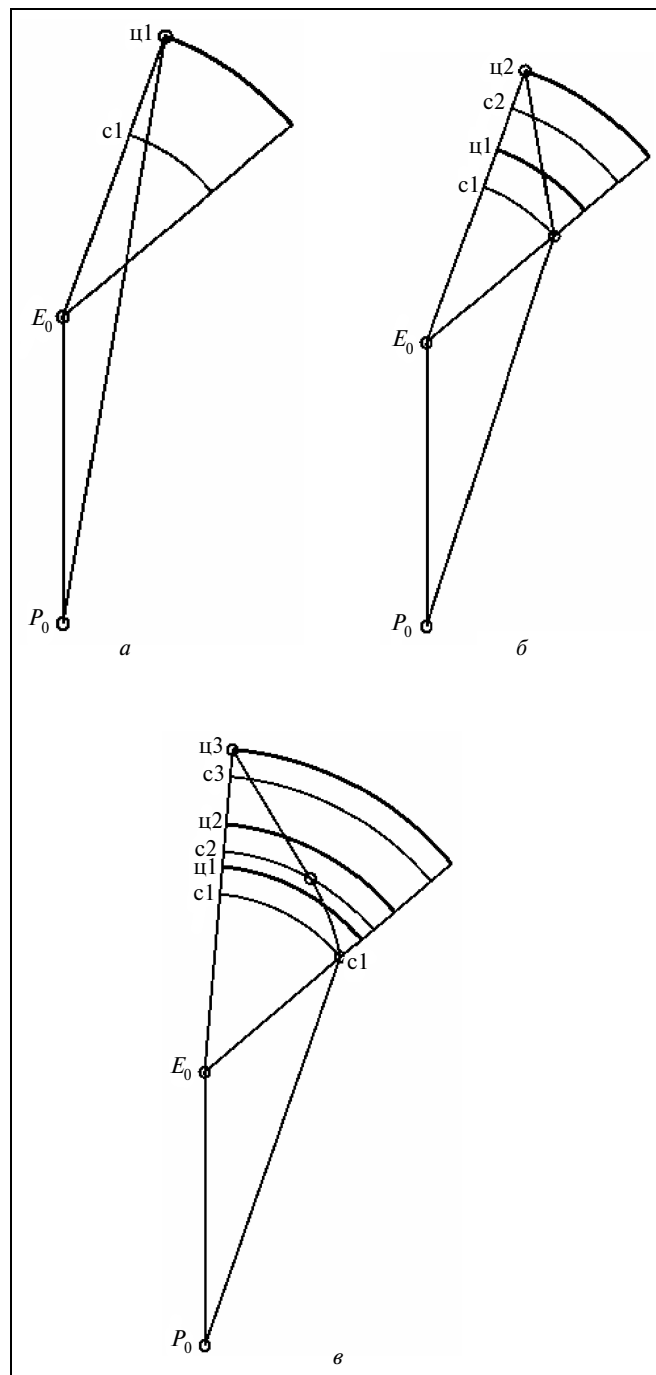


Рис. 4. Оптимальные траектории поиска в секторе типа 2:
 a — прямолинейная поисковая траектория, $\delta = 0,34$, $\psi_H = 20^\circ$, $\psi_K = 50^\circ$; b — двухзвенная поисковая траектория, $\delta = 0,14$, $\psi_H = 20^\circ$, $\psi_K = 50^\circ$; $в$ — трехзвенная поисковая траектория, $\delta = 0,1$, $\psi_H = 5^\circ$, $\psi_K = 50^\circ$

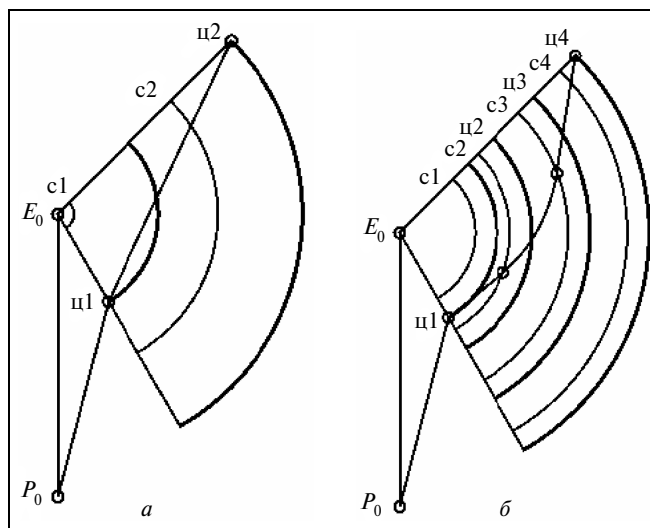


Рис. 5. Оптимальные траектории поиска в секторе типа 3:
 a — двухзвенная поисковая траектория, $\delta = 0,3$, $\psi_H = 45^\circ$, $\psi_K = 150^\circ$; $б$ — четырехзвенная поисковая траектория, $\delta = 0,08$, $\psi_H = 45^\circ$, $\psi_K = 150^\circ$

Если $\varphi_2^R \leq \varphi_2^0$, то оптимальной поисковой траекторией является двухзвенная кусочно-прямолинейная ломаная $P^0 P^0 P^T$.

Время движения по оптимальной двухзвенной траектории вычисляется по формуле

$$T = \theta + \omega = \theta \left[1 + \beta \frac{\sin(\psi_K - \psi_H)}{\sin(\psi_H - \varphi_2^0)} \right],$$

где длительность θ определяется формулой (12), а угол φ_2^0 — уравнением (22).

Если $\varphi_2^0 < \varphi_2^R$, то оптимальная поисковая траектория состоит из четырех участков, включая спиральный участок. Анализ этого случая аналогичен изложенному в п. 1.2.

Результаты проведенного анализа иллюстрируются рис. 3–5. На этих рисунках изображены типовые оптимальные поисковые траектории, соответствующие различным соотношениям между параметрами задачи. Траектории на рис. 3 соответствуют поиску в секторе типа 1; траектории на рис. 4 — поиску в секторе типа 2; траектории на рис. 5 — поиску в секторе типа 3.

На каждом из рисунков изображены начальное и конечное положения объектов P^0 и E^0 и оптимальные поисковые траектории. Дуги окружностей есть геометрические места положений цели (на всех рисунках обозначены буквой «ц» и изображены полужирными линиями) и конца следа (обозначены буквой «с» и изображены тонкими



линиями); номера при обозначениях соответствуют номеру участка поисковой траектории. Точками обозначены позиции преследователя в момент перехода от одного участка поисковой траектории к другому. Все рисунки построены для случая $\beta = v_E/v_P = 0,5$. Значения остальных безразмерных параметров $\delta = R/D_0$, ψ_n и ψ_k указаны на рисунках.

ЗАКЛЮЧЕНИЕ

В настоящей статье (части 1 и 2) даны постановка и решение задачи об оптимизации программной поисковой траектории преследователя, осуществляющего обнаружение подвижного объекта и сближение с ним по информационному признаку «след». Установлена общая структура оптимальной поисковой траектории. Показано, что оптимальная поисковая траектория состоит из одного или нескольких прямолинейных участков и не более, чем одного спирального участка. При этом спиральный участок может отсутствовать, а в том случае, когда спиральный участок есть, он находится между прямолинейными участками траектории. В зависимости от соотношений между параметрами задачи возможны восемь сценариев развития поисковых ситуаций, реализующихся на четырех типах оптимальных поисковых траекторий, содержащих один, два, три и четыре участка.

Авторы выражают признательность рецензенту за сделанные замечания, учет которых позволил существенно улучшить изложение материала статьи.

ЛИТЕРАТУРА

1. Поиск подвижного объекта по информационному признаку «след». Ч. 1. Общая структура оптимальной поисковой траектории / Т.Г. Абрамянц, Ю.А. Беланов, Е.П. Маслов, В.П. Яхно // Проблемы управления. — 2009. — № 5. — С. 61–68.
2. Кинематическая задача поиска подвижного объекта по информационному признаку «след» / Т.Г. Абрамянц, Ю.А. Беланов, Е.П. Маслов, В.П. Яхно // Подводное морское оружие. — 2008. — Вып. 12. — С. 95–102.
3. Вишневецкий Л.С., Меликян А.А. Оптимальное преследование на плоскости при наличии препятствия // Прикладная математика и механика. — 1982. — Т. 46. — Вып. 4. — С. 613–620.
4. Glizer Y.Y. Optimal Planar Interception with Fixed End Conditions: Closed-Form Solution // Journal of Optimization Theory and Applications. — 1996. — Vol. 88, N 3. — P. 503–539.
5. Hybrid Geodesics as Optimal Solutions to the Collision-Free Motion Planning Problem / J. Hu, M. Prandini, R. Johansson, S. Sastry // Lecture Notes in Computer Science. — Berlin: Springer, 2001. — Vol. 234. — P. 305–318.

Статья представлена к публикации членом редколлегии Б.В. Павловым.

Абрамянц Тамара Гургеновна — канд. техн. наук, ст. науч. сотрудник, ☎(495) 334-91-81, ✉abramnc@ipu.ru,

Беланов Юрий Арсеньевич — д-р техн. наук, пенсионер, работал в ЦНИИ автоматики и гидравлики, г. Москва,

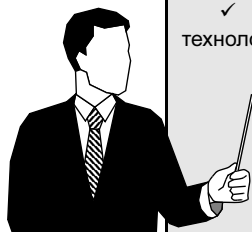
Маслов Евгений Петрович — д-р техн. наук, зав. лабораторией, ☎(495) 334-91-81, ✉maslovep@ipu.ru,

Яхно Виктор Павлович — канд. техн. наук, ст. науч. сотрудник, ☎(495) 334-88-91,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.

Содержание сборника "Управление большими системами", вып. 26, <http://pu.mtas.ru>

- ✓ **Болотова Л.С., Мороз Ю.В.** Игровые подходы к обучению инженера по знаниям
- ✓ **Гусев В.Б., Павельев В.В., Павельев С.В.** Выбор оптимального механизма саморегулирования системы защиты центра обработки данных от аварий и катастроф
- ✓ **Тукубаев З.Б.** Моделирование и исследование алгоритмов динамического управления потоками и очередями сообщений в компьютерных сетях
- ✓ **Гонтарев А.В., Чхартишвили А.Г.** О явных и скрытых коалициях в рефлексивных играх
- ✓ **Горбанева О.И., Угольницкий Г.А.** Модели распределения ресурсов в иерархических системах управления качеством речной воды
- ✓ **Золотова Т.В.** Вопросы согласования интересов в региональной иерархической модели сохранения природных ресурсов
- ✓ **Генкин А.Л.** Комплект алгоритмических и программных модулей для системы управления энергосберегающей технологией при горячей прокатке полос
 - ✓ **Андреевский Б.Р., Фрадков А.Л.** Адаптивное управление летательным аппаратом с идентификацией на скользящих режимах
 - ✓ **Корепанов В.О.** Имитационные модели тактического поведения агентов
 - ✓ **Крайванова В.А.** Модель естественно-языкового интерфейса для систем управления сложными техническими объектами и оценка эффективности алгоритмов на ее основе



НОВЫЕ ИНТЕРВАЛЬНЫЕ БАЙЕСОВСКИЕ МОДЕЛИ НАДЕЖНОСТИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ НЕОДНОРОДНЫХ ПРОЦЕССОВ ПУАССОНА

Л.В. Уткин, С.И. Затенко, Ф. Коолен

Предложен новый класс моделей надежности программного обеспечения, в основу которых положены известные модели, использующие неоднородные процессы Пуассона, например, модели Мусса — Окомото и Джоэл — Окомото. Показано, что основная идея построения моделей заключается в комбинированном применении обобщенного байесовского вывода и метода максимума правдоподобия. Приведены примеры, когда предложенные модели дают более высокое качество прогноза надежности по сравнению с известными.

Ключевые слова: надежность, программное обеспечение, байесовский вывод, распределение вероятностей, неоднородные процессы Пуассона, метод максимума функции правдоподобия, модель.

ВВЕДЕНИЕ

Проявление ошибок в программном обеспечении (ПО) интенсивно изучалось в целях улучшения характеристик программ [1]. Большое число моделей надежности программного обеспечения (МНПО) было разработано в последние десятилетия, но нельзя выбрать одну МНПО, которую можно было бы использовать для самых различных программ. Причина состоит в том, что для многих программ ряд предположений, принятых в МНПО, нереалистичны. Детальный критический обзор известных вероятностных МНПО приведен в работах [2, 3].

Основная идея моделирования надежности ПО состоит в попытке статистически описать процесс отладки программного обеспечения (исправления ошибок и тестового диагностирования до момента проявления очередной ошибки). Обычно предполагается, что после исправления очередной ошибки их общее число почти всегда уменьшается или программа становится лучше в смысле надежности. В данном случае говорят о росте надежности программы. Таким образом, для анализа надежности ПО с использованием статистических моделей их параметры оцениваются на основе имеющихся отладочных данных и данных об отказах. Одна часть параметров модели характеризует рост надежности в результате исправления ошибок, а

другая — случайные изменения времен до отказов или числа отказов в рамках заданного роста надежности.

В работе предлагается новый класс МНПО, в основу которых положены модели, использующие неоднородные процессы Пуассона (МНПП) [4—7] для описания роста надежности и особенностей ПО. Предлагаемые модели будут называться интервальными моделями, так как основные показатели надежности ПО, вычисляемые с их помощью, интервальные. Основная идея построения класса МНПО заключается в комбинированном применении обобщенного байесовского вывода [8], где вместо одного априорного распределения вероятностей рассматривается множество распределений, и метода максимума правдоподобия. При этом каждый из методов ориентирован на определенный набор исходных параметров модели. В разрабатываемых МНПО предполагается: статистическая независимость отказов; равномерность тестовых наборов на множестве всех используемых исходных данных; тесты отражают режим эксплуатации ПО.

Численный анализ предлагаемых МНПО с использованием реальных статистических данных, опубликованных в литературе, наглядно демонстрирует ситуации, когда модели дают более высокое качество прогноза надежности по сравнению с известными моделями надежности.



1. СТАНДАРТНЫЕ МНПО

Одним из инструментов для создания класса МНПО, исходными данными для которых являются значения числа отказов ПО, обозначаемые $N(t)$, за определенный период времени t , служат неоднородные процессы Пуассона. Для любых временных точек $0 < t_1 < t_2 < \dots$ вероятность того, что число отказов между моментами времени t_{i-1} и t_i равно k , может быть записана в виде

$$\begin{aligned} \Pr\{N(t_i) - N(t_{i-1}) = k\} = \\ = \frac{\{m(t_i) - m(t_{i-1})\}}{k!} \exp\{-(m(t_i) - m(t_{i-1}))\}, \\ k = 0, 1, 2, \dots \end{aligned}$$

Здесь $m(t)$ — среднее значение числа отказов до момента времени t . Модели, основанные на неоднородных процессах Пуассона, различаются функцией $m(t)$. Примеры этих функций: $m(t) = at^b$ (модель Дуэйна [6]), $m(t) = a(1 - \exp(-bt))$ (модель Джоэл — Окомото (Д-О) [5]), $m(t) = a \ln(1 + bt)$ (модель Муса — Окомото (М-О) [7]). Приведенные примеры лишь малая часть возможных функций $m(t)$. Их расширенный список можно найти в работе [4], где каждая из функций $m(t)$ наилучшим образом отражает поведение статистических данных об отказах конкретных программ.

Параметры a и b для большинства МНПО имеют схожий смысл. В модели Д-О, например, параметр a интерпретируется, как неизвестное число оставшихся ошибок в ПО, а параметр b — неизвестная интенсивность обнаружения ошибок, которая характеризует рост надежности программы в процессе отладки. Однако не для всех МНПО имеет физическая интерпретация параметров.

Параметры МНПО оцениваются с помощью метода максимума функции правдоподобия, в качестве исходных данных используются значения числа отказов в заданные интервалы тестового диагностирования и отладки. Однако качество прогноза может быть низким вследствие отсутствия достаточного объема статистических данных или их неоднородности. Это можно учитывать в МНПО, рассматривая осторожные или интервальные модели надежности ПО.

2. ИНТЕРВАЛЬНЫЕ МНПО

Пусть $X_i = N(t_i) - N(t_{i-1})$ — случайное число отказов в интервале между t_{i-1} и t_i . Предполагается, что существует функция распределения вероятностей величины X_i , но она неизвестна. Известно только, что она принадлежит множеству $\mathcal{M}_i(\mathbf{d})$, зависящему от множества параметров модели $\mathbf{d}$.

Множество $\mathcal{M}_i(\mathbf{d})$ взаимно однозначно определяется границами, т. е. нижней $\underline{F}_i$ и верхней $\bar{F}_i$ функциями распределения вероятностей

$$\underline{F}_i(k|\mathbf{d}) = \inf_{\mathcal{M}_i(\mathbf{d})} F_i(k), \quad \bar{F}_i(k|\mathbf{d}) = \sup_{\mathcal{M}_i(\mathbf{d})} F_i(k).$$

Отметим, что $\mathcal{M}_i(\mathbf{d})$ — это множество всех функций распределения, ограниченное функциями $\underline{F}_i(k|\mathbf{d})$ и $\bar{F}_i(k|\mathbf{d})$, а не множество параметрических распределений, имеющих ту же форму, что и граничные распределения. Это очень важная особенность предлагаемого в данной работе подхода для комбинирования байесовского метода и метода максимума функции правдоподобия.

Обозначим $\mathbf{K} = (k_1, \dots, k_r)$. Функция правдоподобия записывается в виде:

$$L(\mathbf{K}|\mathbf{d}) = \Pr\{X_1 = k_1, \dots, X_n = k_n\}.$$

Утверждение 1 объясняет, как функция правдоподобия максимизируется по всем распределениям вероятностей, принадлежащим множествам $\mathcal{M}_1(\mathbf{d}), \dots, \mathcal{M}_n(\mathbf{d})$.

Утверждение 1. Если случайные величины $X_1, \dots, X_n$ являются независимыми и дискретными, то

$$\begin{aligned} \max_{\mathcal{M}_1, \dots, \mathcal{M}_n} \Pr\{X_1 = k_1, \dots, X_r = k_r\} = \\ = \prod_{i=1}^n \{\bar{F}_i(k_i) - \underline{F}_i(k_i - 1)\}, \end{aligned} \quad (1)$$

где $\underline{F}_i$ и $\bar{F}_i$ — границы множества $\mathcal{M}_i$, $i = 1, \dots, n$. ♦

Доказательство. Обозначим $\mathbf{M} = (m_1, \dots, m_n)$. Пусть $I_{\{1, \dots, k_i\}}(m)$ — индикаторная функция, принимающая значение 1, если $m \leq k_i$. Верхняя граница совместной вероятности $\Pr\{X_1 = k_1, \dots, X_n = k_n\}$ может быть найдена как решение следующей задачи оптимизации:

$$\max \sum_{m_1=1}^{\infty} \dots \sum_{m_n=1}^{\infty} I_{\{1, \dots, k_n\}}(\mathbf{M}) \prod_{i=1}^n p_i(m_i)$$

при ограничениях $\sum_{m=1}^{\infty} p_i(m) = 1, \quad \underline{F}_i(j) \leq$

$$\leq \sum_{m=1}^{\infty} I_{\{1, \dots, k_i\}}(m) p_i(m) \leq \bar{F}_i(j), \quad i = 1, \dots, n, j = 1, 2, \dots$$

Целевую функцию можно переписать как $\prod_{i=1}^n \sum_{m_i=1}^{\infty} (I_{\{1, \dots, k_i\}}(m_i) - I_{\{1, \dots, k_i-1\}}(m_i) p_i(m_i))$. Введем

новые переменные $F_i(k_i) = \sum_{m_i=1}^{\infty} I_{\{1, \dots, k_i\}}(m_i) p_i(m_i)$.

Тогда задача оптимизации приобретает вид

$$\max \prod_{i=1}^n \{F_i(k_i) - F_i(k_i - 1)\},$$

при ограничениях $\underline{F}_i(k_i) \leq F_i(k_i) \leq \bar{F}_i(k_i)$, $\underline{F}_i(k_i - 1) \leq F_i(k_i - 1) \leq \bar{F}_i(k_i - 1)$, $i = 1, \dots, n$.

Используя известные правила стандартного интервального анализа, получаем (1). ♦

3. ОБОБЩЕННЫЕ БАЙЕСОВСКИЕ МОДЕЛИ ДЛЯ ПОЛУЧЕНИЯ МНОЖЕСТВА $\mathcal{M}$

Один из возможных путей получения множества $\mathcal{M}(\mathbf{d})$ — это использование идеи обобщенного байесовского подхода Уолли [8].

3.1. Стандартный байесовский подход

Если случайная величина имеет распределение вероятностей с вектором неизвестных параметров $\mathbf{b} \in \Omega$, то в соответствии с байесовским подходом [10] эти параметры рассматриваются как случайные величины с априорной плотностью вероятностей $\pi(\mathbf{b}|\mathbf{c})$ с параметрами $\mathbf{c}$. Центральный элемент байесовского подхода состоит в выводе апостериорного распределения неизвестных параметров при получении статистических данных на основе теоремы Байеса. Пусть статистическая информация собрана в виде вектора $\mathbf{K} = (k_1, \dots, k_n)$ наблюдаемых значений случайных величин $X_1, \dots, X_n$. Пусть $p(k)$ — вероятность того, что наблюдаемое значение числа отказов равно k . Тогда апостериорная плотность $\pi(\mathbf{b}|\mathbf{K}, \mathbf{c})$ вычисляется как $\pi(\mathbf{b}|\mathbf{K}, \mathbf{c}) \propto p(k_1) \dots p(k_n) \cdot \pi(\mathbf{b}|\mathbf{c})$.

Априорное распределение зачастую выбирается таким образом, чтобы упростить вычисления. Наиболее эффективный путь в этом направлении — это выбор *согласованных априорных распределений* [10]. Если апостериорное распределение $\pi(\mathbf{b}|\mathbf{K}, \mathbf{c})$ и априорное распределение $\pi(\mathbf{b}|\mathbf{c})$ принадлежат одному семейству распределений, то π и p называются согласованными распределениями.

3.2. Интервальные априорные модели

Если отбросить требование простоты вычислений, то предпочтительное априорное распределение должно минимально влиять на апостериорное распределение, а также учитывать отсутствие априорной информации о параметрах. Априорные распределения, моделирующие отсутствие априорной информации, называются *неинформативными*. Существует немало подходов для выбора неинформативного априорного распределения [10],

имеющих свои достоинства и недостатки. Однако определим не одно априорное распределение, а целый класс $\mathcal{P}$ распределений π , для которого можно найти нижнюю и верхнюю вероятности события A как

$$\underline{P}(A) = \inf\{P_\pi(A): \pi \in \mathcal{P}\},$$

$$\bar{P}(A) = \sup\{P_\pi(A): \pi \in \mathcal{P}\}.$$

Когда априорной информации почти нет, для этого класса $\underline{P}(A) \rightarrow 0$, а $\bar{P}(A) \rightarrow 1$. Это означает, что априори событие A может иметь любую вероятность. К наиболее известным классам априорных распределений следует отнести обобщенную модель Дирихле [8], обобщенные модели экспоненциального семейства распределений [9, 11]. В нашем подходе к построению МНПП множество $\mathcal{P}$ используется для того, чтобы сгенерировать множество $\mathcal{M}$ прогнозируемых распределений с нижней и верхней границами, которые позволят применить утверждение 1.

3.3. Обобщенная модель отрицательного биномиального распределения

Если число отказов имеет распределение Пуассона, то априорное согласованное распределение является гамма-распределением, обозначаемым $G(\alpha, \beta)$ с параметрами $\alpha > 0$, $\beta > 0$. Предположим, что уже было зафиксировано K отказов, наблюдаемых за период времени T . Тогда апостериорным распределением является $G(\alpha + K, \beta + T)$. Отсюда прогнозируемое распределение k отказов за время t при условии, что K отказов были зафиксированы за время T , имеет вид

$$P(k, t) = \int_0^\infty \frac{(\lambda t)^k \exp(-\lambda t)}{k!} \cdot G(\alpha + K, \beta + T) d\lambda =$$

$$= \frac{\Gamma(\alpha + K + k)}{\Gamma(\alpha + K)k!} \left(\frac{\beta + T}{\beta + T + t} \right)^{\alpha + K} \left(\frac{t}{\beta + T + t} \right)^k. \quad (2)$$

Это выражение для отрицательного биномиального распределения. Если взять вектор (α, β) в треугольнике $(0, 0)$, $(s, 0)$, $(0, s)$, то все возможные априорные ожидаемые интенсивности отказов попадают в интервал $(0, \infty)$, так как математическое ожидание случайной величины λ , имеющей гамма-распределение, равно α/β . Множество всех параметров в треугольнике образует обобщенную модель отрицательного биномиального распределения [9]. Параметр $s > 0$ определяет влияние априорного распределения на апостериорное. Он также определяет, как быстро нижняя и верхняя границы вероятностей событий сходятся по мере накопления статистических данных. Меньшие значения s приводят к быстрой сходимости и более строгим заключениям. Большие значения s дают более осторожный статистический вывод. Принимая множество распределений вместо одного



распределения вероятностей, мы можем искать только нижнюю и верхнюю границы вероятностей событий вместо точных значений для одного распределения. Эти границы могут быть получены минимизацией и максимизацией этих вероятностей по всему множеству значений (α, β) в треугольнике.

4. ОБЩАЯ СХЕМА ПОСТРОЕНИЯ ИНТЕРВАЛЬНОЙ МНПП

С одной стороны, большая часть возможных функций среднего значения числа отказов $m(t)$ представляется как $m(t; a, b) = a \cdot \tau(t, b)$. С другой стороны, параметр λ распределения Пуассона в выражении (2) и его аргумент t можно заменить параметром a и дискретным временем $\tau(t_j, b) - \tau(t_{j-1}, b)$ соответственно. Фактически, заменяя λ на a , получаем процесс Пуассона с масштабированным временем тестового диагностирования ПО, т. е. каждый интервал времени $[t_{j-1}, t_j]$ заменяется интервалом $[\tau(t_{j-1}, b), \tau(t_j, b)]$. Используя результаты предыдущих разделов, можно записать функцию распределения вероятностей числа отказов в дискретные интервалы времени между t и t_j ($t > t_j$) после r периодов отладки:

$$P_j(k, t | \alpha, \beta, b) = \frac{\Gamma(\alpha + K + k)}{\Gamma(\alpha + K)k!} \left(\frac{\beta + T_r}{\beta + T_r + \tau(t, b) - \tau(t_j, b)} \right)^{\alpha + K_r} \times \left(\frac{\tau(t, b) - \tau(t_j, b)}{\beta + T_r + \tau(t, b) - \tau(t_j, b)} \right)^k.$$

Здесь T_r — полное масштабированное время для r -го периода, определяемое как

$$T_r = \sum_{i=1}^r [\tau(t_i, b) - \tau(t_{i-1}, b)] = \tau(t_r, b), \\ t_0 = 0, \quad k_0 = 0,$$

K_r — полное число отказов до момента времени за r

периодов отладки, определяемое как $K_r = \sum_{i=1}^r k_i$.

Функция распределения вероятностей для момента времени $t \in [t_{j-1}, t_j]$ после r периодов отладочного процесса имеет вид

$$F_j(k, t | \alpha, \beta, b) = \sum_{l=0}^j \frac{\Gamma(\alpha + K_r + l)}{\Gamma(\alpha + K_r)l!} \left(\frac{\beta + \tau(t_r, b)}{\beta + T_j(t, b) + \tau(t_r, b)} \right)^{\alpha + K_r} \times \left(\frac{T_j(t, b)}{\beta + T_j(t, b) + \tau(t_r, b)} \right)^l.$$

Здесь $T_j(t, b) = \tau(t, b) - \tau(t_{j-1}, b)$.

Из приведенных выражений видно, что модифицированная модель не зависит от параметра a , но зависит от новых параметров априорного гамма-распределения α и β . Кроме того, модель зависит от параметра роста b . Заметим, что альтернативная форма функции распределения вероятностей имеет вид:

$$F_j(k, t | \alpha, \beta, b) = 1 - I(q(j, t, b), k + 1, \alpha + K_r).$$

Здесь

$$q(j, t, b) = \frac{T_j(t, b)}{\beta + \tau(t_r, b) + T_j(t, b)},$$

$B_a(b, c)$ — неполная бета-функция; $B(b, c)$ — бета-функция; $I(a, b, c) = B_a(b, c)/B(b, c)$.

Предлагается следующая схема для построения МНПП и для вычисления функции распределения вероятностей F_{r+1} после r периодов отладочного процесса. Построив обобщенную байесовскую модель и используя ее, получим множество распределений вероятностей $\mathcal{M}_j$ с некоторыми границами $\underline{F}_j$ и $\bar{F}_j$. Далее будет показано, что эти границы не зависят от параметров α и β и определяются только параметром b (параметр s фиксирован). Отсюда следует, что получаем однопараметрическую модель. Тогда из утверждения 1 следует, что максимум функции правдоподобия по множеству $\mathcal{M}_j(b)$ записывается на основе выражения (1). Максимизируя эту функцию по параметру b при фиксированном s , находим оптимальное значение этого параметра. Далее, получив параметр b , функция распределения вероятностей F_{r+1} после r периодов отладки вычисляется посредством построенной обобщенной байесовской модели, но теперь с полученным оптимальным значением b .

Следуя представленной выше схеме, построим сначала обобщенную байесовскую модель. Достаточно просто доказать, что минимум функции $F_j(k, t | \alpha, \beta, b)$ по всем (α, β) в треугольнике $(0, 0)$, $(s, 0)$, $(0, s)$ достигается при $(\alpha, \beta) = (s, 0)$ и ее нижняя граница определяется как

$$\underline{F}_j(k, t | s, b) = 1 - I\left(\frac{T_j(t, b)}{\tau(t_r, b) + T_j(t, b)}, k + 1, s + K_r\right).$$

Максимум функции $F_j(k, t | \alpha, \beta, b)$ достигается при $(\alpha, \beta) = (0, s)$, и ее верхняя граница

$$\bar{F}_j(k, t | s, b) = 1 - I\left(\frac{T_j(t, b)}{s + \tau(t_r, b) + T_j(t, b)}, k + 1, K_r\right).$$

Следующий шаг — использование утверждения 1 и максимизация функции правдоподобия (1) по b . В результате получаем оптимальное значение b^0 :

$$b^0 = \arg \max_b \ln L(\mathbf{K}|s, b) = \\ = \arg \max_b \sum_{j=1}^r \ln \left(I \left(\frac{T_j(t_j, b)}{\tau(t_r, b) + T_j(t_j, b)}, k_j + 1, s + K_r \right) - \right. \\ \left. - I \left(\frac{T_j(t_j, b)}{s + \tau(t_r, b) + T_j(t_j, b)}, k_j + 1, K_r \right) \right).$$

Завершающий шаг — запись границ функций распределения вероятностей числа отказов в интервале времени $[t_r, t]$ после r периодов отладочного процесса

$$\underline{F}_{r+1}(k, t|s, b^0) = \\ = 1 - I \left(\frac{T_{r+1}(t, b^0)}{\tau(t_r, b^0) + T_{r+1}(t, b^0)}, k + 1, s + K_r \right), \\ \bar{F}_{r+1}(k, t|s, b^0) = \\ = 1 - I \left(\frac{T_{r+1}(t, b^0)}{s + \tau(t_r, b^0) + T_{r+1}(t, b^0)}, k + 1, K_r \right).$$

Нижняя $\underline{F}_{r+1}^{(s)} X_{r+1}$ и верхняя $\bar{F}_{r+1}^{(s)} X_{r+1}$ границы ожидаемого числа отказов в интервале $[t_r, t]$ после r периодов отладочного процесса определяются через известные выражения для математического ожидания случайной величины, имеющей отрицательное биномиальное распределение, как

$$\underline{F}_{r+1}^{(s)} X_{r+1} = \frac{K_r T_{r+1}(t, b^0)}{s + \tau(t_r, b^0)}, \\ \bar{F}_{r+1}^{(s)} X_{r+1} = (s + K_r) \frac{T_{r+1}(t, b^0)}{\tau(t_r, b^0)}.$$

5. ЧАСТНЫЕ СЛУЧАИ ИНТЕРВАЛЬНЫХ МНПП

Из приведенных выражений для обобщенных байесовских моделей видно, что они отличаются только функцией $m(t, b)$. Следовательно, используя функции известных МНПП, можно построить интервальные модификации МНПО в рамках предложенного подхода.

Модифицируем модель М-О [7], как одну из наиболее распространенных МНПО. Используя введенные обозначения, можно записать для модели $\tau(t, b) = \ln(1 + bt)$. Отсюда функция правдоподобия

$$L^{(s)}(\mathbf{K}|s, b) = \prod_{j=1}^r \left(I \left(\frac{\ln(1 + bt)}{\ln W}, k_j + 1, s + K_j \right) - \right. \\ \left. - I \left(\frac{\ln(1 + bt)}{s + \ln W}, k_j + 1, K_j \right) \right).$$

Здесь $W = (1 + bt_r)(1 + bt)/(1 + bt_{j-1})$.

Другой МНПО, основанной на неоднородных процессах Пуассона, является модель Д-О [5]. Для нее можно записать $\tau(t, b) = 1 - \exp(-bt)$. Функция правдоподобия тогда имеет вид

$$L^{(s)}(\mathbf{K}|s, b) = \prod_{j=1}^r \left(I \left(\frac{1 - \exp(-bt)}{V}, k_j + 1, s + K_j \right) - \right. \\ \left. - I \left(\frac{1 - \exp(-bt)}{s + V}, k_j + 1, K_j \right) \right).$$

Здесь $V = 1 - \exp(-bt_r) - \exp(-bt) + \exp(-bt_{j-1})$.

6. АНАЛИЗ ИНТЕРВАЛЬНЫХ МОДЕЛЕЙ

Проанализируем предлагаемые интервальные МНПП, используя реальные статистические данные по надежности ПО, опубликованные в работе [7]. Эти данные получены при отладке ПО для управления контроллерами в системе управления. Статистические данные представлены в виде 20-ти тестов и показаны в таблице.

Сначала анализируется предложенная интервальная модификация модели Д-О. Для ее оценки делается прогноз $(i + 1)$ -го среднего (ожидаемого) числа отказов ПО, начиная с $i = 3$. Имея среднее число отказов, можно сравнить этот результат с реальным значением числа отказов в $(i + 1)$ -м тесте, которое дано в таблице. Кроме того, сравниваются результаты моделирования также со стандартной моделью Д-О. Кривая без маркеров на рис. 1 — множество значений числа отказов из таблицы.

Статистические данные по отказам ПО

Номер теста	Время теста, ч	Число отказов
i	$t_i - t_{i-1}$	k_i
1	62,5	9
2	44	4
3	40	7
4	68	6
5	62	5
6	66	3
7	73	2
8	73,5	5
9	92	4
10	71,4	2
11	64,5	4
12	64,7	7
13	36	0
14	54	5
15	39,5	3
16	68	3
17	61	3
18	62,6	4
19	98,7	10
20	25	3

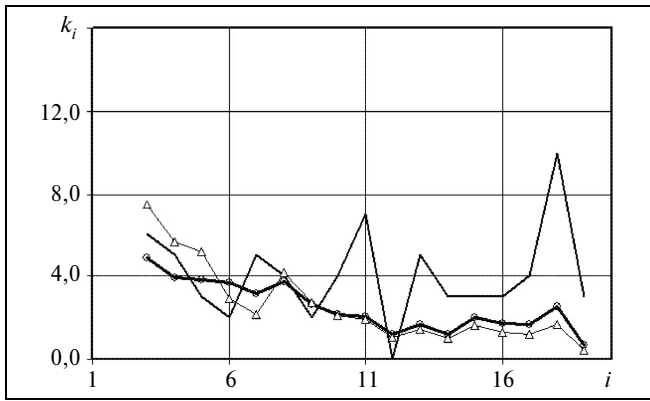


Рис. 1. Прогнозируемые ожидаемые значения числа ошибок для различных моделей, построенных на основе модели Д-О

Жирная кривая с круглыми маркерами — множество прогнозируемых ожидаемых значений числа отказов $E^{(s)}X_{i+1}$, $i = 3, \dots, 20$, при $s = 1$. Здесь значение $E^{(s)}X_{i+1}$ вычисляется с учетом нижней и верхней границ математических ожиданий [8] как $E^{(s)}X_{i+1} = \eta \underline{E}^{(s)}X_{i+1} + (1 - \eta) \bar{E}^{(s)}X_{i+1}$, $\eta = 0,5$. Тонкая кривая с треугольными маркерами изображает множество точных прогнозируемых значений математического ожидания $E X_{i+1}$, вычисленных с использованием модели Д-О. Можно увидеть из приведенных графиков, что интервальная модификация модели Д-О обеспечивает лучший прогноз по сравнению с обычной моделью Д-О.

Введем следующие показатели качества моделей: максимальное отклонение R_1 , среднее значение отклонения R_2 , среднеквадратическое отклонение R_3 .

Используя статистические данные, вычислим соответствующие показатели при $s = 1$ после прогнозирования 17-ти значений числа отказов (от 3-го теста до 19-го теста) для предлагаемой модели $R_1 = 7,491$, $R_2 = 2,054$, $R_3 = 1,786$ и для модели Д-О $R_1^* = 8,311$, $R_2^* = 2,312$, $R_3^* = 1,959$. Из результатов расчета следует, что качество предлагаемой модификации выше по сравнению с обычной моделью Д-О.

Интересно также исследовать, как число тестов влияет на значения тех же самых показателей качества. Вычислим эти показатели после прогнозирования 6-ти значений числа отказов (от 3-го до 8-го теста). В этом случае получаем для новой модели $R_1 = 1,827$, $R_2 = 1,133$, $R_3 = 0,568$ и для модели Д-О $R_1^* = 2,854$, $R_2^* = 1,373$, $R_3^* = 1,007$. Из результатов расчета следует, что качество предлагаемой интервальной модификации повышается по сравнению с обычной моделью Д-О, когда число тестов небольшое. Это видно из рис. 2, где изображены отклонения прогнозируемых ожидаемых зна-

чений числа отказов от реальных данных. Жирная кривая с круглыми маркерами — отклонения, соответствующие предлагаемой модели. Тонкая кривая с треугольными маркерами — отклонения, соответствующие модели Д-О. Очевидно, что новая модель ведет себя лучше при малом числе тестов.

Следующая задача анализа модели заключается в исследовании влияния параметра s на качество прогноза модели. Используя те же самые статистические данные об отказах, вычислим показатели качества при $s = 0,5$ после 17 тестов $R_1 = 7,198$, $R_2 = 1,952$, $R_3 = 1,719$. Те же самые показатели при $s = 0,5$ после 6 тестов приобретают значения: $R_1 = 2,049$, $R_2 = 1,074$, $R_3 = 0,657$. Можно увидеть, что меньшие значения s приводят к более высокому качеству прогноза, когда имеется большой объем данных. Однако качество прогноза падает при малом числе статистических данных. Противоположное заключение можно сделать при анализе случая, когда параметр s возрастает. В частности, при $s = 2$ и после 17-ти тестов получаем $R_1 = 7,734$, $R_2 = 2,134$, $R_3 = 1,852$. Те же самые показатели при $s = 2$ после 6-ти тестов имеют значения $R_1 = 2,026$, $R_2 = 1,172$, $R_3 = 0,547$.

Проанализируем интервальную модификацию модели М-О и сравним ее со стандартной моделью М-О. Соответствующие кривые прогноза для моделей показаны на рис. 3. Отклонения прогнозируемых ожидаемых значений числа отказов от реальных статистических данных изображены на рис. 4. Используя статистические данные, вычисляем показатели качества при $s = 1$ после прогнозирования 17 значений числа отказов (от 3-го до 19-го теста) для предлагаемой интервальной модификации модели М-О $R_1 = 7,817$, $R_2 = 2,134$, $R_3 = 1,852$, и для обычной модели М-О $R_1^* = 7,445$, $R_2^* = 2,222$, $R_3^* = 1,800$. Те же самые показатели после 6-ти тестов (от 3-го до 8-го теста): $R_1 = 1,932$,

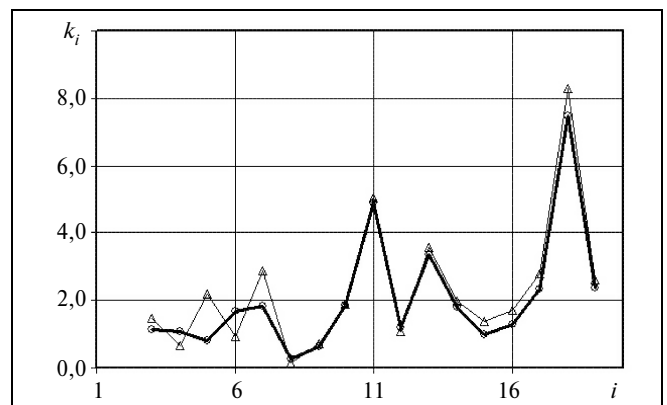


Рис. 2. Отклонения прогнозируемых ожидаемых значений числа ошибок от реальных данных для моделей Д-О

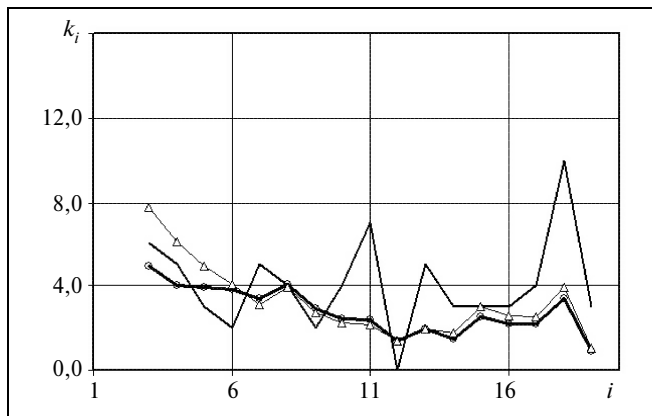


Рис. 3. Прогнозируемые ожидаемые значения числа ошибок для различных моделей, построенных на основе модели М-О

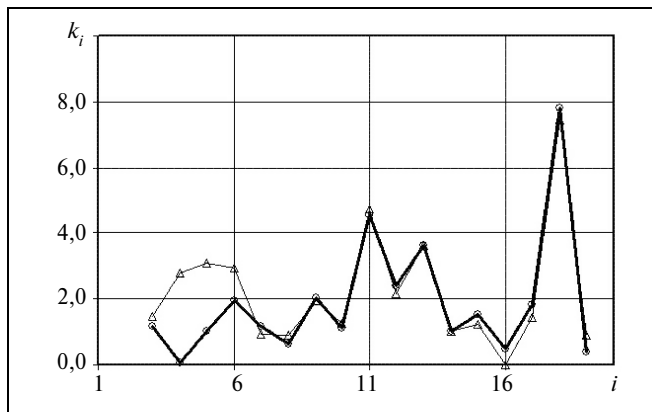


Рис. 4. Отклонения прогнозируемых ожидаемых значений числа ошибок от реальных данных для моделей М-О

$R_2 = 0,990$, $R_3 = 0,620$, $R_1^* = 3,060$, $R_2^* = 2,005$, $R_3^* = 1,020$.

Модель М-О и ее модификации демонстрируют практически те же зависимости показателей качества прогноза от объема выборки и параметра s , что и модель Д-О.

ЗАКЛЮЧЕНИЕ

Предложен подход, направленный на развитие нового класса моделей надежности программного обеспечения. Его основная идея состоит в комбинированном применении обобщенного байесовского подхода и метода максимума функции правдоподобия. Ключевыми элементами подхода являются доказанное утверждение и разделение всех параметров модели на два подмножества, что позволяет уменьшить размерность задачи. Другая важная особенность подхода состоит в возможности получать интервальные оценки надежности программного обеспечения. Точные характеристики надежности, которые обычно стремятся иметь специалисты, далеко не всегда заслуживают доверия,

например, когда статистические данные ограничены и исходная информация бедна. В предложенных моделях результат моделирования представляет собой интервал показателя надежности. Дальнейшее решение принимается исходя из того, насколько высоки требования к надежности анализируемого программного обеспечения.

Предложенный подход носит общий характер. Несмотря на то, что рассмотрены только две частные модели, аналогичным образом можно рассмотреть и другие модификации известных моделей надежности программного обеспечения.

Результаты численного анализа рассмотренных моделей и их сравнение с известными моделями говорят в пользу нового подхода. В то же время, из этих результатов видно, что нельзя выбрать идеальную модель для всех ситуаций. Необходимо пользоваться различными моделями и комбинировать результаты их применения.

ЛИТЕРАТУРА

1. Майерс Г. Надёжность программного обеспечения. — М.: Мир, 1980. — 360 с.
2. Cai K.Y., Wen C.Y., Zhang M.L. A critical review on software reliability modeling // Reliability Engineering and System Safety. — 1991. — Vol. 32, N 3. — P. 357—371.
3. Уткин Л.В., Шубинский И.Б. Нетрадиционные методы оценки надежности информационных систем. — СПб.: Любавич, 2000. — 173 с.
4. Cai K.Y. Towards a conceptual framework of software run reliability modeling // Information Sciences. — 2000. — Vol. 126, N 1—4. — P. 137—163.
5. Goel A.L., Okamoto K. Time dependent error detection rate model for software reliability and other performance measures // IEEE Trans. Reliab. — 1979. — Vol. R-28, N 3. — P. 206—211.
6. Duane J.T. Learning curve approach to reliability monitoring // IEEE Transactions on Aerospace. — 1964. — Vol. AS-2, N 2. — P. 563—566.
7. Misra P.N. Software reliability analysis // IBM Systems Journal. — 1983. — Vol. 22, N 3. — P. 262—270.
8. Walley P. Inferences from multinomial data: Learning about a bag of marbles // Journal of the Royal Statistical Society, Ser. B. — 1996. — Vol. 58, N 1. — P. 3—57.
9. Уткин Л.В. Анализ риска и принятие решений при неполной информации. — СПб.: Наука, 2007. — 404 с.
10. Bernardo J.M., Smith A.F.M. Bayesian Theory. — Chichester: Wiley, 1994.
11. Quaeghebeur E., De Cooman G. Imprecise probability models for inference in exponential families // Proc. of the 4rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'05. — Carnegie Mellon University. Pittsburgh. — 2005. — P. 287—296.

Статья представлена к публикации членом редколлегии П.П. Пархоменко.

Уткин Лев Владимирович — д-р техн. наук, проректор, Санкт-Петербургская государственная лесотехническая академия, ☎(812) 550-02-53, ✉lev.utkin@mail.ru, URL: www.levvu.narod.ru,

Затенко Светлана Ивановна — ст. преподаватель, Санкт-Петербургская государственная лесотехническая академия, ✉s_lana2004@mail.ru,

Коолен Франк (Frank Coolen) — Professor, PhD, Durham University, UK, Durham, ☎(0191 or +44 191) 334-30-48 (direct), 334-30-50 (secr.), ✉Frank.Coolen@durham.ac.uk, URL: http://maths.dur.ac.uk/stats/people/fc/fc.html.



ОРГАНИЗАЦИЯ ДИАГНОСТИРОВАНИЯ ЦИФРОВЫХ СИСТЕМ СО СТРУКТУРОЙ СИММЕТРИЧНОГО ДВУДОЛЬНОГО ГРАФА

В.А. Ведешенков

Отмечено, что особенность цифровых систем со структурой в виде симметричного двудольного графа состоит в использовании равного числа компонентов двух типов: процессоров и блоков памяти. Разработана организация диагностирования систем с такой структурой с точностью до отдельного компонента (процессора, памяти, линии связи). На время диагностирования разнотипные устройства объединяются в проверяемые подсистемы одинакового состава. Рассмотрен пример организации диагностирования системы, состоящей из семи процессоров и семи блоков памяти.

Ключевые слова: цифровые системы, симметричный двудольный граф, компонент, диагностирование, проверяемые подсистемы.

ВВЕДЕНИЕ

Поиск исследователями новых, более эффективных, чем известные, структур связей сложных информационных систем показал, что по ряду параметров структуры в виде симметричных двудольных графов предпочтительнее структур в виде полных графов и полнодоступных двудольных графов [1]. Симметричным двудольным графом называется граф, состоящий из двух подмножеств вершин X и Y равной мощности и ребер, у которых одна из концевых вершин принадлежит подмножеству $X(Y)$, а другая — подмножеству $Y(X)$. Пример такого графа показан на рисунке, где вершины подмножества X обозначены кружками, а подмножества Y — прямоугольниками. В числе возможных областей применения графов указанного семейства называют структуры многопроцессорных вычислительных систем (МВС), где, например, подмножество вершин X представляет совокупность процессорных элементов или вычислительных машин, а подмножество вершин Y — блоки или банки памяти.

Любой метод системного диагностирования технического состояния компонентов в цифровых системах (ЦС) предполагает, что модули, на которые разбивается исходная МВС, однородны и обладают функциональными возможностями, достаточными для реализации необходимых диагностических алгоритмов. Использование устройств

разных типов при построении систем со структурой в виде симметричного двудольного графа приводит к неоднородным системам, что не позволяет непосредственно применить системные методы диагностирования. Однако автоматическое диагностирование состояния компонентов подобных систем представляется необходимым, так как позволяет ускорить время восстановления работоспособности систем после обнаружения факта их неправильного функционирования. Кроме того, различие структур системы, представленной на рисунке, и неоднородной системы обработки радиолокационной информации не позволяет напрямую использовать метод самодиагностирования, предложенный в работе [2].

Далее представлен способ организации диагностирования состояния компонентов ЦС со структурой в виде симметричного двудольного графа. Он основан на подходе, описанном в работе [2]: на этапах контроля и диагностирования связанные разнотипные устройства объединяются в проверяемые подсистемы, обладающие необходимыми функциональными возможностями для реализации диагностических алгоритмов.

Напомним, что в однородных ЦС при выполнении процессов контроля и диагностирования проверяющий модуль подает на входы проверяемого модуля последовательность проверок (тест), получает ответные реакции проверяемого модуля на этот тест, сравнивает их с эталонными, а результат сравнения оценивает в двоичной форме:

0 — результаты правильные, 1 — неправильные. В модели Барси — Грандони — Маестрини (БГМ) [3], используемой в данной работе, предполагается, что нулевая оценка, выдаваемая проверяющим модулем, с большой вероятностью подтверждает исправность проверенного им модуля. Такая предпосылка модели БГМ существенно упрощает процедуру диагностирования, так как для идентификации исправных модулей достаточно 0-путей единичной длины (в терминологии работы [4]).

Как и в большинстве алгоритмов системного диагностирования, будем предполагать, что результаты проверок собираются и обрабатываются (дешифрируются) внешним устройством — диагностическим монитором (ДМ). По результатам дешифрации собранных результатов проверок обслуживающий персонал заменяет отказавшие компоненты и восстанавливает работоспособность МВС.

1. ПОСТАНОВКА ЗАДАЧИ

Исходная МВС со структурой в виде симметричного двудольного графа $Q_{n,s}$ представлена диагностическим графом, у которого n вершин подмножества X представляют процессоры системы, n вершин подмножества Y — блоки памяти, s — степень вершин двудольного графа; т. е. каждая вершина подмножества $X(Y)$ связана дугами с s вершинами подмножества $Y(X)$. (Для графа на рисунке $s = 3$.)

Будем считать, что допускаются устойчивые отказы ограниченного числа вершин и дуг, причем их неисправности таковы, что не влияют на работоспособность смежных модулей. Предполагается также, что в процессе диагностирования новые отказы не возникают. Для описания результатов тестирования компонентов используется модель БГМ. Запуском процессов диагностирования в различных подсистемах ЦС и обработкой полученных результатов занимается исправный ДМ, который находится в модуле, внешнем по отношению к диагностируемой ЦС.

Требуется разработать такой метод организации диагностирования ЦС со структурой в виде симметричного двудольного графа, который обеспечивает ДМ достоверной информацией для диагностирования состояния компонентов (процессоров, памяти, линий связи) системы.

2. СПОСОБ ДИАГНОСТИРОВАНИЯ ЦИФРОВЫХ СИСТЕМ СО СТРУКТУРОЙ В ВИДЕ СИММЕТРИЧНОГО ДВУДОЛЬНОГО ГРАФА

Процесс диагностирования компонентов однородных систем включает в себя такие процедуры [5]:

— исполнение исправным модулем функций проверяющего модуля;

— нахождение пути от одного исправного модуля к другому исправному для передачи ему функций проверяющего модуля;

— дешифрация результатов выполненных проверок.

Для реализации подобных процедур в неоднородных системах со структурой в виде симметричного двудольного графа поступим следующим образом.

Проверяющим модулем m_j назовем устройство активного типа (один из процессоров $m_1, \dots, m_7$ в системе, представленной на рисунке), а в проверяемую подсистему $U_{j,i,k}$ включим следующие компоненты: память p_i , процессор m_k , линию связи l_{i-k} между ними и линию связи l_{j-i} , соединяющую m_j с p_i . (Таким образом, индекс (j, i, k) подсистемы $U_{j,i,k}$ состоит из номера j проверяющего процессора, номера i проверяемой памяти и номера k проверяемого процессора.) Например, для системы на Рисунке в одну из проверяемых подсистем могут входить процессор m_1 в функции проверяющего модуля, память p_8 , процессор m_2 и линии связи l_{8-2} и l_{1-8} .

Наличие s устройств памяти, связанных с одним процессором цифровой системы со структурой симметричного двудольного графа ($s = 3$ для графа на рисунке), делает объединение различных устройств для формирования проверяемых подсистем неоднозначным. Выбор достаточно экономичного подмножества из числа всевозможных проверяемых подсистем является нетривиальной задачей, но здесь не рассматривается.

Тогда отдельная проверка (тест $t_{j,i,k}$) будет выполняться следующим образом. Получив управление в процессе диагностирования, проверяющий модуль (процессор) m_j передает по линии l_{j-i} в память p_i проверяемой подсистемы команды, необходимые для организации тестов самоконтроля в подсистеме $U_{j,i,k}$. Процессор m_k , получив и расшифровав эти команды, осуществляет контроль процессора m_k , линии l_{i-k} и памяти p_i с помощью тестовых программ, хранящихся в памяти p_i . Результаты выполненного теста $t_{j,i,k}$ процессор m_k записывает в ячейки памяти p_i , откуда их после очередного опроса памяти считывает процессор m_j , сравнивает с эталонами из собственной сверхоперативной памяти и формирует оценку $r_{j,i,k} = 0$, если результаты совпали с эталонами, 1 — в противном случае.



На основании полученных оценок процессор m_j выбирает исправную подсистему U_{j,i^*,k^*} , процессору m_{k^*} из состава которой передает управление и необходимую информацию для продолжения процесса диагностирования. Все исправные процессоры находятся в ждущем режиме и периодически опрашивают заранее определенные ячейки памяти. Процессор m_{k^*} , обнаружив в памяти p_{i^*} команду на выполнение процесса диагностирования, выполнит необходимые процедуры, а после их завершения передаст управление следующему процессору с помощью того же механизма (через память одной из проверенной им подсистем).

Таким образом, в предлагаемом способе диагностирования неоднородных ЦС со структурой симметричного двудольного графа предполагается объединение разнотипных устройств исходной системы в проверяемые подсистемы одинакового состава с функциональными возможностями, достаточными для выполнения диагностических процедур.

3. ОРГАНИЗАЦИЯ ДИАГНОСТИРОВАНИЯ КОМПОНЕНТОВ ЦИФРОВОЙ СИСТЕМЫ СО СТРУКТУРОЙ СИММЕТРИЧНОГО ДВУДОЛЬНОГО ГРАФА

Контроль работоспособности однородной цифровой системы заключается в том, что каждый проверяющий модуль должен проверить техническое состояние своих преемников и соответствующих линий связи. Простейший способ выполнения проверок одного модуля разными проверяющими модулями — последовательный. Для его оптимальной реализации на диагностическом графе однородной системы нужно построить гамильтонов цикл, который обеспечит однократный об-

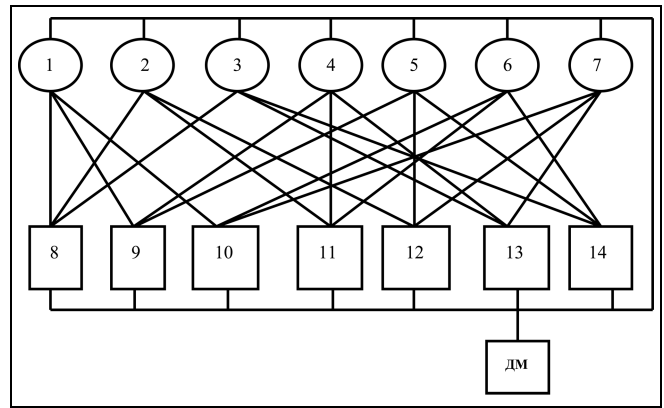


Схема цифровой системы со структурой в виде симметричного двудольного графа $Q_{7,3}$:

1—7 — процессоры; 8—14 — блоки памяти; ДМ — диагностический монитор

ход всех модулей исправной системы. В неоднородных системах функции проверяющих модулей исполняют активные модули (процессоры m_j). Поэтому в неоднородных системах нужно построить цикл, проходящий через все активные модули (процессоры).

Но в системе со структурой в виде симметричного двудольного графа ($Q_{7,3}$ на рисунке) нельзя построить цикл, проходящий только через процессоры, так как между ними нет непосредственных связей. В цикле, выбранном для последующего примера, процессоры, номера которых выделены жирным шрифтом, чередуются с блоками памяти: $5 \rightarrow 12 \rightarrow 7 \rightarrow 10 \rightarrow 6 \rightarrow 14 \rightarrow 3 \rightarrow 13 \rightarrow 4 \rightarrow 11 \rightarrow 2 \rightarrow 8 \rightarrow 1 \rightarrow 9 \rightarrow 5$.

В таблице T_S (табл. 1) показана информация о составе проверяемых подсистем $U_{j,i,k}$ и номерах

Таблица 1

Таблица T_S состава проверяемых подсистем

Строка T_j	Проверяющий модуль	Проверяемые подсистемы $U_{j,i,k}$, оценки результатов их проверок					
	m_j	$U_{j,i1,k1}$	$r_{j,i1,k1}$	$U_{j,i2,k2}$	$r_{j,i2,k2}$	$U_{j,i3,k3}$	$r_{j,i3,k3}$
T_5	5	(5-12), 12, (12-7), 7		(5-9), 9, (9-4), 4		(5-14), 14, (14-6), 6	
T_7	7	(7-10), 10, (10-6), 6		(7-12), 12, (12-2), 2		(7-13), 13, (13-3), 3	
T_6	6	(6-14), 14, (14-3), 3		(6-11), 11, (11-2), 2		(6-10), 10, (10-1), 1	
T_3	3	(3-13), 13, (13-4), 4		(3-14), 14, (14-5), 5		(3-8), 8, (8-2), 2	
T_4	4	(4-11), 11, (11-2), 2		(4-13), 13, (13-3), 3		(4-9), 9, (9-1), 1	
T_2	2	(2-8), 8, (8-1), 1		(2-11), 11, (11-4), 4		(2-12), 12, (12-5), 5	
T_1	1	(1-9), 9, (9-5), 5		(1-8), 8, (8-2), 2		(1-10), 10, (10-7), 7	

проверяющих их модулей m_j , выбранных для последующего примера диагностирования компонентов системы $Q_{7,3}$ (см. рисунок). В таблице семь строк T_j , в каждой из которых для проверяющего модуля m_j ($j = 1, 2, \dots, 7$) записаны номера компонентов, входящих в состав подсистем $U_{j,i1,k1}$, $U_{j,i2,k2}$ и $U_{j,i3,k3}$, проверяемых модулем m_j , в ней также есть поля для записи оценок результатов проверки $r_{j,i1,k1}$, $r_{j,i2,k2}$ и $r_{j,i3,k3}$ этих подсистем модулем m_j . (Отметим, что в клетках строк $T_5, \dots, T_7$ табл. 1 показаны более компактные номера индексов проверяющего модуля и проверяемых подсистем.)

Для отображения местонахождения процессоров в цикле контроля в строке T_j табл. 1 проверяемые подсистемы записаны в таком порядке, что именно процессор m_{k1} , входящий в проверяемую подсистему $U_{j,i1,k1}$, является в построенном цикле соседним справа для проверяющего процессора m_j (после памяти p_{i1}). Например, в строке T_5 процессор m_5 является проверяющим для подсистем $U_{5,12,7}$, $U_{5,9,4}$ и $U_{5,14,6}$, и в следующей строке T_7 в качестве проверяющего модуля записан процессор m_7 , являющийся в цикле контроля соседом справа для процессора m_5 (после памяти p_{12}).

Поскольку задачи, возлагаемые на системы диагностирования рассматриваемых ЦС и неоднородных ЦС, рассмотренных в статье [2], во многом сходны, то в организации системы диагностирования рассматриваемых здесь ЦС можно использовать большинство диагностических процедур, описанных в статье [2]. Поэтому в последующем тексте будем опускать некоторые подробности, заменяя их необходимыми ссылками.

Традиционно процесс контроля запускается диагностическим монитором через сравнительно небольшие интервалы времени, что препятствует накоплению отказавших компонентов в системе.

Назначим в качестве начального модуля процесса контроля один из процессоров, обозначим его m_n .

По сигналу ДМ о начале контроля модуль m_n выполняет следующие действия:

- из таблицы T_5 выделяет строку T_n , находит в ней индексы $(n, i1, k1)$, $(n, i2, k2)$, ... проверяемых подсистем $U_{n,i1,k1}$, $U_{n,i2,k2}$, ...;

- передает процессору m_{k1} (m_{k2} , ...) команду «Выполнить самоконтроль подсистемы $U_{n,i1,k1}$ ($U_{n,i2,k2}$, ...) и передать результаты проверок в модуль m_n ».

Получив эту команду, модуль m_{k1} (m_{k2} , ...) выполняет самоконтроль подсистемы $U_{n,i1,k1}$ ($U_{n,i2,k2}$, ...)

и передает полученные результаты проверок в модуль m_n .

Модуль m_n формирует оценку $r_{n,i1,k1}$ ($r_{n,i2,k2}$, ...), равную 0, если результаты тестов совпали с эталонами, и равную 1 при несовпадении; записывает ее в соответствующие клетки строки T_n , затем анализирует полученные оценки:

- если среди полученных оценок есть как единичные, так и нулевые, то передает процессору одной из подсистем с нулевой оценкой управление для продолжения процесса контроля;

- если тесты самоконтроля всех проверявшихся подсистем дали единичные оценки, то ДМ должен запустить процесс контроля с другого проверяющего модуля.

Модуль m_{k1} (m_{k2} , ...) и каждый из последующих «соседей» по циклу контроля выполняют аналогичные действия в случае получения нулевых оценок. Этот процесс проверок и «хождения» по вершинам выбранного цикла заканчивается, когда управление возвращается к начальному модулю m_n . Если же на одном из этапов контроля (кроме начального) получена хотя бы одна единичная оценка, свидетельствующая об обнаружении одного или нескольких отказавших компонентов, то процесс контроля прерывается, и система приступает к процессу диагностирования отказавших компонентов.

На основе информации, полученной в процессе контроля, ДМ запускает процесс диагностирования с того модуля m_j , который первым получил хотя бы одну единичную оценку при проверке состояний связанных с ним подсистем $U_{j,i1,k1}$, $U_{j,i2,k2}$, ...; естественно, модуль m_j должен быть исправным и проверенным в данном «сеансе» контроля, в частности, он не может быть начальным модулем процесса контроля m_n .

Начальный модуль процесса диагностирования m_j выполняет следующие действия:

- формирует промежуточную таблицу PT_j , переписывая в нее строку T_j с оценками, полученными в процессе контроля (в индексе символа данной таблицы указан индекс проверяющего модуля m_j);

- по нулевым оценкам в строке T_j выделяет подмножество индексов исправных модулей $\{q_j\}$ и одному из этих модулей (например, первому в строке T_j , если он исправен) передает таблицу PT_j и управление процессом диагностирования.

Если среди модулей, проверенных модулем m_j , нет ни одного исправного, то для получения диагностических результатов ДМ должен повторить запуск процесса контроля с другого начального модуля.



Очередной исправный модуль m_v , получив управление процессом диагностирования, выполняет следующие действия:

1) из таблицы состава T_S выделяет строку T_v , переписывает ее в таблицу PT_j , полученную от предшественника — модуля m_j (m_{v-1});

2) в строке T_v выделяет индексы $(v, i1, k1)$, $(v, i2, k2)$... проверяемых подсистем $U_{v,i1,k1}$, $U_{v,i2,k2}$, ..., выдает команды на самоконтроль этих подсистем, получает результаты этих проверок и в клетки $r_{v,i1,k1}$, $(r_{v,i2,k2}, \dots)$, ... строки T_v записывает оценки результатов выполненных тестов;

3) по нулевым оценкам проверок в строке T_v выделяет подмножество $\{q_v\}$ индексов исправных «преемников» — проверенных модулей, сравнивает индексы из подмножества $\{q_v\}$ с индексами строк в таблице PT_j ;

— если один индекс (или больше) из подмножества $\{q_v\}$ отсутствует в списке индексов строк в таблице PT_j , то передает модулю с этим индексом таблицу PT_j и управление процессом диагностирования;

— если список $\{q_v\}$ пуст или все индексы $\{q_v\}$ — в списке индексов строк в таблице PT_j , то переход к п. 4;

4) находит один из исправных модулей, еще не исполнявших функции проверяющего, и передает ему таблицу PT_j и управление диагностированием.

Для нахождения одного из исправных модулей, еще не исполнявших функций проверяющего модуля, модуль m_v находит индексы $\{q_{v*}\}$ исправных модулей строки T_{v*} и сравнивает их с индексами строк в таблице PT_j . Если при сравнении списков индексов окажется, что

— один индекс (или больше) из подмножества $\{q_{v*}\}$ отсутствует в списке индексов строк в таблице PT_j , то модуль m_v передает модулю с этим индексом $(v+1)$ таблицу PT_j и управление диагностированием;

— список $\{q_{v*}\}$ пуст или все индексы $\{q_{v*}\}$ — в списке индексов строк в таблице PT_j , то модуль m_v переходит к анализу предыдущей строки T_{v*-1} .

Эти действия выполняются поочередно снизу вверх, начиная с нижней строки T_v . При нахождении исправного модуля m_{v+1} в строке T_{v*} управление диагностированием и таблица PT_j передаются этому модулю снизу вверх через проверяющие модули $m_v, \dots, m_{v*}$. Если при достижении верхней строки T_j окажется, что и в этой строке нет исправ-

ного модуля, не исполнявшего функций проверяющего модуля, то поиск прекращается, и модуль m_v пересылает в ДМ таблицу PT_j для обработки полученных результатов проверок.

Процедура обработки полученных результатов проверок достаточно полно изложена в работе [2]. Поэтому в данном тексте эта процедура будет показана на примере.

4. ПРИМЕР

Особенности разработанной организации диагностирования покажем на примере ЦС со структурой симметричного двудольного графа отказов (система $Q_{7,3}$ на рисунке).

В цикле, выбранном для данного примера, процессоры, номера которых выделены жирным шрифтом, чередуются с блоками памяти: $5 \rightarrow 12 \rightarrow 7 \rightarrow 10 \rightarrow 6 \rightarrow 14 \rightarrow 3 \rightarrow 13 \rightarrow 4 \rightarrow 11 \rightarrow 2 \rightarrow 8 \rightarrow 1 \rightarrow 9 \rightarrow 5$. Таблица состава T_S , предназначенная для организации тестирования в системе $Q_{7,3}$, приведена в табл. 1.

Предположим, что из-за большого интервала между циклами контроля в рассматриваемой ЦС накопились отказы двух модулей 6 и 8 и двух линий связи (4—13) и (5—12).

По сигналу ДМ о начале контроля начальный модуль 5 выполнит следующие действия: из таблицы T_S (см. табл. 1) выделит строку T_5 , найдет в ней индексы (5, 12, 7) (5, 9, 4), (5, 14, 6) проверяемых подсистем $U_{5,12,7}$, $U_{5,9,4}$, $U_{5,14,6}$, передаст модулю 7 (4, 6) команду «Выполнить самоконтроль подсистемы $U_{5,12,7}$ ($U_{5,9,4}$, $U_{5,14,6}$) и вернуть результаты проверок в модуль 5».

Получив эту команду, модуль 7 (4, 6) выполнит самоконтроль подсистемы $U_{5,12,7}$ ($U_{5,9,4}$, $U_{5,14,6}$) и передаст полученные результаты проверок модулю 5.

Модуль 5 формирует оценки $r_{5,12,7} = 1$, $r_{5,9,4} = 0$, $r_{5,14,6} = 1$, записывает их в соответствующие клетки строки T_5 (для сокращения числа таблиц будем записывать результаты проверок, находящиеся в разных модулях, в общую табл. 2), анализирует полученные оценки: так как среди полученных оценок есть две единичные и одна нулевая, то передает модулю 4 подсистемы $U_{5,9,4}$ управление для продолжения процесса контроля. (В строке T_1 и в последующих полужирным шрифтом в результатах проверки выделен 0 того исправного модуля, к которому передается управление для продолжения процесса контроля или диагностирования.)

Модуль 4 выполнит аналогичные действия:

— передаст в подсистемы $U_{4,13,3}$, $U_{4,11,6}$, $U_{4,9,1}$ команду «Выполнить самоконтроль и вернуть результаты проверок в модуль 4»;

— получит оценки результатов контроля $r_{4,13,3} = 1$, $r_{4,11,6} = 1$, $r_{4,9,1} = 0$, проанализирует их: так как среди полученных оценок есть единичные ($r_{4,13,3} = 1$, $r_{4,11,6} = 1$), то заведомо исправный модуль 4 перейдет к процессу диагностирования отказавших компонентов и выполнит следующие действия:

— образует промежуточную таблицу PT_4 и переписывает в нее строку T_4 с полученными оценками результатов проверок;

— по нулевым оценкам в строке T_4 выделит индекс исправной подсистемы $U_{4,9,1}$, в котором найдет номер исправного модуля $\{q_4 = 1\}$, которому передаст таблицу PT_4 и управление процессом диагностирования.

В свою очередь, модуль 1 выполнит следующие действия:

— из таблицы T_5 (табл. 1) выделит строку T_1 , переписывает ее в таблицу PT_4 , в строке T_1 найдет индексы проверяемых подсистем $U_{1,8,2}$, $U_{1,9,5}$ и $U_{1,10,7}$, в которые передаст команду «Выполнить самоконтроль и вернуть результаты проверок в модуль 1»;

— получит оценки результатов контроля $r_{1,8,2} = 1$, $r_{1,9,5} = 0$, $r_{1,10,7} = 0$;

— по нулевым оценкам в строке T_1 выделит индексы исправных подсистем $U_{1,9,5}$ и $U_{1,10,7}$, в которых найдет номера исправных модулей $\{q_1 = 5, 7\}$ и первому из них (модулю 5) передаст таблицу PT_4 и управление процессом диагностирования.

Модуль 5 вместо проверки подсистем $U_{5,12,7}$, $U_{5,9,4}$ и $U_{5,14,6}$ выберет из строки T_5 результаты их проверок (на этапе контроля) $r_{5,12,7} = 1$, $r_{5,9,4} = 0$, $r_{5,14,6} = 1$ и определит, что исправен модуль 4. Но так как модуль 4 уже есть в числе проверяющих модулей на этапе диагностирования, то модуль 5 должен искать исправный модуль для исполнения функций проверяющего модуля среди ранее проверенных. В данном случае в предыдущей строке T_1 есть нулевые оценки проверок модулей 5 и 7. Таким образом, модуль 5 через модули 9, 1 и 10

должен передать модулю 7 таблицу PT_4 и управление дальнейшим процессом диагностирования.

Модуль 7 выполнит необходимые проверки подсистем $U_{7,10,6}$, $U_{7,12,2}$ и $U_{7,13,3}$, определит, что среди его «преемников» есть два исправных модуля: 2 и 3, и первому из них (модулю 2) передаст таблицу PT_4 и управление дальнейшим процессом диагностирования. В свою очередь, модуль 2 после проверок своих «преемников» — подсистем $U_{2,11,4}$, $U_{2,8,1}$ и $U_{2,12,5}$ выяснит, что в этих подсистемах один исправный модуль 4, который уже есть в числе проверяющих модулей на этапе диагностирования. Поэтому, как и модуль 5, модуль 2 должен искать исправный модуль для исполнения функций проверяющего модуля среди ранее проверенных. В данном случае в предыдущей строке T_7 есть нулевые оценки проверок модулей 2 и 3. Таким образом, модуль 2 через модули 12, 7 и 13 должен передать модулю 3 таблицу PT_4 и управление дальнейшим процессом диагностирования.

В свою очередь, модуль 3 после проверок своих «преемников» — подсистем $U_{3,14,5}$, $U_{3,13,4}$ и $U_{3,8,2}$ выяснит, что исправный модуль 5 уже есть в числе проверяющих модулей на этапе диагностирования. Поэтому, как и модули 5 и 2, модуль 3 будет искать исправный модуль для исполнения функций проверяющего модуля. Не обнаружив такого модуля в расположенных выше строках таблицы PT_4 , так как все исправные модули уже исполняли подобные функции, модуль 3 закончит процесс диагностирования и перешлет в ДМ таблицу PT_4 для обработки полученных результатов проверок.

Для обработки и дешифрации результатов ДМ переписывает их из таблицы PT_4 в заранее заготовленную

Таблица 2

Информация к Примеру

Строка T_j	Проверяющий модуль	Проверяемые подсистемы $U_{j,i,k}$, оценки их проверок					
	m_j	$U_{j,i1,k1}$	$r_{j,i1,k1}$	$U_{j,i2,k2}$	$r_{j,i2,k2}$	$U_{j,i3,k3}$	$r_{j,i3,k3}$
Строки с результатами контроля							
T_5	5	(5-12), 12, (12-7), 7	1	(5-9), 9, (9-4), 4	0	(5-14), 14, (14-6), 6	1
T_4	4	(4-13), 13, (13-3), 3	1	(4-11), 11, (11-6), 6	1	(4-9), 9, (9-1), 1	0
Таблица РТ ₄							
T_4	4	(4-13), 13, (13-3), 3	1	(4-11), 11, (11-6), 6	1	(4-9), 9, (9-1), 1	0
T_1	1	(1-8), 8, (8-2), 2	1	(1-9), 9, (9-5), 5	0	(1-10), 10, (10-7), 7	0
T_5	5	(5-12), 12, (12-7), 7	1	(5-9), 9, (9-4), 4	0	(5-14), 14, (14-6), 6	1
T_7	7	(7-10), 10, (10-6), 6	1	(7-12), 12, (12-2), 2	0	(7-13), 13, (13-3), 3	0
T_2	2	(2-11), 11, (11-4), 4	0	(2-8), 8, (8-1), 1	1	(2-12), 12, (12-5), 5	1
T_3	3	(3-14), 14, (14-5), 5	0	(3-13), 13, (13-4), 4	1	(3-8), 8, (8-2), 2	1



диагностическую таблицу T_D [2]. В табл. 3, а и 3, б представлены две части таблицы T_D после обработки результатов из таблицы PT_4 (см. табл. 2). Поясним структуру этих таблиц. В клетках левого столбца перечислены компоненты проверяемых подсистем $U_{j,i,k}$, взятые из соответствующих клеток таблицы состава T_S (см. табл. 1), а клетки следующего столбца предназна-

чены для записи оценок $r_{j,i,k}$ проверок этих подсистем, полученных в процессе диагностирования. Последующие столбцы T_D обозначены символами одиночных компонентов системы $Q_{7,3}$: процессоры — символами u с номерами 1 — 7, блоки памяти — символами l с номерами 8 — 14, линии связи — символами l с номерами начальной и конечной вершин. Нетрудно заметить,

Таблица 3а

Диагностическая таблица T_D

Проверяемые подсистемы $U_{j,i,k}$	$r_{j,i,k}$	Компоненты системы																	
		$u5$	$l5-12$	$u12$	$l12-7$	$u7$	$l7-10$	$u10$	$l10-6$	$u6$	$u8$	$l1-10$	$l1-8$	$u2$	$l2-8$	$u11$	$l11-2$	$l11-4$	$u4$
(5-12), 12, (12-7), 7	1		1	1	1	1													
(5-9), 9, (9-4), 4	0																		0
(5-14), 14, (14-6), 6	1								1	1									
(7-10), 10, (10-6), 6	1						1	1	1	1									
(7-12), 12, (12-2), 2	0			0	0									0					
(7-13), 13, (13-3), 3	0																		
(6-14), 14, (14-3), 3	—																		
(6-11), 11, (11-2), 2	—													*		*	*		
(6-10), 10, (10-1), 1	—							*	*			*							
(3-13), 13, (13-4), 4	1																		1
(3-14), 14, (14-5), 5	0	0																	
(3-8), 8, (8-2), 2	1										1			1	1				
(4-11), 11, (11-2), 2	0													0		0	0	0	
(4-13), 13, (13-3), 3	1																		
(4-9), 9, (9-1), 1	0																		
(2-8), 8, (8-1), 1	1										1		1		1				
(2-11), 11, (11-4), 4	0															0	0	0	0
(2-12), 12, (12-5), 5	1	1	1	1															
(1-9), 9, (9-5), 5	0	0																	
(1-8), 8, (8-2), 2	1										1		1	1	1				
(1-10), 10, (10-7), 7	0					0	0	0				0							
A_0		0	—	0	0	0	0	0	—	—	—	0	—	0	—	0	0	0	0
A_1		—	1	—	—	—	—	—	—	1	1	—	—	—	—	—	—	—	—
A_Z		—	—	—	—	—	—	—	Z	—	—	—	Z	—	Z	—	—	—	—



что в столбце $U_{j,i,k}$ проверяемых подсистем каждая линия показана дважды: в прямом и инверсном направлении, например (5—12) и (12—5), (12—7) и (7—12) и т. д. Однако в списке отдельных компонентов каждая из линий записана один раз для сокращения ширины табл. 3, а и 3, б.

Вхождение компонента u_k (l_{j-i}) в состав подсистемы $U_{j,i,k}$, приведенной в левом столбце таблицы T_D , отме-

чается знаком «*», поставленным в клетку на пересечении столбца u_k (l_{j-i}) и строки $U_{j,i,k}$. (Так как в табл. 3, а и 3, б показаны результаты обработки информации из PT_4 , то знаки «*» остались только в тех трех строках, для которых нет оценок проверок этих подсистем.)

В нижней части таблицы T_D находятся три строки A_0 , A_1 и A_Z , предназначенные для записи результатов обработки полученной информации.

Таблица 3б

Диагностическая таблица T_D

Проверяемые подсистемы $U_{j,i,k}$	$r_{j,i,k}$	Компоненты системы																
		$u13$	$l4-13$	$u3$	$l6-11$	$l3-14$	$u14$	$l14-5$	$l3-8$	$u9$	$l9-1$	$l3-13$	$u1$	$l4-9$	$l5-9$	$l6-14$	$l2-12$	$l7-13$
(5-12), 12, (12-7), 7	1																	
(5-9), 9, (9-4), 4	0									0				0	0			
(5-14), 14, (14-6), 6	1						1	1								1		
(7-10), 10, (10-6), 6	1																	
(7-12), 12, (12-2), 2	0																0	
(7-13), 13, (13-3), 3	0	0		0								0						0
(6-14), 14, (14-3), 3	—			*		*	*									*		
(6-11), 11, (11-2), 2	—				*													
(6-10), 10, (10-1), 1	—											*						
(3-13), 13, (13-4), 4	1	1	1									1						
(3-14), 14, (14-5), 5	0					0	0	0										
(3-8), 8, (8-2), 2	1								1									
(4-11), 11, (11-2), 2	0																	
(4-13), 13, (13-3), 3	1	1	1	1								1						
(4-9), 9, (9-1), 1	0									0	0		0	0				
(2-8), 8, (8-1), 1	1												1					
(2-11), 11, (11-4), 4	0																	
(2-12), 12, (12-5), 5	1																1	
(1-9), 9, (9-5), 5	0									0	0				0			
(1-8), 8, (8-2), 2	1												1					
(1-10), 10, (10-7), 7	0																	
A_0		0	—	0	—	0	0	0	—	0	0	0	0	0	0	—	0	0
A_1		—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
A_Z		—	—	—	Z	—	—	—	Z	—	—	—	—	—	—	Z	—	—



Обработка записей из таблицы PT_4 заключается в следующем:

- из строк T_4 , T_1 и T_3 таблицы PT_4 выделены оценки результатов проверок $r_{4,13,3} = 1$, $r_{4,11,6} = 1$, $r_{4,9,1} = 0$, ..., $r_{3,14,5} = 0$, $r_{3,13,4} = 1$, $r_{3,8,2} = 1$ и записаны в соответствующие клетки столбца $r_{j,i,k}$ табл. 3, а и 3, б, три клетки этого столбца заполнены прочерками «—»; в этих строках и сохранились знаки «*»;
- значения 0 и 1, записанные в столбце $r_{j,i,k}$, переписаны в клетки той же строки $U_{j,i,k}$, отмеченные знаком «*»;
- заполнены прочерками «—» клетки строк A_0 , A_1 , A_Z ;
- после этого в клетки этих строк записаны результаты диагностирования в соответствии со следующими правилами:
 - в k -ю клетку строки A_0 записан нуль, если в вышележащих клетках этого столбца есть хотя бы один нуль;
 - в клетки на пересечении строки A_1 и столбцов, озаглавленных символами вершин u_6 и u_8 , записаны единицы, так как в вышележащих клетках этих столбцов есть хотя бы одна единица и нет ни одного нуля;
 - в клетки на пересечении строки A_1 и столбцов, озаглавленных символами дуг $l(5-12)$, $l(4-13)$ записаны единицы, так как в этих столбцах нет ни одного нуля и каждая из вершин, которые связывают эти линии, исправна (в соответствующих клетках строки A_0 записан нуль),
 - в шести клетках строки A_Z , озаглавленных символами дуг $l(j-i)$, записан символ Z , так как хотя бы одна из вершин (j или i), которые связывает каждая из этих линий, подозревается в отказе (в соответствующей клетке строки A_1 записана единица);
- расшифровка полученных записей в строках A_0 , A_1 и A_Z дает следующие диагностические заключения:
 - исправны 25 компонентов, номера которых определяются местоположением нулевых записей в строке A_0 ;
 - в соответствии с местоположением единиц в строке A_1 подозреваются в отказе процессор 6, блок памяти 8, линии связи (5—12) и (4—13), что совпадает с номерами компонентов в исходной отказовой ситуации;
 - в соответствии с местоположением символов Z в строке A_Z остались неопределенными состояния линий (10—6), (1—8), (2—8), (6—11), (3—8), (6—14), так как их проявления «маскируются» неисправным состоянием одной из вершин (6 или 8), которые они соединяют.

Таким образом, идентифицированы состояния 29-ти из 35-ти компонентов системы $Q_{7,3}$.

ЗАКЛЮЧЕНИЕ

Особенность цифровых систем (ЦС) со структурой в виде симметричного двудольного графа состоит в наличии компонентов разных типов, на-

пример, процессоров и блоков памяти. Неоднородность компонентов ЦС подобной структуры не позволяет напрямую применять методы системного диагностирования, разработанные для однородных ЦС.

Предложен способ диагностирования компонентов ЦС со структурой в виде симметричного двудольного графа, в основе которого лежит объединение связанных разнотипных устройств на этапах контроля и диагностирования в проверяемые подсистемы одинакового состава. С учетом полученной таким образом одинаковости состава проверяемых подсистем разработана организация диагностирования отказавших компонентов в ЦС подобной структуры.

Показан пример организации диагностирования отказавших компонентов в системе, состоящей из семи процессоров и семи блоков памяти; требуемая точность диагностирования — до отдельного компонента ЦС: процессор, память и линия связи между ними. В данном примере получен правильный результат: диагноз совпадает с исходной отказовой комбинацией из четырех компонентов. Но можно подобрать такие сочетания неисправных компонентов, диагноз которых даст неточный или неправильный результат. Работы по определению меры диагностируемости (числа правильно диагностируемых отказавших компонентов, расположенных произвольным образом) для ЦС подобной структуры неизвестны.

ЛИТЕРАТУРА

1. Каравай М.Ф., Пархоменко П.П., Подлазов В.С. Комбинаторные методы построения двудольных однородных минимальных квазиполных графов (симметричных блок-схем) // Автоматика и телемеханика. — 2009. — № 2. — С. 153—170.
2. Ведешенков В.А. Способ самодиагностирования неоднородных цифровых систем // Проблемы управления. — 2005. — № 4. — С. 33—40.
3. Barsi F., Grandoni F., Maestrini P. A theory of diagnosability of digital systems // IEEE Trans. Comput. — 1976. — Vol. C-25, № 6. — P. 585—593.
4. Пархоменко П.П. Определение технического состояния многопроцессорных вычислительных систем путем анализа графа синдромов // Автоматика и телемеханика. — 1999. — № 5. — С. 126—135.
5. Ведешенков В.А. Организация самодиагностирования технического состояния цифровых систем // Автоматика и телемеханика. — 2003. — № 11. — С. 165—182.

Статья представлена к публикации членом редколлегии чл.-корр. РАН П.П. Пархоменко.

Ведешенков Виктор Алексеевич — д-р техн. наук, гл. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН,
☎(495) 335-75-90, ✉vva@ipu.ru.

ОПТИМИЗАЦИЯ ОДНОРОДНЫХ НЕМАРКОВСКИХ СЕТЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ

В.Н. Задорожный

Предложен новый эффективный аналитико-имитационный метод оптимизации немарковских сетей массового обслуживания. Экспериментально оценена скорость сходимости и точность метода. Даны практические рекомендации по его применению.

Ключевые слова: сеть массового обслуживания, оптимизация, аналитико-имитационное моделирование.

ВВЕДЕНИЕ

Производительность организационно-технических систем, предназначенных для обработки или обслуживания дискретных потоков каких-либо однотипных единиц (заявок), часто оценивается по времени прохождения заявок через эти системы. Унифицированным формализованным представлением подобных систем выступает сеть массового обслуживания (СеМО) со статистически однородными заявками — однородная сеть [1].

В виде СеМО традиционно представляются, например, информационно-вычислительные системы [1—5]. Заявки рассматриваются как передаваемые сообщения или как пользовательские запросы, обрабатываемые ресурсами системы. Время прохождения заявки через СеМО будем называть временем ответа. Среднее время ответа E зависит от распределения имеющихся ресурсов сети между ее узлами.

В последние годы успешно развиваются методы оптимизации марковских [1, 2] СеМО — сетей с экспоненциальными распределениями времени обслуживания заявок и (если сеть не замкнута) с пуассоновскими входными потоками. Задача оптимального распределения ресурса замкнутой марковской сети сводится к системе нелинейных уравнений, эффективно решаемой численными методами [1]. Решается задача оптимизации замкнутых марковских сетей с несколькими классами сообщений [6]. В работах [1, 7] для немарковских сетей рассматриваются методы приближенного расчета, основанные на аппроксимации произвольных распределений распределениями, допускающими рациональное преобразование Лапласа. Однако этот путь сопряжен со значительными вычислительными трудностями даже для одновариантного расчета сетей с небольшим числом узлов (особенно если приходится аппроксимировать

распределения случайных величин, ограниченных узкими диапазонами возможных значений [8]). В работе [9] предпринята попытка решения задач оптимизации немарковских СеМО путем аппроксимации их узлов аналитическими выражениями, учитывающими первые два момента времени обслуживания и интервалов поступления заявок в узлы, с учетом межузловых взаимодействий. Однако до решения оптимизационных задач эта работа не доводится.

Таким образом, в общем случае для оптимизации немарковских сетей приходится использовать имитационное моделирование (ИМ). При этом если задача оптимизации содержит более двух-трех независимых переменных, то ее решение становится практически невозможным без привлечения градиентных методов. Но расчет градиентов в ИМ существенно затрудняется стохастической по-

решностью вычисляемых оценок $\hat{E}$ отклика E [7, 10]. Известные методы [7, 11—13] решения этой проблемы либо пригодны только при ИМ изолированных систем массового обслуживания [7, 12, 13] и не распространяются на СеМО, либо их теоретически возможное применение к СеМО на практике приводит к значительным трудностям. Например, применению методов, опирающихся на имитацию большого числа периодов регенерации [11—14] препятствует то обстоятельство, что в СеМО, как правило, эти периоды бывают практически бесконечными. Успешные применения подобных методов ограничиваются классами СеМО, учитывающими специфику конкретных сетевых объектов при конкретных диапазонах их параметров [3].

Для задачи оптимального распределения ресурса по узлам однородной немарковской СеМО в статье предлагается метод, позволяющий эффективно решать проблему градиентов путем приме-



нения простой сепарабельной аппроксимации целевой функции.

Далее формулируется решаемая задача.

1. ЗАДАЧА ОПТИМИЗАЦИИ ОДНОРОДНОЙ НЕМАРКОВСКОЙ СЕТИ

Рассмотрим сначала *открытую* сеть, в которую поступает рекуррентный поток заявок с интенсивностью Λ . Интервалы поступления заявок — независимые случайные величины (сл. в.) с функцией распределения вероятностей (ф. р. в.) $A(t)$. Заявка из входного потока сети с вероятностью p_{0i} попадает в i -й узел, $i = \overline{1, n}$. В любом из K_i каналов i -го узла время обслуживания заявки (независимая сл. в.) имеет ф. р. в. $B_i(t)$. После обслуживания в i -м узле заявка случайно и независимо, в соответствии с заданными переходными вероятностями p_{ij} , выбирает один из узлов j для продолжения своего маршрута или, с вероятностью p_{i0} , уходит из сети. Вероятности p_{ij} ($i, j = \overline{0, n}$) задаются неразложимой стохастической матрицей $\mathbf{P} = \|p_{ij}\|$.

Стационарное среднее время E прохождения заявки через сеть (среднее время ответа) можно представить в виде:

$$E = \sum_{i=1}^n \alpha_i u_i = \sum_{i=1}^n \alpha_i (w_i + b_i) = \sum_{i=1}^n \alpha_i \left(w_i + \frac{1}{\mu_i} \right), \quad (1)$$

где α_i — среднее число посещений i -го узла заявкой за время ее прохождения через сеть, u_i — среднее время пребывания заявки в i -м узле, w_i — среднее время ожидания заявки в очереди i -го узла, b_i — среднее время обслуживания заявки в i -м узле, $\mu_i = b_i^{-1}$ — интенсивность обслуживания заявки каналом i -го узла.

Коэффициенты α_i однозначно определяются из системы уравнений баланса:

$$\alpha_i = \sum_{j=0}^n \alpha_j p_{ji}, \quad i = \overline{0, n}, \quad \alpha_0 \equiv 1.$$

Через коэффициенты α_i последовательно определяются интенсивности $\lambda_i = \Lambda \cdot \alpha_i$ входных потоков узлов, их коэффициенты загрузки $\rho_i = \lambda_i / (\mu_i K_i)$, и проверяются условия стационарности $\rho_i \leq 1$ (или $\mu_i \geq \lambda_i / K_i$), $i = \overline{1, n}$. Значения w_i в формуле (1) определяются посредством ИМ.

Если сеть *замкнута*, то входной поток заявок отсутствует (параметр Λ не задается), и в сети циркулирует заданное постоянное число заявок R . Завершением цикла обслуживания заявки считается

ее переход по некоторой «терминальной» дуге с «выхода» на «вход» сети. Среднее время ответа E замкнутой сети (среднее время цикла обслуживания) также определяется формулой (1) и может вычисляться посредством ИМ.

Теперь рассмотрим следующую обобщенную версию сформулированной в работе [1] задачи оптимизации однородной замкнутой экспоненциальной (марковской) сети.

Стоимость (ресурс) M однородной сети как функция вектора $\vec{\mu} = (\mu_1, \dots, \mu_n)$ интенсивностей обслуживания в узлах $i = \overline{1, n}$ задается в виде

$M(\vec{\mu}) = \sum_{i=1}^n c_i \mu_i^{\beta_i}$, где c_i — стоимостные коэффициенты, $\beta_i > 0$ — коэффициенты нелинейности. Требуется найти вектор $\vec{\mu} = \vec{\mu}_{\text{opt}}$, доставляющий минимум функции $E = E(\vec{\mu})$:

$$E(\vec{\mu}) = \sum_{i=1}^n \alpha_i \left(w_i(\vec{\mu}) + \frac{1}{\mu_i} \right) \rightarrow \min_{\vec{\mu}}, \quad (2)$$

и принадлежащий следующей области допустимых решений (ОДР):

$$M(\vec{\mu}) = \sum_{i=1}^n c_i \mu_i^{\beta_i} = M^*, \quad \mu_i \geq \mu_{\min}, \quad i = \overline{1, n}, \quad (3)$$

где для открытой сети $\mu_{\min} = \lambda_i / K_i$ (граница области стационарности), а для замкнутой $\mu_{\min} = 0$. Для ресурса M^* в выражении (3) должно выполняться условие $M^* \geq M_{\min}$, где для открытой сети $M_{\min} = \sum_{i=1}^n c_i \mu_{\min}^{\beta_i} = \sum_{i=1}^n c_i (\lambda_i / K_i)^{\beta_i}$, а для замкнутой $M_{\min} = 0$.

В задаче (2), (3) имеется в виду, что изменение любой интенсивности μ_i приводит к изменению среднего $b_i = \mu_i^{-1}$ и к соответствующему масштабному изменению ф. р. в. $B_i(t)$. Вид ф. р. в. $B_i(t)$ не изменяется, поскольку ассоциируется со случайной трудоемкостью заявок, тогда как варьируемый параметр μ_i определяется производительностью каналов i -го узла.

2. ОБЩАЯ СТРУКТУРА И ОПОРНЫЕ ЭЛЕМЕНТЫ МЕТОДА «НАПРАВЛЯЮЩИХ ГИПЕРБОЛ»

Предлагаемый метод решения задачи (2), (3) состоит из двух этапов.

Этап I: ускоренный градиентный поиск точки $\vec{\mu}_{\text{opt}}$ методом «направляющих гипербол» (НГ), использующим ИМ сети и сепарабельную аппроксимацию целевой функции (название метода отражает роль и вид функций одного переменного — слагаемых этой аппроксимации).

Этап II (не обязательный): уточнение найденного решения методом циклического покоординатного спуска [15] (модифицированным), не использующим аппроксимаций.

Предлагаемый в настоящей статье метод решения задачи (2), (3) использует уточненную версию предложенной в работе [16] аппроксимации функции $E(\vec{\mu})$. Аппроксимация $E^{ap}(\vec{\mu})$ среднего времени ответа $E(\vec{\mu})$ на каждой итерации $k \geq 2$ поиска оптимального решения $\vec{\mu}_{opt}$ формируется по результатам ИМ сети в точках $\vec{\mu} = \vec{\mu}^{k-1}$ и $\vec{\mu} = \vec{\mu}^k$ и применяется для определения следующей точки $\vec{\mu} = \vec{\mu}^{k+1}$.

Определим опорные элементы метода НГ.

Центр $\vec{\mu}_c$ ОДР (3) при ресурсе $M^* > M_{min}$ определяется условием равной загрузки узлов: $\rho_i = \lambda_i / (\mu_i K_i) = \alpha_i \lambda_0 / (\mu_i K_i) = \rho_c$, $i = \overline{1, n}$, где λ_0 — интенсивность на терминальной дуге (для открытой сети $\lambda_0 = \Lambda$). Отсюда $\mu_i = (\alpha_i \lambda_0) / (K_i \rho_c)$, $\mu_i / \mu_1 = (\alpha_i / K_i) (K_1 / \alpha_1)$ и $\mu_i = (\alpha_i / K_i) (K_1 / \alpha_1) \mu_1$. Подставляя последнее выражение μ_i в формулу (3), имеем:

$$\sum_{i=1}^n c_i \left(\frac{\alpha_i}{K_i} \cdot \frac{K_1}{\alpha_1} \cdot \mu_1 \right)^{\beta_i} = M^*, \quad \mu_i = (\alpha_i / K_i) (K_1 / \alpha_1) \mu_1, \quad i = \overline{2, n}. \quad (4)$$

Отсюда численными методами легко находится единственный положительный корень μ_1 , определяющий все остальные координаты μ_i центра $\vec{\mu}_c$ ОДР. Если сеть открытая, то в центре $\vec{\mu}_c$ сразу определяются все $\rho_i = \rho_c \leq 1$, $i = \overline{1, n}$.

Если все $\beta_i = 1$, то из формулы (4) координаты центра ОДР можно выразить явно:

$$\mu_i = M^* (\alpha_i / K_i) \left(\sum_{j=1}^n c_j \alpha_j / K_j \right)^{-1}, \quad i = \overline{1, n}.$$

Открытая сеть имеет в центре $\vec{\mu}_c$ максимум пропускной способности $V(\vec{\mu})$, определяемой для нее как $V(\vec{\mu}) = \max\{\Lambda: \rho_i = \alpha_i \Lambda / (\mu_i K_i) \leq 1, i = \overline{1, n}\}$.

Диаметр D ОДР определим как длину максимального из диапазонов варьирования переменных μ_i : $D = \max\{l_i\}$, где $l_i = \mu_{imax} - \mu_{imin}$ и, согласно

формуле (3), $\mu_{imax} = \left[\left(M^* - \sum_{j \neq i} c_j \mu_{jmin}^{\beta_j} \right) c_i^{-1} \right]^{1/\beta_i}$, $i = \overline{1, n}$.

Малый шаг размером, например, $D \cdot 10^{-4}$, будем применять при построении и сканировании траекторий на поверхности ограничений, определяемой уравнением (3). Шаг выбирается с учетом требований к точности оптимизации.

Аппроксимация $E^{ap}(\vec{\mu})$ целевой функции $E(\vec{\mu})$, используемая для приближенного вычисления градиента, представляет собой сепарабельную функцию варьировемых переменных μ_i :

$$E^{ap}(\vec{\mu}) = \sum_{i=1}^n \alpha_i \left(W_i(\mu_i) + \frac{1}{\mu_i} \right), \quad \text{где } W_i(\mu_i) = \begin{cases} \frac{R_i}{\mu_i - S_i}, & \text{если } \hat{w}_i^k \neq \hat{w}_i^{k-1}, \\ \hat{w}_i^k, & \text{если } \hat{w}_i^k = \hat{w}_i^{k-1}, \end{cases} \quad (5)$$

и на каждом шаге k оптимизации заново настраивается (посредством коэффициентов R_i и S_i) по оценкам $\hat{w}_i^{k-1}$ и $\hat{w}_i^k$ среднего времени ожидания, найденным для узлов $i = \overline{1, n}$ с помощью ИМ сети в точках $\vec{\mu} = \vec{\mu}^{k-1}$ и $\vec{\mu} = \vec{\mu}^k$. При $\hat{w}_i^k \neq \hat{w}_i^{k-1}$ выражение $R_i / (\mu_i - S_i)$ в формуле (5) аппроксимирует соответствующую функцию $w_i(\vec{\mu})$ в выражении (2) так, что его значение в точках $\vec{\mu} = \vec{\mu}^{k-1}$ и $\vec{\mu} = \vec{\mu}^k$ совпадает с оценками $\hat{w}_i^{k-1} \approx w_i(\vec{\mu}^{k-1})$ и $\hat{w}_i^k \approx w_i(\vec{\mu}^k)$. Таким образом, имеем $R_i / (\mu_i^{k-1} - S_i) = \hat{w}_i^{k-1}$ и $R_i / (\mu_i^k - S_i) = \hat{w}_i^k$, откуда

$$S_i = \frac{\hat{w}_i^k \mu_i^k - \hat{w}_i^{k-1} \mu_i^{k-1}}{\hat{w}_i^k - \hat{w}_i^{k-1}}, \quad R_i = \hat{w}_i^{k-1} (\mu_i^{k-1} - S_i), \quad i = \overline{1, n}, \quad (6)$$

(верхний индекс здесь везде соответствует шагу оптимизации).

При $\hat{w}_i^k = \hat{w}_i^{k-1}$ расчет значений R_i и S_i в формуле (5) не нужен, но для определения *валидной части* $[L]$ (см. далее) полагаем $S_i \rightarrow \infty$. Такая настройка функции $E^{ap}(\vec{\mu})$ обеспечивает ее совпадение с целевой функцией $E(\vec{\mu})$ в точках $\vec{\mu}^{k-1}$ и $\vec{\mu}^k$ (с точностью до стохастической погрешности оценок ИМ). В других точках $\vec{\mu}$ точность аппроксимации $E^{ap}(\vec{\mu})$ тем хуже, чем они дальше от точек $\vec{\mu}^{k-1}$ и $\vec{\mu}^k$, и чем «менее сепарабельна» $E(\vec{\mu})$, т. е. чем сильнее изменение интенсивностей μ_i в одних узлах влияет на среднее время w_j в других ($i \neq j$).

Градиент $\nabla E^{ap}(\vec{\mu})$ в точке $\vec{\mu} = \vec{\mu}^k$ есть аппроксимация градиента $\nabla E(\vec{\mu})$ в этой точке, и вычисляется с помощью выражения, полученного дифференцированием выражения (5):

$$\nabla E^{ap}(\vec{\mu}^k) = \left(\alpha_1 \frac{\partial W_1}{\partial \mu_1} - \frac{\alpha_1}{(\mu_1)^2}, \dots, \alpha_n \frac{\partial W_n}{\partial \mu_n} - \frac{\alpha_n}{(\mu_n)^2} \right), \quad (7)$$



$$\text{где } \frac{\partial W_i}{\partial \mu_i} = \begin{cases} \frac{-R_i}{(\mu_i^k - S_i)^2}, \hat{w}_i^k \neq \hat{w}_i^{k-1}, \\ 0, \hat{w}_i^k = \hat{w}_i^{k-1}. \end{cases} \quad i = \overline{1, n}.$$

Валидная часть $[L]$ исходящей из точки $\vec{\mu}^k$ траектории L поиска точки $\vec{\mu}^{k+1}$ лежит между $\vec{\mu}^k$ и первой на L точкой $\vec{\mu}$, у которой какая-либо координата μ_i достигает границы ОДР $\mu_i = \mu_{i\min}$ или полюса $\mu_i = S_i$ аппроксимации (5).

Приведем пошаговое описание метода НГ с расчетными формулами.

3. МЕТОД «НАПРАВЛЯЮЩИХ ГИПЕРБОЛ»

Начальная фаза оптимизации. Задаем число итераций $N > 2$ (выбирается с учетом допустимых затрат машинного времени) и две точки $\vec{\mu}^1 = \vec{\mu}_c$ и $\vec{\mu}^2 \neq \vec{\mu}^1$, принадлежащие ОДР (3). С помощью ИМ вычисляем в этих точках оценки среднего времени ответа $\hat{E}^1$ и $\hat{E}^2$, и, соответственно, оценки среднего времени ожидания $(\hat{w}_1^1, \dots, \hat{w}_n^1)$, $(\hat{w}_1^2, \dots, \hat{w}_n^2)$ в узлах $1, \dots, n$. Полагаем $k = 2$.

Основной цикл. Известны точки $\vec{\mu}^{k-1}$ и $\vec{\mu}^k$ с оценками $(\hat{w}_1^{k-1}, \dots, \hat{w}_n^{k-1})$, $\hat{E}^{k-1}$ и $(\hat{w}_1^k, \dots, \hat{w}_n^k)$, $\hat{E}^k$ откликов $(w_1^{k-1}, \dots, w_n^{k-1})$, E^{k-1} и $(w_1^k, \dots, w_n^k)$, E^k .

1. Используя оценки $(\hat{w}_1^{k-1}, \dots, \hat{w}_n^{k-1})$ и $(\hat{w}_1^k, \dots, \hat{w}_n^k)$, находим по формулам (6) коэффициенты R_i и S_i , $i = \overline{1, n}$, аппроксимации $E^{ap}(\vec{\mu})$.

Вычисляем градиент (7) функции E^{ap} . Направление $-\nabla E^{ap}(\vec{\mu}^k)$ наискорейшего убывания функции $E^{ap}(\vec{\mu})$ проецируем на поверхность ограничений (3). Если $\beta_i = 1$ для всех $i = \overline{1, n}$, то поверхность (3) является гиперплоскостью, и проекция L на нее направления вектора $-\nabla E^{ap}(\vec{\mu}^k)$ есть направление вектора $-\nabla E_{pr}^{ap}(\vec{\mu}^k) = -\nabla E^{ap}(\vec{\mu}^k) + \vec{n}(\vec{n} \nabla E^{ap}(\vec{\mu}^k)) = \vec{e}$, где $\vec{n} = \vec{c}/|\vec{c}|$ — нормаль к гиперплоскости ограничений, $\vec{c} = (c_1, \dots, c_n)$ — вектор стоимостных коэффициентов, $|\vec{x}|$ — длина вектора $\vec{x}$. Валидная часть $[L]$ проекции L ограничена точками $\vec{\mu}^k$ и $\vec{\mu} = \vec{\mu}^k + h\vec{e}$, где $h = \min\{h_1, h_2\}$,

$$h_1 = \min\{h_{1i}; h_{1i} > 0; i = \overline{1, n}\},$$

$$h_2 = \min\{h_{2i}; h_{2i} > 0; i = \overline{1, n}\},$$

$$h_{1i} = -(\mu_i^k - \mu_{i\min})/e_{1i}, \quad h_{2i} = -(\mu_i^k - S_i)/e_{2i}, \\ i = \overline{1, n}.$$

Если не все β_i равны 1, то проекцию L направления антиградиента строим пошагово, как исходящую из точки $\vec{\mu}^k$ ломаную, узлы $\vec{\mu}$ которой суть проекции на поверхность (3) равноотстоящих с малым шагом точек направления $-\nabla E^{ap}(\vec{\mu}^k)$. Для каждого очередного узла $\vec{\mu}$ проверяются условия его валидности $\mu_i > \mu_{i\min}$ и $(\mu_i - S_i)(\mu_i^k - S_i) > 0$, $i = \overline{1, n}$. Равенство знаков разностей $(\mu_i - S_i)$ и $(\mu_i^k - S_i)$ означает, что координата μ_i текущего узла $\vec{\mu}$ траектории L и координата μ_i^k начальной ее точки $\vec{\mu}^k$ находятся по одну сторону от полюса S_i аппроксимации. Построение $[L]$ завершается получением и отбрасыванием первого невалидного узла либо выяснением, что очередная точка в направлении антиградиента $-\nabla E^{ap}(\vec{\mu}^k)$ уже не имеет проекции на поверхность ограничений.

2. В качестве следующей точки $\vec{\mu}^{k+1}$ выбираем решение (получаемое методом сканирования) задачи одномерной оптимизации $E^{ap}(\vec{\mu}) \rightarrow \min, \vec{\mu} \in [L]$.

3. Полагаем $k = k + 1$. С помощью ИМ вычисляем оценки $(\hat{w}_1^k, \dots, \hat{w}_n^k)$ и $\hat{E}^k$. Если $k < N$, то переходим к шагу 1, иначе — к шагу 4.

4. Точку $\vec{\mu}^* \in \{\vec{\mu}^1, \dots, \vec{\mu}^N\}$ с оценкой $\hat{E}(\vec{\mu}^*) = \min\{\hat{E}^1, \dots, \hat{E}^N\}$ принимаем в качестве приближенного решения задачи. *Конец алгоритма.*

4. УСКОРЕННЫЙ ПОКООРДИНАТНЫЙ СПУСК (ВТОРОЙ ЭТАП ОПТИМИЗАЦИИ)

Погрешность первого этапа (метода НГ) включает в себя две составляющие: *стохастическую* и *детерминированную*. Стохастическая составляющая контролируется расчетом доверительных интервалов для оценок $(\hat{w}_1^k, \dots, \hat{w}_n^k)$ и $\hat{E}^k$, и ее можно снижать путем удлинения прогонов модели. Детерминированная составляющая обусловлена применением сепарабельной аппроксимации $E^{ap}(\vec{\mu})$ для несепарабельной (в общем случае) функции $E(\vec{\mu})$. Это приводит к тому, что решение $\vec{\mu}^*$, найденное методом НГ, отличается от искомого $\vec{\mu}_{\text{opt}}$ даже при условии полного устранения стохастических погрешностей оценок ИМ. Поэтому решение $\vec{\mu}^*$ в общем случае целесообразно уточнять (или проверять) методом, не использующим аппроксимацию $E^{ap}(\vec{\mu})$.

Для этого на втором этапе оптимизации в окрестности решения $\vec{\mu}^*$ определяется $2(n-1)$ пробных точек, каждая из которых отличается от $\vec{\mu}^*$ лишь одной из $(n-1)$ «свободных» координат μ_i

(например, одной из координат $\mu_1, \dots, \mu_{n-1}$) на величину $\pm \Delta\mu$. «Связанная» координата пробной точки (например, μ_n) определяется через известные «свободные» ее координаты путем решения уравнения (3), чтобы обеспечить принадлежность всех пробных точек ОДР. Отклонение $\Delta\mu$ выбирается с учетом допустимой погрешности решения и с учетом возможности надежного сравнения откликов $E(\bar{\mu})$ в соответствующих точках по их оценкам $\hat{E}(\bar{\mu})$. Далее решение $\bar{\mu}^*$ уточняется путем ИМ сети в пробных точках и замены $\bar{\mu}^*$ тем пробным решением $\bar{\mu}$, для которого $E(\bar{\mu}) < E(\bar{\mu}^*)$. Если такое не находится, процесс завершается, иначе для нового решения цикл уточнения повторяется. Для исключения закливания возврат к пройденным решениям запрещается. Для ускоренного сравнения откликов $E(\bar{\mu})$ и $E(\bar{\mu}^*)$ с заданной надежностью по их оценкам $\hat{E}(\bar{\mu})$ и $\hat{E}(\bar{\mu}^*)$ применяется метод «общих случайных чисел» [10].

5. ИСПЫТАНИЯ МЕТОДА «НАПРАВЛЯЮЩИХ ГИПЕРБОЛ» НА ОТКРЫТЫХ СЕТЯХ

Подробно охарактеризовать возможности метода НГ можно на примере оптимизации открытой версии тестовой СеМО-1 (рис. 1). Суммарный ресурс $M = 30$ распределяется здесь при $\vec{c} = (c_1, \dots, c_n) = (K_1, \dots, K_9) = (1, 2, 1, 1, 1, 1, 1, 3, 1)$, т. е. для каждой системы массового обслуживания стоимостный коэффициент равен числу ее каналов. Типы распределений $B_i(t)$ для узлов $i = 1, \dots, 9$ определены как $R, R, R, M, M, E^2, E^2, E^2, R$ соответственно, где M — экспоненциальное распределение, R — равномерное (на отрезке от 0 до двух средних), E^m — эрланговское распределение m -го порядка. Входной поток СеМО-1 пуассоновский и имеет интенсивность $\Lambda = 1$. Переходные вероятности указаны на рис. 1.

Тестовая сеть СеМО-1 применялась для сравнения метода НГ с базовым алгоритмом оптимизации, в котором градиент рассчитывался методом малых приращений. В каждом прогоне модели СеМО-1 имитировалось прохождение через сеть около 1 млн. заявок. На рис. 2 приведена типичная траектория изменений целевой функции в процессе оптимизации, выполняемой методом НГ. Приближение к точке оптимума происходит за 7—11 шагов, т. е. число эффективных итераций близко к размерности n факторного пространства. Вблизи точки оптимума растут стохастические ошибки аппроксимации, перенастраиваемой по двум сближающимся точкам факторного пространства, и происходит «отбрасывание» очередного приближения $\bar{\mu}^{k+1}$ от искомой точки опти-

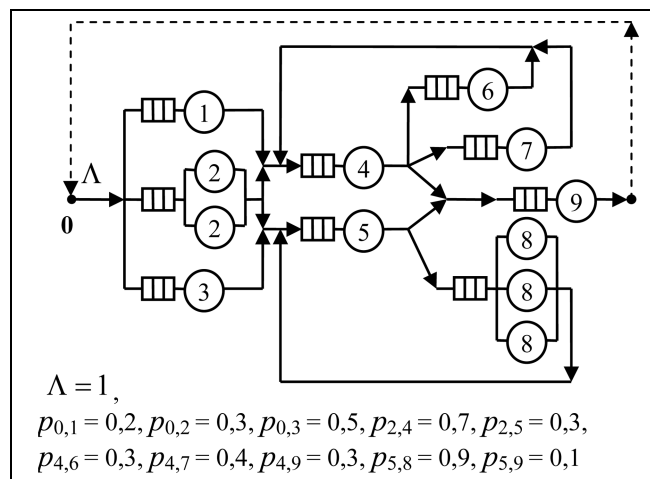


Рис. 1. Тестовый пример СеМО-1. Штриховая дуга соответствует замкнутой версии сети

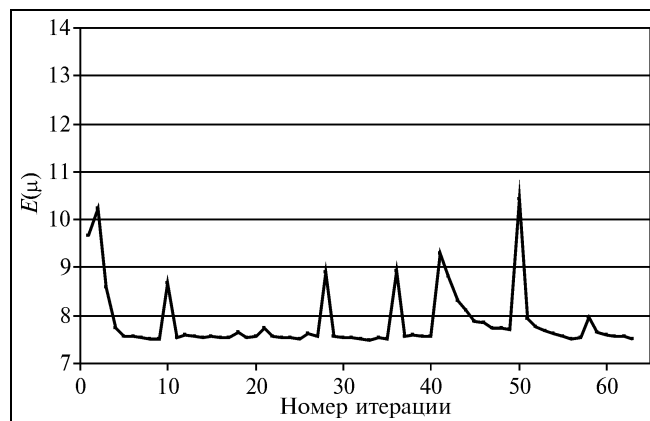


Рис. 2. Изменения отклика E в ходе поиска оптимума методом «направляющих гипербол»

му, вследствие чего значение целевой функции резко возрастает. Затем процесс вновь возвращается к точке оптимума. На рис. 3 показана типичная траектория значений целевой функции, получаемая при использовании базового метода. Заметно выигрывая у этого метода в глубине оптимизации, метод НГ превосходит его и в быстродействии. Это достигается как благодаря резкому снижению числа итераций, так и потому, что для вычисления градиента на каждой итерации в методе НГ используется лишь один прогон имитационной модели, а не $(n + 1)$ прогонов, как в базовом методе. Решение, определяемое для СеМО-1 методом НГ за 7—11 итераций, обеспечивает среднее время ответа $E \approx 7,49 \dots 7,51$. Базовый метод за 100...140 итераций (т. е. за 1000...1400 прогонов модели) дает более слабый результат $E \approx 9,3 \dots 9,5$.

Тестовая задача решалась также с помощью оптимизатора OptQuest [17]. Программой OptQuest за



750...1000 прогонов модели, т. е. при стократной «форе» по времени, достигается среднее время ответа $E \approx 7,91$, также проигрывающее результату метода НГ.

В более полный набор тестов включены открытые сети СеМО-2, ..., СеМО-5.

Сеть СеМО-2 отличается от СеМО-1 лишь равномерным на отрезке $[0,5; 1,5]$ распределением интервалов поступления заявок в сеть. Сеть СеМО-3 состоит из 30-ти узлов и представляет собой каскадное (последовательное) соединение трех сетей, подобных СеМО-1, усложненное возвратами заявок с выходов всех каскадов на их входы и на входы предшествующих каскадов. Сеть СеМО-4 содержит 20 узлов, ее маршрутная матрица сформирована с помощью датчика случайных чисел при соблюдении требования связности сети. Число каналов в каждом узле также разыгрывалось как случайное и не превышало шести. Сеть СеМО-5 содержит 100 узлов и имеет структуру переходов, представленную на рис. 4. Соответствующие переходные вероятности, число каналов в узлах и вид законов распределения $B_i(t)$ определялись датчиком случайных чисел. Кроме распределений, использованных в СеМО-2, в СеМО-5 применялись

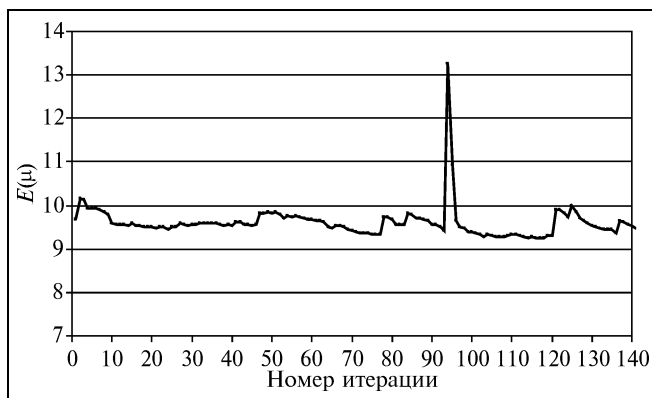


Рис. 3. Изменения отклика E при применении базового метода малых приращений

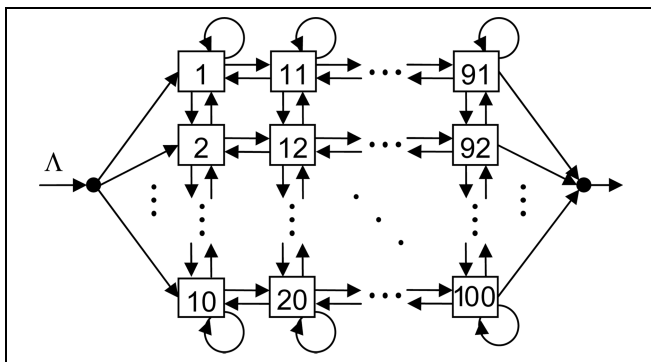


Рис. 4. Структура переходов в СеМО-5

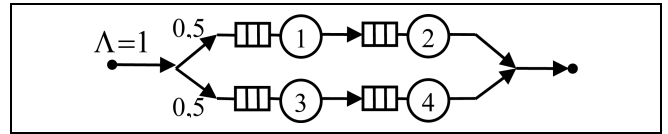


Рис. 5. Тестовая двухлинейная система

следующие распределения: эрланговское 3-го порядка; треугольное; гамма-распределение с параметром формы $\alpha = 2,5$; гиперэкспоненциальные распределения с коэффициентами вариации (к. в.) $C = 2$ и $C = 3$ и распределения Вейбулла с к. в. $C = 2$, $C = 3$ и $C = \sqrt{5}$. Последнему распределению принадлежат и интервалы поступления заявок в СеМО-4 и СеМО-5.

В табл. 1 приведены результаты оптимизации этих четырех тестовых сетей с помощью метода НГ. Каждая сеть оптимизировалась дважды: при распределении ресурса $M \approx 1,5M_{\min}$ и при $M \approx 2M_{\min}$. Время T_M итерации (практически совпадающее со временем прогона модели) указано в минутах для двухядерной ПЭВМ с тактовой частотой процессора 2 ГГц и объемом оперативной памяти 0,97 ГБ.

Как видно из табл. 1, при числе узлов и каналов, достигающем сотен (СеМО-5), когда высокие затраты времени на имитацию существенно ограничивают возможность увеличения числа итераций, метод НГ позволяет достигать значительного сокращения времени E при числе итераций N , меньшем размерности n факторного пространства.

6. СПЕЦИАЛЬНЫЕ ИСПЫТАНИЯ МЕТОДА «НАПРАВЛЯЮЩИХ ГИПЕРБОЛ»

Для оценки качества приближенного решения $\bar{\mu}^*$ введем три показателя: эффект оптимизации $Q = (E_c - E^*)/E_c$, фазовую погрешность $\delta = |\bar{\mu}_{\text{opt}} - \bar{\mu}^*|/|\bar{\mu}_{\text{opt}}|$ и упущенный эффект $\varepsilon = (E^* - E_{\text{opt}})/E_{\text{opt}}$, где $E_c = E(\bar{\mu}_c)$, $E^* = E(\bar{\mu}^*)$, $E_{\text{opt}} = E(\bar{\mu}_{\text{opt}})$.

Сравнивая выражения (2) и (5), можно видеть, что детерминированная составляющая погрешности, обусловленная использованием сепарабельной аппроксимации (5) для несепарабельной функции (2), должна быть более заметной при оптимизации тех сетей, в которых изменение интенсивностей обслуживания μ_i в одних узлах сильнее влияет на среднее время ожидания w_j в других. Но большое влияние μ_i на w_j возникает тогда, когда заявки переходят из узла i непосредственно в узел j , и при этом к. в. распределения $B_j(t)$ значительно отличается от к. в. распределения $B_i(t)$ и (или) от к. в. интервалов поступления заявок в i -й узел [1, 9].

С учетом этих факторов для специальных испытаний точности метода НГ сконструирована

«неудобная» для него двухлинейная система (ДС) (рис. 5). Интервалы поступления заявок в ДС распределены равномерно на отрезке $[0, 2]$. Время обслуживания в узлах 1 и 4 детерминированное (к. в. $C_1 = C_4 = 0$). В узлах 2 и 3 время обслуживания разыгрывается как функция x стандартной сл. в. η : $x = d[(\eta + \varepsilon)^{-1} - (1 + \varepsilon)^{-1}]$. Здесь сл. в. η распределена равномерно на отрезке $[0, 1]$, масштабный коэффициент d определяет требуемое среднее $M(x) = b$, а параметр $\varepsilon > 0$ позволяет формировать требуемый к. в. $C_x = C$ ($C \rightarrow \infty$ при $\varepsilon \rightarrow 0$). Если для узлов 2 и 3 взять $\varepsilon = 0,216$, $d = 1,106$, то получим в них $b = 1$ и $C = 1$ (вариант ДС1 — *ординарный тест*). При $\varepsilon = 0,019$, $d = 0,332$ имеем $b = 1$, $C = 2$ (вариант ДС2 — *усложненный тест*), а при $\varepsilon = 0,001$, $d = 0,169$ получаем $b = 1$, $C = 5,22$ (вариант ДС3 — *жесткий тест*). Все коэффициенты c_i и β_i взяты равными единице.

В табл. 2 представлены решения, полученные методом НГ (этап I) и последующим покоординатным спуском (этап II) для тестов ДС1, ДС2 и ДС3. На этапе II результат, полученный методом

НГ, предварительно округлялся до сотых долей (с учетом равенства $\mu_1 + \mu_2 + \mu_3 + \mu_4 = M$) и затем уточнялся ускоренным методом покоординатного спуска при отклонениях свободных координат $\Delta\mu_i = \pm 0,01$ (N' — число получившихся шагов). Средние значения времени ответа E_I и E_{II} , полученные на этапах I и II, приведены со всеми точно установленными значащими цифрами. Известные интервалы между сравниваемыми пробными точками позволили оценить фазовую погрешность δ сверху. Как видно из табл. 2, погрешность δ метода НГ выходит за пределы 1 % лишь при весьма больших перепадах к. в. C_i . При $C_2 = C_3 \approx 5$ ($C_1 = C_4 = 0$) она может достигать 3 %. При этом упущенный эффект оптимизации ε не превышает половины процента. Значение $\varepsilon < 0$ в табл. 2 получено для случая, когда решение $\bar{\mu}^*$ на этапе II ухудшилось из-за округления.

Можно предположить, что в сетях с ветвящимися и сливающимися маршрутами точность метода НГ будет не хуже, поскольку рассеивание и смешивание потоков обуславливает тенденцию их

Таблица 1

Результаты испытаний алгоритма на тестовых сетях СеМО-2, ..., СеМО-5

Число узлов n , общее число каналов K	СеМО-2 $n = 9, K = 12$		СеМО-3 $n = 30, K = 134$		СеМО-4 $n = 20, K = 52$		СеМО-5 $n = 100, K = 260$	
Число итераций $N \times$ время итерации T_M	20×2		60×10		40×5		50×94	75×94
Распределемый ресурс M	24	32	96	128	39	52	437	582
Отклик $\hat{E}_c$ в центре ОДР	13,98	8,043	49,92	27,88	78,08	38,19	278,1	163,3
Отклик $\hat{E}^* = \hat{E}(\bar{\mu}^*)$	10,83	6,225	41,53	22,97	55,12	29,94	249,6	143,8
Эффект оптимизации $(\hat{E}_c - \hat{E}^*)/\hat{E}_c \cdot 100 \%$	22,5	22,6	17	18	29	22	10	12

Таблица 2

Результаты оптимизации тестовых ДС

Вариант теста	M	ρ_c	Этап	$N; N'$	Найденное приближение к точке оптимума				$E_I; E_{II}$	$\delta \cdot 100\%$	$\varepsilon \cdot 100\%$
					μ_1	μ_2	μ_3	μ_4			
Ординарный $C_2 = C_3 = 1$	4	0,5	I II	48 0	1,105 1,11	0,946 0,95	0,890 0,88	1,059 1,06	2,925	0,6	0,00
Усложненный $C_2 = C_3 = 2$			I II	62 0	1,198 1,20	0,913 0,90	0,772 0,78	1,117 1,12	4,320 4,319	0,8	0,02
Жесткий $C_2 = C_3 = 5,22$			I II	122 5	1,282 1,33	0,837 0,84	0,646 0,63	1,235 1,20	12,921 12,871	3,0	0,4
Ординарный $C_2 = C_3 = 1$	2,5	0,8	I II	4 0	0,654 0,66	0,603 0,60	0,599 0,59	0,644 0,65	9,790 9,766	1,0	0,2
Усложненный $C_2 = C_3 = 2$			I II	29 0	0,692 0,69	0,596 0,60	0,557 0,56	0,655 0,65	19,972 19,987	0,6	-0,1
Жесткий $C_2 = C_3 = 5,22$			I II	8 1	0,687 0,70	0,604 0,60	0,533 0,53	0,676 0,67	90,16 89,76	1,2	0,4



Таблица 3

Оценка точности метода НГ для СеМО-1

Этап	N; N'	Найденное приближение к точке оптимума									E _I ; E _{II}	$\delta \cdot 100 \%$	$\varepsilon \cdot 100 \%$
		μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9			
I	33	0,818	0,597	1,405	3,025	9,490	1,272	1,682	2,965	2,219	7,514	0,4	0,01
II	5	0,79	0,60	1,43	3,02	9,49	1,28	1,69	2,96	2,22	7,513		

приближения к экспоненциальным, и, соответственно, форма $E(\bar{\mu})$ должна в таких открытых сетях приближаться к сепарабельной. Это подтверждается данными о точности метода НГ в табл. 3, полученной при двухэтапной оптимизации СеМО-1.

7. ОПТИМИЗАЦИЯ ЗАМКНУТЫХ СЕТЕЙ

Оптимизация замкнутой сети методом НГ характеризуется более высокой детерминированной составляющей погрешности, так как функция $E(\bar{\mu})$ здесь «сильно» несепарабельна. Изменение интенсивности μ_j в любом узле замкнутой сети приводит к изменению интенсивностей λ_j входных потоков всех узлов $j, j = \overline{1, n}$, и, тем самым, непосредственно изменяет время w_j во всех узлах. Испытания метода НГ для замкнутой версии СеМО-1 показывают, что при линейном ограничении (3) значение δ может достигать 0,03 (в открытой версии СеМО-1 $\delta = 0,004$). При нелинейном ограничении (3) (с разбросом коэффициентов β_j в пределах от 0,5 до 1,5) δ равно 0,06...0,08. Но упущенный эффект при этом остается небольшим ($\varepsilon < 0,01$). В этих испытаниях длина прогонов составляла около 1 млн. заявок, а число N итераций лежало в пределах нескольких десятков. Таким образом, и при оптимизации замкнутых сетей на практике часто можно ограничиваться применением только первого этапа оптимизации — метода НГ.

ЗАКЛЮЧЕНИЕ

Для оптимизации немарковских СеМО в общем случае приходится применять ИМ, которое характеризуется высокой сложностью вычисления градиентов, обусловленной стохастическим характером получаемых путем ИМ оценок. При решении задачи оптимального распределения ресурса проблема градиентов эффективно решается предлагаемым двухэтапным методом оптимизации, ядром которого составляет метод «направляющих гипербола». Он характеризуется относительно хорошей точностью и приемлемой вычислительной трудоемкостью, позволяющей рекомендовать его для практического применения при проектировании или модернизации сетей массового обслуживания, содержащих десятки и сотни узлов.

ЛИТЕРАТУРА

1. Вишневский В.М. Теоретические основы проектирования компьютерных сетей. — М.: Техносфера, 2003. — 512 с.
2. Клейнрок Л. Вычислительные системы с очередями. — М.: Мир, 1979. — 600 с.
3. Вишневский В.М., Пороцкий С.М. Моделирование ведомственных систем электронной почты // Автоматика и телемеханика. — 1996. — № 12. — С. 48–57.
4. Жожикашвили В.А., Вишневский В.М. Сети массового обслуживания. Теория и применение к сетям ЭВМ. — М.: Радио и связь, 1988. — 192 с.
5. Феррари Д. Оценка производительности вычислительных систем. — М.: Мир, 1981. — 576 с.
6. Герасимов А.И. Оптимизация замкнутых сетей массового обслуживания с несколькими классами сообщений // Проблемы передачи информации. — 1994. — Т. 30, № 1. — С. 85–96.
7. Рыжиков Ю.И. Имитационное моделирование. Теория и технологии. — СПб.: КОРОНА принт; М.: Альтекс-А, 2004. — 384 с.
8. Клейнрок Л. Теория массового обслуживания. — М.: Машиностроение, 1979. — 432 с.
9. Gabriel R. Bitran, Reinaldo Morabito. Open Queueing Networks: Optimization and Performance Evaluation Models for Discrete Manufacturing Systems / Сайт Массачусетского технологического института. — Режим доступа: <http://dspace.mit.edu/bitstream/handle/1721.1/2537/SWP-3743-31904719.pdf?sequence=1>.
10. Клейнен Дж. Статистические методы в имитационном моделировании. — М.: Статистика, 1978. — Вып. 1. — 221 с.
11. Johnson M.E., Jackson J. Infinitesimal Perturbation Analysis: a Tool for Simulation // J. of the Operational Res. Soc. — 1989. — Vol. 40, N 3. — P. 134–160.
12. Rubinstein R.Y. Sensitivity analysis of computer simulation models via the efficient score // Oper. Res. — 1989. — Vol. 37. — P. 72–81.
13. Suri R, Zazanis M. Perturbation Analysis Gives Strongly Consistent Sensitivity Estimates for the M[G]1 Queue // Mgmt Science. — 1988. — Vol. 34. — P. 39–64.
14. Иглхарт Д.Л., Шедлер Д.С. Регенеративное моделирование сетей массового обслуживания. — М.: Радио и связь, 1984. — 135 с.
15. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы. — М.: Мир, 1982. — 583 с.
16. Задорожный В.Н. Методы двухуровневого моделирования систем с очередями // Тр. VII междунар. конф. «Идентификация систем и задачи управления» SICPRO'08. — Москва, 28–31 января 2008 / ИПУ РАН. — М., 2008. — С. 1484–1563.
17. Карнов Ю.Г. Имитационное моделирование систем. Введение в моделирование с AnyLogic-5. — СПб.: БХВ-Петербург, 2005. — 400 с.

Статья представлена к публикации членом редколлегии А.С. Манделем.

Задорожный Владимир Николаевич — канд. техн. наук, доцент, Омский государственный технический университет, ☎(3812) 65-20-84, ✉zwn@yandex.ru.

ЭФФЕКТИВНОЕ ПРИМЕНЕНИЕ МНОГОЯДЕРНЫХ МИКРОПРОЦЕССОРОВ ДЛЯ РЕАЛИЗАЦИИ АЛГОРИТМОВ УПРАВЛЕНИЯ СТОХАСТИЧЕСКИМ ОБЪЕКТОМ¹

В.Н. Фетисов

Рассмотрена проблема применения многоядерных микропроцессоров для реализации алгоритма управления с прогнозирующей моделью марковским объектом. Показано, что для различных операционных систем реализовать параллельные вычисления можно довольно просто, если использовать соответствующие системные функции. Рассмотрены особенности применения предложенного подхода. В частности, указана область исходных данных, когда организация параллельных вычислений может оказаться неэффективной. Показана более высокая эффективность предлагаемого способа вычислений по сравнению с применением языка программирования OpenMP.

Ключевые слова: марковский объект, управление с прогнозирующей моделью, параллельные вычисления, многоядерный микропроцессор.

ВВЕДЕНИЕ

Традиционно словосочетание «параллельные вычисления и задачи управления» ассоциируется с проблемами, которые решаются с применением суперкомпьютеров или хотя бы с компьютерами, оснащенными десятками микропроцессоров. Конечно, методы и средства данного широкого научного направления не предназначены для управления объектами в реальном времени, но эффективны для использования на стадии проектирования. Данная работа нарушает установившуюся традицию. В статье предложено использовать параллельные вычисления непосредственно в алгоритме управления на стадии реализации. Сравнительно недорогие технические средства для этого уже имеются [1]. (Пока примитивные — на базе микропроцессоров всего с двумя ядрами). Но корпорация «Intel» обещает в скором времени разрабо-

тать микропроцессоры с ядрами более четырех для применения в промышленности.

Цель данной работы — представление простого и надежного метода использования параллельных вычислений при реализации алгоритмов управления стохастическим объектом.

Описание предлагаемого метода дано на примере алгоритма управления с прогнозирующей моделью для марковского объекта.

1. ПОСТАНОВКА ЗАДАЧИ

Главное достоинство алгоритмов управления с прогнозирующей моделью — гибкость перестройки контроллера к изменениям математической модели объекта, а также простота реализации алгоритма при наличии ограничений на управление и выходную координату [2—6]. Недостаток этих алгоритмов в случае стохастического объекта заключается в громоздкости вычислений на каждом шаге управления, что может привести к невозможности практической реализации. Поэтому задача использования параллельных вычислений для алгоритма управления марковским объектом довольно актуальна [7, 8].

¹ Содержание статьи (в кратком изложении) было представлено на междунар. конф. «Параллельные вычисления и задачи управления» РАСО'2008, г. Москва, ИПУ РАН.



Сформулируем задачу управления. Пусть требуется минимизировать критерий качества

$$I(u) = \mathbf{M}\{Q(u_1, \dots, u_N, \dots)\} = \\ = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{M}\left\{\sum_{n=1}^N g(x_n, u_n)\right\},$$

где $\mathbf{M}\{\}$ — символ математического ожидания, x_n и u_n удовлетворяют уравнению

$$x_n = \phi(x_{n-1}, u_n) + \sigma \varepsilon_n, \quad x_0 = c,$$

где σ — положительное число, ε_n — независимые случайные величины с нулевым математическим ожиданием и единичной дисперсией, функции g и ϕ дифференцируемы по своим аргументам. Выходная координата объекта управления x_n наблюдается, $u_1, \dots, u_N$ — управления.

В дальнейшем математическое ожидание критерия качества будем аппроксимировать с помощью метода Монте-Карло. Тогда алгоритм управления с прогнозирующей моделью определяется минимизацией на каждом шаге управления вспомогательного (скользящего) критерия качества вида

$$I_{x_{n-1}}(v) = \frac{1}{R} \sum_{r=1}^R \sum_{t=1}^T g(y_t^r, v_t) \quad (1)$$

по переменным v_t . Величины y_t и v_t связаны между собой с помощью уравнения, которое будем называть прогнозирующей моделью

$$y_t^r = \phi(y_{t-1}^r, v_t) + \sigma \varepsilon_t^r, \quad t_0^r = x_{n-1}.$$

Числа T и R в формуле (1) являются наименьшими, для которых выполняется неравенство

$$\mathbf{P}\left\{\left|\frac{1}{R} \sum_{r=1}^R \left[\sum_{t=1}^T g(y_t^r, v_t) - \lim_{N \rightarrow \infty} \sum_{t=1}^T g(y_t^r, v_t)\right]\right| > \alpha\right\} \leq \delta,$$

где величины α и δ заданы.

Обозначим $v_1^*, \dots, v_N^*$ управление, дающее минимум критерия (1). В качестве управляющего воздействия на n -м шаге положим $u_n^* = v_1^*(x_{n-1})$. На шаге $n+1$ после минимизации вспомогательного критерия положим $u_{n+1}^* = v_1^*(x_n)$ и на следующих шагах будем действовать по аналогичному правилу. Последовательность $\{u_n^*\}$ и определяет алгоритм управления с прогнозирующей моделью.

Процедура минимизации семейства функционалов $I_{x_{n-1}}(v)$, $n = 1, 2, \dots$, должна осуществляться в реальном времени непосредственно в процессе управления объектом. В отличие от синтеза систе-

мы управления, когда множество управлений принадлежит функциональному пространству, в системах управления с прогнозирующей моделью множество управлений определено в конечномерном линейном пространстве, что существенно упрощает решение задачи управления.

Главная проблема минимизации функционала (1) связана с вычислительным аспектом задачи, поскольку теперь все вычисления должны уложиться во временной интервал допустимого по требованиям точности шага управления.

Решение задачи управления (1) обычно невозможно получить в аналитическом виде. Поэтому применяются рекуррентные методы оптимизации. Без нарушения общности далее будет рассмотрен метод градиента [8].

Для упрощения математических выражений и без нарушения общности рассмотрим оптимизацию критерия (1) на одном временном интервале. Тогда получим

$$I(v) = \frac{1}{R} \sum_{r=1}^R \sum_{t=1}^T g(y_t^r, v_t), \quad (2)$$

$$y_t^r = \phi(y_{t-1}^r, v_t) + \sigma \varepsilon_t^r, \quad t_0^r = c.$$

Далее индекс n будет обозначать номер шага рекуррентной оптимизации. Имеем

$$V_n = V_{n-1} - \mu \text{grad}_n, \quad V_0 = 0,$$

где μ — положительное число, V_n — вектор управлений размерности T , а компоненты вектора grad_n^k , $k = 1, \dots, T$, определяются формулой

$$\text{grad}_n^k \approx [I(v_{1n}^1, \dots, v_n^k + \Delta, \dots, v_n^T) - \\ - I(v_n^1, \dots, v_n^k, \dots, v_n^T)]/\Delta, \quad (3)$$

где Δ — положительное число.

Известно [8], что имеет место равенство

$$\lim_{\Delta \rightarrow 0, R \rightarrow \infty, n \rightarrow \infty} V_n = V.$$

Задача рекуррентной минимизации критерия качества (2) довольно громоздка в вычислительном аспекте.

Идеи использования решения громоздких в вычислительном аспекте задач на компьютерах с несколькими процессорами появились давно [9]. В работе [10] была предложена архитектура микропроцессора с несколькими модулями операций с плавающей запятой (FPU), позволяющая осуществлять параллельные вычисления для широкого класса задач управления. Корпорация «Intel» для реализации параллельных вычислений выбрала другой (более общий) путь и в 2006 г. объявила

о создании микропроцессора с двумя «ядрами» (а несколько позже с четырьмя), в составе которых оказались почти все составляющие обычного микропроцессора. В настоящее время корпорация «Intel» приступила к разработке микропроцессоров с числом ядер исчисляемым десятками. Многоядерная архитектура совместно с некоторыми изменениями системных функций ряда операционных систем привела к возможности использовать стандартные языки программирования для реализации параллельных вычислений. Корректное написание программы позволяет почти в два раза для двух ядер и почти в четыре раза для четырех ядер увеличить скорость вычислений.

Далее будет предложен простой способ программной реализации алгоритма управления с марковской прогнозирующей моделью для микропроцессора с несколькими ядрами. Рассмотрен пример решения такой задачи с помощью четырехъядерного процессора и показана степень эффективности параллельного программирования в зависимости от размерности вектора градиента.

2. РЕАЛИЗАЦИЯ АЛГОРИТМА УПРАВЛЕНИЯ ДЛЯ МНОГояДЕРНЫХ МИКРОПРОЦЕССОРОВ

Вычисление вектора градиента в алгоритме управления занимает более 99 % от всего расчетного времени. Используя закон Амдала [9] для расчета коэффициента увеличения производительности β :

$$\beta \leq \frac{1}{\omega + (1 - \omega)n^{-1}},$$

где ω — доля программного кода без параллельных расчетов и n — число ядер, получим для четырехъядерного процессора предельное увеличение производительности в 3,9 раза. Как будет показано далее, увеличение производительности часто может быть более низким.

Перепишем вектор (3) в детализированном виде (назовем такое представление факторизацией вектора градиента)

$$\text{grad}_n^k \approx \frac{1}{R} \sum_{r=1}^R \left[\sum_{t=1}^T g(y_{in}^r, v_n^k + \Delta, \dots, v_n^T) - \sum_{t=1}^T g(y_{in}^r, v_{in}^k) \right] / \Delta, \quad (4)$$

где переменные y_{in}^r и v_n^k связаны уравнением

$$y_{in}^r = \phi(y_{(t-1)n}^r, v_{in}^k + \xi) + \sigma \varepsilon_t^r, \quad t_0^r = c.$$

Здесь вектор ξ размерности T имеет вид

$$\xi_t = \begin{cases} \Delta, & t = k, \\ 0, & t \neq k. \end{cases}$$

Как видно из формулы (4), слагаемые суммы по индексу r (выражения в квадратных скобках) не связаны между собой и могут быть рассчитаны параллельно при наличии многоядерного микропроцессора (или нескольких обычных микропроцессоров).

Пусть M ($M \leq R$) обозначает число ядер микропроцессора. Разобьем сумму по r на части. Всего таких частей будет R/M , если R делится на M без остатка, и $[R/(M-1)] = 1$ в противном случае. Таким образом, распределив вычисления полученных частичных сумм по ядрам, теоретически можно в M раз увеличить расчетную скорость.

Существует ряд языков параллельного программирования высокого уровня [11–14], которые разработаны для различных операционных систем — ОС (Windows, LUNIX и др.) Один из таких языков, OpenMP, наиболее распространен. (Официально он называется «стандартом программирования» и первоначально предназначался для программирования на суперкомпьютерах.)

На рис. 1 [14] показан график увеличения производительности двухъядерного процессора благодаря использованию параллельных вычислительных потоков с применением языка OpenMP при вычислении суммы вида

$$x = \sum_{i=1}^R (y_{i-1} + y_{i+1})/2,$$

где последовательность $\{y_i\}$ независима. По оси абсцисс отложена величина R , по оси ординат — относительное увеличение производительности.

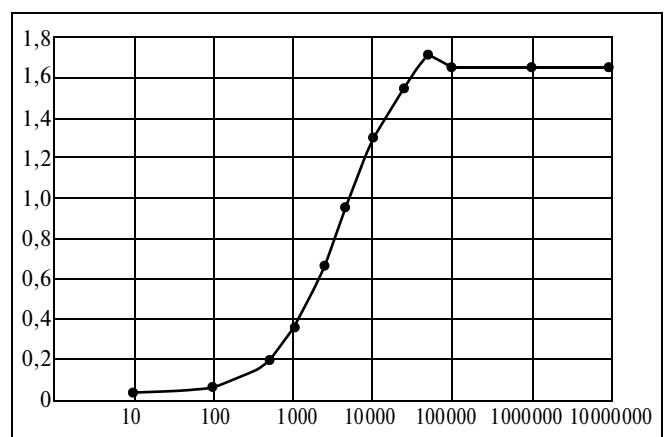


Рис. 1. Улучшение производительности с использованием языка OpenMP



Теперь покажем, как можно организовать параллельные вычисления, используя язык программирования Си и системные функции современных операционных систем (функции API в случае ОС «Windows»).

Далее представлен фрагмент программы реализации параллельных потоков, который был использован при решении задачи оптимизации с применением четырехъядерного микропроцессора (операционные системы x64 «Microsoft»):

```
HANDLE hThread[MAX_THREADS];
LPVOID npTTYInfo = NULL;
DWORD dwThreadID[MAX_THREADS];
//MAX_THREADS обозначает число ядер//
//Инициализируем параллельные потоки://
hThread[0]=CreateThread(
    (LPSECURITY_ATTRIBUTES) NULL,
    0,
    (LPTHREAD_START_ROUTINE)M0,
    (LPVOID) npTTYInfo,
    0, &dwThreadID[0]);
hThread[0]=CreateThread(
    (LPSECURITY_ATTRIBUTES) NULL,
    0,
    (LPTHREAD_START_ROUTINE)MAX,
    (LPVOID) npTTYInfo,
    0, &dwThreadID[0]);
//Ждем окончания вычисления в потоках://
WaitForMultipleObjects(MAX_THREADS-1,
    hThread, TRUE, INFINITE);
//Далее реализация вычислений в
//параллельных потоках//
DWORD WINAPI M0(LPVOID npTTYInfo)
{
    //Вычисление градиента для условия 0//
    return 0;
}
DWORD WINAPI MAX(LPVOID npTTYInfo)
{
    //Вычисление градиента для условия MAX//
    return MAX_THREADS-1;
}
```

В случае операционной системы UNIX вместо системной функции WaitForMultipleObjects используется следующий фрагмент программы:

```
for(k = 0; k < MAX_THREADS; k++)
pthread_join(dwThreadID[k], NULL).
```

Таким образом, в случае применения предложенного метода, для организации параллельных вычислений независимо от применяемой ОС используются всего две специальные системные функции.

При программировании нескольких параллельных вычислительных потоков желательно соблюдать следующие правила:

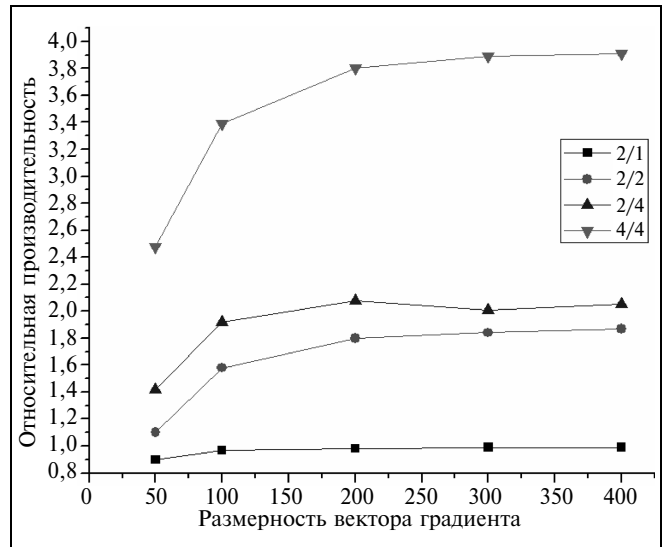


Рис. 2. Результаты моделирования для исследования относительной производительности

- глобальные переменные для различных потоков должны иметь уникальные идентификаторы;
- по возможности не пользоваться стандартными математическими библиотеками;
- прототипы математических функций определять со спецификатором «inline».

Рассмотренный в данной работе подход к организации параллельных вычислений был реализован на примере задачи управления нелинейным марковским объектом второго порядка вида

$$y_t = a_1 y_{t-1} + a_{21} y_{t-2} + a_{22} y_{t-2}^2 + u_t + \sigma \varepsilon_t, \\ y_0 = c_0, \quad y_1 = c_1$$

с квадратичным критерием качества.

Здесь y_t — наблюдаемая выходная координата, t — дискретное время, u_t — управление, ε_t — последовательность независимых случайных величин, a_1 , a_{21} , a_{22} и σ — константы.

На рис. 2 представлен график, показывающий увеличение относительной производительности вычисления вектора градиента в зависимости от его размерности при использовании предложенного подхода реализации параллельных вычислений.

Здесь отношение чисел, указанное для каждой кривой, означает отношение числа параллельных вычислительных потоков к числу ядер. Для малых значений размерности вектора градиента низкая производительность параллельных вычислений объясняется громоздкостью выполнения системных функций, организующих параллельные потоки.

Добавим, что имеет смысл повышать производительность именно в ситуации, когда произво-

дятся громоздкие вычисления. При малом объеме вычислений проблема повышения производительности вычислителя обычно не столь важна. Из сравнения рис. 1 и 2 можно сделать вывод, что метод системных функций более эффективен, чем использование языка OpenMP.

Для ОС Vista, Red Hat Enterprise 5 и Solaris 10 результаты моделирования практически совпали. Все эксперименты проводились на 64-разрядном микропроцессоре «Intel Core 2 Quad» с четырьмя ядрами.

ЗАКЛЮЧЕНИЕ

Решение задачи управления с прогнозирующей моделью марковским объектом весьма громоздкое в вычислительном аспекте. Показано, что алгоритм оптимизации на каждом шаге управления может быть реализован с помощью параллельных вычислительных потоков. Параллельные вычислительные потоки можно реализовать различными способами. Один из альтернативных способов организации параллельных вычислений предложен в работе [10].

Представлен способ реализации алгоритма оптимизации на компьютере с многоядерным микропроцессором. Рассмотрен простой и наиболее естественный способ организации параллельных вычислений для различных операционных систем.

Анализируется, в каких случаях организация параллельных вычислений приводит к эффективному решению для многоядерного микропроцессора, и в каких случаях параллельные вычисления перестают быть эффективными. Показана более высокая эффективность предлагаемого способа вычислений по сравнению с применением языка программирования OpenMP.

ЛИТЕРАТУРА

1. *ARK-3399* — High-performance Compact Embedded Computer with Intel® Core™ Duo/Core™ 2 Duo. URL: <http://www.advantech.com/products/> (дата обращения: 29.10.2008).
2. *Буков В.Н.* Адаптивные прогнозирующие системы управления полетом. — М.: Наука, 1987.
3. *Filatov N.M., Unbehauen H.* Adaptive predictive control policy for nonlinear stochastic systems // *IEEE Trans. on automatic control*. — 1995. — Vol. 40, N 11. — P. 1943–1949.
4. *Camacho E.F., Bordons C.* Model Predictive Control. — London: Springer-Verlag, 1999.
5. *Constrained model predictive control: Stability and optimality (Survey Paper)* / D.Q. Mayne, J.B. Rawlings, C.V. Rao, O.M. Scokaert // *Automatica*. — 2000. — Vol. 36. — P. 789–814.
6. *Фетисов В.Н.* Точность стохастической системы управления с детерминированной моделью в качестве прогнозатора // *Автоматика и телемеханика*. — 2002. — № 12. — С. 49–58.
7. *Параллельные вычисления и задачи управления (аналитический обзор)* / А.Л. Бунич, К.С. Гинсберг, А.В. Добровидов и др. // *Автоматика и телемеханика*. — 2002. — № 12. — С. 3–23.
8. *Поляк Б.Т.* Введение в оптимизацию. — М.: Наука, 1983.
9. *Amdahl G.* Validity of the single processor approach to achieving large scale computing capabilities, AFIPS Conference Proceedings. — 1967. — Vol. 30. — P. 483.
10. *Фетисов В.Н.* Перспектива разработки и использования микропроцессоров с параллельной архитектурой устройства обработки чисел с плавающей запятой // Тр. междунар. конф. «Параллельные вычисления и задачи управления» РАСО' 2004. Москва 2 — 4 октября 2004 г. / ИПУ РАН. — М., 2004. — С. 182–188.
11. *Левин М.П.* Параллельное программирование с использованием OpenMP. — М.: Бином, 2008.
12. *Raхул В. Патил и Бобби Джордж.* Средства и приемы для выявления проблем параллельного выполнения // *MSDN Magazine*. — 2008. — N 8.
13. *Intel Threading Building Blocks (Tutorial)*, Intel Corporation, 2007. URL: <http://www.intel.com>. (дата обращения: 16.01.2008).
14. *Kang Su Gatlin, Pete Isensee.* Reap the Benefits of Multithreading without All the Work // *MSDN Magazine*. — 2005. — N 10.

Статья представлена к публикации членом редколлегии В.Д. Малюгиным

Фетисов Вячеслав Николаевич — д-р техн наук, профессор, Московский технический университет связи и информатики, ✉ vfetisov@stream.ru.

Новая книга

Дартау Л.А., Мизерницкий Ю.Л., Стефанюк А.Р. Здоровье человека и качество жизни: проблемы и особенности управления. — М.: СИНТЕГ, 2009. — 400 с.

Проблемы медико-социального благополучия населения, здоровья и качества жизни рассмотрены с позиций классической теории автоматического управления. Обоснована ключевая роль индивида в сохранении здоровья. Показано, что непрерывное обучение здоровью и эпизодическое лечение являются основой «воздействия на объект» в цепи прямой связи, а перманентное выявление факторов риска заболеваний — основой для коррекции как управляющих воздействий, так и образа жизни индивида в этико-правовых рамках нового общественного договора о социальном партнерстве. Сделан вывод, что ответственность за конечный результат (индивидуальное и популяционное здоровье и качество жизни) должна быть поделена между всеми субъектами управления и основана на межсекторальном сотрудничестве, которое обеспечивается администрацией в местах проживания, работы или учебы. На основе анализа документов Всемирной организации здравоохранения, касающихся современного состояния здоровья населения развитых стран, предложен возможный вариант реализации ее рекомендаций в России.

Руководителям структур государственного, территориального и муниципального управления, организаторам здравоохранения и образования, а также широкому кругу читателей, заинтересованных в информации о современных возможностях и ограничениях в управлении здоровьем и качеством жизни.

Владимир Николаевич БУРКОВ (к 70-летию со дня рождения)

17 ноября 2009 года исполнилось 70 лет известному ученому в области управления большими системами, создателю теории активных систем, лауреату Государственной премии СССР, премии Совета Министров СССР, премии Дж. фон Неймана, заслуженному деятелю науки Российской Федерации, академику Российской академии естественных наук, доктору технических наук, члену редколлегии журнала «Проблемы управления», профессору Владимиру Николаевичу Буркову.



В 1967 году В.Н. Бурков окончил аспирантуру Московского физико-технического института под руководством А.Я. Лернера и остался работать в его лаборатории в Институте автоматизации и телемеханики (ныне Институт проблем управления им. В.А. Трапезникова РАН), где и сейчас работает в качестве заведующего лабораторией активных систем. Ровно 40 лет назад — в 1969 году, во время стажировки в Париже (в Обществе экономики и прикладной математики) у Владимира Николаевича возникла идея теории активных систем, основоположником которой он является. В рамках этой теории еще в Советском Союзе были разработаны и применены на практике уникальные механизмы согласованного управления, распределения ресурсов, противозатратного налогообложения и многие другие. И сейчас теория дает интересные и значимые результаты. Изменение экономического курса России с плановой экономики на конкурентную привело к созданию технологий реформирования и реструктуризации предприятий, городов и целых областей, механизмов обеспечения экологической безопасности регионов.

Теория активных систем с самого своего зарождения начала и продолжает накапливать фундаментальные теоретические результаты. Школы, проводимые раньше, научные конференции по активным системам и управлению большими системами, которые проводятся и сейчас, собирали и

продолжают собирать ученых из разных городов и стран, оставаясь базой для обмена новыми результатами и достижениями.

Более 15 учеников В.Н. Буркова защитили докторские диссертации, более 70 стали кандидатами наук. Один из его учеников — Д.А. Новиков, самый молодой доктор в Институте, избран членом-корреспондентом РАН.

Начало нового века привело к новому повороту в научной биографии В.Н. Буркова. Совместно с дочерью, Ириной Бурковой, им был разработан новый метод решения задач нелинейной оптимизации — метод сетевого программирования, который позволил на единой основе получить новые алгоритмы решения целого ряда задач нелинейной и дискретной оптимизации. На основе этого метода продолжают ставиться и решаться новые задачи — как самим Бурковым, так и его учениками, недостатка которых у Владимира Николаевича никогда не было и, дай Бог, не будет. Ведь многие из них остаются не только коллегами, но и друзьями, которые продолжают считать В.Н. Буркова своим учителем во всех отношениях — и в беззаветной преданности науке, и в необыкновенной порядочности по отношению к людям, и в любви к своей семье и Родине, и в хранении традиций и уважении к ним, и в умении быть увлеченным человеком, воспринимающим жизнь во всем многообразии ее проявлений.

В.Н. Бурков известен и уважаем во многих научных коллективах нашей страны и многих стран мира — Австрии, Великобритании, США, Германии, Японии, Турции, Франции, Польше, Болгарии, Венгрии... Впрочем, все перечислить сложно.

Беззаветно преданный лучшим традициям Института проблем управления (ИПА), Владимир Николаевич продолжает активно участвовать в их сохранении и развитии и в нынешнее непростое время, отдавая накопленный опыт и знания новым организаторам науки управления.

**Дорогой Владимир Николаевич! Поздравляем Вас с юбилеем и желаем здоровья на многие годы!
Творческих Вам успехов!**

Сотрудники Института проблем управления им. В.А. Трапезникова РАН
Редакционная коллегия
Редакция