Российская Академия Наук

Автоматика и телемеханика

Журнал основан в 1936 году

Выходит 12 раз в год

1

ЯНВАРЬ

Москва 2020

Учредители журнала:
Отделение энергетики, машиностроения, механики и процессов управления РАН,
Институт проблем управления им. В.А. Трапезникова РАН (ИПУ РАН),
Институт проблем передачи информации им. А.А. Харкевича РАН (ИППИ РАН)

Адрес редакции: 117997, Москва, Профсоюзная ул., 65
Тел./факс: (495) 334-87-70
Электронная почта: redacsia@ipu.ru

Зав. редакцией *Е.А. Мартехина*

Москва
ООО «ИКЦ «АКАДЕМКНИГА»

Линейные системы

© 2020 г. В.Н. ОВЧАРЕНКО, д-р техн. наук (owcharenko.v@yandex.ru)
(Московский авиационный институт
(национальный исследовательский университет))

СТРУКТУРНО-ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ С ПОСТОЯННЫМИ ПАРАМЕТРАМИ

Рассматривается задача структурно-параметрической идентификации стационарных линейных динамических систем. Предложен новый метод структурно-параметрической идентификации динамической системы, основанный на построении множества слабоэквивалентных систем возрастающей сложности. Под структурно-параметрической идентификацией понимается оценка порядка дифференциальных уравнений, всех коэффициентов, удовлетворяющих некоторым ограничениям, неизвестных начальных условий и смещения измерителя. Решение задачи структурно-параметрической идентификации получается за конечное число шагов. На тестовом примере показано, что предложенный метод имеет высокую чувствительность в условиях интенсивных измерительных ошибок.

Ключевые слова: структурная идентификация, параметрическая идентификация, линейные динамические системы.

DOI: 10.31857/S0005231020010018

1. Введение

Большое количество работ по структурной идентификации математических моделей, появившихся в трудах различных конференций, семинаров, монографиях и журналах указывает на значительный интерес специалистов к этой проблеме и актуальность самой проблемы. Необходимость решения проблемы структурной идентификации обусловлена различными причинами, которые встречаются в процессе создания сложных технических систем [1]. Одной из таких причин является повышение точности вычисления оценок неизвестных параметров по результатам ограниченного числа экспериментов. Вычисленные оценки параметров сравниваются с их априорными значениями, полученными либо расчетным путем, либо в полунатурных экспериментах в полностью контролируемых условиях их проведения. Неприемлемые ошибки оценок неизвестных параметров, существенное отличие оценок от их априорных значений указывают на несоответствие экспериментальных данных структуре математической модели, принятой на этапе обработки результатов эксперимента. Причиной несоответствия экспериментальных данных структуре математической модели могут быть скрытые, ненаблюдаемые переменные, влияющие на отклик динамической системы. Ненаблюдаемые переменные могут быть обусловлены изменением условий эксперимента

и их присутствие в каждом эксперименте не обязательно. Таким образом, возникает проблема совместного определения параметров и структуры математической модели по результатам натурного эксперимента, т.е. проблема структурно-параметрической идентификации.

Л. Заде [2] связывает структурную идентификацию с понятием эквивалентности систем из выбранного класса систем. Одной из первых монографий, посвященных структурной идентификации линейных динамических систем, была, по-видимому, [3], в которой был разработан метод типовой табличной идентификации устойчивых динамических систем. Основой метода является анализ оценок корреляционных функций пары вход — выход. Этот метод давал решение задачи структурной идентификации для динамических систем до третьего порядка включительно. Эквивалентность структур динамических систем рассматривалась в терминах корреляционных функций пары вход — выход. В [4, 5] с общих позиций рассматривается анализ проблемы структурной идентификации и установлено, что ее успешное решение сводится к разработке и обоснованию соответствующих критериев и алгоритмов. В монографии [6] предлагается метод оценки порядка линейной динамической системы, основанный на расширении входного пространства системы введением вспомогательных переменных. В [7] анализируется структура модели в целях дальнейшего решения задачи параметрической идентификации.

Решение задачи параметрической идентификации динамических систем, заданных дифференциальными уравнениями, традиционно проводится во временной области и связано с необходимостью численного интегрирования дифференциальных уравнений. Это может привести к дополнительным трудностям решения задачи параметрической идентификации, обусловленным динамическими характеристиками объекта и информационно-измерительной системы. Поэтому иногда задачу параметрической идентификации целесообразно решать частотными методами [8]. Частотные методы идентификации были одними из первых методов идентификации математических моделей динамических систем. Свойства стационарной линейной динамической системы можно описать в терминах отношения амплитуд и сдвига фаз гармонических сигналов на входе и выходе изучаемого объекта. К особенностям частотных методов относятся: а) простота вычислений частотных характеристик наблюдаемых переменных в присутствии измерительных шумов; б) возможность непараметрической идентификации структуры математической модели в виде частотных характеристик; в) возможность независимого выбора точек частотного диапазона для каждой пары входного и выходного сигналов; г) возможность идентификации временных запаздываний в экспериментальных данных; д) возможность обобщения вычислительных алгоритмов идентификации на многоканальные системы вход — выход и на нелинейные динамические системы; е) возможность применения к идентификации неустойчивых динамических систем.

В [9–11] предложен частотно-временной метод параметрической идентификации, сочетающий в себе вычисления в частотной и во временной области и позволяющий применить его к решению задачи структурно-параметрической идентификации линейных систем с постоянными коэффициентами. Новый метод структурно-параметрической идентификации основан на сквозном

применении частотно-временного метода как для решения задачи параметрической идентификации, так и для решения проблемы структурной идентификации. В данной статье предложен новый метод структурной идентификации, основанный на анализе свойств слабой эквивалентности линейных систем [12] различной структуры. Проведенный анализ позволил предложить алгоритм определения порядка системы дифференциальных уравнений на множестве слабоэквивалентных динамических систем по наблюдаемой паре вход — выход на ограниченном интервале времени в присутствии измерительных помех. Порядок динамической системы определяется за конечное число шагов.

На иллюстративном примере показано, что предложенный метод имеет высокую чувствительность и позволяет выявить скрытые ненаблюдаемые входные сигналы в условиях интенсивных инструментальных помех.

2. Постановка задачи структурно-параметрической идентификации

Рассмотрим устойчивую линейную стационарную непрерывную систему, заданную дифференциальным уравнением n-го порядка на интервале времени $t \in [0, T]$

$$\begin{aligned} x^{(n)} + a_{n-1}x^{(n-1)} + a_{n-2}x^{(n-2)} + \ldots + a_1 x^{(1)} + a_0 x = \\ = b_{n-1}u^{(n-1)} + b_{n-2}u^{(n-2)} + \ldots + b_1 u^{(1)} + b_0 u, \end{aligned} \tag{1}$$

где x — скалярный выходной сигнал; u — скалярный входной сигнал, такой что $u(t) \neq \text{const}$; $x^{(k)}, u^{(k)}$ — k-е производные; начальные условия $x_0^{(k)}$; $k = 0, 1, \ldots, n-1$ известны полностью или частично.

Наблюдается скалярная функция времени $y(t)$, линейно связанная с процессом $x(t)$:

$$y(t) = x(t) + b_y + \eta(t), \tag{2}$$

где $\eta(t)$ — стационарный случайный процесс с нулевым средним, описывает измерительный шум; b_y — неизвестное смещение измерителя на интервале $[0, T]$, обусловленное смещением нуля измерителя и/или конечным временным интервалом измерений (на этом интервале измеритель "шумит" несимметрично).

Обозначим через $\alpha = (a_{n-1}, \ldots, a_0)$, $\beta = (b_{n-1}, \ldots, b_0)$, $\chi = (x_0^{(n-1)}, \ldots, x_0)$ – n-мерные векторы параметров и начальных условий уравнения (1). На входном сигнале $u(t)$ наблюдается выход $y(t) = y(t; \alpha, \beta, \chi)$, зависящий от векторов параметров, начальных условий и смещения, т.е. наблюдаемый выход $y(t)$ является траекторией в $(3n+1)$-мерном параметрическом пространстве. В общем случае на входной сигнал и параметры наложены ограничения $u \in U$, $(\alpha, \beta) \in \Theta$.

Под структурно-параметрической идентификацией будем понимать решение следующей задачи.

Задача 1. На интервале $[0, T]$ *выполняется одиночный эксперимент. Требуется по наблюдениям пары вход — выход* $(u(t), y(t))$, $t \in [0, T]$ *определить порядок* n*; все коэффициенты* $(\alpha, \beta) \in \Theta$*; неизвестные начальные условия* χ *системы* (1) *и смещение* b_y *в* (2).

Дальнейший анализ ограничен системами конечного порядка.

3. Принцип оценки порядка динамической системы

Рассмотрим уравнения (1) и (2) при условиях $b_y = 0$, $\eta(t) \equiv 0$, а в качестве пары вход — выход — (u, x). Принцип оценки порядка динамической системы (1) по результатам единственного эксперимента на интервале $[0, T]$ основан на фундаментальных свойствах линейных систем, к которым относится свойство эквивалентности двух и более систем [12]. Обозначим через $\mathbb{S}_n$ линейную динамическую систему (1) порядка n; векторы α, β, χ имеют размерности, согласованные с порядком системы.

Определение 1. Система $\mathbb{S}_n$ является следствием системы $\mathbb{S}_m$ на интервале $[0, T]$ (вложением в систему $\mathbb{S}_m$), или $\mathbb{S}_n \subset \mathbb{S}_m$, если каждая пара вход — выход (u, x) системы $\mathbb{S}_n$ является также парой вход — выход (u, x) системы $\mathbb{S}_m$.

Определение 2. Если система $\mathbb{S}_n$ является следствием системы $\mathbb{S}_m$ на интервале $[0, T]$, а система $\mathbb{S}_m$ является следствием системы $\mathbb{S}_n$ на том же интервале, то системы $\mathbb{S}_n$ и $\mathbb{S}_m$ слабоэквивалентны $\mathbb{S}_n \equiv \mathbb{S}_m$ (в условиях одиночного эксперимента) на интервале $[0, T]$.

Определение 3. Системы слабоэквивалентны $\mathbb{S}_n \equiv \mathbb{S}_m$ (в условиях одиночного эксперимента) на интервале $[0, T]$ тогда и только тогда, когда для каждого входа $u(t)$ и $(3n + 1)$-мерного параметрического состояния $\gamma_n = (\alpha, \beta, \chi)_n$ системы $\mathbb{S}_n$ найдется $(3m + 1)$-мерное параметрическое состояние $\gamma_m = (\alpha, \beta, \chi)_m$ (зависящее от $u(t)$ и γ_n) системы $\mathbb{S}_m$ такое, что реакция на $u(t)$ системы $\mathbb{S}_n$, находящейся в состоянии γ_n, совпадает с реакцией на $u(t)$ системы $\mathbb{S}_m$, находящейся в состоянии γ_m, и наоборот.

В символической записи

$$\{\mathbb{S}_n \equiv \mathbb{S}_m\} \Leftrightarrow \{\forall\gamma_n \forall u \exists\gamma_m [\bar{S}_n(\gamma_n; u) = \bar{S}_m(\gamma_m; u)]\};$$
$$\{\mathbb{S}_m \equiv \mathbb{S}_n\} \Leftrightarrow \{\forall\gamma_m \forall u \exists\gamma_n [\bar{S}_m(\gamma_m; u) = \bar{S}_n(\gamma_n; u)]\}.$$

Здесь порядок кванторов указывает на зависимость начального состояния одной системы от начального состояния другой системы; $\bar{S}_n(\gamma_n; u)$, $\bar{S}_m(\gamma_m; u)$ — реакции систем $\mathbb{S}_n, \mathbb{S}_m$ на входной сигнал $u(t)$ и параметры γ_n, γ_m (различной размерности) соответственно.

Теорема 1. Пусть наблюдаемая на интервале $[0, T]$ пара вход — выход (u, x) порождена динамической системой $\mathbb{S}_n$ порядка n. Тогда не существует динамической системы $(\mathbb{S}_k, k < n)$ меньшего порядка, слабоэквивалентной системе $\mathbb{S}_n$.

Доказательство теоремы 1. Действительно, выход $x(t)$ системы $\mathbb{S}_n$ порядка n определяется $(3n + 1)$-мерным вектором параметров γ_n, тогда как реакция системы $(\mathbb{S}_k, k < n)$ порядка k для всех входных сигналов $u(t)$ определяется $(3k+1)$-мерным вектором параметров γ_k меньшей размерности. Поэтому любой отклик системы $\mathbb{S}_k$ порядка k является проекцией выхода $x(t)$ на пространство параметров меньшей размерности. Следовательно, система $\mathbb{S}_k$ не может быть вложением в систему $\mathbb{S}_n$. $\square$

Пусть на интервале $[0, T]$ в единичном эксперименте с динамической системой $\mathbb{S}_{n_0}$ порядка n_0 получена пара вход — выход (u, x). Предположим, что для $\forall k, m$ таких, что $n_0 < k < m$, можно построить слабоэквивалентные системы

$\mathbb{S}_n \subset \mathbb{S}_k$ и $\mathbb{S}_k \subset \mathbb{S}_m$ и выполняется свойство транзитивности $\mathbb{S}_n \subset \mathbb{S}_m$. Тогда существует последовательность вложений $\mathbb{S}_{n_0} \subset \mathbb{S}_{n_0+1} \subset \cdots \subset \mathbb{S}_m \subset \cdots$. Отсюда следует, что для данной пары вход — выход (u, x) в условиях единичного эксперимента n_0 — наименьший порядок, начиная с которого существуют слабоэквивалентные системы большего порядка.

Порядок n динамической системы (1) определяется размерностью векторов параметров α и β и не зависит от вектора начальных условий χ. Поэтому необходимо рассмотреть некоторые общие свойства вложенных систем $\mathbb{S}_{n_0} \subset \mathbb{S}_{n_0+1} \subset \cdots \subset \mathbb{S}_m \subset \cdots$ на наборах эквивалентных начальных условий.

Определение 4. Динамические системы $\mathbb{S}_n$ и $\mathbb{S}_m$ слабоэквивалентны на интервале $[0, T]$ при нулевых начальных условиях (в единичном эксперименте), если каждому нулевому начальному состоянию системы $\mathbb{S}_n$ соответствует эквивалентное нулевое начальное состояние системы $\mathbb{S}_m$.

Теорема 2. Пусть слабоэквивалентные системы $\mathbb{S}_n \equiv \mathbb{S}_m, n \neq m$ также слабоэквивалентные при нулевом входном сигнале $u(t) \equiv 0$ на интервале $[0, T]$

$$x(t, \gamma_n; u = 0) = x(t, \gamma_m; u = 0).$$

Тогда динамические системы $\mathbb{S}_n$ и $\mathbb{S}_m$ слабоэквивалентны при нулевых начальных условиях

$$x(t; \alpha_n, \beta_n, \chi_n = 0; u) = x(t; \alpha_m, \beta_m, \chi_m = 0; u),$$

а их передаточные функции равны $W_n(p) = W_m(p)$.

Доказательство теоремы 2 следует из свойства разложения реакции линейной динамической системы на сумму свободного и вынужденного движений. □

Таким образом, анализ вложений динамической системы сводится к анализу передаточных функций и их частотных характеристик в условиях единичного эксперимента.

Следствие 1. Пусть n_0 — наименьший порядок, начиная с которого существуют слабоэквивалентные системы большего порядка при нулевых начальных условиях. Тогда для вложений $\mathbb{S}_{n_0} \subset \mathbb{S}_{n_0+1} \subset \cdots \subset \mathbb{S}_m \subset \cdots$ последовательные отношения частотных характеристик слабоэквивалентных динамических систем возрастающего порядка не зависят от частоты

$$\frac{W_{n_0+1}(j\omega)}{W_{n_0}(j\omega)} = \frac{W_{n_0+2}(j\omega)}{W_{n_0+1}(j\omega)} = \cdots = \frac{W_{n_0+k}(j\omega)}{W_{n_0+k-1}(j\omega)} = \cdots = 1. \tag{3}$$

Доказательство следствия вытекает из теоремы 2 путем очевидных последовательных подстановок частотных характеристик динамических систем возрастающего порядка. Амплитудно-фазовые частотные характеристики (3) имеют значения $(1, \pm 2\pi k, k = 0, 1, \ldots)$ для всех частот, согласованных с интервалом $[0, T]$ (см. далее раздел 4). □

Полученные результаты приводят к следующему принципу оценки порядка динамической системы: по наблюдениям на интервале $[0, T]$ пары вход — выход (u, x) вычисляются последовательность оценок передаточных функций возрастающего порядка и отношение их частотных характеристик (3); наименьшее значение n_0, для которого выполняются условия (3), принимается за

оценку порядка динамической системы (1); оценка порядка n_0 динамической системы выполняется за $(n_0 + 1)$ шагов.

Рассмотрим условия слабой эквивалентности динамических систем в присутствии измерительного шума $\eta(t) \neq 0,\ t \in [0, T]$. Этот случай сводится к предыдущему, если принять

$$y(t) = \hat{x}(t),$$

где $\hat{x}(t) = x(t) + \eta(t)$ — оценка выходного сигнала системы (1). Вычисляя математическое ожидание оценки $\hat{x}(t)$, получим

$$M[\hat{x}(t)] = x(t),$$

т.е. необходимо потребовать выполнение условия несмещенности $\hat{x}(t)$. Если условие несмещенности оценки $\hat{x}(t)$ выполняется, то свойства слабой эквивалентности динамических систем следует записать относительно математического ожидания оценки выходного сигнала $M[\hat{x}(t)]$. Кроме того, из свойства несмещенности оценки выхода $\hat{x}(t)$ следует несмещенность оценки частотной характеристики

$$M[\widehat{W}(j\omega)] = W(j\omega)$$

и выражение (3) принимает вид

$$\frac{M[\widehat{W}_{n_0+1}(j\omega)]}{M[\widehat{W}_{n_0}(j\omega)]} = \frac{M[\widehat{W}_{n_0+2}(j\omega)]}{M[\widehat{W}_{n_0+1}(j\omega)]} = \cdots = \frac{M[\widehat{W}_{n_0+k}(j\omega)]}{M[\widehat{W}_{n_0+k-1}(j\omega)]} = \cdots = 1.$$

Необходимо отметить следующие особенности применения принципа оценки порядка к данным натурного эксперимента:

1) амплитудно-фазовые частотные характеристики отношений (3) имеют значения $(1, \pm 2\pi k)$, $k = 0, 1, \ldots$ для всех $n > n_0$ и на всех частотах;
2) если условия (3) выполняются начиная с некоторого $(n_0 + 1)$ для всех практически важных частот, то это указывает на линейную динамическую систему;
3) если условия (3) выполняются в ограниченном частотном диапазоне, то это указывает на исходную нелинейную систему, но которая проявляет себя как линейная система в этом частотном интервале;
4) частотный диапазон для проверки условий (3) должен быть согласован с частотным спектром входного сигнала и не пересекаться с частотным спектром измерительного шума.

4. Параметрическая идентификация динамической системы

Идентификация частотных характеристик пары вход — выход (u, x) для систем возрастающего порядка по наблюдениям системы (1), (2) на интервале $[0, T]$ является нетривиальной задачей. Трудности ее решения обусловлены

влиянием на оценки частотных характеристик измерительного шума, неизвестного смещения и неизвестных начальных условий как функций неизвестного порядка динамической системы. Поэтому в настоящей работе предлагается сначала определить все неизвестные параметры $(\alpha, \beta) \in \Theta$, χ, b_y динамической системы (1), (2) для различных возрастающих порядков, а затем на оценках параметров при нулевых начальных условиях $\chi = 0$ вычислить частотные характеристики. Для решения этой задачи запишем динамическую систему (1), (2) в форме Фробениуса:

$$\dot{z} = A(\alpha)z + B(\beta)u, \tag{4}$$

где

$$A(\alpha) = \begin{bmatrix} -a_{n-1} & 1 & 0 & \cdots & 0 \\ -a_{n-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -a_1 & 0 & . & . & 1 \\ -a_0 & 0 & . & . & 0 \end{bmatrix}; \quad B(\beta) = \begin{bmatrix} b_{n-1} \\ b_{n-2} \\ \vdots \\ b_1 \\ b_0 \end{bmatrix}; \quad z = \begin{bmatrix} x \\ z_1 \\ \vdots \\ z_{n-1} \end{bmatrix};$$

$(z_1, \ldots, z_{n-1})^{\mathrm{T}}$ — вектор ненаблюдаемых вспомогательных переменных.

Наблюдаемая скалярная функция (2) примет вид

$$y(t) = Hz + b_y + \eta(t), \tag{5}$$

где $H = [1, 0, \cdots 0]$ — матрица размера $1 \times n$.

Математические модели динамической системы, записанные уравнениями (1), (2) и в форме Фробениуса (4), (5), эквивалентны по параметрам (α, β) на паре вход — выход (u, x) и не эквивалентны по вектору начальных условий $(x_0^{(n-1)}, \ldots, x_0) \neq (x_0, z_{10}, \ldots, z_{(n-1)0})$. Неэквивалентность по вектору начальных условий не оказывает влияния на оценку частотных характеристик, которые вычисляются на оценках параметров (α, β).

Задача 2. Пусть задан порядок n динамической системы (4), (5). *Требуется по наблюдениям пары вход — выход* (u, y) *на интервале* $[0, T]$ *вычислить оценки неизвестных параметров* $(\alpha, \beta) \in \Theta$, *начальных условий* $\chi = (x_0, z_{10}, \ldots, z_{(n-1)0})$ *и смещения* b_y.

Для решения этой задачи применим частотно-временной метод идентификации [9–11], который имеет ряд следующих достоинств: 1) переход в частотную область выполняется с помощью финитного преобразования Фурье, которое вычисляется только один раз; 2) общая задача идентификации разделяется на задачу оценки параметров $(\alpha, \beta) \in \Theta$ (решается в частотной области) и задачу оценки начальных условий (решается во временной области на оценках параметров); 3) если ограничиться только вычислением частотных характеристик системы (1), то необходимость решения второй задачи отсутствует.

Определим дискретное множество частот

$$\Omega = \left\{ \omega_k : \omega_k = \frac{2\pi}{T} k, k = 1, \ldots, K \right\}, \tag{6}$$

где $K \leqslant T f_N$; $f_N = 1/h$ — частота Найквиста; h — шаг измерений, и вычислим на Ω финитное преобразования Фурье уравнений (4), (5):

$$j\omega_k Z_T(j\omega_k) - z(0) + z(T) = A(\alpha) Z_T(j\omega_k) + B(\beta) U_T(j\omega_k);$$
$$Y_T(j\omega_k) = H Z_T(j\omega_k) + \eta_T(j\omega_k).$$

Здесь $(U_T(j\omega_k), Z_T(j\omega_k), Y_T(j\omega_k), \eta_T(j\omega_k))$ — финитные преобразования Фурье функций $(u(t), z(t), y(t), \eta(t))$:

$$(U_T(j\omega_k), Z_T(j\omega_k), Y_T(j\omega_k), \eta_T(j\omega_k), 0) =$$
$$= \int_0^T (u(t), z(t), y(t), \eta(t), b_y) e^{-j\omega_k t} dt; \;\; j = \sqrt{-1}.$$

Выполняя элементарные алгебраические преобразования, получим

$$Y_T(j\omega_k) = H[j\omega_k E_n - A(\alpha)]^{-1} [\Delta z + B(\beta) U_T(j\omega_k)] + \eta_T(j\omega_k), \tag{7}$$

где E_n — единичная матрица порядка n; $\Delta z = z(0) - z(T)$. Отсюда видно, что $Y_T(j\omega_k)$ зависит от параметров (α, β) и от разности граничных условий Δz, которые не наблюдаются и нуждаются в оценке.

Запишем критерий метода наименьших квадратов на дискретном множестве частот (6) в виде

$$J(\alpha, \beta, \Delta z) = \sum_{k=1}^{K} \varepsilon(-j\omega_k) \varepsilon(j\omega_k), \tag{8}$$

где $\varepsilon(j\omega_k) = Y_T(j\omega_k) - H[j\omega_k E_n - A(\alpha)]^{-1} [\Delta z + B(\beta) U_T(j\omega_k)]$.

Оценки параметров $(\hat{\alpha}, \hat{\beta}) \in \Theta$ и разности граничных условий $\Delta \hat{z}$ вычисляются минимизацией критерия (8) численными методами нелинейного программирования:

$$(\hat{\alpha}, \hat{\beta}, \Delta \hat{z}) = \arg \min_{(\hat{\alpha}, \hat{\beta}) \in \Theta, \Delta z} J(\alpha, \beta, \Delta z). \tag{9}$$

Вычисление интегралов Фурье целесообразно выполнять по формуле Филона [11, 13], так как в этом случае точность вычислений не зависит от частотного спектра измеренных данных. Оценки $(\hat{\alpha}, \hat{\beta}) \in \Theta$ вычисляются алгебраическими методами. Необходимость численного интегрирования уравнений (4) отсутствует. Поэтому требование устойчивости систем (1) и (4) учитывается неявно через множество допустимых значений неизвестных параметров Θ.

Оценки вектора начальных условий $\hat{\chi}$ и смещения $\hat{b}_y$ определяются на оценках параметров $(\hat{\alpha}, \hat{\beta})$ во временной области методом наименьших квадратов

$$(\hat{\chi}, \hat{b}_y) = \arg \min_{(\chi, b_y)} \int_0^T e^2(t, \hat{\alpha}, \hat{\beta}; \chi, b_y) dt, \tag{10}$$

где $e(t,\hat{\alpha},\hat{\beta};\chi,b_y)=y(t)-\hat{x}(t,\hat{\alpha},\hat{\beta};\chi)-b_y$; $\hat{x}$ — оценка выходного сигнала системы (1) или (4), вычисляется на входном сигнале $u(t)$ многократным интегрированием уравнений n-го порядка системы (4) на оценках параметров $(\hat{\alpha},\hat{\beta})$ и начальных условиях χ.

Результаты, полученные решением задач (9) и (10), дают полное решение задачи 2 параметрической идентификации системы (1) в частотно-временной области.

5. Алгоритм идентификации порядка системы (1)

Рассмотрим алгоритм определения порядка системы (1), если измерительные шумы отсутствуют $\eta(t)\equiv 0$.

1. Задают начальное приближение частотной характеристики $W_0(j\omega)$ и полагают $n:=0$.
2. Увеличивают порядок системы на единицу $n:=n+1$ и решают задачу 2 параметрической идентификации.
3. На оценках параметров $(\hat{\alpha},\hat{\beta})\in\Theta$ вычисляют частотные характеристики системы (1)

$$W_n(j\omega_k)=\frac{\hat{b}_{n-1}(j\omega_k)^{n-1}+\hat{b}_{n-2}(j\omega_k)^{n-2}+\ldots+\hat{b}_1(j\omega_k)+\hat{b}_0}{(j\omega_k)^n+\hat{a}_{n-1}(j\omega_k)^{n-1}+\hat{a}_{n-2}(j\omega_k)^{n-2}+\ldots+\hat{a}_1(j\omega_k)+\hat{a}_0}.$$

4. Вычисляют отношение частотных характеристик (3) $W_{n+1}(j\omega)/W_n(j\omega)$.
5. Строят диаграммы Боде отношений (3) на множестве частот Ω и запоминают $W_n(j\omega)$, найденную на шаге 3.
6. Переходят к шагу 2 и повторяют все вычисления.
7. За оценку порядка динамической системы n_0 принимается наименьшее n, начиная с которого диаграммы Боде отношений (3) на всех частотах Ω принимают значения $(0,\pm 2\pi k, k=0,1,\ldots)$. На этом процесс решения задачи структурно-параметрической идентификации заканчивается.

Алгоритм оценивания порядка системы (1) в присутствии измерительных шумов $\eta(t)\neq 0$, $t\in[0,T]$ состоит из аналогичной последовательности шагов. Однако отношение частотных характеристик на шаге 4 вычисляется на оценках частотных характеристик возрастающего порядка

$$\widehat{W}_{n+1}(j\omega)/\widehat{W}_n(j\omega).$$

Очевидно, что в этом случае условие (3) будет выполняться только в среднем и шаг 7 формулируется в следующем виде

8. За оценку порядка динамической системы n_0 принимается наименьшее n, начиная с которого диаграммы Боде отношений (3) на всех частотах Ω принимают средние по частоте значения $(0,\pm 2\pi k, k=0,1,\ldots)$.

Следует отметить высокую вычислительную эффективность метода частотно-временной идентификации по сравнению с методами идентификации динамических систем только во временной области.

6. Пример

Рассмотрим пример решения задачи структурно-параметрической идентификации двух динамических систем.

Первая динамическая система описывается системой дифференциальных уравнений второго порядка

$$\begin{aligned} \dot{x} &= a_1 x + z_1 + b_1 u; \\ \dot{z}_1 &= a_0 x + b_0 u. \end{aligned} \tag{11}$$

Вторая динамическая система описывается системой дифференциальных уравнений третьего порядка

$$\begin{aligned} \dot{x} &= \tilde{a}_2 x + z_1 + \tilde{b}_2 u; \\ \dot{z}_1 &= \tilde{a}_1 x + z_2 + \tilde{b}_1 u; \\ \dot{z}_2 &= \tilde{a}_0 x + \tilde{b}_0 u. \end{aligned} \tag{12}$$

Рассмотрим динамические системы (11), (12) такие, что ($a_0 = \tilde{a}_1$, $a_1 = \tilde{a}_2$, $b_0 = \tilde{b}_1$, $b_1 = \tilde{b}_2$). Входной сигнал $u(t)$ в системах (11), (12) является одним и тем же. В этих условиях систему (12) можно рассматривать как систему (11), в которой переменная $z_1(t)$ возбуждается дополнительным ненаблюдаемым сигналом $z_2(t)$, зависящим от наблюдаемой пары (u, x). Множество допустимых значений параметров определяется условиями устойчивости систем (11), (12) $\Theta = \{(a_0, a_1, \tilde{a}_0, \tilde{a}_1, \tilde{a}_2) < 0\}$.

Отклик систем (11), (12) наблюдается одним и тем же измерителем

$$y(t) = x(t) + b_y + \eta(t),$$

где $\eta(t) \in [-1, 1]$ — случайная последовательность с нулевым средним, распределенная по равномерному закону; $b_y = 1$ — смещение измерителя. Здесь уровень измерительных шумов и величина смещения приняты нереально большими.

Можно ли в экспериментах над динамическими системами (11) и (12), выполненными в одинаковых условиях, различить структуры этих систем?

Данные одиночного "эксперимента" получены численным решением системы (11) с параметрами ($a_0 = -4$, $a_1 = -1$, $b_0 = 10$, $b_1 = 0$) и системы (12) с параметрами ($\tilde{a}_0 = -4$, $\tilde{a}_1 = -1$, $\tilde{a}_2 = -1$, $\tilde{b}_2 = 0$, $\tilde{b}_1 = 10$, $\tilde{b}_0 = -1$). Начальные значения для первых двух переменных систем (11) и (12) приняты равными $x(0) = 1$, $z_1(0) = -1$, $z_2(0) = 0$. Эксперимент выполняется на интервале времени $t \in [0, T]$, $T = 30$ и на одном и том же входном сигнале переменной частоты

$$u(t) = 1 + 2\sin((\omega_0 + \dot{\omega}t)t), \quad \dot{\omega} = \frac{2 - \omega_0}{T}; \quad \omega_0 = \frac{2\pi}{T}.$$

Уравнения (11) и (12) интегрировались с шагом $h = 0{,}001$. Псевдослучайная последовательность $\eta(t)$ получена генератором случайных чисел RANDOM пакета MATLAB.

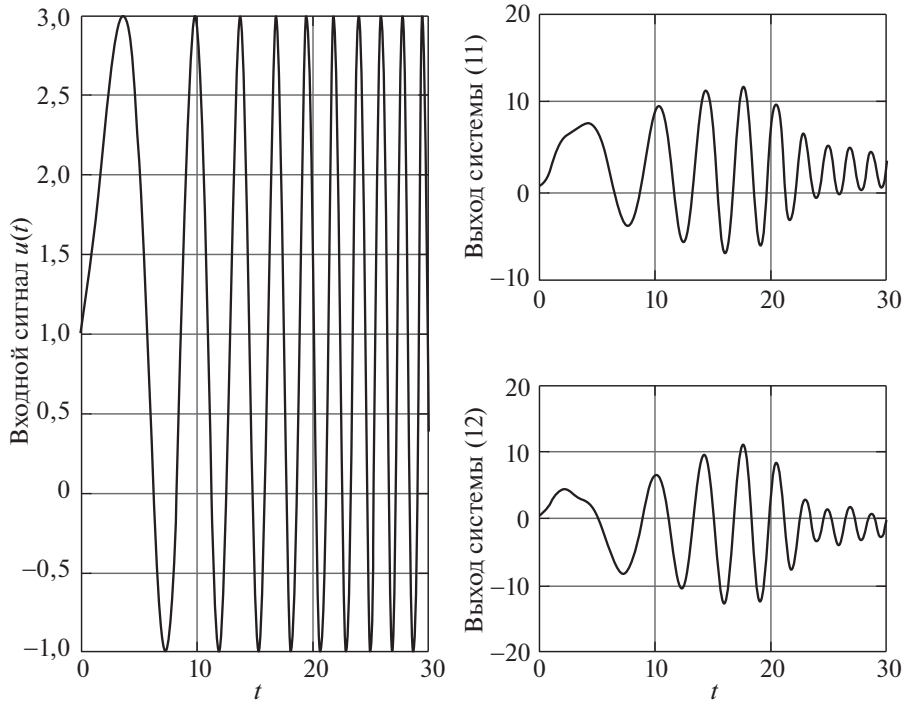


Рис. 1. Входной и выходные сигналы систем (11) и (12).

На рис. 1 показаны входной сигнал и выходные сигналы систем (11) и (12) (для удобства восприятия графики приведены без измерительного шума $\eta(t) \equiv 0$). Видно, что выходы систем (11) и (12) почти не отличаются друг от друга.

Задача параметрической идентификации решалась частотно-временным методом на множестве частот $\Omega = \{\omega_k : \omega_k = k\omega_0, k = 1, \ldots, 50\}$. В задаче идентификации порядка начальное приближение частотной характеристики было принято $W_0(j\omega) = 1$. В процессе решения задачи структурно-параметрической идентификации порядок динамических систем возрастал от $n = 1$ до $n = 5$. Полученные результаты показаны на рис. 2 и 3. Видно, что для системы (11) слабоэквивалентные системы появляются начиная с $n = 2$, а для системы (12) — начиная с порядка $n = 3$, т.е. порядки систем (11) и (12) вычислены правильно. Получены следующие оценки параметров систем (11) и (12), оценки начальных условий и оценка смещения: для системы (11) $\hat{a}_1 = -1$; $\hat{a}_0 = -3{,}99$; $\hat{b}_1 = 10$; $\hat{b}_0 = 0{,}01$; $\hat{x}(0) = 0{,}996$; $\hat{z}_1(0) = -1{,}036$; $\hat{b}_y = 0{,}993$; для системы (12) $\hat{a}_2 = -1$; $\hat{a}_1 = -4$; $\hat{a}_0 = -1$; $\hat{b}_2 = -0{,}005$; $\hat{b}_1 = 9{,}987$; $\hat{b}_0 = -1{,}017$; $\hat{x}(0) = 1{,}039$; $\hat{z}_1(0) = -0{,}941$; $\hat{z}_2(0) = 0{,}123$; $\hat{b}_y = 1{,}011$. Очевидна близость этих оценок и априорных значений. Для сравнения приведем значения оценок параметров системы (11), тогда как “экспериментальные” данные были получены интегрированием системы (12): $\hat{a}_1 = -0{,}643$; $\hat{a}_0 = -3{,}714$; $\hat{b}_1 = 1{,}141$; $\hat{b}_0 = 8{,}833$; $\hat{x}(0) = 0{,}959$;

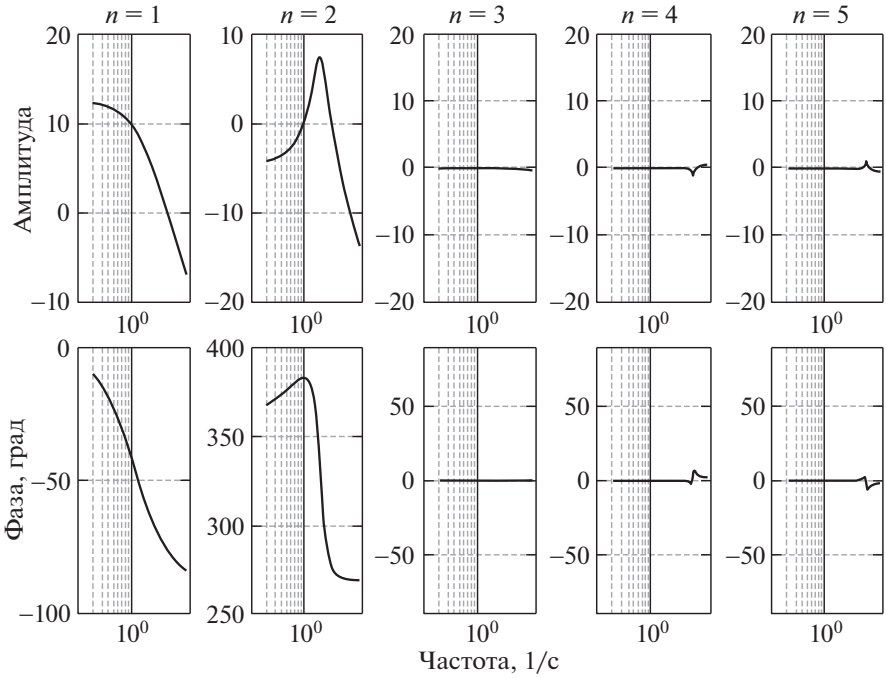


Рис. 2. Диаграммы Боде отношений W_{n+1}/W_n для системы (11).

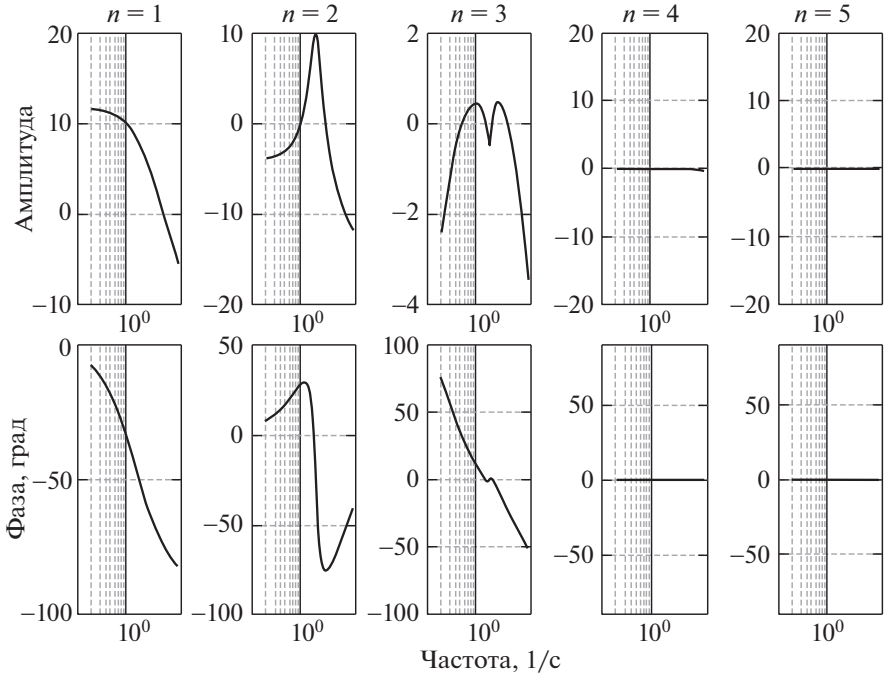


Рис. 3. Диаграммы Боде отношений W_{n+1}/W_n для системы (12).

$\hat{z}_1(0) = 3{,}076$; $\hat{b}_y = -2{,}451$. Эти результаты получены на некоторой реализации измерительного шума и являются типичными и для других реализаций.

Численное решение примера получено в пакете компьютерной математики MATLAB с применением программ: LSIM для решения систем (11) и (12); FMINCON для решения задач (9) и (10); BODE для вычисления частотных характеристик. Финитные интегралы Фурье вычислялись по формуле Филона программой, приведенной в [11].

Таким образом, предложенный метод корректно решает задачу структурно-параметрической идентификации (в условиях смещения нуля измерителя и большой интенсивности измерительного шума).

7. Заключение

Предложен новый метод структурно-параметрической идентификации динамической системы, заданной линейными дифференциальными уравнениями с постоянными коэффициентами. Под структурно-параметрической идентификацией понимаются: оценка порядка дифференциальных уравнений; всех коэффициентов, удовлетворяющих некоторым ограничениям (например, по условиям устойчивости); неизвестных начальных условий и смещения измерителя. Метод основан на свойстве слабой эквивалентности двух и более систем. Решение задачи структурной идентификации ищется на множестве слабоэквивалентных систем. Для рассматриваемого класса динамических систем определение порядка линейных дифференциальных уравнений сводится к анализу отношений частотных характеристик последовательно усложняемых структур. За оценку порядка динамической системы принимается наименьший порядок динамической системы, начиная с которого усложнение структуры динамической системы приводит только к последовательности новых слабоэквивалентных систем. Оценка порядка динамической системы определяется за конечное число шагов. Учет влияния измерительного шума на решение задачи структурной идентификации приводит к замене частотных характеристик динамических систем на математическое ожидание этих частотных характеристик. Однако в условиях одиночного эксперимента (или с малым числом реализаций) это не приводит к изменению алгоритма структурной идентификации.

Решению задачи структурной идентификации предшествует необходимость решения задачи параметрической идентификации, которое ищется частотно-временным методом. В процессе решения задачи параметрической идентификации вычисляются оценки параметров динамической системы в частотной области, а затем во временной области определяются оценки начальных условий и смещение измерителя. На оценках параметров вычисляются частотные характеристики динамической системы. В результате решения задачи структурно-параметрической идентификации получают оценки порядка динамической системы, оценки параметров и оценку смещения измерителя.

На иллюстративном примере (рис. 2 и 3) показано, что предложенный новый метод структурно-параметрической идентификации имеет высокую чувствительность и позволяет разделить две динамические системы различной

структуры в условиях одиночного эксперимента и больших измерительных ошибок, переходные процессы которых визуально близки.

Разработанный метод полностью решает задачу структурно-параметрической идентификации стационарной линейной системы по наблюдениям пары вход — выход на ограниченном временном интервале в условиях измерительных ошибок.

СПИСОК ЛИТЕРАТУРЫ

1. Труды XII Всероссийского совещания по проблемам управления. М.: ИПУ, 2014.
2. *Эйкхофф П.* Основы идентификации систем управления. М.: Мир, 1975.
3. Типовые линейные модели объектов управления / Под ред. Н.С. Райбмана. М.: Энергоатомиздат, 1983. 264 с.
4. *Гинсберг К.С.* Новый подход к проблеме структурной идентификации. II // АиТ. 2002. № 6. С. 85–98.

   Ginsberg K.S. A New Approach to the Problem of Structural Identification. II // Autom. Remote Control. 2002. V. 63. No. 6. C. 946–959.
5. *Гинсберг К.С.* К основам методологии структурной идентификации для цели проектирования технических систем // Тр. XIII Всерос. сов. по проблемам управления. М.: ИПУ, 2019.
6. *Карабутов Н.Н.* Структурная идентификация систем: анализ информационных структур. М.: Книжный дом “ЛИБРОКОМ”, 2009. 176с.
7. *Novara C., Vincent T., Hsu K., Milanese M., Poolla K.* Parametric identification of structured nonlinear systems // Automatica. 2011. № 47. С. 711–721.
8. *Корсун О.Н.* Алгоритм идентификации динамических систем с функционалом в частотной области // АиТ. 2003. № 5. С. 111–121.

   Korsun O.N. An Identification Algorithm for Dynamic Systems with a Functional in the Frequency Domain // Autom. Remote Control. 2003. T. 64. № 5. С. 772–781.
9. *Овчаренко В.Н.* Идентификация аэродинамических характеристик воздушных судов по полетным данным. М.: Изд-во МАИ, 2017.
10. *Danilevich E.V., Evstratov A.R., Kukharenko N.I., Ovcharenko V.N., Poplavskii B.K.* Identification of Constant Parameters of Dynamic Systems by a Time-Frequency Method // J. Comput. Syst. Sci. Int. 2018. V. 57. No. 4. P. 3–13.
11. *Овчаренко В.Н.* Аэродинамические характеристики летательных аппаратов: идентификация по полетным данным. М.: ЛЕНАНД, 2019. 236 с.
12. *Заде Л., Дезоер Ч.* Теория линейных систем. М.: Наука, ГРФ.-М; Л., 1970.
13. *Бахвалов Н.С.* Численные методы. М.: Наука, 1973.

Статья представлена к публикации членом редколлегии Н.Н. Бахтадзе.

Поступила в редакцию 20.03.2019

После доработки 25.06.2019

Принята к публикации 18.07.2019

© 2020 г. Е.Н. РОЗЕНВАССЕР, д-р техн. наук (fishka33@mail.ru)
(Государственный морской технический университет, Санкт-Петербург),
Б.П. ЛЯМПЕ, д-р инженерии,
В. ДРЕВЕЛОВ, д-р инженерии (wolfgang.drewelow@uni-rostok.de),
Т. ЯЙНШ, д-р инженерии (torsten.jeinsch@uni-rostok.de)
(Университет Росток, ФРГ)

СТАНДАРТИЗИРУЕМОСТЬ И H_2-ОПТИМИЗАЦИЯ ОДНОКОНТУРНОЙ МНОГОМЕРНОЙ ИМПУЛЬСНОЙ СИСТЕМЫ С МНОЖЕСТВЕННЫМИ ЗАПАЗДЫВАНИЯМИ

Изучается одноконтурная многомерная система с тремя звеньями чистого запаздывания. Приводятся достаточные условия стандартизируемости системы S_0, при выполнении которых задача H_2-оптимизации системы S_0 сводится к более простой задаче H_2-оптимизации некоторой эквивалентной импульсной системы с одним звеном чистого запаздывания. Строится множество фиксированных полюсов H_2-оптимальной системы S_0.

Ключевые слова: одноконтурная многомерная импульсная система с запаздыванием, стандартизируемость, H_2-оптимизация, фиксированные полюса.

DOI: 10.31857/S000523102001002X

1. Введение

Проблема учета запаздывания играет важную роль при решении задач анализа и синтеза импульсных систем. Различным аспектам этой проблемы посвящена значительная литература. Разнообразные подходы к решению указанной проблемы содержатся в публикациях [1–20] и цитированных там источниках. Анализ существующей литературы показывает, что теоретические и вычислительные трудности, связанные с исследованием импульсных систем с запаздыванием, существенно возрастают с увеличением количества элементов чистого запаздывания в структуре изучаемой системы. В связи с этим несомненный теоретический и практический интерес представляет задача построения для заданной импульсной системы с множественными запаздываниями эквивалентной системы с меньшим числом элементов чистого запаздывания.

В [20] на основе концепции параметрической передаточной матрицы (ППМ) сформирован класс многомерных импульсных систем с множественными запаздываниями, которые названы стандартизируемыми. При этом показано, что задача H_2-оптимизации стандартизируемой системы сводится к решению аналогичной задачи для эквивалентной стандартной импульсной системы S_τ, содержащей только один элемент чистого запаздывания. Поэтому решение задачи H_2-оптимизации для стандартизируемой импульсной

системы может быть получено с помощью алгоритма, описанного в [19]. В настоящей статье, которая является непосредственным продолжением [20], общие результаты [20] применяются к практически важному типу импульсных систем: многомерной одноконтурной импульсной системе S_0 с несколькими запаздываниями. Для конкретности рассматривается ситуация, когда число элементов чистого запаздывания в контуре управления равно трем, а число составляющих наблюдаемого вектора выхода, имеющих различные временны́е сдвиги, равно двум. Однако предлагаемый подход без существенных изменений распространяется на общий случай. В статье приведены достаточные условия стандартизируемости системы S_0, обычно выполняющиеся в приложениях, и построена соответствующая стандартная система S_τ. Кроме того, определена совокупность стационарных элементов, которые порождают полюса замкнутой H_2-оптимальной системы, не зависящие от вида используемого преобразователя «цифра-аналог».

2. Обобщенная стандартная импульсная система с множественными запаздываниями

Приведем используемые в последующем изложении общие свойства обобщенной стандартной импульсной системы с запаздыванием S_g, установленные в [20]. В определении обобщенной стандартной импульсной системы предполагается, что управляемый стационарный объект описывается уравнениями состояния:

$$\begin{gathered}\frac{dv(t)}{dt} = Av(t) + B_1x(t-\tau_1) + Bu(t-\tau_2),\\ y(t) = Cv(t),\end{gathered} \tag{2.1}$$

где $v(t)$ – вектор состояния объекта, $y(t)$ – вектор управляемого выхода, $x(t)$ – вектор входа, $u(t)$ – вектор управления и A, B, B_1, C – постоянные матрицы соответствующих размеров. Предполагается, что пара A, B полностью управляема и пара A, C полностью наблюдаема. Кроме того, в (2.1) τ_1 и τ_2 – неотрицательные постоянные.

Предполагается, что объект (2.1) управляется импульсным регулятором (ИР), который имеет период квантования T и описывается системой уравнений (2.2)–(2.4)

$$\xi_k = y(kT), \quad k = 0, \pm, 1, \dots, \tag{2.2}$$

$$\begin{gathered}\alpha_0\psi_k + \alpha_1\psi_{k-1} + \ldots + \alpha_\rho\psi_{k-\rho} =\\ = \beta_0\xi_k + \beta_1\xi_{k-1} + \ldots + \beta_\rho\xi_{k-\rho}, \quad \det\alpha_0 \neq 0,\end{gathered} \tag{2.3}$$

$$u(t) = h(t-kT)\psi_k, \quad kT < t < (k+1)T, \tag{2.4}$$

где α_i, β_i – постоянные матрицы соответствующих размеров и $h(t)$ – матрица, элементы которой имеют ограниченную вариацию на интервале $0 \leqslant t \leqslant T$, а условие $\det\alpha_0 \neq 0$ является условием каузальности дискретного регулятора.

Используя оператор обратного сдвига ζ [3], уравнение алгоритма управления (2.3) можно записать в полиномиальной форме

$$\alpha(\zeta)\psi_k = \beta(\zeta)\xi_k, \tag{2.5}$$

где $\alpha(\zeta)$ и $\beta(\zeta)$ – полиномиальные матрицы вида

$$\begin{aligned} \alpha(\zeta) &= \alpha_0 + \alpha_1\zeta + \ldots + \alpha_\rho\zeta^\rho, \\ \beta(\zeta) &= \beta_0 + \beta_1\zeta + \ldots + \beta_\rho\zeta^\rho. \end{aligned} \tag{2.6}$$

Далее рациональную матрицу

$$W_d(\zeta) \stackrel{\Delta}{=} \alpha^{-1}(\zeta)\beta(\zeta) \tag{2.7}$$

будем называть передаточной матрицей алгоритма управления.

Кроме того, в определении системы S_g предполагается, что для любой пары различных собственных чисел матрицы A p_1 и p_2 выполняется условие непатологичности периода квантования [3]

$$e^{p_1 T} \neq e^{p_2 T}. \tag{2.8}$$

В качестве наблюдаемого выхода системы S_g рассматривается вектор

$$\bar{z}'(t) = \left[z_1'(t) \ \ldots \ z_\gamma'(t)\right], \tag{2.9}$$

где штрих – оператор транспонирования и $z_i(t)$ – наблюдаемые парциальные векторы, заданные соотношениями

$$z_i(t) = C_i v(t - \tau_{3i}) + D_i u(t - \tau_2 - \tau_{3i}), \tag{2.10}$$

где τ_{3i}, $i = 1, \ldots, \gamma$, – вещественные постоянные, которые могут иметь значения произвольного знака.

В совокупности соотношения (2.1)–(2.10) определяют систему дифференциально разностных уравнений, которую, при выполнении всех указанных выше условий, назвали в [20] обобщенной стандартной системой с множественными запаздываниями S_g. Частный случай системы S_g, соответствующий значению $\gamma = 1$, рассмотрен в [19] и назван там стандартной импульсной системой S_τ.

По отношению к парциальному выходу $z_i(t)$ система S_g сводится к стандартной системе $S_{\tau i}$, которой в соответствии с [19] может быть сопоставлена параметрическая передаточная матрица (ППМ) $W_i(s,t)$, определяемая формулой

$$W_i(p,t) = \phi_{L_{\tau i}\mu}(T,p,t)\tilde{R}_N(p)M_\tau(p) + K_{\tau i}(p), \quad i = 1, \ldots, \gamma. \tag{2.11}$$

Здесь

$$\phi_{L_{\tau i}\mu}(T,p,t)=\frac{1}{T}\sum_{k=-\infty}^{\infty}L_{\tau i}(p+kj\omega)\mu(p+kj\omega)e^{kj\omega t},\quad \omega=\frac{2\pi}{T},$$

$$\mu(p)=\int_0^T h(t)e^{-pt}dt,$$

$$\tilde{R}_N(p)=\tilde{W}_d(p)\left[I-\tilde{D}_{N\mu}(T,p,-\tau_2)\tilde{W}_d(p)\right]^{-1},\tag{2.12}$$

$$\tilde{D}_{N\mu}(T,p,-\tau_2)=\frac{1}{T}\sum_{k=-\infty}^{\infty}N(p+kj\omega)\mu(p+kj\omega)e^{-(p+kj\omega)\tau_2},$$

$$\tilde{W}_d(p)=W_d(\zeta)\,\big|_{\zeta=e^{-pT}}\ .$$

В (2.11) и (2.12) использованы обозначения

$$K_{\tau i}(p)=K_i(p)e^{-p\tau_{K_i}},\quad L_{\tau i}(p)=L_i(p)e^{-p\tau_{Li}},\tag{2.13}$$

где $K_i(p)$, $L_i(p)$ – рациональные матрицы вида:

$$K_i(p)=C_i(pI-A)^{-1}B_1,\quad L_i(p)=C_i(pI-A)^{-1}B+D_i.\tag{2.14}$$

Помимо этого, в (2.11) и далее

$$M_\tau(p)=M(p)e^{-p\tau_M},\quad N_\tau(p)=N(p)e^{-p\tau_N},\tag{2.15}$$

где

$$M(p)=C(pI-A)^{-1}B_1,\quad N(p)=C(pI-A)^{-1}B.\tag{2.16}$$

Фигурирующие в (2.13), (2.15) постоянные τ_{Ki}, τ_{Li}, τ_M, τ_N определяются формулами:

$$\tau_{Ki}=\tau_M+\tau_{3i},\quad \tau_{Li}=\tau_N+\tau_{3i},$$
$$\tau_M=\tau_1,\quad \tau_N=\tau_2.\tag{2.17}$$

Из (2.9) следует, что при входе $x(t)$ и наблюдаемом выходе $\bar{z}(t)$ ППМ системы S_g имеет вид

$$W_{\bar{z}x}(p,t)=\begin{bmatrix}W_1(p,t)\\ \dots\\ W_\gamma(p,t)\end{bmatrix},\tag{2.18}$$

что с учетом (2.11) приводит к соотношению

$$W_{\bar{z}x}(p,t)=\phi_{\bar{L}_{\tau\mu}}(T,p,t)\tilde{R}_N(p)M_\tau(p)+\bar{K}_\tau(p),\tag{2.19}$$

где

$$\bar{K}'_\tau(p) = \left[\ K'_{\tau 1}(p)\ \ \ldots\ \ K'_{\tau\gamma}(p)\ \right], \quad \bar{L}'_\tau(p) = \left[\ L'_{\tau 1}(p)\ \ \ldots\ \ L'_{\tau\gamma}(p)\ \right] \tag{2.20}$$

и

$$\phi_{\bar{L}_{\tau\mu}}(T,p,t) = \frac{1}{T}\sum_{k=-\infty}^{\infty} \bar{L}_\tau(p+kj\omega)\mu(p+kj\omega)e^{kj\omega t}. \tag{2.21}$$

Далее правую часть формулы (2.19) будем называть стандартной формой ППМ для импульсной системы с множественными запаздываниями.

3. Стандартизируемость одноконтурной импульсной системы с множественными запаздываниями

Определение 1. Далее импульсную систему S^0 произвольной структуры со входом $x(t)$ и выходом $\bar{z}(t)$, состоящую из линейных стационарных элементов и импульсного регулятора вида (2.2)–(2.4)*, будем называть структурно стандартизируемой, если у нее существует ППМ $W^0_{\bar{z}x}(p,t)$ от входа $x(t)$ к выходу $\bar{z}(t)$, имеющая стандартную форму* (2.19)*.*

Определение 2. Структурно стандартизируемую систему S^0 с ППМ $W^0_{\bar{z}x}(p,t)$ будем называть стандартизируемой, если существует обобщенная стандартная система $\tilde{S}_g$, ППМ которой $\tilde{W}_{\bar{z}x}(p,t)$ совпадает с ППМ $W^0_{\bar{z}x}(p,t)$.

Далее обобщенную стандартную систему $\tilde{S}_g$ будем называть порождающей для стандартизируемой системы S^0. Ниже системы $\bar{S}_g$ и S^0 считаются эквивалентными. При этом на систему S^0 переносятся все качественные особенности системы $\tilde{S}_g$, установленные в [20].

В данном разделе изучается вопрос стандартизируемости многомерной одноконтурной системы S_0 с тремя запаздываниями, структура которой изображена на рис. 1.

На рис. 1 И.Р. – импульсный регулятор (2.2)–(2.4), $W_i(p)$, $i = 1,2,3$, – рациональные матрицы, свойства которых будут оговорены далее, χ^2 – положительная постоянная, τ_i, $i = 1,2,3$, – неотрицательные постоянные. В качестве

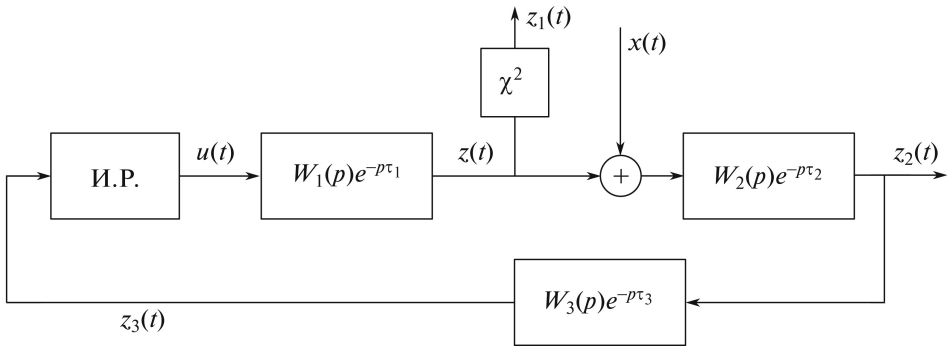


Рис. 1.

вектора наблюдаемого выхода будем рассматривать вектор

$$\bar{z}(t) = \begin{bmatrix} \chi^2 z(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix}. \tag{3.1}$$

Теорема 1. Для стандартизируемости системы S^0 при выборе наблюдаемого выхода в виде (3.1) *достаточно выполнения следующих условий*:

1) *рациональные матрицы*

$$N(p) \stackrel{\Delta}{=} W_3(p)W_2(p)W_1(p), \quad M(p) = W_3(p)W_2(p), \qquad \bar{K}(p) \stackrel{\Delta}{=} \begin{bmatrix} 0 \\ W_2(p) \end{bmatrix} \tag{3.2}$$

– *строго правильные;*

2) *рациональная матрица*

$$\bar{L}(p) \stackrel{\Delta}{=} \begin{bmatrix} \chi^2 W_1(p) \\ W_2(p)W_1(p) \end{bmatrix} \tag{3.3}$$

– *по меньшей мере правильная;*

3) *выполнено условие*

$$\mathrm{M\,deg} N(p) = \mathrm{M\,deg} W_1(p) + \mathrm{M\,deg} W_2(p) + \mathrm{M\,deg} W_3(p), \tag{3.4}$$

где M deg – *обозначение степени Мак-Миллана* [20, 21];

4) *для всех различных полюсов матрицы $N(p)$ выполнены условия вида* (2.8).

Доказательства теоремы 1 и последующих теорем 2 и 3 приведены в Приложении.

Замечание 1. Условие (3.4) означает отсутствие внутренних сокращений в произведении $W_3(p)W_2(p)W_1(p)$. В скалярном случае условие (3.4) равносильно несократимости этого произведения.

Замечание 2. Условия теоремы 1 не зависят от величин запаздываний τ_i, $i = 1, 2, 3$.

Если для системы S_0 не выполняется хотя бы одно из условий 1–3 теоремы 1, то она не является стандартизируемой.

4. H_2-оптимизация и фиксирующие полюса стандартизируемой системы S_0

Теорема 2. Пусть для системы S_0 выполнены условия теоремы 1. Тогда передаточная матрица $W_d^0(\zeta)$ H_2-оптимального алгоритма управления совпадает с передаточной матрицей H_2-оптимального алгоритма управления для одноконтурной системы $\hat{S}_0$, изображенной на рис. 2, где выполнено условие

$$\tau = \tau_1 + \tau_2 + \tau_3 \tag{4.1}$$

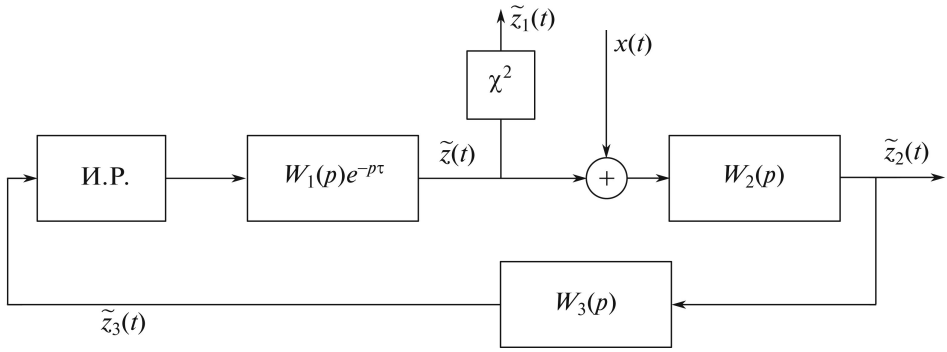


Рис. 2.

и вектор наблюдаемого выхода $\tilde{z}(t)$ *выбран в виде*

$$\tilde{z}(t) = \begin{bmatrix} \chi^2 \tilde{z}(t) \\ \tilde{z}_2(t) \end{bmatrix} = \begin{bmatrix} \tilde{z}_1(t) \\ \tilde{z}_2(t) \end{bmatrix}. \tag{4.2}$$

При выполнении условий теоремы 2 задача H_2-оптимизации системы $\tilde{S}_0$ сводится к решению аналогичной задачи для соответствующей порождающей расширенной стандартной системы $\bar{S}_\tau$, что может быть выполнено на основе результатов [19].

В соответствии с указанным в [20] на системы S_0 и $\tilde{S}_0$ распространяются все качественные особенности H_2-оптимальной порождающей системы $\bar{S}_\tau$. В частности, системы S_0 и $\tilde{S}_0$ имеют совпадающие множества фиксирующих полюсов.

Теорема 3. Пусть для системы S_0 *выполнены условия теоремы 1. Обозначим через* $\mathcal{M}_0$ *объединение множеств полюсов матриц* $W_1(p)$ *и* $W_3(p)$. *Также обозначим через* $\mathcal{M}_1$ *множество различных полюсов из множества* $\mathcal{M}_0$. *Пусть* $p_1, \ldots, p_\lambda$ *— элементы множества* $\mathcal{M}_1$. *Тогда при* $\operatorname{Re} p_i < 0$ *число* $\zeta_i = e^{-p_i T}$ *является полюсом* H_2*-оптимальной системы* S_0, *а при* $\operatorname{Re} p_i > 0$ *аналогичным свойством обладает число* $\zeta_i = e^{p_i T}$.

В соответствии с терминологией [21] $\mathcal{M}_1$ – это множество фиксирующих полюсов системы S_0, а множество чисел ζ_i – это множество фиксированных полюсов H_2-оптимальной системы. Из теоремы 3 следует практически важный вывод о том, что при проектировании системы S_0 на основе методов H_2-оптимизации необходимо накладывать определенные ограничения на свойства полюсов матриц $W_1(p)$ и $W_3(p)$. Если среди этих полюсов имеются полюса, лежащие вблизи мнимой оси, то степень устойчивости процессов в H_2-оптимальной системе может оказаться неудовлетворительной. Если же среди полюсов матриц $W_1(p)$ и $W_3(p)$ имеются полюса, лежащие на мнимой оси, то проектирование на основе методов H_2-оптимизации оказывается невозможным, поскольку H_2-оптимальная система оказывается на границе области устойчивости.

5. Заключение

В статье описан подход к решению задач стандартизируемости и H_2-оптимизации многомерных одноконтурных импульсных систем с несколькими запаздываниями, основанный на переходе к эквивалентной системе с одним запаздыванием. Рассмотрен способ построения множества фиксированных полюсов для H_2-оптимальной системы.

ПРИЛОЖЕНИЕ

Доказательство теоремы 1. Доказательство проводится в несколько этапов.

Покажем, прежде всего, что при выполнении условий теоремы 1 для системы S_0 существует ППМ $W_0(p,t)$, которая имеет стандартную формулу. Это означает, что система S_0 структурно стандартизируема. Для этого в соответствии с общим подходом [16–21] предполагаем, что на вход системы S_0 поступает матричный экспоненциальный сигнал

$$x(t) = e^{pt} I, \tag{П.1}$$

и находим режим функционирования системы S_0, в котором

$$z_i(p,t) = e^{pt} W_i(p,t), \quad i = 1,2,3, \tag{П.2}$$

где матрицы

$$W_i(p,t) = W_i(p,t+T), \quad i = 1,2,3, \tag{П.3}$$

— это ППМ от входа $x(t)$ к выходам $z_i(t)$, $i = 1,2,3$. В условиях теоремы 1 элементы матрицы $W_3(p,t)$ непрерывны относительно t. Поэтому, используя стробоскопическое свойство аналого-цифрового преобразователя (АЦП) (2.2), можно перейти к рассмотрению разомкнутой системы на рис. 3.

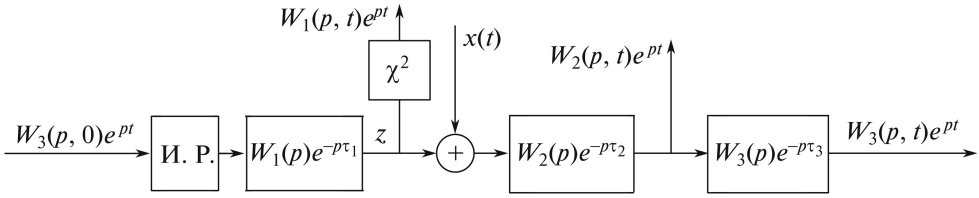


Рис. 3.

Используя (П.1)–(П.3), из рис. 3 можно получить после сокращения на e^{pt}

$$W_3(p,t) = \phi_{N_\tau \mu}(T,p,t)\tilde{W}_d(p)W_3(p,0) + W_3(p)W_2(p)e^{-p(\tau_2+\tau_3)}, \tag{П.4}$$

где

$$\phi_{N_\tau \mu}(T,p,t) = \frac{1}{T}\sum_{k=-\infty}^{\infty} N(p+kj\omega)\mu(p+kj\omega)e^{-(p+kj\omega)\tau}e^{kj\omega t}. \tag{П.5}$$

Кроме того,

$$N_\tau(p) = N(p)e^{-p\tau}, \tag{П.6}$$

где, как и ранее,

$$\tau = \tau_1 + \tau_2 + \tau_3. \tag{П.7}$$

Поскольку матрица (П.5) при всех p, не являющихся ее полюсами, непрерывна относительно t, то в левой и правой частях (П.4) можно положить $t = 0$. В результате приходим к равенству

$$W_3(p,0) = \phi_{N_\tau\mu}(T,p,0)\tilde{W}_d(p)W_3(p,0) + W_3(p)W_2(p)e^{-p(\tau_2+\tau_3)}. \tag{П.8}$$

Если учесть, что

$$\begin{gathered}\phi_{N_\tau\mu}(T,p,0) = \\ = \frac{1}{T}\sum_{k=-\infty}^{\infty} N(p+kj\omega)M(p+kj\omega)e^{-(p+kj\omega)\tau} = \tilde{D}_{N_\mu}(T,p,-\tau),\end{gathered} \tag{П.9}$$

то от (П.8) приходим к равенству

$$W_3(p,0) = \left[I - \tilde{D}_{N\mu}(T,p,-\tau)\tilde{W}_d(p)\right]^{-1} W_3(p)W_2(p)e^{-p(\tau_3+\tau_2)}. \tag{П.10}$$

Возвращаясь к рис. 3, с учетом (П.10) получаем

$$\begin{gathered}W_1(p,t) = \chi^2\phi_{W_{1\tau}\mu}(T,p,t)\tilde{R}_N(p)W_3(p)W_2(p)e^{-p(\tau_3+\tau_2)}, \\ W_2(p,t) = \phi_{W_{2\tau}W_{1\tau}\mu}(T,p,t)\tilde{R}_N(p)W_3(p)W_2(p)e^{-p(\tau_2+\tau_3)} + W_2(p)e^{-p\tau_2}, \\ \tilde{R}_N(p) = \tilde{W}_d(p)\left[I - \tilde{D}_{N\mu}(T,\beta,-\tau)\tilde{W}_d(p)\right]^{-1}.\end{gathered} \tag{П.11}$$

Используя (П.11) и (2.18), после несложных преобразований можно установить, что ППМ системы S_0 от входа $x(t)$ к выходу (3.1) имеет стандартную форму (2.19) с матрицами $\bar{K}(p)$, $\bar{L}(p)$, $M(p)$, $N(p)$, которые определены формулами (3.2), (3.3), а постоянные τ_N, τ_M, τ_{Li}, τ_{Ki} равны

$$\begin{gathered}\tau_N = \tau_1 + \tau_2 + \tau_3 = \tau, \quad \tau_M = \tau_2 + \tau_3, \\ \tau_{L1} = \tau_1, \quad \tau_{L2} = \tau_1 + \tau_2, \\ \tau_{K1} = 0, \quad \tau_{K2} = \tau_2.\end{gathered} \tag{П.12}$$

Из доказанного следует, что система S_0 структурно стандартизируема.

Для доказательства стандартизируемости системы S_0 покажем, что при выполнении условий теоремы 1 выполнены общие необходимые и достаточные условия стандартизируемости, приведенные в теореме 1 из [20]. Условия а) и б) этой теоремы выполняются очевидным образом. Выполнение условий (2.17) сразу следует из (П.12).

Условие 1 выполнено по определению. Остается доказать, что условие в) теоремы 1 из [20] также выполняется, т.е. что из условия (3.4) вытекает равенство

$$\text{(П.13)} \qquad \mathrm{M}\deg N(p) = \mathrm{M}\deg \begin{bmatrix} \bar{K}(p) & \bar{L}(p) \\ M(p) & N(p) \end{bmatrix} \stackrel{\Delta}{=} \mathrm{M}\deg W_0(p).$$

Для этого отметим, что с учетом (3.2) и (3.3) имеем

$$\text{(П.14)} \qquad W_0(p) = \left[\begin{array}{c|c} \bar{K}(p) & \bar{L}(p) \\ \hline M(p) & N(p) \end{array}\right] = \left[\begin{array}{c|c} 0 & \chi^2 W_1(p) \\ W_2(p) & W_2(p)W_1(p) \\ \hline W_3(p)W_2(p) & W_3(p)W_2(p)W_1(p) \end{array}\right].$$

Ведем обозначения:

$$\text{(П.15)} \qquad \mathrm{M}\deg W_i(p) \stackrel{\Delta}{=} \psi_i, \quad i = 1,2,3, \quad \psi_1 + \psi_2 + \psi_3 \stackrel{\Delta}{=} \psi.$$

Из свойств степени Мак-Миллана, (П.14) и (3.4) сразу следует, что

$$\text{(П.16)} \qquad \begin{gathered} \mathrm{M}\deg W_0(p) \geqslant \mathrm{M}\deg N(p) = \\ = \mathrm{M}\deg W_1(p) + \mathrm{M}\deg W_2(p) + \mathrm{M}\deg W_3(p) = \psi. \end{gathered}$$

С другой стороны, матрицу $W_0(p)$ можно представить в виде произведения

$$\text{(П.17)} \qquad W_0(p) = \bar{W}_3(p)\bar{W}_2(p)\bar{W}_1(p),$$

где

$$\text{(П.18)} \qquad \begin{gathered} \bar{W}_3(p) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & W_3(p) \end{bmatrix}, \quad \bar{W}_2(p) = \begin{bmatrix} 0 & \chi^2 I \\ W_2(p) & W_2(p) \\ W_2(p) & W_2(p) \end{bmatrix}, \\ \bar{W}_1(p) = \begin{bmatrix} I & 0 \\ 0 & W_1(p) \end{bmatrix}. \end{gathered}$$

Покажем, что справедливы равенства

$$\text{(П.19)} \qquad \mathrm{M}\deg \bar{W}_i(p) = \psi_i, \quad i = 1,2,3.$$

Для этого отметим, что из (П.18) имеем

$$\text{(П.20)} \qquad \mathrm{M}\deg \bar{W}_i(p) \geqslant \psi_i, \quad i = 1,2,3.$$

С другой стороны, из (П.15) вытекает существование представлений MFD (Matrix fraction description) [22]

$$\text{(П.21)} \qquad W_i(p) = a_i^{-1}(p) b_i(p), \quad i = 1,2,3,$$

где

$$\deg\det a_i(p) = \psi_i, \quad i = 1, 2, 3. \tag{П.22}$$

С помощью (П.21) и (П.18) при $i = 3$ находим MFD

$$\bar{W}_3(p) = \bar{a}_3^{-1}(p)\bar{b}_3(p), \tag{П.23}$$

где

$$\bar{a}_3(p) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & a_3(p) \end{bmatrix}, \quad \bar{b}_3(p) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & b_3(p) \end{bmatrix}. \tag{П.24}$$

Поскольку

$$\deg\det \bar{a}_3(p) = \deg\det a_3(p) = \psi_3, \tag{П.25}$$

то из свойств непонижаемых MFD [22] следует, что

$$\mathrm{M}\deg\bar{W}_3(p) \leqslant \psi_3. \tag{П.26}$$

Сопоставляя (П.26) и (П.20), приходим к равенству (П.19) при $i = 3$. Для случая $i = 1$ доказательство (П.20) проводится аналогично. Остается рассмотреть случай $i = 2$. Очевидно, что имеем

$$\bar{W}_2(p) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & I \\ 0 & 0 & I \end{bmatrix} \bar{W}_4(p), \tag{П.27}$$

где

$$\bar{W}_4(p) = \begin{bmatrix} 0 & \chi^2 I \\ 0 & 0 \\ W_2(p) & W_2(p) \end{bmatrix}. \tag{П.28}$$

Из (П.27) следует, что

$$\mathrm{M}\deg\bar{W}_2(p) = \mathrm{M}\deg W_4(p), \tag{П.29}$$

так как первый сомножитель в (П.27) – постоянная несингулярная матрица. Используя MFD (П.21) при $i = 2$, можно получить MFD

$$\bar{W}_2(p) = \bar{a}_2^{-1}\bar{b}_2(p), \tag{П.30}$$

где

$$\deg\det \bar{a}_2(p) = \psi_2. \tag{П.31}$$

Из (П.31), (П.30) следует, что

$$\mathrm{M}\deg\bar{W}_2(p) \leqslant \psi_2. \tag{П.32}$$

Сопоставляя (П.32) и (П.20), получаем, что

$$\mathrm{M}\deg\bar{W}_2(p) = \psi_2, \tag{П.33}$$

что завершает доказательство формул (П.19). Из (П.19) и (П.17) вытекает, что

$$\mathrm{M}\deg W_0(p) \leqslant \psi_1 + \psi_2 + \psi_3 = \psi. \tag{П.34}$$

С другой стороны, из (П.14), (П.15) находим, что

$$\mathrm{M}\deg W_0(p) \geqslant \mathrm{M}\deg N(p) = \psi_1 + \psi_2 + \psi_3 = \psi. \tag{П.35}$$

Сопоставление (П.34) и (П.35) приводит к равенству

$$\mathrm{M}\deg W_0(p) = \mathrm{M}\deg N(p) = \psi, \tag{П.36}$$

что завершает доказательство теоремы 1.

Доказательство теоремы 2. С помощью выкладок, аналогичных использованным при доказательстве теоремы 1, устанавливается, что ППМ системы $\tilde{S}_0$ от входа $x(t)$ к выходам $\tilde{z}_i(t)$ имеют вид:

$$\begin{gathered} W_{\tilde{z}_1 x}(p,t) = \chi^2 \psi_{W_{1\tau}\mu}(T,p,t)\tilde{R}_N(p)W_3(p)W_2(p), \\ W_{\tilde{z}_2 x}(p,t) = \chi^2 \psi_{W_{21\tau}\mu}(T,p,t)\tilde{R}_N(p)W_3(p)W_2(p) + W_2(p), \end{gathered} \tag{П.37}$$

где использованы обозначения:

$$\begin{gathered} W_{1\tau}(p) = W_1(p)e^{-p\tau}, \quad W_{21\tau}(p) = W_2(p)W_1(p)e^{-p\tau}, \\ \tilde{R}_N(p) = \tilde{W}_d(p)\left[I - \tilde{D}_{N\mu}(T,p,-\tau)\tilde{W}_d(p)\right]^{-1}. \end{gathered} \tag{П.38}$$

Поэтому ППМ $W_{\tilde{z}x}(p,t)$ от входа $x(t)$ к выходу $\tilde{z}(t)$ (4.2) имеет стандартную форму

$$W_{\tilde{z}x}(p,t) = \phi_{\bar{L}_\tau\mu}(T,p,t)\tilde{R}_N(p)M(p) + \bar{K}(p). \tag{П.39}$$

Здесь

$$\bar{L}_\tau(p) = \bar{L}(p)e^{-p\tau} \tag{П.40}$$

и аналогично предыдущему

$$\bar{K}(p) = \begin{bmatrix} K_1(p) \\ K_2(p) \end{bmatrix}, \quad \bar{L}(p) = \begin{bmatrix} L_1(p) \\ L_2(p) \end{bmatrix}, \tag{П.41}$$

где

$$K_1(p) = 0, \quad K_2(p) = W_2(p),$$

$$L_1(p) = \chi^2 W_1(p), \quad L_2(p) = W_2(p)W_1(p), \tag{П.42}$$

$$M(p) = W_3(p)W_2(p), \quad N(p) = W_3(p)W_2(p)W_1(p).$$

Из приведенных соотношений и теоремы 4 из [20] следует, что системы S_0 и $\bar{S}_0$ – H_2-эквивалетны в смысле [20] и им соответствует единая порождающая расширенная стандартная система $\bar{S}_\tau$.

Доказательство теоремы 3. Пусть имеем непонижаемые левые и правые MFD (ILMFD и IRMFD) [22]

$$W_i(p) = a_{\ell i}^{-1}(p) b_{\ell i}(p) = b_{ri}(p) a_{ri}^{-1}(p), \quad i = 1, 2, 3, \tag{П.43}$$

где в силу (П.15)

$$\deg \det a_{\ell i}(p) = \deg \det a_{ri}(p) = \psi_i. \tag{П.44}$$

Из (П.42) и (П.43) следует

$$M(p) = a_{\ell 3}^{-1}(p) b_{\ell 3}(p) a_{\ell 2}^{-1}(p) b_{\ell 2}(p). \tag{П.45}$$

Из (3.4) вытекает

$$\mathrm{M}\deg M(p) = \psi_2 + \psi_3.$$

Поэтому существует ILMFD

$$b_{\ell 3}(p) a_{\ell 2}^{-1}(p) = a_{\ell 4}^{-1}(p) b_{\ell 4}(p), \tag{П.46}$$

в котором

$$\det a_{\ell 4}(p) \sim \det a_{\ell 2}(p), \tag{П.47}$$

где символ $\sim$ означает эквивалентность полиномиальных матриц [22, 23] и, в частности, скалярных полиномов. Поэтому справедливо равенство

$$\deg \det a_{\ell 4}(p) = \deg \det a_{\ell 2}(p) = \psi_2. \tag{П.48}$$

С помощью (П.46) получаем MFD

$$M(p) = a_M^{-1}(p) B_M(p), \tag{П.49}$$

где

$$a_M(p) = a_{\ell 4}(p) a_{\ell 3}(p), \quad b_M(p) = b_{\ell 4}(p) b_{\ell 2}(p), \tag{П.50}$$

причем

$$\deg\det a_M(p) = \psi_2 + \psi_3 = \mathrm{M}\deg M(p). \tag{П.51}$$

Равенство (П.51) означает, что правая часть (П.49) – это ILMFD. Продолжая вычисления, используем равенство

$$N(p) = W_3(p)W_2(p)W_1(p) = M(p)W_1(p). \tag{П.52}$$

С помощью (П.43) при $i = 1$ и (П.49) получаем, что

$$N(p) = a_M^{-1}(p)b_M(p)a_1^{-1}(p)b_1(p). \tag{П.53}$$

При выполнении (3.4) существует ILMFD

$$b_M(p)a_1^{-1}(p) = a_{\ell 5}^{-1}(p)b_{\ell 5}(p), \tag{П.54}$$

где

$$\begin{gathered}\det a_{\ell 5}(p) \sim \det a_{\ell 1}(p),\\ \deg\det a_{\ell 5}(p) = \deg\det a_{\ell 1}(p) = \psi_1.\end{gathered} \tag{П.55}$$

С учетом (П.54) из (П.52) получается ILMFD

$$N(p) = a_{\ell N}^{-1}(p)b_{\ell N}(p), \tag{П.56}$$

где

$$a_{\ell N}(p) = a_{\ell 5}(p)a_M(p), \quad b_{\ell N}(p) = b_{\ell 5}(p)b_M(p), \tag{П.57}$$

причем по построению

$$\deg\det a_{\ell N}(p) = \psi_1 + \psi_2 + \psi_3 = \mathrm{M}\deg N(p). \tag{П.58}$$

Из (П.55) и (П.57) в силу доказанного в [21] следует, что полюса матрицы $W_1(p)$ являются фиксирующими.

Для доказательства утверждения, относящегося к матрице $W_3(p)$, используем (3.3) в виде

$$\bar{L}(p) = \begin{bmatrix} \chi^2 I \\ W_2(p) \end{bmatrix} W_1(p). \tag{П.59}$$

Аналогично предыдущему с помощью (П.59) устанавливается, что

$$\mathrm{M}\deg\bar{L}(p) \leqslant \psi_2 + \psi_1. \tag{П.60}$$

В то же время из (3.3) имеем

$$\mathrm{M}\deg\bar{L}(p) \geqslant \psi_1 + \psi_2. \tag{П.61}$$

Сопоставляя (П.60) и (П.61), находим, что

$$\mathrm{M}\deg\bar{L}(p) = \psi_1 + \psi_2. \tag{П.62}$$

Из (П.62) вытекает существование IRMFD

$$\bar{L}(p) = b_{r\bar{L}}(p)a^{-1}_{r\bar{L}}(p), \tag{П.63}$$

где

$$\deg\det a_{r\bar{L}}(p) = \psi_1 + \psi_2. \tag{П.64}$$

Из (П.63) следует, что произведение

$$\bar{L}(p)a_{r\bar{L}}(p) = \begin{bmatrix} \chi^2 W_1(p) \\ W_2(p)W_1(p) \end{bmatrix} a_{r\bar{L}}(p) \tag{П.65}$$

— полиномиальная матрица. Следовательно, произведение

$$W_2(p)W_1(p)a_{r\bar{L}}(p) \stackrel{\Delta}{=} b_L(p) \tag{П.66}$$

— тоже полиномиальная матрица. Из (П.66) находим правое MFD

$$W_2(p)W_1(p) = b_L(p)a^{-1}_{r\bar{L}}(p), \tag{П.67}$$

которое в силу (3.4) непонижаемо. С помощью (П.43) и (П.67) находим

$$N(p) = b_{r3}(p)a^{-1}_{r3}(p)b_L(p)a^{-1}_{r\bar{L}}(p). \tag{П.68}$$

При выполнении (3.4) имеем

$$a^{-1}_{r3}(p)b_L(p) = b_{r4}(p)a^{-1}_{r4}(p), \tag{П.69}$$

где правая часть – IRMFD. Подставляя (П.69) в (П.68), приходим к IRMFD

$$N(p) = b_{rN}(p)a^{-1}_{rN}(p), \tag{П.70}$$

где

$$a_{rN}(p) = a_{r\bar{L}}(p)a_{r4}(p), \quad b_{rN}(p) = b_{\bar{L}}(p)b_{r4}(p). \tag{П.71}$$

Сопоставляя IRMFD (П.66) и (П.70) на основании результатов [21] получаем, что полюса матрицы $W_3(p)$ являются фиксирующими.

СПИСОК ЛИТЕРАТУРЫ

1. *Kwakernaak H., Sivan R.* Linear Optimal Control Systems. N.Y.: Wiley-Interscience, a Division of John Wiley & Sons, Inc., 1972.
2. *Ackermann J.* Abtastregelung. Berlin: Springer-Verlag, 3 ed., 1988.
3. *Astrom K.J., Wittenmark B.* Computer Controlled Systems: Theory and Design. Englewood Cliffs, NJ: Prentice-Hall, 3rd ed., 1997.
4. *Chen T., Francis B.A.* Optimal sampled-data control systems. Berlin–Heidelberg–N.Y.: Springer-Verlag, 1995.
5. *Fridman E., Shaked U.* Sampled-Data $\mathcal{H}_\infty$ State Feedback Control of Systems with State Delays // Int. J. Control. 2000. V. 73. No. 12. P. 1115–1128.
6. *Khargonekar P.P., Yamamoto J.* Delayed Signal Reconstruction Using Sampled-Data Control // Proc. 35th IEEE Conf. on Decision Contr. Kyoto. 1996. P. 1259–1263.
7. *Yamamoto Y., Hara S.* Performance Lower Bound for a Sampled-Data Signal Reconstruction / V. Blondel, E. Sontag, M. Vidyasagar, J. Willems eds. Open Problems in Mathematical Systems and Control Theory. London: Springer-Verlag, 1998. P. 277–279.
8. *Lennartson B.* Sampled-Data Control for Time-Delayed Plants // Int. J. Control. 1989. V. 49. P. 1601–1614.
9. *Hara S., Fujioka H., Kabamba P.T.* A Hybrid State-Space Approach to Sampled-Data Feedback Control // Linear Algebra Appl. 1994. P. 679–712.
10. *Wittenmark B.* Sampling of a System with Time Delay // IEEE Trans. Autom. Control. May 1985. V. AC-30. P. 507–510.
11. *Jugo J.* Discretization of Continuous Time-Delay Systems // Proc. 15th IFAC Triennial World Congr. V. Linear systems/Time-delay systems. P. REG1450, Barcelona, 2002.
12. *Polyakov K.Yu.* $\mathcal{H}_2$-optimal Sampled-Data Control of Plants with Multiple Input and Output Delays // Asian J. Control. June 2006. V. 8. No. 2. P. 107–116.
13. *Emilia Fridman* Introduction to Time-Delay Systems. Analysis and Control // Cham–Heidelberg–N.Y.–Dordrecht–London: Springer, 2014.
14. *Mirkin L., Shima T., Tadmor G.* Analog Loop Shifting in $\mathcal{H}_2$ Optimization of Input-Delay Sampled-Data Systems // 52nd IEEE Conf. on Decision and Control. December 10–13, 2013, Florence, Italy.
15. *Mirkin L., Shima T., Tadmor G.* Sampled-Data $\mathcal{H}_2$ Optimization of Systems with I/O Delays via Analog Loop Shifting // IEEE Trans. Autom. Control. March 2014. V. 59. No. 3. P. 787–791.
16. *Розенвассер Е.Н.* Линейная теория цифрового управления в непрерывном времени. М.: Наука, 1994.
17. *Rosenwasser E.N., Lampe B.P.* Digitale Regelung in kontinuierlicher Zeit – Analyse und Entwurf im Frequenzbereich. B.G. Teubner, Stuttgart, 1997.
18. *Rosenwasser E.N., Lampe B.P.* Computer Controlled Systems – Analysis and Design with Process-orientated Models. London–Berlin–Heidelberg: Springer-Verlag, 2000.
19. *Лямпе Б.П., Розенвассер Е.Н.* H_2-оптимизация импульсных систем с запаздыванием на основе метода параметрической передаточной матрицы // АиТ. 2010. № 1. С. 49–69.

    Lampe B.P., Rosenwasser E.N. H_2-optimization of Time-Delayed Sampled-Data Systems on the Basis of the Parametric Transfer Matrix Method // Autom. Remote Control. 2010. V. 71. No. 1. P. 49–69.

20. *Розенвассер Е.Н., Лямпе Б.П., Древело В., Яйнш Т.* Стандартизируемость и H_2-оптимизация импульсных систем с множественными запаздываниями // АиТ. 2019. № 3. С. 26–44.

Rosenwasser E.N., Lampe B.P., Drewelow W., Jeinsch T. Standardizability and H_2-Optimization of Sampled-Data Systems with Multiple Delays // Autom. Remote Control. 2019. V. 80. No. 3. P. 413–428.

21. *Rosenwasser E.N., Lampe B.P.* Multivariable Computer-Controlled Systems — A Transfer Function Approach. London: Springer, 2006.

22. *Kailath T.* Linear Systems. Englewood Cliffs, NJ: Prentice Hall, 1980.

23. *Гантмахер Ф.Р.* Теория матриц. М.: ГИТТЛ, 1954.

Статья представлена к публикации членом редколлегии А.А. Бобцовым.

Поступила в редакцию 01.03.2019

После доработки 29.04.2019

Принята к публикации 18.07.2019

Нелинейные системы

© 2020 г. М.Г. ЮМАГУЛОВ, д-р физ.-мат. наук (yum_mg@mail.ru)
(Башкирский государственный университет, Уфа),
М.Ф. ФАЗЛЫТДИНОВ (fazlitdin_marat@mail.ru)
(ООО Газпромнефть НТЦ, Санкт-Петербург)

ПРИБЛИЖЕННЫЕ ФОРМУЛЫ И АЛГОРИТМЫ ПОСТРОЕНИЯ ЦЕНТРАЛЬНЫХ МНОГООБРАЗИЙ ДИНАМИЧЕСКИХ СИСТЕМ

Предлагаются новые подходы, позволяющие получить аппроксимации второго и более высоких порядков центральных многообразий негиперболических точек равновесия динамических систем с непрерывным и дискретным временем. Полученные формулы приводят к новым конструктивным алгоритмам построения центральных многообразий. Предлагаемые формулы и алгоритмы носят общий характер в том смысле, что они позволяют строить центральные многообразия в терминах исходных уравнений и применимы к ситуациям, когда матрица линеаризации имеет произвольный порядок вырождения.

Ключевые слова: динамические системы, точка равновесия, центральное многообразие, бифуркация, аппроксимация, приближенные формулы, алгоритмы.

DOI: 10.31857/S0005231020010031

1. Введение и постановка задачи

Важным объектом изучения нелинейных динамических систем является центральное многообразие негиперболической точки равновесия. В фазовом пространстве системы это многообразие локально инвариантно для ее траекторий; оно содержит точку равновесия и касается в ней соответствующего подпространства линеаризации системы. В естественном смысле вся нетривиальная динамика системы в окрестности точки равновесия сосредоточена на центральном многообразии. Этот факт следует из теоремы о центральном многообразии (см. [1–7]) и принципа сведения А.Н. Шошитайшвили [8], позволяющего при изучении динамики системы избавиться от “гиперболических переменных”, оказывающих вполне предсказуемое влияние, и свести задачу к исследованию системы на центральном многообразии.

Теория центрального многообразия находит многочисленные приложения во многих задачах теории динамических систем, нелинейной динамики, механики, теории управления и др. Ограничимся здесь ссылками на работы [9–11], в которых методы этой теории использовались в задачах моделирования управляемых процессов, в задачах параметрической идентификации и др. Сочетание теории центрального многообразия с теорией нормальных

форм Пуанкаре [2, 4, 12] широко используется и для изучения задач о бифуркациях в динамических системах [2–4, 7, 13, 14].

Так как центральное многообразие и динамика системы на нем, как правило, не могут быть точно рассчитаны, то актуальными являются разработки соответствующих аппроксимаций. При этом практический интерес, как правило, представляют аппроксимации второго и третьего порядков. Имеющиеся в литературе подходы, позволяющие получать приближенное представление центрального многообразия, как правило, направлены на решение конкретных задач. В [7] предложены общие подходы и формулы в ситуациях, когда матрица линеаризации имеет порядок вырождения, равный 1 или 2. Использование этих формул для исследования конкретных уравнений, как правило, требует предварительного преобразования исходных уравнений, что далеко не всегда является тривиальной задачей. Следует также отметить подходы, основанные на применении современной компьютерной техники и пакетов символьных вычислений (см., например, [15]). Эти подходы позволили существенно продвинуться в построении центральных многообразий, в частности в задаче вычисления аппроксимаций высоких порядков.

В настоящей работе предлагается общая схема, позволяющая получить новые приближенные формулы для центральных многообразий динамических систем в терминах исходных уравнений. Предлагаемая схема носит общий характер и в том смысле, что она применима к ситуациям, когда матрица линеаризации имеет произвольный порядок вырожденности.

Основными объектами исследования в статье являются динамические системы с непрерывным временем:

$$\frac{dx}{dt} = Ax + a(x), \quad x \in R^N, \tag{1.1}$$

и динамические системы с дискретным временем:

$$x_{n+1} = Ax_n + a(x_n), \quad n = 0, 1, 2, \ldots, \quad x_n \in R^N, \tag{1.2}$$

в которых A — квадратная матрица, а функция $a(x)$ является C^m-гладкой ($m \geqslant 2$) и удовлетворяет равенствам $a(0) = 0$ и $a'(0) = 0$. Предполагается, что точки равновесия $x = 0$ систем (1.1) и (1.2) являются негиперболическими: для системы (1.1) матрица A имеет одно или несколько чисто мнимых собственных значений, а для системы (1.2) матрица A имеет одно или несколько собственных значений, равных 1 по модулю.

2. Динамические системы с непрерывным временем

Рассмотрим сначала систему (1.1). Если ее точка равновесия $x = 0$ является гиперболической (т.е. матрица A не имеет собственных значений на мнимой оси), то по теореме Гробмана – Хартмана (см., например, [1, 3]) в малой окрестности точки $x = 0$ фазовый портрет нелинейной системы (1.1) топологически эквивалентен фазовому портрету линейной системы

$$x' = Ax, \quad x \in R^N. \tag{2.1}$$

Другими словами, качественное поведение решений нелинейной автономной системы (1.1) в окрестности гиперболической точки равновесия $x = 0$ полностью определяется поведением решений соответствующей линейной системы (2.1).

2.1. Теорема о центральном многообразии

Всюду ниже будем предполагать, что точка равновесия $x = 0$ системы (1.1) является негиперболической, т.е. матрица A имеет одно или несколько чисто мнимых собственных значений. В этом случае теорема Гробмана – Хартмана уже неприменима. Другими словами, для описания поведения траекторий нелинейной системы (1.1) вблизи точки $x = 0$ недостаточно анализа только линеаризованной системы (2.1). Здесь необходимо учитывать нелинейные члены системы (1.1).

Пусть спектр σ матрицы A состоит из двух непустых частей: $\sigma = \sigma_0 \cup \sigma^0$, где σ_0 содержит собственные значения, вещественные части которых равны нулю, а σ^0 — остальные собственные значения. Множество σ^0 также состоит из двух частей (одна из которых может быть пустым множеством): $\sigma^0 = \sigma_- \cup \sigma^+$, где множество σ_- содержит собственные значения, вещественные части которых отрицательны, а σ^+ — собственные значения, вещественные части которых положительны. Обозначим через E_0, E_- и E^+ — корневые подпространства матрицы A, отвечающие соответственно частям σ_0, σ_- и σ^+ ее спектра. Пусть k_0, k_- и k^+ — это размерности подпространств E_0, E_- и E^+; тогда $k_0 + k_- + k^+ = N$ и $1 \leqslant k_0 \leqslant N - 1$. Пространство R^N представляется в виде прямой суммы $R^N = E_0 \bigoplus E_- \bigoplus E^+$ инвариантных для оператора $A : R^N \to R^N$ подпространств E_0, E_- и E^+.

Имеет место следующая *теорема о центральном многообразии* (см. [1–7]).

Теорема 1. Существует δ_0-окрестность $T(0, \delta_0)$ точки $x = 0$ такая, что система (1.1) *имеет в шаре $T(0, \delta_0)$:*

— единственные C^m-гладкие инвариантные k_--мерное устойчивое многообразие W_s и k^+-мерное неустойчивое многообразие W_u;

— C^m-гладкое инвариантное k_0-мерное многообразие W_c.

Эти многообразия пересекаются только в точке $x = 0$ и касаются в ней подпространств E_-, E^+ и E_0 соответственно.

Инвариантность многообразий W_s и W_u для системы (1.1) означает, что если некоторая ее траектория в некоторый момент времени находится на многообразии W_s (или W_u), то она будет находиться на W_s (или W_u) и во все последующие моменты времени до тех пор, пока эта траектория остается в шаре $T(0, \delta_0)$. Устойчивость многообразия W_s означает, что все траектории системы (1.1), начинающиеся на W_s, стремятся при $t \to +\infty$ к точке $x = 0$. Соответственно, неустойчивость многообразия W_u означает, что все траектории системы (1.1), начинающиеся в ненулевой точке многообразия W_u, за конечное время покидают шар $T(0, \delta_0)$, а при $t \to -\infty$ стремятся к точке $x = 0$.

Многообразие W_c называют *центральным многообразием* системы (1.1). Инвариантность многообразия W_c для системы (1.1) означает, что если некоторая ее траектория в некоторый момент времени находится на многообра-

зии W_c, то она будет находиться на W_c и во все последующие моменты времени до тех пор, пока эта траектория остается в шаре $T(0,\delta_0)$.

2.2. Вспомогательные сведения

Приведем некоторые вспомогательные сведения относительно свойств центрального многообразия (см., например, [1–7]).

Положим $E^0 = E_- \bigoplus E^+$, т.е. E^0 — это корневое подпространство матрицы A, отвечающее части σ^0 ее спектра. Размерность подпространства E^0 равна $k^0 = k_- + k^+$. Пространство R^N представляется в виде прямой суммы $R^N = E_0 \bigoplus E^0$ инвариантных для оператора $A : R^N \to R^N$ подпространств E_0 и E^0. Обозначим, наконец, через $P_0 : R^N \to E_0$ и $P^0 : R^N \to E^0$ соответствующие операторы проектирования.

Центральное многообразие W_c системы (1.1) может быть задано уравнением вида $v = \psi(u)$, где $u \in E_0$, $v \in E^0$, а функция $\psi(u)$ является гладкой и удовлетворяет равенствам $\psi(0) = 0$, $\psi'(0) = 0$. Другими словами, центральное многообразие W_c может быть локально (в малой окрестности точки $x = 0$) описано равенством

$$W_c = \Big\{x :\, x = u + \psi(u) \;\mid\; u \in E_0,\;\; \psi(u) \in E^0,\;\; \psi(0) = 0, \psi'(0) = 0\Big\}. \tag{2.2}$$

Упомянутый выше принцип сведения А.Н. Шошитайшвили здесь состоит в том, что задача о поведении решений N-мерного уравнения (1.1) в окрестности точки $x = 0$ может быть сведена к аналогичной задаче для k_0-мерного уравнения:

$$u' = Au + P_0 a(u + \psi(u)), \quad u \in E_0. \tag{2.3}$$

Уравнение (2.3) содержит все основные особенности, присущие исходному уравнению (1.1).

Замечание 1. Центральное многообразие W_c системы (1.1), вообще говоря, не является единственным. Однако все возможные центральные многообразия имеют совпадающие тейлоровские разложения соответствующих функций $v = \psi(u)$ в точке $u = 0$, т.е. все эти функции имеют одинаковые производные $\psi^j(0)$ $(j = 0, 1, \ldots, k)$. Другими словами, все центральные многообразия мало отличаются друг от друга. При этом каждое из этих многообразий содержит все ограниченные решения системы (1.1), содержащиеся в шаре $T(0,\delta_0)$. В частности, они содержат все точки равновесия, периодические, гомо- и гетероклинические орбиты, лишь бы их траектории располагались в малой окрестности точки $x = 0$. Поэтому для изучения задачи о таких решениях можно выбирать любое из центральных многообразий.

Замечание 2. Если правая часть системы (1.1) является аналитической (а именно, когда функция $a(x)$ в некоторой окрестности точки $x = 0$ представима в виде сходящегося ряда Тейлора, начинающегося со второй степени), то эта система не может иметь более одного аналитического центрального многообразия. Если при этом аналитического центрального многообразия

система не имеет, то ряд Тейлора для функции $v = \psi(u)$, определяющей центральное многообразие W_c и вычисленный в точке $u = 0$, расходится в любой окрестности этой точки. Тем не менее в силу теоремы 1 частичные суммы этого ряда могут давать хорошую аппроксимацию центрального многообразия W_c.

2.3. Схема построения центрального многообразия

Перейдем к задаче приближенного построения центрального многообразия (2.2) системы (1.1), а именно, функции $v = \psi(u)$. Ниже будет использоваться следующее утверждение (см., например, [4]).

Теорема 2. Функция $v = \psi(u)$ (такая, что $\psi(0) = 0$, $\psi'(0) = 0$) описывает центральное многообразие W_c системы (1.1) *тогда и только тогда, когда она в некоторой окрестности точки $u = 0$ подпространства E_0 является решением уравнения*

$$\psi'(u)[Au + P_0 a(u + \psi(u))] = A\psi(u) + P^0 a(u + \psi(u)). \tag{2.4}$$

Функцию $v = \psi(u)$ будем строить в виде

$$\psi(u) = \psi_2(u) + \psi_3(u) + \ldots + \psi_s(u) + \widehat{\psi}(u), \tag{2.5}$$

где $\psi_j(u)$ — однородные функции порядка j, определенные в малой окрестности точки $u = 0$ подпространства E_0 и принимающие значения в E^0, а функция $\widehat{\psi}(u)$ является C^m-гладкой и удовлетворяет соотношению $\|\widehat{\psi}(u)\| = O(\|u\|^{s+1})$ при $u \to 0$.

Ограничимся приведением схемы построения функций $\psi_2(u)$ и $\psi_3(u)$; построение последующих функций $\psi_j(u)$ проводится по той же схеме. В этой связи будем считать, что нелинейность в правой части уравнения (1.1) представима в виде $a(x) = a_2(x) + a_3(x) + \widehat{a}_4(x)$, где $a_2(x)$ содержит квадратичные по x слагаемые, $a_3(x)$ — слагаемые третьей степени, а $\widehat{a}_4(x)$ является C^m-гладкой и удовлетворяет соотношению $\|\widehat{a}_4(x)\| = O(\|x\|^4)$ при $x \to 0$.

Имеет место следующая

Лемма 1. Функции $\psi_2(u)$ и $\psi_3(u)$ являются решениями уравнений

$$\psi_2'(u)Au - A\psi_2(u) = P^0 a_2(u), \tag{2.6}$$

$$\psi_3'(u)Au - A\psi_3(u) = -\psi_2'(u)P_0 a_2(u) + P^0[a_2'(u)\psi_2(u) + a_3(u)]. \tag{2.7}$$

Справедливость этого утверждения устанавливается простым подсчетом путем подстановки (2.5) в (2.4).

С целью изучения вопроса о разрешимости уравнений (2.6) и (2.7) обозначим через через F_p множество однородных порядка p (p — натуральное число) функций $\psi(u)$, определенных в подпространстве E_0 и принимающих значения в подпространстве E^0, т.е.

$$F_p = \Big\{\psi(u) \mid \ \psi : E_0 \to E^0, \ \psi(\alpha u) \equiv \alpha^p \psi(u)\Big\}. \tag{2.8}$$

Для каждого p множество F_p образует линейное пространство с обычными операциями сложения элементов и умножения на вещественные числа. Далее, через L обозначим действующий в пространстве F_p линейный оператор, сопоставляющий каждой функции $\psi(u) \in F_p$ функцию $L\psi(u) \in F_p$, определенную равенством

$$L\psi(u) = \psi'(u)Au - A\psi(u). \tag{2.9}$$

Конечно, оператор L будет зависеть от p. Однако для простоты будем использовать одно и то же обозначение L для всех действующих в пространствах F_p операторов (2.9) независимо от значения p.

Уравнения (2.6) и (2.7) однотипны, имея вид

$$L\psi(u) = b(u), \tag{2.10}$$

относительно неизвестной функции $\psi(u) \in F_p$; здесь L — оператор (2.9). При этом, например, для (2.6) имеем: $\psi(u) \in F_2$, $b(u) = P^0 a_2(u) \in F_2$, а L — оператор, действующий в пространстве F_2.

Лемма 2. Определенный равенством (2.9) *линейный оператор* $L : F_p \to \to F_p$ *обратим.*

Доказательство этой леммы вынесено в Приложение.

Из леммы 2 следует однозначная разрешимость уравнений (2.6) и (2.7):

$$\psi_2(u) = L^{-1}P^0 a_2(u), \tag{2.11}$$

$$\psi_3(u) = L^{-1}P^0[-\psi_2'(u)P_0 a_2(u) + a_2'(u)\psi_2(u) + a_3(u)], \tag{2.12}$$

где через L^{-1} обозначен обратный оператор для (2.9); точнее, в (2.11) L^{-1} — это обратный для оператора $L : F_2 \to F_2$, а в (2.12) — для оператора $L : F_3 \to \to F_3$.

Для вычисления функций (2.11) и (2.12) необходимо знание обратного оператора L^{-1} и операторов проектирования P_0 и P^0 на подпространства E_0 и E^0 соответственно. Вопрос о построении оператора L^{-1} обсуждается ниже. Для случая, когда подпространство E_0 является одномерным или двумерным, формулы для операторов проектирования также приводятся ниже. Задача построения операторов проектирования в общей ситуации может быть решена стандартными методами спектральной теории операторов (см., например, [16]).

2.3.1. Алгоритм построения обратного оператора L^{-1}. Задача построения обратного оператора L^{-1} равносильна задаче решения уравнения (2.10). Ограничимся рассмотрением этой задачи в пространстве F_2. В общем случае задача может быть решена по той же схеме.

Напомним, что k_0 и k^0 — это размерности подпространств E_0 и E^0 такие, что $k_0 + k^0 = N$ и $1 \leqslant k_0 \leqslant N - 1$. Для простоты обозначений положим $k = k_0$.

На первом этапе предлагаемого алгоритма выберем в подпространстве E_0 некоторый базис $e_1, \ldots, e_k$. Каждый вектор $u \in E_0$ единственным образом

представляется в виде $u = u_1e_1 + \ldots + u_ke_k$. Тогда каждый вектор $b(u) \in F_2$ единственным образом представляется в виде $b(u) = \sum\limits_{i,j=1}^{k} u_iu_jb_{ij}$, в котором $b_{ij} = b_{ji} \in E^0$. Решение $\psi(u) \in F_2$ уравнения (2.10) будем искать в виде $\psi(u) = \sum\limits_{i,j=1}^{k} u_iu_jg_{ij}$, в котором векторы $g_{ij} = g_{ji} \in E^0$ требуют определения.

На втором этапе вычислим значение оператора $L\psi(u)$, определенного левой частью уравнения (2.10). Имеем

$$Au = \sum_{i=1}^{k} u_iAe_i, \quad A\psi(u) = \sum_{i,j=1}^{k} u_iu_jAg_{ij}. \tag{2.13}$$

Несложно убедиться в том, что для $h = h_1e_1 + \ldots + h_ke_k \in E_0$ имеет место равенство

$$\psi'(u)h = 2(u_1B_1 + \ldots + u_kB_k)h, \tag{2.14}$$

где $B_j : E_0 \to E^0$ — линейные операторы, задаваемые равенствами $B_jh = \sum\limits_{i=1}^{k} h_ig_{ij}$.

Равенства (2.13) и (2.14) в совокупности с равенством $b(u) = \sum\limits_{i,j=1}^{k} u_iu_jb_{ij}$ позволяют представить уравнение (2.10) (путем приравнивания коэффициентов при одинаковых выражениях u_iu_j) как систему линейных уравнений с неизвестными векторами g_{ij}. Эта система однозначно разрешима в силу леммы 2. Решение полученной системы представляет собой третий (заключительный) этап предлагаемого алгоритма вычисления функции (2.11).

Предлагаемый алгоритм может быть доведен до программ вычисления коэффициентов $\psi_j(u)$ центрального многообразия (2.2) системы (1.1) с применением современной компьютерной техники и пакетов символьных вычислений.

Приведем для иллюстрации алгоритм вычисления функции (2.11) в ситуации, когда матрица A имеет пару простых чисто мнимых собственных значений $\pm\omega_0 i$ и не имеет других чисто мнимых собственных значений.

В рассматриваемом случае подпространство E_0 является двумерным. Обозначим через $e, g \in R^N$ ненулевые векторы такие, что $A(e + ig) = i\omega_0(e + ig)$; векторы e и g образуют базис в E_0. Каждый вектор $u \in E_0$ единственным образом представим в виде $u = u_1e + u_2g$, а каждая функция $b(u) \in F_2$ — в виде

$$b(u) = u_1^2b_{11} + 2u_1u_2b_{12} + u_2^2b_{22}, \tag{2.15}$$

где $b_{11}, b_{12}, b_{22} \in E^0$. Решение $\psi(u) \in F_2$ уравнения (2.10) будем искать в виде $\psi(u) = u_1^2g_{11} + u_2^2g_{22} + 2u_1u_2g_{12}$, в котором $g_{11}, g_{12}, g_{22} \in E^0$ требуют определения.

Подставляя в уравнение (2.10) равенства (2.13) и (2.14) (с учетом рассматриваемого случая), а также равенство (2.15), получим уравнение

$$2\omega_0\left[\left(u_2^2-u_1^2\right)g_{12}+u_1u_2(g_{11}-g_{22})\right]-\left(u_1^2Ag_{11}+u_2^2Ag_{22}+2u_1u_2Ag_{12}\right)=\\=u_1^2b_{11}+u_2^2b_{22}+2u_1u_2b_{12}.$$

Приравнивая затем в этом уравнении коэффициенты при одинаковых выражениях u_iu_j, получим систему трех линейных уравнений относительно трех неизвестных векторов $g_{11}, g_{12}, g_{22} \in E^0$. Полученная система однозначно разрешима:

$$\begin{aligned} g_{12} &= -(A^2+4\omega_0^2I)^{-1}[\omega_0(b_{11}-b_{22})+Ab_{12}],\\ g_{11} &= -A^{-1}(2\omega_0g_{12}+b_{11}),\\ g_{22} &= A^{-1}(2\omega_0g_{12}-b_{22}). \end{aligned} \tag{2.16}$$

Здесь для простоты через $(A^2+4\omega_0^2I)^{-1}$ и A^{-1} обозначены обратные операторы для операторов $(A^2+4\omega_0^2I) : E^0 \to E^0$ и $A : E^0 \to E^0$; обратимость последних операторов следует из предположения, что оператор $A : E^0 \to E^0$ не имеет чисто мнимых собственных значений.

Таким образом, в рассматриваемом случае уравнение (2.10) имеет единственное решение

$$\psi(u) = L^{-1}b(u) = u_1^2g_{11}+u_2^2g_{22}+2u_1u_2g_{12}, \tag{2.17}$$

где g_{11}, g_{12}, g_{22} — векторы (2.16).

Предлагаемый алгоритм позволяет строить решение уравнения (2.10) в общей ситуации. Формулы, определяющие это решение, в свою очередь, позволяют (в соответствии с (2.5)) получить представление центрального многообразия W_c системы (1.1) до любого порядка. Продемонстрируем эффективность предлагаемого подхода в задачах построения центрального многообразия в случаях, когда: а) матрица A имеет простое собственное значение 0; б) матрица A имеет пару простых собственных значений $\pm\omega_0 i$.

2.4. Формулы для центрального многообразия: случай нулевого собственного значения

Пусть матрица A имеет простое собственное значение 0 и не имеет других чисто мнимых собственных значений. В этом случае существуют собственные векторы e и g матрицы A и транспонированной матрицы A^* соответственно, отвечающие простому собственному значению 0 и удовлетворяющие равенствам

$$\|e\| = 1, \quad (e,g) = 1. \tag{2.18}$$

Подпространство E_0 является одномерным; оно содержит вектор e. Операторы проектирования P_0 и P^0 здесь определяются равенствами

$$P_0x = (x,g)e, \quad P^0 = I - P_0. \tag{2.19}$$

Так как подпространство E_0 является одномерным, то векторы $u \in E_0$ можно задавать равенством $u = \varepsilon e$, где $\varepsilon \in (-\infty, \infty)$. Соответственно, произвольный вектор $x \in R^N$ единственным образом представляется в виде суммы $x = \varepsilon e + v$, так что $\varepsilon = (x, g)$ и $v = P^0 x$. Формулу (2.2) для описания центрального многообразия W_c системы (1.1) в рассматриваемом случае можно представить равенством

$$W_c = \{x : x = \varepsilon e + \psi(\varepsilon)\}, \tag{2.20}$$

в котором функция $\psi(\varepsilon)$ принимает свои значения в подпространстве E^0, при этом она является C^m-гладкой и удовлетворяет равенствам $\psi(0) = 0$ и $\psi'(0) = 0$. Наконец, формулу (2.5) для приближенного представления функции $\psi(\varepsilon)$ здесь можно представить в виде

$$\psi(\varepsilon) = \varepsilon^2 \psi_2 + \varepsilon^3 \psi_3 + \widehat{\psi}_4(\varepsilon), \tag{2.21}$$

где $\psi_2, \psi_3 \in E^0$ — требующие определения коэффициенты, а принимающая свои значения в подпространстве E^0 функция $\widehat{\psi}_4(\varepsilon)$ является гладкой и удовлетворяет соотношению $\|\widehat{\psi}_4(\varepsilon)\| = O(\varepsilon^4)$, $\varepsilon \to 0$.

Положим $B_0 = -A + P_0$. По построению оператор $B_0 : R^N \to R^N$ обратим, причем подпространства E_0 и E^0 инвариантны для него. Положим далее для краткости

$$a_2 = a_2(e), \quad a_3 = a_3(e), \quad a_2' = a_2'(e). \tag{2.22}$$

Теорема 3. Пусть матрица A имеет простое собственное значение 0, *а вещественные части остальных ее собственных значений не равны нулю. Тогда центральное многообразие W_c системы* (1.1) *может быть описано равенством* (2.20), *в котором $\psi(\varepsilon)$ — функция* (2.21), *а коэффициенты ψ_2 и ψ_3 определяются равенствами*

$$\psi_2 = B_0^{-1} P^0 a_2, \quad \psi_3 = B_0^{-1} P^0 [-2(a_2, g)\psi_2 + a_2' \psi_2 + a_3]. \tag{2.23}$$

Доказательство этой теоремы вынесено в Приложение.

Таким образом, в условиях теоремы 3 система (1.1) имеет одномерное центральное многообразие и оно представимо в виде (2.20):

$$W_c = \left\{ x : x = \varepsilon e + \varepsilon^2 \psi_2 + \varepsilon^3 \psi_3 + \widehat{\psi}_4(\varepsilon) \right\},$$

где ψ_2 и ψ_3 — коэффициенты (2.23), а функция $\widehat{\psi}_4(\varepsilon)$ удовлетворяет соотношению $\|\widehat{\psi}_4(\varepsilon)\| = O(\varepsilon^4)$ при $\varepsilon \to 0$.

2.5. Формулы для центрального многообразия: случай пары чисто мнимых собственных значений

Пусть теперь матрица A имеет пару простых собственных значений $\pm\omega_0 i$ и не имеет других чисто мнимых собственных значений. В этом случае существуют векторы $e, g, e^*, g^* \in R^N$, удовлетворяющие равенствам $A(e + ig) =$

$= \omega_0 i(e + ig)$ и $A^*(e^* + ig^*) = -\omega_0 i(e^* + ig^*)$, где A^* — транспонированная матрица. Эти векторы можно нормировать равенствами

$$(e, e^*) = (g, g^*) = 1, \quad (e, g^*) = (g, e^*) = 0, \quad \|e\| = \|g\| = 1. \tag{2.24}$$

Подпространство E_0 является двумерным; оно содержит векторы e и g. Операторы проектирования P_0 и P^0 здесь определяются равенствами

$$P_0 x = (x, e^*)e + (x, g^*)g, \quad P^0 = I - P_0. \tag{2.25}$$

Так как подпространство E_0 является двумерным, то векторы $u \in E_0$ можно задавать равенством $u = u_1 e + u_2 g$, где u_1, $u_2 \in (-\infty, \infty)$. Соответственно, произвольный вектор $x \in R^N$ единственным образом представляется в виде суммы $x = u + v$, так что $u = u_1 e + u_2 g$, $u_1 = (x, e^*)$, $u_2 = (x, g^*)$ и $v = P^0 x$. Формулу (2.2) для описания центрального многообразия W_c системы (1.1) в рассматриваемом случае можно представить равенством

$$W_c = \{x : x = u_1 e + u_2 g + \psi(u)\}, \tag{2.26}$$

в котором функция $\psi(u)$ принимает свои значения в подпространстве E^0, при этом она является C^m-гладкой и удовлетворяет равенствам $\psi(0) = 0$ и $\psi'(0) = 0$. Формулу (2.5) для приближенного представления функции $\psi(u)$ здесь можно представить в виде

$$\psi(u) = \psi_2(u) + \psi_3(u) + \widehat{\psi}_4(u), \tag{2.27}$$

где $\psi_2(u), \psi_3(u) \in E^0$ — требующие определения однородные по u функции порядка 2 и 3 соответственно. Функция $\widehat{\psi}_4(u) \in E^0$ является гладкой и удовлетворяет соотношению $\|\widehat{\psi}_4(u)\| = O(\|u\|^4)$, $u \to 0$.

Из лемм 1 и 2, а также из формул (2.11) и (2.12) следует, что верна

Теорема 4. Пусть матрица A имеет пару простых собственных значений $\pm\omega_0 i$, а вещественные части остальных ее собственных значений не равны нулю. Тогда центральное многообразие W_c системы (1.1) *может быть описано равенством* (2.26), *в котором $u = u_1 e + u_2 g$, $\psi(u)$ — функция* (2.27), *а $\psi_2(u)$ и $\psi_3(u)$ определяются равенствами* (2.11) *и* (2.12).

Пример 1. Рассмотрим систему Лэнгфорда (см., например, [7]) вида

$$\begin{cases} x_1' = -x_2 + x_1 x_3 \\ x_2' = x_1 + x_2 x_3 \\ x_3' = k x_3 - (x_1^2 + x_2^2 + x_3^2), \end{cases} \tag{2.28}$$

где $k \neq 0$. Эта система представима в виде (1.1) при

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & k \end{bmatrix}, \qquad a(x) = a_2(x) = \begin{bmatrix} x_1 x_3 \\ x_2 x_3 \\ -x_1^2 - x_2^2 - x_3^2 \end{bmatrix}.$$

Так как матрица A имеет собственные значения $\lambda_{1,2} = \pm i$ и $\lambda_3 = k$, то точка равновесия $x = 0$ системы (2.28) является негиперболической. Поэтому в силу теоремы 4 система (2.28) имеет в окрестности точки $x = 0$ фазового пространства R^3 двумерное центральное многообразие W_c. Используя вышеприведенный подход определим приближенные формулы для этого многообразия. Ограничимся вычислением квадратичного приближения, т.е. в формуле (2.27) ограничимся вычислением слагаемого $\psi_2(u)$.

В соответствии с равенством (2.11) имеем $\psi_2(u) = L^{-1}P^0a_2(u)$. Далее в качестве векторов, удовлетворяющих равенствам (2.24), здесь можно взять векторы $e = e^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $g = g^* = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$. Тогда $u = u_1e + u_2g = \begin{bmatrix} u_1 \\ -u_2 \\ 0 \end{bmatrix}$, $a_2(u) = \begin{bmatrix} 0 \\ 0 \\ -u_1^2 - u_2^2 \end{bmatrix}$, $P^0a_2(u) = \begin{bmatrix} 0 \\ 0 \\ -u_1^2 - u_2^2 \end{bmatrix}$; здесь P^0 определяется в (2.25).

Полученная квадратичная функция $b(u) = P^0a_2(u)$ представляется в виде (2.15) при $b_{11} = b_{22} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$, $b_{12} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Тогда из (2.16) получим $g_{11} = g_{22} = \begin{bmatrix} 0 \\ 0 \\ 1/k \end{bmatrix}$, $g_{12} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Следовательно, функция $\psi_2(u) = L^{-1}P^0a_2(u)$ определяется (см. (2.17)) равенством $\psi_2(u) = u_1^2g_{11} + u_2^2g_{22} + 2u_1u_2g_{12} = \begin{bmatrix} 0 \\ 0 \\ (u_1^2 + u_2^2)/k \end{bmatrix}$, а искомое центральное многообразие W_c — равенством

$$W_c = \left\{ x : x = \begin{bmatrix} x_1 \\ x_2 \\ (x_1^2 + x_2^2)/k \end{bmatrix} + O(\|x\|^3) \right\}. \tag{2.29}$$

3. Динамические системы с дискретным временем

3.1. О центральном многообразии

Рассмотрим теперь динамическую систему (1.2) с дискретным временем. Качественное поведение решений этой системы в окрестности гиперболической точки равновесия $x = 0$ (когда матрица A не имеет собственных значений, равных 1 по модулю) по теореме Гробмана — Хартмана (см., например, [1, 3]) полностью определяется поведением решений соответствующей линейной системы $x_{n+1} = Ax_n$. Если же точка равновесия $x = 0$ системы (1.2) является негиперболической, то для описания поведения траекторий нелинейной системы (1.2) вблизи точки $x = 0$ недостаточно анализа только указанной линейной системы. Здесь необходимо учитывать нелинейные члены системы (1.2).

Ниже будем считать, что нелинейность системы (1.2) представима в виде $a(x) = a_2(x) + a_3(x) + \tilde{a}_4(x)$, где $x \in R^N$, функции $a_2(x)$ и $a_3(x)$ являются соответственно квадратичной и кубической по x слагаемым, а функция $\tilde{a}_4(x)$ является гладкой и удовлетворяет соотношению: $\|\tilde{a}_4(x)\| = O(\|x\|^4)$, $x \to 0$. Общий случай, когда $a(x) = a_2(x) + \ldots + a_s(x) + \tilde{a}_{s+1}(x)$, может быть рассмотрен по той же схеме.

Предполагается, что спектр σ матрицы A состоит из двух непустых частей: $\sigma = \sigma_0 \cup \sigma^0$, где σ_0 содержит собственные значения, равные 1 по модулю, а σ^0 — остальные собственные значения. Обозначим через E_0 и E^0 корневые подпространства матрицы A, отвечающие соответственно частям σ_0 и σ^0 ее спектра. Пусть k_0 и k^0 — это размерности подпространств E_0 и E^0. Пространство R^N представляется в виде прямой суммы $R^N = E_0 \bigoplus E^0$ инвариантных для оператора $A : R^N \to R^N$ подпространств E_0 и E^0. Обозначим через $P_0 : R^N \to E_0$ и $P^0 : R^N \to E^0$ соответствующие операторы проектирования.

Имеет место следующий аналог теоремы 1 о центральном многообразии (см., например, [1, 3]).

Теорема 5. Существует δ_0-окрестность $T(0, \delta_0)$ точки $x = 0$ такая, что система (1.2) в этой окрестности имеет C^m-гладкое инвариантное k_0-мерное многообразие W_c, содержащее точку $x = 0$ и касающееся в ней подпространства E_0.

Эта теорема может быть дополнена утверждениями о существовании устойчивого и неустойчивого многообразий W_s и W_u.

Инвариантность многообразия W_c для системы (1.2) означает, что если $x \in W_c \bigcap T(0, \delta_0)$ и $F(x) \in T(0, \delta_0)$, то $F(x) \in W_c$; здесь $F(x) = Ax + a(x)$. Многообразие W_c называют *центральным* для отображения (1.2) в окрестности неподвижной точки $x = 0$.

Центральное многообразие W_c системы (1.2) может быть задано уравнением вида $v = \psi(u)$, где $u \in E_0$, $v \in E^0$, а функция $\psi(u)$ является гладкой и удовлетворяет равенствам $\psi(0) = 0$, $\psi'(0) = 0$. Другими словами, центральное многообразие W_c может быть локально (в малой окрестности точки $x = 0$) описано равенством вида (2.2).

Принцип сведения А.Н. Шошитайшвили здесь состоит в том, что задача о поведении решений N-мерного уравнения (1.2) в окрестности точки $x = 0$ может быть сведена к аналогичной задаче для k_0-мерного уравнения:

$$u_{n+1} = Au_n + P_0 a(u_n + \psi(u_n)), \quad u_n \in E_0, \tag{3.1}$$

где $u = P_0 x$. Уравнение (3.1) содержит все основные особенности, присущие исходному уравнению (1.2).

3.2. Схема построения центрального многообразия

Имеет место следующий аналог теоремы 2.

Теорема 6. Функция $v = \psi(u)$ (такая, что $\psi(0) = 0$, $\psi'(0) = 0$) описывает центральное многообразие W_c системы (1.2) тогда и только тогда,

когда она в некоторой окрестности точки $u = 0$ подпространства E_0 является решением уравнения

$$\psi(Au + P_0 a(u + \psi(u))) = A\psi(u) + P^0 a(u + \psi(u)). \tag{3.2}$$

Перейдем к задаче приближенного построения функции $v = \psi(u)$. С этой целью представим ее в виде (2.5): $\psi(u) = \psi_2(u) + \psi_3(u) + \dots$

Лемма 3. Функции $\psi_2(u)$ и $\psi_3(u)$ являются решениями уравнений

$$\psi_2(Au) - A\psi_2(u) = P^0 a_2(u), \tag{3.3}$$

$$\psi_3(Au) - A\psi_3(u) = -\psi_2'(Au)P_0 a_2(u) + P^0[a_2'(u)\psi_2(u) + a_3(u)]. \tag{3.4}$$

Справедливость этого утверждения устанавливается простым подсчетом путем подстановки (2.5) в (3.2).

Как и выше, через F_p будем обозначать линейное пространство однородных порядка p функций $\psi(u)$, определенных в подпространстве E_0 и принимающих значения в подпространстве E^0 (см. (2.8)). Через J обозначим линейный оператор, действующий в пространстве F_p и сопоставляющий каждой функции $\psi(u) \in F_p$ функцию

$$J\psi(u) = \psi(Au) - A\psi(u). \tag{3.5}$$

При этом для простоты будем использовать одно и то же обозначение J для всех действующих в пространствах F_p операторов (3.5) независимо от значения p.

Имеет место следующий аналог леммы 2.

Лемма 4. Определенный равенством (3.5) *линейный оператор $J : F_p \to \to F_p$ обратим.*

Доказательство этой леммы вынесено в Приложение.

Из леммы 4 следует однозначная разрешимость уравнений (3.3) и (3.4):

$$\psi_2(u) = J^{-1}P^0 a_2(u), \tag{3.6}$$

$$\psi_3(u) = J^{-1}P^0[-\psi_2'(Au)P_0 a_2(u) + a_2'(u)\psi_2(u) + a_3(u)], \tag{3.7}$$

где через J^{-1} обозначен обратный оператор для (3.5); точнее, в (3.6) J^{-1} — это обратный для оператора $J : F_2 \to F_2$, а в (3.7) — для оператора $J : F_3 \to \to F_3$.

3.3. Формулы для центрального многообразия

Перейдем к задаче построения функций (3.6) и (3.7). Здесь ограничимся рассмотрением ситуаций, когда матрица A имеет:

— **P1)** простое собственное значение 1;

— **P2)** простое собственное значение -1;

— **P3)** пару простых собственных значений $e^{\pm\varphi_0 i}$, где $0 < \varphi_0 < \pi$.

При этом предполагается, что остальные собственные значения матрицы A не равны по модулю единице.

Случай P1.

Этот случай почти дословно повторяет аналогичный случай для динамической системы с непрерывным временем (1.1), рассмотренный в разделе 2.4. Поэтому укажем здесь лишь то, что присуще системе с дискретным временем (1.2).

Обозначим через e и g собственные векторы матриц A и A^* соответственно, отвечающие простому собственному значению 1 и удовлетворяющие равенствам (2.18). В отличие от раздела 2.4, оператор B_0 здесь определим равенством $B_0 = I - A + P_0$. По построению оператор $B_0 : R^N \to R^N$ обратим, причем подпространства E_0 и E^0 инвариантны для него.

Теорема 7. Пусть матрица A имеет простое собственное значение 1, *а модули остальных ее собственных значений не равны единице. Тогда центральное многообразие W_c системы* (1.2) *может быть описано равенством* (2.20), *в котором $\psi(\varepsilon)$ — функция* (2.21), *а коэффициенты ψ_2 и ψ_3 определяются равенствами* (2.23).

Случай P2.

В этом случае существуют собственные векторы e и g матриц A и A^* соответственно, отвечающие простому собственному значению -1 и удовлетворяющие равенствам (2.18). Подпространство E_0 здесь также (как и в случае P1) является одномерным; оно содержит вектор e. Наконец, операторы проектирования P_0 и P^0 определяются теми же равенствами (2.19).

Как и в случае P1, здесь уравнение центрального многообразия W_c можно искать в виде (2.20). Положим

$$B_1 = I - A, \quad B_2 = -I - A + P_0. \tag{3.8}$$

По построению операторы $B_1 : R^N \to R^N$ и $B_2 : R^N \to R^N$ обратимы, причем подпространства E_0 и E^0 инвариантны для них.

Теорема 8. Пусть матрица A имеет простое собственное значение -1, а модули остальных ее собственных значений не равны единице. Тогда центральное многообразие W_c системы (1.2) *может быть описано равенством* (2.20), *в котором $\psi(\varepsilon)$ — функция* (2.21), *а коэффициенты ψ_2 и ψ_3 определяются равенствами*

$$\psi_2 = B_1^{-1} P^0 a_2, \quad \psi_3 = B_2^{-1} P^0 [2(a_2, g)\psi_2 + a_2' \psi_2 + a_3]. \tag{3.9}$$

Доказательство теоремы 8 вынесено в Приложение.

Пример 2. Рассмотрим модель Хенона (см., например, [2]) вида

$$\begin{cases} x_{n+1} = y_n, \\ y_{n+1} = \dfrac{1}{3} x_n - \frac{2}{3} y_n - y_n^2, \end{cases} \tag{3.10}$$

т.е. систему (1.2) при $N = 2$, $A = \begin{bmatrix} 0 & 1 \\ 1/3 & -2/3 \end{bmatrix}$, $a(w) = a_2(w) = \begin{bmatrix} 0 \\ -y^2 \end{bmatrix}$, где $w = (x, y)$. Матрица A имеет собственные значения $\lambda_1 = -1$ и $\lambda_2 = 1/3$. Рассмотрим вопрос о построении центрального многообразия W_c системы (3.10).

Вычисления по формулам (2.18), (2.19), (3.8) и (3.9) приводят к равенствам

$$e = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad g = \frac{\sqrt{2}}{4} \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \psi_2 = -\frac{3}{16} \begin{bmatrix} 3 \\ 1 \end{bmatrix}. \tag{3.11}$$

Поэтому искомое центральное многообразие W_c определяется равенством $W_c = \{w : w = \varepsilon e + \varepsilon^2 \psi_2 + O(\varepsilon^3)\}$, в котором e и ψ_2 — это векторы из (3.11).

Случай P3.

Этот случай имеет смысл рассматривать только при $N \geqslant 3$.

Так как матрица A имеет пару простых собственных значений $e^{\pm i\varphi_0}$, то найдутся ненулевые векторы $e, g, e^*, g^* \in R^N$ такие, что выполняются равенства $A(e + ig) = e^{i\varphi_0}(e + ig)$, $A^*(e^* + ig^*) = e^{-i\varphi_0}(e^* + ig^*)$; здесь A^* — транспонированная матрица. Векторы e, g, e^*, g^* можно считать нормированными в соответствии с равенствами (2.24).

Подпространство E_0 — это корневое подпространство оператора A, отвечающее простым собственным значениям $e^{\pm i\varphi_0}$. Пространство E_0 является двумерным; в качестве его базиса могут использоваться векторы e и g. Пространство R^N может быть представлено в виде $R^N = E_0 \oplus E^0$, где E^0 — дополнительное инвариантное для A подпространство размерности $N - 2$.

Равенство $R^N = E_0 \oplus E^0$ определяет операторы проектирования $P_0 : R^N \to \to E_0$ и $P^0 : R^N \to E^0$ так, что $P^0 = I - P_0$, а оператор P_0 может быть представлен в виде $P_0 x = (x, e^*)e + (x, g^*)g$; последнее следует из того, что по предположению векторы e, g, e^*, g^* выбраны в соответствии с равенствами (2.24).

Центральное многообразие W_c системы (1.2) в рассматриваемом случае P3 естественно строить в виде равенства (2.26) (применительно к рассматриваемой ситуации), в котором функция $\psi(u)$ определяется равенством (2.27).

Из лемм 3 и 4, а также из формул (3.6) и (3.7) следует, что верна

Теорема 9. Пусть матрица A имеет пару простых собственных значений $e^{\pm i\varphi_0}$, где $0 < \varphi_0 < \pi$, а модули остальных ее собственных значений не равны единице. Тогда центральное многообразие W_c системы (1.2) *может быть описано равенством* (2.26)*, в котором $\psi(u)$ — функция* (2.27)*, а функции $\psi_2(u)$ и $\psi_3(u)$ определяются равенствами* (3.6) *и* (3.7)*.*

Полученные в настоящей статье результаты относительно исследования системы (1.2) являются развитием результатов, приведенных в [14] относительно построения центральных многообразий дискретных динамических систем. Отметим в этой связи, что в указанной работе, в частности, приведена детальная схема расчета функций (3.6) и (3.7).

4. Заключение

В статье предложены новые приближенные формулы и алгоритмы построения центральных многообразий в окрестностях негиперболических точек равновесия динамических систем с непрерывным и дискретным временем (системы (1.1) и (1.2)). Предлагаемые подходы позволяют получить аппроксимации центрального многообразия в терминах исходных систем. Предлагаемая схема носит общий характер и в том смысле, что она применима к

ситуациям, когда матрица линеаризации имеет произвольный порядок вырождения. Основные утверждения для непрерывных динамических систем приведены в теоремах 3 и 4, для дискретных динамических систем — в теоремах 7–9.

ПРИЛОЖЕНИЕ

В Приложении приводятся доказательства лемм 2 и 4, а также теорем 3 и 8.

Доказательство леммы 2.

Справедливость этой леммы можно установить как следствие аналогичного утверждения, приведенного в [12, стр. 211–212]. Пусть для простоты все собственные значения λ_j матрицы A являются полупростыми. Будем рассматривать оператор (2.9) в более широком (чем F_p) линейном пространстве H_p однородных порядка p функций $h(x)$, определенных и принимающих значения в R^N. В [12] показано, что спектр оператора $L:\ H_p \to H_p$ совпадает с множеством чисел вида

$$\mu = p_1\lambda_1 + \ldots + p_N\lambda_N - \lambda_j, \tag{П.1}$$

где $p = p_1 + \ldots + p_N$. Оператор $L:\ F_p \to F_p$ является сужением оператора $L:\ H_p \to H_p$. Несложно показать, что для оператора $L:\ F_p \to F_p$ в (П.1) в сумме $p_1\lambda_1 + \ldots + p_N\lambda_N$ следует брать только слагаемые, отвечающие чисто мнимым собственным значениям матрицы A, а в качестве λ_j — только остальные собственные значения матрицы A. Тогда $\mu \neq 0$ и, следовательно, оператор $L:\ F_p \to F_p$ обратим.

Доказательство теоремы 3.

В условиях этой теоремы векторы $u \in E_0$ можно задавать равенством $u = \varepsilon e$, где $Ae = 0$ и, следовательно, $Au = 0$. Поэтому действующий в пространстве F_p линейный оператор (2.9) здесь принимает вид $L\psi(u) = -A\psi(u)$. Далее, произвольный элемент $\psi(u) \in F_p$ здесь представим в виде $\psi(u) = \varepsilon^p v$, где $v = \psi(e) \in E^0$. Несложно видеть, что для $\psi(u) = \varepsilon^p v$ имеет место равенство $L^{-1}\psi(u) = \varepsilon^p B_0^{-1} v$ (напомним, что в силу леммы 2 оператор L обратим). Отсюда и из общих формул (2.11) и (2.12) получим равенства (2.23).

Доказательство леммы 4.

Для доказательства этой леммы приведем сначала вспомогательные построения (аналоги построений, приведенных в [12, стр. 211–212]. Рассмотрим линейный оператор (3.5) в более широком (чем F_p) линейном пространстве H_p однородных порядка p функций $h(x)$, определенных и принимающих значения в R^N. Пусть для простоты оператор A диагональный и $\Lambda = \{\lambda_1, \ldots, \lambda_N\}$ — множество его собственных значений; через e_j обозначим собственный вектор, отвечающий собственному значению λ_j.

Пусть $p = p_1 + \ldots + p_N$, где p_j — целые неотрицательные числа. Для вектора $x = (x_1, x_2, \ldots, x_N) \in R^N$ определим многочлен $x^p = x_1^{p_1} x_2^{p_2} \ldots x_N^{p_N}$.

Лемма 5. Спектр оператора $J:\ H_p \to H_p$ совпадает с множеством чисел вида

$$\mu = \lambda_1^{p_1}\lambda_2^{p_2} \ldots \lambda_N^{p_N} - \lambda_j, \tag{П.2}$$

а векторы $x^p e_j$ являются соответствующими собственными векторами.

Справедливость этого утверждения вытекает из равенства

$$Jh(x) = h(Ax) - Ah(x) = (\lambda_1^{p_1}\lambda_2^{p_2}\dots\lambda_N^{p_N} - \lambda_j)x^p e_j,$$

в котором $h(x) = x^p e_j$.

Вернемся к доказательству леммы 4. Оператор $J : F_p \to F_p$ является сужением оператора $J : H_p \to H_p$. Несложно показать, что для оператора $J : F_p \to F_p$ в равенстве (П.2) в произведении $\lambda_1^{p_1}\lambda_2^{p_2}\dots\lambda_N^{p_N}$ следует брать только сомножители, отвечающие собственным значениям оператора A, равным единице по модулю. При этом в качестве λ_j следует брать только остальные собственные значения оператора A. Тогда $\mu \neq 0$ и, следовательно, оператор $J : F_p \to F_p$ обратим.

Доказательство теоремы 8.

В условиях этой теоремы векторы $u \in E_0$ можно задавать равенством $u = \varepsilon e$, где $Ae = -e$. Произвольный элемент $\psi(u) \in F_p$ здесь представим в виде $\psi(u) = \varepsilon^p v$, где $v = \psi(e) \in E^0$. Поэтому действующий в пространстве F_2 линейный оператор (3.5) здесь принимает вид $J\psi(u) = \varepsilon^2(I - A)v$, а в пространстве F_3 — вид $J\psi(u) = -\varepsilon^3(I + A)v$. Следовательно, для $\psi(u) = \varepsilon^2 v \in F_2$ имеет место равенство $J^{-1}\psi(u) = \varepsilon^2 B_1^{-1}v$, а для $\psi(u) = \varepsilon^3 v \in F_3$ — равенство $J^{-1}\psi(u) = \varepsilon^3 B_2^{-1}v$. Отсюда и из формул (3.6) и (3.7) получим равенства (3.9).

СПИСОК ЛИТЕРАТУРЫ

1. *Шильников Л.П., Шильников А.Л., Тураев Д.В. и др.* Методы качественной теории в нелинейной динамике. Ч. 1. М.-Ижевск: Ин-т компьют. исслед., 2004.
2. *Шильников Л.П., Шильников А.Л., Тураев Д.В. и др.* Методы качественной теории в нелинейной динамике. Ч. 2. М.–Ижевск: НИЦ “Регулярная и хаотическая динамика”, Ин-т компьют. исслед., 2009.
3. *Гукенхеймер Дж., Холмс Ф.* Нелинейные колебания, динамические системы и бифуркации векторных полей. М.-Ижевск: Ин-т компьют. исслед., 2002.
4. *Ван Д., Ли Ч., Чоу Ш.Н.* Нормальные формы и бифуркации векторных полей на плоскости. М.: МЦНМО, 2005.
5. *Kelley A.* The stable, center-stable, center-unstable, unstable manifolds // J. Diff. Equat. 1967. No. 3. P. 546–570.
6. *Vanderbauwhede A.* Centre manifolds, normal forms and elementary bifurcations // Dynam. Reported. 1989. V. 2. P. 89–169.
7. *Kuznetsov Yu.A.* Elements of Applied Bifurcation Theory. N.Y.: Springer. 1998.
8. *Шошитайшвили А.Н.* Бифуркации топологического типа векторного поля вблизи особой точки // Тр. семинаров им. И.Г. Петровского. 1975. Вып. 1. С. 279–309.
9. *Никульчев Е.В.* Качественное исследование управляемых систем с нелинейной динамикой на центральном многообразии // Вестн. МГАПИ. Естеств. и техн. науки. 2006. № 1. С. 150–161.
10. *Никульчев Е.В.* Геометрический подход к моделированию нелинейных систем по экспериментальным данным. М.: МГУП, 2007.
11. *Hamzi B., Kang W., Krener A.J.* Control of center manifolds // Proc. 42nd IEEE Conf. Decision and Control. V. 3. Maui, HI, 2003. P. 2065–2070.

12. *Арнольд В.И.* Геометрические методы в теории обыкновенных дифференциальных уравнений. Ижевск: Регулярная и хаотическая динамика, 2000.
13. *Юмагулов М.Г., Гусарова Н.И., Муртазина С.А., Фазлытдинов М.Ф.* Операторные методы вычисления ляпуновских величин в задачах о локальных бифуркациях динамических систем // Уфим. мат. журн. 2018. Т. 10. № 1. С. 25–49.
14. *Юмагулов М.Г., Фазлытдинов М.Ф.* Бифуркационные формулы и алгоритмы построения центральных многообразий дискретных динамических систем // Изв. вузов. Математика. 2019. № 3. С. 71–89.
15. *Qesmi R., Ait Babram M., Hbid M.L.* Symbolic computation for center manifolds and normal forms of Bogdanov bifurcation in retarded functional differential equations // Nonlinear Anal. 2007. V. 66. P. 2833–2851.
16. *Kato T.* Теория возмущений линейных операторов. М.: Мир, 1975.

Статья представлена к публикации членом редколлегии А.Н. Соболевским.

Поступила в редакцию 31.12.2018

После доработки 04.05.2019

Принята к публикации 18.07.2019

Стохастические системы

© 2020 г. М.А. ГОРЕЛОВ, канд. физ.-мат. наук (griefer@ccas.ru),
Ф.И. ЕРЕШКО, д-р техн. наук (fereshko@yandex.ru)
(Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, Москва)

ИНФОРМИРОВАННОСТЬ И ДЕЦЕНТРАЛИЗАЦИЯ УПРАВЛЕНИЯ (СТОХАСТИЧЕСКИЙ СЛУЧАЙ)

Рассматривается задача принятия решений в условиях риска. Предполагается, что лицо, принимающее решение, может обработать лишь ограниченный объем информации о неопределенном факторе и ориентируется на математическое ожидание своего выигрыша. Сравнивается два способа управления. В одном из них решение принимается централизованно. Во втором оперирующая сторона передоверяет часть своих полномочий по выбору решений нескольким агентам. При этом предполагается, что оперирующая сторона знает интересы агентов и рассчитывает на их рациональное поведение, а в остальном осторожна по отношению к неопределенности их выбора.

Ключевые слова: принятие решений в условиях риска, информационная теория иерархических систем, децентрализация управления.

DOI: 10.31857/S0005231020010043

1. Введение

Актуальность задачи выбора оптимальной структуры системы управления сложным объектом (производством, транспортом, войсками и т.п.) вряд ли может вызывать сомнение. Поэтому неудивительно и стремление построить математические модели, описывающие такой выбор. Было предложено несколько альтернативных подходов к моделированию [1–5], однако приходится констатировать, что пока вопросов больше, чем ответов. Видимо, это в значительной степени связано с тем, что пока не удается отделить существенные черты моделируемого объекта от многих второстепенных деталей.

В данной статье исследуется зависимость оптимальной структуры системы управления от объема доступной информации о внешнем неопределенном факторе. Впервые на эту зависимость обратили внимание Ю.Б. Гермейер и Н.Н. Моисеев в начале семидесятых годов прошлого века [6–8]. Первая формальная модель такого рода была построена в [9]. Прикладной смысл этого исследования — создание математического аппарата для анализа упрощенных моделей, позволяющих делать качественные выводы и, что самое главное, формировать на модельном уровне представление о предмете исследований у лиц, принимающих решения. Более подробно о деталях используемого подхода и об исходных содержательных предпосылках можно прочесть во введении к цитированной статье.

Построенная модель позволяет получить качественные выводы, хорошо согласующиеся с содержательными представлениями об управлении в условиях неопределенности. Это дает основание говорить о том, что модель верно отражает какие-то существенные черты описываемого объекта. Однако эти выводы получены при довольно жестких предположениях. Поэтому естественно возникает вопрос о том, насколько полученные выводы зависят от сделанных предположений. Есть и другой, чисто утилитарный, вопрос: насколько зависит от этих предположений возможность математического исследования построенной модели? Этим двум вопросам и посвящена в значительной степени данная статья.

Основные гипотезы статьи [9] сохранены. Изменено лишь предположение об отношении лица, принимающего решения, к неопределенности. В [9] оно предполагалось осторожным. Далее считается, что на множестве неопределенных факторов задана вероятностная мера и лицо, принимающее решение, склонно ориентироваться на математическое ожидание своего выигрыша по этой мере. Такое предположение в прикладных исследованиях, разумеется, нуждается в дополнительном обосновании, однако оно весьма распространено, а в западной литературе, пожалуй, даже более популярно, чем принцип максимального гарантированного результата. Отметим, что в данной статье не используются результаты типа закона больших чисел, поэтому все вероятности можно рассматривать как субъективные, что существенно облегчает обоснование адекватности подобного рода моделей. Исследование моделей в новых предположениях оказывается более сложным и требует привлечения иного математического аппарата. Однако решить соответствующие задачи удается в достаточной общности.

Как и в [9], задача поиска оптимальной структуры иерархической системы не ставится. Вместо этого на качественном уровне производится сравнение двух схем управления: централизованной и децентрализованной.

2. Объект управления

Рассмотрим следующую модель управляемой системы. Оперирующая сторона может по своему усмотрению выбирать любое управление w из множества W. Помимо этого выбора на результат управления влияет еще некий неопределенный фактор α из множества A, значение которого оперирующая сторона не контролирует. Эффективность управления оценивается значением $g(w, \alpha)$ функции $g : W \times A \to \mathbb{R}$ (как обычно $\mathbb{R}$ — множество действительных чисел).

Будем считать, что на множестве A задана вероятностная мера $\wp$, известная оперирующей стороне. В дальнейшем будем предполагать, что оперирующая сторона риск-нейтральна по отношению к этой неопределенности, т.е. ориентируется на математическое ожидание своего выигрыша.

Примем еще одно предположение, отражающее представление "технологической структурированности" рассматриваемой управляемой системы. Будем считать, что множество W представимо в виде декартова произведения $W = U \times V^1 \times \ldots \times V^n$. Тогда всякий элемент $w \in W$ может быть записан в

виде $w = (u, v^1, \ldots, v^n)$, где $u \in U$, $v^i \in V^i$, $i = 1, \ldots, n$. Такую форму записи там, где она удобна, будем использовать без особых оговорок.

Сделаем следующие стандартные предположения. Будем предполагать, что на множествах $u \in U$, $v^i \in V^i$, $i = 1, \ldots, n$, и A заданы топологии, в которых эти множества компактны. Функцию g будем считать непрерывной в топологии декартова произведения $U \times V^1 \times \ldots \times V^n \times A$. Меру $\wp$ будем считать борелевской.

Замечание 1. Вероятно, эти предположения можно ослабить без потери всех результатов, полученных далее. Однако это приводит к необходимости более аккуратных и, как следствие, более длинных рассуждений. Поскольку не очень понятно, могут ли найтись интерпретации данной модели, в которых эти предположения будут ограничительными, вдаваться в эти технические детали пока не станем.

Топологии на множествах $u \in U$, $v^i \in V^i$, $i = 1, \ldots, n$ индуцируют топологию на их произведении $W = U \times V^1 \times \ldots \times V^n$. В дальнейшем, когда речь зайдет о топологии на множестве W, будем иметь в виду именно топологию произведения.

Согласно теореме Тихонова [10, стр. 217], множество W будет компактным.

3. Модель централизованного управления

Допустим, оперирующая сторона не имеет никакой дополнительной информации о реализовавшемся значении неопределенного фактора α. Если она зафиксирует управление $w \in W$, то математическое ожидание ее выигрыша составит

$$\mathrm{M}g(w, \alpha) = \int\limits_A g(w, \alpha) \wp(d\alpha).$$

При оптимальном выборе управления w она получит результат равный

$$R_0(0) = \max_{w \in W} \int\limits_A g(w, \alpha) \wp(d\alpha).$$

Рассмотрим другой крайний случай. Предположим, в момент принятия решения оперирующей стороне становится точно известно реализовавшееся значение неопределенного фактора α. Тогда она может выбрать управление w так, чтобы получить выигрыш, равный $\max\limits_{w \in W} g(w, \alpha)$. Соответственно математическое ожидание выигрыша составит

$$R_0(\infty) = \int\limits_A \max_{w \in W} g(w, \alpha) \wp(d\alpha).$$

В данной статье основной интерес будет представлять промежуточный случай. Пусть оперирующая сторона имеет возможность получать информа-

цию о реализовавшемся значении неопределенного фактора, но объем информации, которую она способна получить и своевременно обработать, ограничен. А именно, будем считать, что оперирующая сторона может использовать l бит информации и других ограничений на использование информации нет.

Формализуется сказанное следующим образом. Введем обозначение. Здесь и далее $\Phi(X,Y)$ будет обозначать семейство всех функций, отображающих множество X в множество Y.

Сделанное предположение означает, что вся информация о неопределенном факторе, доступная оперирующей стороне, может быть закодирована словами $s=(s_1,\ldots,s_l)$ из нулей и единиц длины l. Множество $\{0,1\}^l$ (декартову степень множества $\{0,1\}$) обозначим буквой S. Поскольку ограничений на доступ к информации о неопределенном факторе у оперирующей стороны нет, выбор "способа кодировки" $P:A\to S$ — это ее прерогатива. Кроме того, в зависимости от полученной информации $s\in S$ оперирующая сторона вправе выбрать любое управление $w\in W$. Т.е., по сути, она может выбирать функцию $w_*:S\to W$. Если оперирующая сторона зафиксирует способ кодировки $P\in\Phi(A,S)$ и правило выбора управления $w_*\in\Phi(S,W)$ и реализуется значение неопределенного фактора $\alpha\in A$, то оперирующая сторона получит сообщение $P(\alpha)$, выберет управление $w_*(P(\alpha))$ и ее выигрыш составит $g(w_*(P(\alpha)),\alpha)$.

В таком случае математическое ожидание выигрыша будет равно $\int\limits_A g(w_*(P(\alpha)),\alpha)\,\wp(d\alpha)$, а при наилучшем выборе стратегии (w_*,P) из множества $\Phi(S,W)\times\Phi(A,S)$ результат составит

$$R_0(l)=\sup_{(w_*,P)\in\Phi(S,W)\times\Phi(A,S)}\int\limits_A g(w_*(P(\alpha)),\alpha)\,\wp(d\alpha).$$

Замечание 2. Можно представить себе ситуацию, когда функция w_* выбрана так, что функция $g(w_*(P(\alpha)),\alpha)$ будет неизмеримой. Поэтому данная постановка задачи требует некоторого уточнения. Возможны, по меньшей мере, два способа такого уточнения: либо можно понимать интеграл в определении величины $R_0(l)$ как нижний интеграл, либо можно ограничить класс стратегий оперирующей стороны множеством измеримых функций (по отношению к алгебре всех подмножеств конечного множества S). И тот, и другой подход методологически оправдан. Из дальнейшего будет видно, что при обоих подходах и в задаче этого раздела, и в задаче раздела 4 получается один и тот же результат. Это можно рассматривать как некий аргумент в пользу рассматриваемых постановок. В дальнейшем, дабы не уклоняться от сути дела, на подобного рода технических проблемах, если они решаются стандартными способами, акцент делаться не будет.

Упростим формулу, определяющую величину $R_0(l)$. Фиксируем функцию $w_*\in\Phi(S,W)$. Она принимает $m=2^l$ различных значений. Пусть множество этих значений есть $\{w_0,w_1,\ldots,w_{m-1}\}$. Сообщение $s=(s_1,\ldots,s_l)$ можно рассматривать как двоичную запись $s_1\ldots s_l$ натурального числа из множества $\{0,1,\ldots,m-1\}$. Имея в виду такое отождествление, можно, не ограничивая общности, считать, что $w_*(s)=w_s$.

Если функция $w_* \in \Phi(S,W)$ фиксирована, то способ кодировки информации $P \in \Phi(A,S)$ разумно выбирать так, чтобы при каждом значении $\alpha \in A$ сообщение $r = P(\alpha)$ удовлетворяло условию $g(w_r,\alpha) = \max\limits_{s=0,1,\ldots,m-1} g(w_s,\alpha)$. При таком выборе стратегии (w_*,P) математическое ожидание выигрыша будет равно $\int\limits_A \max\limits_{s=0,1,\ldots,m-1} g(w_s,\alpha)\wp(d\alpha)$. А при наилучшем выборе функции w_* можно рассчитывать на получение ожидаемого результата $\max\limits_{(w_0,w_1,\ldots,w_{m-1})\in W^m} \int\limits_A \max\limits_{s=0,1,\ldots,m-1} g(w_s,\alpha)\wp(d\alpha)$.

Понятно, что приведенные рассуждения обратимы, поэтому справедлива следующая

Теорема 1. Имеет место равенство

$$R_0(l) = \max_{(w_0,w_1,\ldots,w_{m-1})\in W^m} \int_A \max_{s=0,1,\ldots,m-1} g(w_s,\alpha)\wp(d\alpha).$$

Замечание 3. Из стандартных теорем анализа следует, что функция $\max\limits_{s=0,1,\ldots,m-1} g(w_s,\alpha)$ непрерывно зависит от $w_0,w_1,\ldots,w_{m-1}$ и α, а функция $\int\limits_A \max\limits_{s=0,1,\ldots,m-1} g(w_s,\alpha)\wp(d\alpha)$ непрерывно зависит от $w_0,w_1,\ldots,w_{m-1}$. Поэтому внешний максимум в результирующем выражении для $R_0(l)$ достигается. Значит, достигается и верхняя грань в определении величины $R_0(l)$. Соответствующие технические рассуждения опускаем.

4. Модель децентрализованного управления

Рассмотрим другой способ управления той же системой.

Предположим, оперирующая сторона передает право выбора управлений v^i n агентам: агент с номером i получает право выбора управления $v^i \in V^i$ $(i = 1,\ldots,n)$. Выбор управления $u \in U$ оперирующая сторона (Центр) оставляет за собой.

Появление у агента i права влиять на ситуацию неизбежно влечет появление у него собственных целей. Процесс формирования этих целей сложен и мало изучен. В данной модели эти цели считаются заданными экзогенно. Будем предполагать, что цель агента i описывается стремлением к максимизации значения функции $h^i\left(u,v^i,\alpha\right)$. Существенным является то, что эта функция зависит от его собственного управления, управления Центра и неопределенного фактора, но не зависит от выборов остальных агентов.

Будем считать, что Центр по-прежнему имеет возможность получить и обработать l бит информации о неопределенном факторе α. Таким образом, стратегией центра является пара $(u_*,P) \in \Phi(S,U) \times \Phi(A,S)$ функций $u_* : S \to U$ и $P : A \to S$ (смысл этих конструкций тот же, что и в модели раздела 3). Предположим, что каждый из агентов в момент принятия решений имеет точную информацию об этом неопределенном факторе.

Допустим, Центр оставляет за собой право первого хода, т.е. он первым выбирает свою стратегию (u_*,P) и сообщает ее всем агентам.

В этих условиях агент i принимает решение в условиях полной определенности: он знает реализовавшееся значение неопределенного фактора α и управление $u_*(P(\alpha))$, которое должен будет выбрать Центр. Поэтому если Центр знает функцию выигрыша h^i агента i, то он может рассчитывать на то, что в случае, когда реализуется значение неопределенного фактора α, этот агент выберет свое управление из множества

$$BR^i(u_*, P, \alpha) = \left\{ v^i \in V^i : \; h^i(u_*(P(\alpha)), v^i, \alpha) = \max_{v^i \in V^i} h^i(u_*(P(\alpha)), v^i, \alpha) \right\}.$$

Поскольку для i-го агента все выборы из этого множества равноценны, дальше уточнить множество возможных выборов агента у Центра нет.

Если Центр осторожен по отношению к такого рода неопределенностям, то при выборе стратегии (u_*, P) и реализовавшемся значении неопределенного фактора α он должен рассчитывать на выигрыш

$$\min_{v^1 \in BR^1(u_*, P, \alpha)} \dots \min_{v^n \in BR^n(u_*, P, \alpha)} g\left(u_*\left(P\left(\alpha\right)\right), v^1, \dots, v^n, \alpha\right).$$

Математическое ожидание этого выигрыша равно

$$\int\limits_A \min_{v^1 \in BR^1(u_*, P, \alpha)} \dots \min_{v^n \in BR^n(u_*, P, \alpha)} g\left(u_*\left(P\left(\alpha\right)\right), v^1, \dots, v^n, \alpha\right) \wp\left(d\alpha\right),$$

и если считать Центр риск-нейтральным, он будет ориентироваться именно на этот результат. Тогда при оптимальном выборе своей стратегии он получит выигрыш

$$R_1(l) =$$

$$= \sup_{(u_*, P) \in \Phi(S,U) \times \Phi(A,S)} \int\limits_A \min_{v^1 \in BR^1(u_*, P, \alpha)} \dots \min_{v^n \in BR^n(u_*, P, \alpha)} g(u_*(P(\alpha)), v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

Упростим и эту формулу с тем, чтобы избавиться от верхней грани по функциональному пространству. Вновь фиксируем множество значений $\{u_0, u_1, \dots, u_{m-1}\}$ функции u_* и будем считать, что эти значения перенумерованы так, что $u_*(s) = u_s$.

Если агент i знает, что Центр выберет управление $u_s \in U$ и реализовалось значение неопределенного фактора α, то он выберет свое управление из множества

$$E^i(u_s, \alpha) = \left\{ v^i \in V^i : \; h^i(u_s, v^i, \alpha) = \max_{\omega^i \in V^i} h^i(u_s, \omega^i, \alpha) \right\}.$$

В таком случае Центр может гарантированно рассчитывать на получение выигрыша равного

$$\min_{v^1 \in E^1(u_s, \alpha)} \dots \min_{v^n \in E^n(u_s, \alpha)} g(u_s, v^1, \dots, v^n, \alpha).$$

Естественно выбирать способ кодировки P так, чтобы каждому значению $\alpha \in A$ он ставил в соответствие сообщение u_r, удовлетворяющее условию

$$\min_{v^1 \in E^1(u_r,\alpha)} \dots \min_{v^n \in E^n(u_r,\alpha)} g(u_r, v^1, \dots, v^n, \alpha) =$$

$$= \max_{s=0,1,\dots,m-1} \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha).$$

Справедливо следующее утверждение.

Лемма 1. При любом наборе $u_0, u_1, \dots, u_{m-1}$ функция

$$\varphi(\alpha) = \max_{s=0,1,\dots,m-1} \min_{v^1 \in E^1(u_s,\alpha)} \min_{v^2 \in E^2(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha)$$

$\wp$-измерима.

Доказательство леммы приведено в Приложении.

Лемма 1 позволяет заключить, что при фиксированной функции u_* со значениями $\{u_0, u_1, \dots, u_{m-1}\}$ риск-нейтральный Центр должен ориентироваться на результат

$$\int_A \max_{s=0,1,\dots,m-1} \min_{v^1 \in E^1(u_s,\alpha)} \min_{v^2 \in E^2(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

А поскольку выбор значений $\{u_0, u_1, \dots, u_{m-1}\}$ — это тоже право Центра, он может получить выигрыш, равный

$$\max_{(u_0,u_1,\dots,u_{m-1}) \in U^m} \int_A \max_{s=0,1,\dots,m-1} \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

Анализируя приведенные рассуждения, несложно убедиться, что на больший выигрыш Центр не может рассчитывать. Поэтому справедлива следующая теорема.

Теорема 2. Имеет место равенство

$$R_1(l) =$$

$$= \max_{(u_0,u_1,\dots,u_{m-1}) \in U^m} \int_A \max_{s=0,1,\dots,m-1} \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

Для удобства сравнения централизованной и децентрализованной схем управления рассмотрим те же два крайних случая, что и в разделе 3.

Случай, когда Центр не имеет информации о неопределенном факторе, вполне вкладывается в рассмотренную схему. Если Центр получает $l = 0$ бит информации, то количество разных сообщений, которое он может получить, равно $m = 2^l = 1$, и множество $S = \{0,1\}^0$ состоит из одной точки. При этих соглашениях определение величины $R_1(0)$ и утверждение теоремы 2 оказываются корректными и для величины $R_1(0)$ можно получить выражение

$$R_1(0) = \max_{u \in U} \int_A \min_{v^1 \in E^1(u,\alpha)} \dots \min_{v^n \in E^n(u,\alpha)} g(u, v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

Случай, когда Центр имеет точную информацию о реализовавшемся значении неопределенного фактора, формально не вкладывается в рассмотренную схему, но может быть рассмотрен аналогично.

В этом случае стратегиями Центра являются произвольные функции $u_{\#} : A \to U$. При фиксированной стратегии $u_{\#} \in \Phi(A, U)$ и значении неопределенного фактора $\alpha \in A$ агент i выберет свое управление из множества

$$Br^i(u_{\#}, \alpha) = \left\{ v^i \in V^i : \; h^i\left(u_{\#}(\alpha), v^i, \alpha\right) = \max_{v^i \in V^i} h^i\left(u_{\#}(\alpha), v^i, \alpha\right) \right\}.$$

Поэтому осторожный Центр может рассчитывать на результат

$$\min_{v^1 \in Br^1\left(u_{\#}, \alpha\right)} \dots \min_{v^n \in Br^n\left(u_{\#}, \alpha\right)} g\left(u_{\#}(\alpha), v^1, \dots, v^n, \alpha\right),$$

математическое ожидание которого равно

$$\int_A \min_{v^1 \in Br^1\left(u_{\#}, \alpha\right)} \dots \min_{v^n \in Br^n\left(u_{\#}, \alpha\right)} g\left(u_{\#}(\alpha), v^1, \dots, v^n, \alpha\right) \wp(d\alpha),$$

и для величины $R_1(\infty)$ получим определение

$$R_1(\infty) = \sup_{u_{\#} \in \Phi(A,U)} \int_A \min_{v^1 \in Br^1(u_{\#}, \alpha)} \dots \min_{v^n \in Br^n(u_{\#}, \alpha)} g(u_{\#}(\alpha), v^1, \dots, v^n, \alpha) \wp(d\alpha).$$

Рассуждения, вполне аналогичные доказательству теоремы 2, позволяют получить для этой величины выражение

$$R_1(\infty) = \int_A \max_{u \in U} \min_{v^1 \in E^1(u,\alpha)} \dots \min_{v^n \in E^n(u,\alpha)} g\left(u, v^1, \dots, v^n, \alpha\right) \wp(d\alpha).$$

5. Сравнение централизованного и децентрализованного способов управления

Для любого l имеет место неравенство $R_0(l) \leqslant R_0(\infty)$.

В самом деле, рассмотрим отображение $\psi : \Phi(S, W) \times \Phi(A, S) \to \Phi(A, W)$, ставящее в соответствие паре функций (w_*, P) их композицию $w_{\#} = w_* \circ P$. Тогда для любого $\alpha \in A$ будем иметь $g\left(w_*\left(P(\alpha)\right), \alpha\right) = g\left(w_{\#}(\alpha), \alpha\right)$ и, следовательно, $\int\limits_A g\left(w_*\left(P(\alpha)\right), \alpha\right) \wp(d\alpha) = \int\limits_A g\left(w_{\#}(\alpha), \alpha\right) \wp(d\alpha)$. Фиксируем произвольное $\varepsilon > 0$ и выберем стратегию (w_*, P) так, что

$$\int_A g\left(w_*\left(P(\alpha)\right), \alpha\right) \wp(d\alpha) \geqslant \sup_{(w_*, P) \in \Phi(S,W) \times \Phi(A,S)} \int_A g\left(w_*\left(P(\alpha)\right), \alpha\right) \wp(d\alpha) - \varepsilon.$$

Тогда для соответствующей функции $w_{\#} = w_* \circ P$ будем иметь

$$R_0(l) - \varepsilon = \sup_{(w_*, P) \in \Phi(S,W) \times \Phi(A,S)} \int_A g\left(w_*\left(P(\alpha)\right), \alpha\right) \wp(d\alpha) - \varepsilon \leqslant$$

$$\leqslant \int\limits_A g\left(w_*\left(P\left(\alpha\right)\right),\alpha\right)\wp\left(d\alpha\right) = \int\limits_A g\left(w_{\#}\left(\alpha\right),\alpha\right)\wp\left(d\alpha\right) \leqslant$$

$$\leqslant \sup_{w_{\#}\in\Phi(A,U)} \int\limits_A g\left(w_{\#}\left(\alpha\right),\alpha\right)\wp\left(d\alpha\right) = R_0\left(\infty\right).$$

В силу произвольности ε отсюда получается неравенство $R_0\left(l\right) \leqslant R_0\left(\infty\right)$.

Далее, для любого l справедливо неравенство $R_0\left(l\right) \leqslant R_0\left(l+1\right)$.

Для доказательства рассмотрим отображения

$$\vartheta : \Phi\left(\left\{0,1\right\}^l, U\right) \to \Phi\left(\left\{0,1\right\}^{l+1}, U\right) \text{ и } \Theta : \Phi\left(A, \left\{0,1\right\}^l\right) \to \Phi\left(A, \left\{0,1\right\}^{l+1}\right),$$

определенные следующим образом. Пусть отображение ϑ ставит в соответствие функции $w_* : \left\{0,1\right\}^l \to U$ такую функцию $w_{**} : \left\{0,1\right\}^{l+1} \to U$, что при любом $s \in \left\{0,1\right\}^l$ выполняются равенства $w_{**}(s,0) = w_{**}(s,1) = w_*(s)$. А отображение Θ ставит в соответствие функции $P : A \to \left\{0,1\right\}^l$ такую функцию $P_* : A \to \left\{0,1\right\}^{l+1}$, что для любого $\alpha \in A$ имеет место равенство $P_*(\alpha) = (P(\alpha), 0)$. Непосредственно проверяется, что тогда для любого α справедливо равенство $g(w_{**}(P_*(\alpha)), \alpha) = g(w_*(P(\alpha)), \alpha)$, а значит,

$$\int\limits_A g(w_{**}(P_*(\alpha)), \alpha)\wp(d\alpha) = \int\limits_A g(w_*(P(\alpha)), \alpha)\wp(d\alpha).$$

Фиксируем произвольное $\varepsilon > 0$. Выберем стратегию (w_*, P) так, что

$$\int\limits_A g\left(w_*\left(P\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) \geqslant$$

$$\geqslant \sup_{(w_*,P)\in\Phi\left(\left\{0,1\right\}^l,W\right)\times\Phi\left(A,\left\{0,1\right\}^l\right)} \int\limits_A g\left(w_*\left(P\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) - \varepsilon.$$

Тогда

$$R_0\left(l\right) - \varepsilon = \sup_{(w_*,P)\in\Phi\left(\left\{0,1\right\}^l,W\right)\times\Phi\left(A,\left\{0,1\right\}^l\right)} \int\limits_A g\left(w_*\left(P\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) - \varepsilon \leqslant$$

$$\leqslant \int\limits_A g\left(w_*\left(P\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) = \int\limits_A g\left(w_{**}\left(P_*\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) \leqslant$$

$$\leqslant \sup_{(w_{**},P_*)\in\Phi\left(\left\{0,1\right\}^{l+1},W\right)\times\Phi\left(A,\left\{0,1\right\}^{l+1}\right)} \int\limits_A g\left(w_{**}\left(P_*\left(\alpha\right)\right),\alpha\right)\wp(d\alpha) = R_0\left(l+1\right).$$

В силу произвольности ε отсюда немедленно следует нужное неравенство $R_0\left(l\right) \leqslant R_0\left(l+1\right)$.

Практически дословно повторяя те же рассуждения, можно показать, что для любого l выполняются неравенства $R_1(l) \leqslant R_1(\infty)$ и $R_1(l) \leqslant R_1(l+1)$.

Установим еще одно неравенство. Очевидно, для любого $\alpha \in A$ выполняется неравенство

$$\max_{w \in W} g(w, \alpha) = \max_{u \in U} \max_{v^1 \in V^1} \dots \max_{v^n \in V^n} g\left(u, v^1, \dots, v^n, \alpha\right) \geqslant$$
$$\geqslant \max_{u \in U} \min_{v^1 \in E^1(u,\alpha)} \dots \min_{v^n \in E^n(u,\alpha)} g\left(u, v^1, \dots, v^n, \alpha\right).$$

Следовательно,

$$\int_A \max_{u \in U} \max_{v^1 \in V^1} \dots \max_{v^n \in V^n} g\left(u, v^1, \dots, v^n, \alpha\right) \wp(d\alpha) \geqslant$$
$$\geqslant \int_A \max_{u \in U} \min_{v^1 \in E^1(u,\alpha)} \dots \min_{v^n \in E^n(u,\alpha)} g\left(u, v^1, \dots, v^n, \alpha\right) \wp(d\alpha).$$

Таким образом, $R_0(\infty) \geqslant R_1(\infty)$.

Справедливо следующее утверждение.

Лемма 2. Имеет место равенство $\lim\limits_{l \to \infty} R_0(l) = R_0(\infty)$.

Доказательство леммы приведено в Приложении.

Из леммы 2 следует, что если $R_0(\infty) > R_1(\infty)$, то при достаточно большом l выполняется неравенство $R_0(l) > R_1(\infty)$, а значит, $R_0(l) > R_1(l)$.

Рассмотрим два частных случая.

Начнем со случая, когда интересы Центра и агентов "идеально согласованы", т.е. имеет место равенство $g\left(u, v^1, \dots, v^n, \alpha\right) = \sum\limits_{i=1}^{n} h^i\left(u, v^i, \alpha\right)$. В этом случае в силу теоремы 2 имеем

$$R_1(l) = \max_{(u_0, u_1, \dots, u_{m-1}) \in U^m} \int_A \max_{s=0,1,\dots,m-1} \max_{v_s^q \in V^1} \dots \max_{v_s^n \in V^n} g(u_s, v_s^1, \dots, v_s^n, \alpha) \wp(d\alpha),$$

или

$$R_1(l) = \max_{(u_0, u_1, \dots, u_{m-1}) \in U^m} \int_A \max_{\left(v_0^1, v_1^1, \dots, v_{m-1}^0\right) \in (V^1)^m} \dots$$
$$\dots \max_{\left(v_0^n, v_1^n, \dots, v_{m-1}^n\right) \in (V^n)^m} \max_{s=0,1,\dots,m-1} g(u_s, v_s^1, \dots, v_s^n, \alpha) \wp(d\alpha).$$

А результат теоремы 1 может быть переписан в виде

$$R_0(l) = \max_{(u_0, u_1, \dots, u_{m-1}) \in U^m} \max_{\left(v_0^1, v_1^1, \dots, v_{m-1}^1\right)} \dots \max_{\left(v_0^n, v_1^n, \dots, v_{m-1}^n\right)} \int_A \max_{s=0,1,\dots,m-1} g(w_s, \alpha) \wp(d\alpha).$$

Отсюда видно, что в данном случае $R_0(l) \leqslant R_1(l)$ (так как "максимум суммы меньше суммы максимумов").

Рассмотрим противоположный случай, когда агенты “враждебны Центру”, т.е. справедливо равенство $g\left(u, v^1, \ldots, v^n, \alpha\right) = -\sum\limits_{i=1}^{n} h^i\left(u, v^i, \alpha\right)$. Тогда в силу теоремы 1

$$R_1(l) =$$

$$= \max_{(u_0,u_1,\ldots,u_{m-1})\in U^m} \int\limits_A \max_{s=0,1,\ldots,m-1} \min_{v_s^q\in V^1} \min_{v_s^2\in V^2} \ldots \min_{v_s^n\in V^n} g(u_s, v_s^1, \ldots, v_s^n, \alpha)\wp\left(d\alpha\right).$$

Значит,

$$R_1(l) = \max_{(u_0,u_1,\ldots,u_{m-1})\in U^m} \int\limits_A \min_{\left(v_0^1,v_1^1,\ldots,v_{m-1}^0\right)\in(V^1)^m} \ldots$$

$$\ldots \min_{\left(v_0^n,v_1^n,\ldots,v_{m-1}^n\right)\in(V^n)^m} \max_{s=0,1,\ldots,m-1} g(u_s, v_s^1, \ldots, v_s^n, \alpha)\wp\left(d\alpha\right) \leqslant$$

$$\leqslant \max_{(u_0,u_1,\ldots,u_{m-1})\in U^m} \min_{\left(v_0^1,v_1^1,\ldots,v_{m-1}^0\right)\in(V^1)^m} \ldots$$

$$\ldots \min_{\left(v_0^n,v_1^n,\ldots,v_{m-1}^n\right)\in(V^n)^m} \int\limits_A \max_{s=0,1,\ldots,m-1} g(u_s, v_s^1, \ldots, v_s^n, \alpha)\wp\left(d\alpha\right) \leqslant$$

$$\leqslant \max_{(u_0,u_1,\ldots,u_{m-1})\in U^m} \max_{\left(v_0^1,v_1^1,\ldots,v_{m-1}^0\right)\in(V^1)^m} \ldots$$

$$\ldots \max_{\left(v_0^n,v_1^n,\ldots,v_{m-1}^n\right)\in(V^n)^m} \int\limits_A \max_{s=0,1,\ldots,m-1} g(u_s, v_s^1, \ldots, v_s^n, \alpha)\wp\left(d\alpha\right) = R_0\left(l\right).$$

Понятно, что все доказанные в этом разделе неравенства могут обращаться в равенство (например, если функция g постоянна). Однако “в типичном случае” все они обращаются в строгие неравенства.

Полученные результаты можно суммировать следующим образом.

Теорема 3. При фиксированном объеме доступной информации l могут выполняться как неравенство $R_0\left(l\right) > R_1(l)$, так и противоположное неравенство $R_1\left(l\right) > R_0\left(l\right)$. Однако при любом $\varepsilon > 0$ при достаточно больших l имеет место неравенство $R_0\left(l\right) > R_1(l) - \varepsilon$, а в типичном случае при достаточно больших l справедливо и неравенство $R_0\left(l\right) > R_1(l)$.

Таким образом, при малых объемах доступной оперирующей стороне информации в случае, когда интересы агентов “хорошо согласованы” с интересами Центра, выгоднее децентрализованный способ управления, а в случае, когда интересы агентов “плохо согласованы” с интересами Центра, выгодна централизация управления. А при больших объемах доступной оперирующей стороне информации всегда выгоднее централизованный способ управления.

6. Заключение

Вполне естественные предположения, принятые при построении модели, приводят к хорошо интерпретируемым качественным выводам. А получившиеся математические задачи имеют достаточно конструктивные решения.

Разумеется, трудно говорить о непосредственном применении построенных моделей на практике. Но, благодаря своей абстрактности, эти модели имеют высокую степень общности. И в описанную схему можно вложить многие более конкретные модели. Кроме того, какие-то из принятых гипотез можно варьировать, что открывает широкое поле для дальнейших исследований. Кроме направлений, отмеченных в заключении в [9], стохастическая постановка допускает еще одно видоизменение, быть может даже более интересное. Речь идет о следующем.

В данной статье использовалось осреднение выигрыша оперирующей стороны и жесткое гарантированное ограничение на объем обрабатываемой ею информации. В самом деле, по постановке оперирующая сторона независимо от реализовавшегося значения неопределенного фактора получает сообщение длины l. Наверное, даже более естественно требовать, чтобы все сообщения кодировались словами, длина которых не превосходит l. Но при таком ограничении это почти ничего не меняет по существу.

Можно ограничение сформулировать иначе. Допустим, что сообщения кодируются словами разной длины, и требуется, чтобы математическое ожидание длины получаемого оперирующей стороной сообщения не превосходило заданной величины. Тогда можно существенно выиграть за счет того, что часто встречающиеся сообщения будут кодироваться короткими словами, а длинные слова будут использоваться для сообщения о редких событиях.

Эта идея впервые была использована в теории передачи информации К. Шенноном. Пионерские работы Шеннона дали толчок большому числу исследований. В результате этого были получены важные результаты, далеко выходящие за рамки теории передачи информации. В задачах принятия решений аналогичные постановки, по-видимому, до сих пор не исследовались.

ПРИЛОЖЕНИЕ

Доказательство леммы 1. Для удобства будем использовать терминологию из книги [11] и только те утверждения, которые явно сформулированы в ней на с. 272 и с. 283.

Нужно доказать, что при любом c измеримо множество $\{\alpha \in A : \varphi(\alpha) < c\}$. Но

$$\{\alpha \in A : \varphi(\alpha) < c\} = \bigcap_{s=0}^{m-1} \left\{ \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha) < c \right\}.$$

Поэтому достаточно доказать, что при любом c измеримо множество

$$\left\{ \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha) < c \right\},$$

т.е. при любом $u_s \in U$ измерима функция

$$\psi(\alpha) = \min_{v^1 \in E^1(u_s,\alpha)} \dots \min_{v^n \in E^n(u_s,\alpha)} g(u_s, v^1, \dots, v^n, \alpha).$$

Для упрощения формул введем обозначения $v = \left(v^1, \dots, v^n\right)$, $V = \prod_{i=1}^{n} V^i$ и рассмотрим функцию $h(u, v, \alpha) = \sum_{i=1}^{n} h^i\left(u, v^i, \alpha\right)$. В силу специального вида функции h выполняется равенство $\psi(\alpha) = \min_{v \in E(u_s,\alpha)} g(u_s, v, \alpha)$, где

$$E(u_s, \alpha) = \left\{ v \in V : \ h(u_s, v, \alpha) = \max_{\omega \in V} h(u_s, \omega, \alpha) \right\}.$$

Функция $\psi(\alpha)$ измерима тогда и только тогда, когда измерима функция $-\psi(\alpha)$, следовательно, достаточно доказать, что при любом c измеримо множество $C = \{\alpha \in A : -\psi(\alpha) < -c\} = \{\alpha \in A : \psi(\alpha) > c\}$. А поскольку мера $\wp$ предполагается борелевской, достаточно доказать, что это множество является открытым. Допустим противное. Тогда существует элемент $\alpha \in A$ и сходящаяся к нему последовательность $\alpha_1, \alpha_2, \dots$ такие, что $\alpha \in C$, но $\alpha_k \notin C$, $k = 1, 2, \dots$. Функция h является непрерывной, а множество $E(u_s, \alpha_k)$ задается условием типа равенства, поэтому оно замкнуто. А поскольку пространство V компактно, его замкнутое подмножество $E(u_s, \alpha_k)$ тоже является компактным. Значит, в некоторой точке $v_k \in E(u_s, \alpha_k)$ достигается минимум $\min_{v \in E(u_s,\alpha_k)} g(u_s, v, \alpha_k)$. Так как по предположению $\alpha_k \notin C$, выполняется неравенство $g(u_s, v_k, \alpha_k) \leqslant c$.

Пространство V является компактным, поэтому последовательность $v_1, v_2, \dots$ можно считать сходящейся к некоторому элементу $v_0 \in V$ (в противном случае можно перейти к подпоследовательности).

Фиксируем произвольное $\omega \in V$. Поскольку $v_k \in E(u_s, \alpha_k)$, выполняется неравенство $h(u_s, v_k, \alpha_k) \geqslant h(u_s, \omega, \alpha_k)$. Переходя в этом неравенстве к пределу при $k \to \infty$, получим $h(u_s, v_0, \alpha) \geqslant h(u_s, \omega, \alpha)$. В силу произвольности ω отсюда следует, что $v_0 \in E(u_s, \alpha)$. А переходя к пределу в неравенстве $g(u_s, v_k, \alpha_k) \leqslant c$, получим $g(u_s, v_0, \alpha) \leqslant c$ и тем более $\psi(\alpha) = \min_{v \in E(u_s,\alpha)} g(u_s, v, \alpha) \leqslant c$, что противоречит условию $\alpha \in C$.

Полученное противоречие доказывает лемму.

Замечание 4. Для строгого обоснования результатов раздела 3 нужно доказать, что при всех $w_0, w_1, \dots, w_{m-1}$ измерима функция $\phi(\alpha) = \max_{s=0,1,\dots,m-1} g(w_s, \alpha)$. Это утверждение является частным случаем только что доказанной леммы. Действительно, рассмотрим модель, в которой множество V^i состоит из одной точки v^i, $i = 1, \dots, n$. Тогда $\varphi(\alpha) = \max_{s=01,\dots,m-1} g(u_s, v^1, \dots, v^n, \alpha)$, и нужный результат только обозначением отличается от уже доказанного.

Доказательство леммы 2. Фиксируем произвольное $\varepsilon > 0$.

Для каждого $\alpha \in A$ можно выбрать $\omega(\alpha) \in W$ так, что $g(\omega(\alpha), \alpha) = \max\limits_{\varpi \in W} g(\varpi, \alpha)$ и, значит, $g(\omega(\alpha), \alpha) > \max\limits_{\varpi \in W} g(\varpi, \alpha) - \varepsilon$. Поэтому открытые множества

$$O(\omega) = \left\{ \alpha \in A : g(\omega, \alpha) > \max_{\varpi \in W} g(\varpi, \alpha) - \varepsilon \right\}$$

покрывают компактное множество A. Следовательно, можно выбрать конечный набор $\omega_0, \omega_1, \ldots, \omega_k$ элементов множества W так, что множества $O(\omega_0), O(\omega_1), \ldots, O(\omega_k)$ будут по-прежнему покрывать множество A. Значит, для любого $\alpha \in A$ будет выполняться условие

$$\max_{s=0,1,\ldots,k} g(\omega_s, \alpha) > \max_{\varpi \in W} g(\varpi, \alpha) - \varepsilon.$$

Если n выбрано так, что $m = 2^n \geqslant k$, то можно положить $w_s = \omega_s$ при $s \leqslant k$ и $w_s = \omega_k$ при $s > k$, и тогда будет

$$\max_{s=0,1,\ldots,m-1} g(w_s, \alpha) > \max_{\varpi \in W} g(\varpi, \alpha) - \varepsilon.$$

Следовательно,

$$\int_A \max_{s=0,1,\ldots,m-1} g(w_s, \alpha)\, \wp(d\alpha) > \int_A \max_{\varpi \in W} g(\varpi, \alpha)\, \wp(d\alpha) - \varepsilon = R_0(\infty) - \varepsilon$$

и тем более

$$R_0(l) = \max_{(w_0, w_1, \ldots, w_{m-1}) \in W^m} \int_A \max_{s=0,1,\ldots,m-1} g(w_s, \alpha) \wp(d\alpha) > R_0(\infty) - \varepsilon.$$

В силу произвольности ε отсюда следует утверждение леммы.

СПИСОК ЛИТЕРАТУРЫ

1. *Месарович М., Мако Д., Такакхара И.* Теория иерархических многоуровневых систем. М.: Мир, 1973.
2. *Новиков Д.А.* Институциональное управление организационными системами. М.: ИПУ РАН, 2003.
3. *Воронин А.А., Мишин С.П.* Оптимальные иерархические структуры. М.: ИПУ РАН, 2003.
4. *Alonso R., Dessein W., Matouschek N.* When Does Coordination Require Centralization? // Amer. Econom. Rev. 2008. V. 98. P. 145–179.
5. *Melamud N., Mookherjee D., Reichelstein S.* Hierarchical Decentralization of Incentive Contracts // Rand J. Econom. 1995. V. 26. P. 654–672.
6. *Гермейер Ю.Б., Моисеев Н.Н.* О некоторых задачах теории иерархических систем / Пробл. прикл. мат. и механики. М.: Наука, 1971. С. 30–43.
7. *Моисеев Н.Н.* Математические задачи системного анализа. М.: Наука, 1981.

8. *Моисеев Н.Н.* Иерархические структуры и теория игр // Изв. АН СССР. Сер. Техн. кибернетика. 1973. № 6. С. 1–11.
9. *Горелов М.А., Ерешко Ф.И.* Информированность и децентрализация управления // АиТ. 2019. № 6. С. 156–172.

   Gorelov V.A., Ereshko F.I. Awareness and Control Decentralization // Autom. Remote Control. 2019. V. 80. No. 6. P. 1063–1076.
10. *Энгелькинг Р.* Общая топология. М.: Мир, 1986.
11. *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. М.: Наука, 1981.

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

Поступила в редакцию 14.02.2019

После доработки 11.04.2019

Принята к публикации 25.04.2019

© 2020 г. Е.С. ПАЛАМАРЧУК, канд. физ.-мат. наук (e.palamarchuck@gmail.com)
(Центральный экономико-математический институт РАН, Москва)

ОПТИМАЛЬНЫЙ РЕГУЛЯТОР ДЛЯ НЕАВТОНОМНОЙ ЛИНЕЙНОЙ СТОХАСТИЧЕСКОЙ СИСТЕМЫ С ДВУСТОРОННИМ ЦЕЛЕВЫМ ФУНКЦИОНАЛОМ[1]

Рассматривается задача стохастического линейного регулятора на бесконечном интервале времени с двусторонним целевым функционалом и переменной матрицей диффузии. В двустороннем квадратичном целевом функционале пределы интегрирования имеют противоположный знак и зависят от длины интервала планирования. Показано, что при ограничениях на рост матрицы диффузии известный закон управления в виде линейной обратной связи по состоянию будет являться оптимальным по критерию обобщенного долговременного среднего и его потраекторного аналога. Также проводится анализ вероятностного поведения оптимальной траектории развития системы.

Ключевые слова: стохастический линейный регулятор, двусторонний целевой функционал, переменная матрица диффузии.

DOI: 10.31857/S0005231020010055

1. Введение

Стохастические линейные регуляторы относятся к классу систем управления, имеющих важное теоретическое и практическое значение, см. [1, гл. 3]. При этом их динамика обычно рассматривается на положительной полуоси изменения параметра времени $t \in [t_0, T]$ и горизонте планирования $[t_0, T] \subseteq [0, +\infty)$. Вместе с тем, в теоретико-операторной перспективе, т.е. при возникновении бесконечномерных пространств состояний, см., например, [2], анализ эволюции систем может проводиться на всей числовой прямой, т.е. при $t \in (-\infty, +\infty)$, и постановка задач управления осуществляется на интервалах $[t_0 - T, t_0 + T]$, где $T \geqslant 0$, и затем $T \to +\infty$, см. [3, 4]. Кроме того, как подчеркивается в [5], существуют области приложений (обработка сигналов, статистическое оценивание, передача информации и др.), моделирование в которых также предполагает возможность значений независимой переменной $t \in (-\infty, +\infty)$. Опишем систему управления, исследуемую в данной статье. Пусть на полном вероятностном пространстве $\{\Omega, \mathcal{F}, \mathbf{P}\}$ задан n-мерный случайный процесс X_t, $t \in \mathbb{R}$, $\mathbb{R}$ — множество действительных чисел, описываемый уравнением

$$dX_t = A_t X_t dt + B_t U_t dt + G_t dw_t, \tag{1}$$

[1] Работа выполнена в рамках НИР "Теория и методы для компьютерного и математического моделирования и анализа общественных систем и процессов", номер государственной регистрации АААА-Ф18-118021990120-2.

где A_t, B_t — ограниченные матрицы с зависящими от времени элементами; шумовые воздействия моделируются с помощью так называемого двустороннего винеровского процесса w_t, $t \in \mathbb{R}$, задаваемого обычным образом как $w_t = w_t^{(1)}$, $t \geqslant 0$, и $w_t = w_{-t}^{(2)}$, $t < 0$, при двух независимых d-мерных стандартных винеровских процессах $w_t^{(1)}$, $w_t^{(2)}$, $t \geqslant 0$, см., например, [6, с. 7]; множество допустимых управлений $\mathcal{U}$ состоит из k-мерных квадратично интегрируемых случайных процессов U_t, $t \in \mathbb{R}$, согласованных с фильтрацией $\{\mathcal{F}_t\}_{t\in\mathbb{R}}$, $\mathcal{F}_t = \sigma\{w_s, s \leqslant t\}$ ($\sigma(\cdot)$ — знак σ-алгебры), таких что существует решение уравнения (1), т.е., см., например, [3], процесс X_t, $t \in \mathbb{R}$, для которого почти наверное (п.н.) выполняется $X_t = X_s + \int_s^t A_\tau X_\tau \, d\tau + \int_s^t B_\tau U_\tau \, d\tau + \int_s^t G_\tau \, dw_\tau$ при всех $s \leqslant t$; G_t — матрица диффузии, о предположениях относительно ее элементов будет сказано далее, а здесь отметим, что в рассмотрение могут включаться ситуации как ограниченных параметров возмущений (например, постоянных $G_t \equiv G$ или затухающих $\|G_t\| \to 0$), так и нарастающих $\|G_t\| \to \infty$, $t \to \pm\infty$ ($\|\cdot\|$ — матричная евклидова норма).

Для $T > 0$ в качестве двустороннего целевого функционала на $[-T, T]$ определим случайную величину

$$J_{2T}(U) = \int_{-T}^{T} (X_t' Q_t X_t + U_t' R_t U_t) \, dt, \tag{2}$$

где $U \in \mathcal{U}$ — допустимое управление; $Q_t \geqslant qI$, $R_t \geqslant \rho I$, $t \in \mathbb{R}$, — ограниченные симметричные матрицы, $q, \rho > 0$ — некоторые константы ($'$ — знак транспонирования, запись $A \geqslant B$ для матриц означает, что разность $A - B$ неотрицательно определена, I — единичная матрица).

Ранее задачи стохастического линейного регулятора на бесконечном интервале времени ($T \to +\infty$) с функционалом вида (2) рассматривались в [7] при управлении передачей информации в сетях, для приложений в инженерных системах — см. [8; 9, часть 13.2.10]. При этом в качестве критерия оптимальности использовалось долговременное среднее, т.е. $\limsup\limits_{T\to+\infty}\{\mathrm{E}J_{2T}/(2T)\} \to \inf\limits_{U\in\mathcal{U}}$. Очевидно, что при таком подходе не учитывается специфика изменения матрицы диффузии G_t во времени, например, ее неограниченность на бесконечности, как, например, в когнитивной модели [10], или ее вырождение, см. случай диффузии [11]. В данной статье для определения управлений, оптимальных в среднем на бесконечном интервале времени, предлагается критерий, который обобщает приведенный выше:

$$\limsup_{T\to+\infty} \frac{\mathrm{E}J_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2 \, dt} \to \inf_{U\in\mathcal{U}}. \tag{3}$$

Более сильным (в вероятностном смысле) критерием, чем долговременное среднее, является потраекторное эргодическое, когда задача

$$\limsup_{T\to+\infty}\{J_{2T}/(2T)\} \to \inf_{U\in\mathcal{U}}$$

решается с вероятностью единица, см. [3]. При учете фактора воздействия на динамику системы переменной матрицы диффузии можно использовать потраекторное обобщенное долговременное среднее, когда ставится задача

$$\limsup_{T\to+\infty} \frac{J_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} \to \inf_{U\in\mathcal{U}} \text{ с вероятностью единица.} \tag{4}$$

Следует отметить, что обобщенные долговременные средние также вводились в [12–14] для стохастического линейного регулятора с односторонним целевым функционалом, т.е. при интегрировании на $[0,T]$ в (2). Задачи с двусторонними функционалами рассматривались в [3, 4], а используемые там критерии оптимальности являлись стандартными для систем с ограниченными коэффициентами (упомянутые выше долговременное среднее и потраекторное эргодическое). При этом для элементов множества допустимых управлений предполагались выполненными условия конечности моментов соответствующих им процессов, точнее, $\sup_{t\in\mathbb{R}}(\mathrm{E}\|X_t\|^2 + \mathrm{E}\|U_t\|^2) < \infty$, см. [4], или же $\sup_{t\in\mathbb{R}}(\mathrm{E}\|X_t\|^4 + \mathrm{E}\|U_t\|^4) < \infty$, см. [3], а также эргодического среднего $\limsup_{T\to\infty}\{(2T)^{-1}\int_{-T}^{T}\|U_t\|^2\,dt\} < \infty$ в [4]. По сравнению с анализом, проведенным в [3, 4], в настоящей статье представляется ряд обобщений (для случая конечномерных систем управления). Во-первых, включается ситуация неограниченного изменения во времени матрицы диффузии ($\|G_t\| \to \infty$, $t \to \pm\infty$), и применяются новые критерии обобщенных долговременных средних (см. (3), (4)), учитывающие этот факт. Во-вторых, задачи (3) и (4) решаются для гораздо более широкого класса управляющих воздействий, чем было сделано в [3, 4]: достаточно потребовать существования решения (1) и квадратичной интегрируемости управлений, т.е. $\int_s^t \|U_\tau\|^2\,d\tau < \infty$, $-\infty < s \leqslant t < +\infty$. Важно подчеркнуть, что известная форма оптимальной стратегии в виде линейной обратной связи по состоянию, структура которой также включает решение уравнения Риккати (см., например, [1, 3, 4]), сохраняется и в рассматриваемом случае для (3) и (4). Для оптимальной траектории, соответствующей такому управлению, в [3] было выявлено свойство глобальной асимптотической устойчивости в среднем квадратичном. В данной статье будут получены более точные оценки изменений этого процесса во времени как в среднем квадратичном смысле, так и с вероятностью единица, в зависимости от коэффициентов матрицы диффузии, что представляется обобщением результата [15], где изучался скалярный стационарный процесс. Таким образом, цель данной статьи — нахождение управления U_t^*, оптимального в задачах (3) и (4), и исследование свойств соответствующей ему оптимальной траектории X_t^* при $t \to \pm\infty$. Дальнейшее изложение организовано следующим образом. В разделе 2 вводятся основные предположения о параметрах системы управления (1)–(2) и решается задача (3). Раздел 3 посвящен проблеме потраекторной оптимальности U^* для задачи (4) и стохастическому анализу динамики траектории X^*. Кроме того, в разделе 3 приводятся примеры различных классов функций, которые могут описывать изменение матрицы диффузии G_t в рамках основных предположений. Заключение содержит выводы и информацию о направлении дальнейших исследований.

2. Оптимальность в среднем на бесконечном интервале времени

Сначала сформулируем предположения о коэффициентах (1)–(2), в рамках которых будут получены основные результаты.

Предположение $\mathcal{AB}$. *Пара матриц* (A_t, B_t) *является стабилизируемой при* $t \in \mathbb{R}$.

Стабилизируемость пары (A_t, B_t), см., например, [2, 4], означает существование ограниченной матрицы K_t с кусочно-непрерывными элементами, при которой матрица $\mathcal{A}_t = A_t + B_t K_t$ является экспоненциально устойчивой, $t \in \mathbb{R}$, т.е. соответствующая ей фундаментальная матрица $\Phi(t,s)$ допускает оценку $\|\Phi(t,s)\| \leqslant \kappa_0 e^{-\kappa(t-s)}$, $s \leqslant t$, $\kappa_0, \kappa > 0$ — константы. При этом, как известно, фундаментальная матрица определяется из решения задачи $\dfrac{\partial \Phi(t,s)}{\partial t} = \mathcal{A}_t \Phi(t,s)$, $\Phi(s,s) = I$. Далее формулируется предположение относительно параметров возмущений, т.е. матрицы G_t, $t \in \mathbb{R}$. Введем множество $\mathcal{T} = \{-\infty; +\infty; \pm\infty\}$ и запись $t \to \mathcal{T}$ будем использовать для сокращенного обозначения ситуаций $t \to -\infty$, $t \to +\infty$ или $t \to \pm\infty$.

Предположение $\mathcal{G}$. *Для элементов матрицы диффузии* G_t *выполняется одно из следующих условий*:

1) G_t — *ограничена при* $t \to \mathcal{T}$;

2) $\|G_t\| \to +\infty$, G_t — *дифференцируема, при этом* $d \ln \|G_t\|/dt \to 0$, $t \to \mathcal{T}$.

Необходимо подчеркнуть, что возможность выполнения условий 1, 2 для матрицы G_t зависит от того, на какой полуоси (положительной или отрицательной) изменяется параметр $t \in \mathbb{R}$. В частности, для $\|G_t\| = e^{\sqrt[m]{t}}$, где m — нечетное число, имеет место условие 1 при $t \to -\infty$ и условие 2 при $t \to +\infty$.

В условиях предположения $\mathcal{AB}$, см. [2, 4], существует управление

$$U_t^* = -R_t^{-1} B_t' \Pi_t X_t^*, \tag{5}$$

где ограниченная симметричная матрица $\Pi_t \geqslant pI$, $p > 0$ — константа, удовлетворяет уравнению Риккати

$$\dot{\Pi}_t + \Pi_t A_t + A_t' \Pi_t - \Pi_t B_t R_t^{-1} B_t' \Pi_t + Q_t = 0. \tag{6}$$

При подстановке (5) в (1) становится очевидно, что процесс X_t^*, $t \in \mathbb{R}$, является решением линейного стохастического дифференциального уравнения (СДУ)

$$dX_t^* = (A_t - B_t R_t^{-1} B_t' \Pi_t) X_t^* dt + G_t dw_t \tag{7}$$

и представляет собой аналог процесса Орнштейна–Уленбека при $t \in \mathbb{R}$ в случае СДУ с переменными коэффициентами. При этом матрица $A_t^* = A_t - B_t R_t^{-1} B_t' \Pi_t$ — экспоненциально устойчива, см. [2, 4], а ряд других свойств X_t^*, $t \in \mathbb{R}$, устанавливается в лемме 1.

Лемма 1. Пусть выполнены предположения $\mathcal{AB}$ *и* $\mathcal{G}$. *Тогда решением* (7) *является процесс вида* $X_t^* = \int_{-\infty}^t \Phi(t,s) G_s dw_s$, *где* $\Phi(t,s)$ — *фундаментальная матрица, соответствующая экспоненциально устойчивой матрице* $A_t^* = A_t - B_t R_t^{-1} B_t' \Pi_t$. *При этом существует константа* $c_G > 0$, *такая что* $E\|X_t^*\|^2 \leqslant c_G \max\{1, \|G_t\|^2\}$, $t \in \mathbb{R}$.

Доказательство леммы 1, а также всех последующих утверждений вынесено в приложение. В следующей далее теореме 1 приводится результат об оптимальности в среднем на бесконечном интервале времени управления U^*.

Теорема 1. Пусть выполнены предположения $\mathcal{AB}$ и $\mathcal{G}$. Тогда закон управления U^*, задаваемый (5)–(7)*, является решением задачи*

$$\limsup_{T\to+\infty} \frac{EJ_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} \to \inf_{U\in\mathcal{U}}, \tag{8}$$

при этом

$$0 < \limsup_{T\to+\infty} \frac{EJ_{2T}(U^*)}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} = \limsup_{T\to+\infty} \frac{\int\limits_{-T}^{T} tr(G_t'\Pi_t G_t)\,dt}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} < \infty, \tag{9}$$

где $\mathrm{tr}(\cdot)$ *— след матрицы.*

3. Потраекторная стохастическая оптимальность

В приводимых далее леммах 2 и 3 характеризуются асимптотические свойства траекторий процесса X_t^*, $t\in\mathbb{R}$. Знание этих свойств оказывается необходимым при исследовании стохастической оптимальности управления U^* в задаче (4).

Лемма 2. Пусть выполнены предположение $\mathcal{AB}$ и п. 2 предположения $\mathcal{G}$. Тогда существует константа $\bar{c}>0$, такая что

$$\limsup_{t\to\mathcal{T}} \frac{\|X_t^*\|^2}{\|G_t\|^2 \ln|t|} < \bar{c} < \infty \quad \textit{с вероятностью единица,}$$

где $|\cdot|$ *— модуль скалярной переменной.*

Приведенная в лемме 2 функция $h_t = \|G_t\|^2 \ln|t|$ является мажорантой, т.е. верхней функцией процесса X_t^*, см. [16, определение 1], при условии выполнения п. 2 предположения $\mathcal{G}$. Для ограниченной G_t, $t\geqslant 0$, результат о виде h_t был получен ранее в [16], частный случай скалярного стационарного процесса рассмотрен в [15].

Лемма 3. Пусть выполнены предположение $\mathcal{AB}$ и предположение $\mathcal{G}$. Если в п. 2 предположения $\mathcal{G}$ также $d\ln\|G_t\|/dt\cdot\ln|t|\to 0$, $t\to\mathcal{T}$, то

$$\lim_{T\to+\infty} \frac{\|X_{-T}^*\|^2 + \|X_T^*\|^2}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} = 0 \quad \textit{с вероятностью единица.}$$

Соотношение в лемме 3 утверждает, что нормировка при помощи $\Gamma_T = \sqrt{\int_{-T}^{T} \|G_t\|^2\, dt}$ как прошлых (X^*_{-T}), так и последующих (X^*_T) значений траектории обеспечит стремление результирующего процесса к нулю п.н. с ростом длины "окна" рассматриваемых наблюдений. Заданная таким образом функция Γ_T определяет среднеквадратичное отклонение компонент вектора интегральных шумовых воздействий за период $[-T, T]$, точнее, берется $\mathcal{Z}_T = \int_{-T}^{T} G_t dw_t$ и тогда $\mathrm{E}\|\mathcal{Z}_T\|^2 = \int_{-T}^{T} \|G_t\|^2\, dt$.

При анализе задачи (4) для случая $\|G_t\| \to \infty$, $t \to \mathcal{T}$, потребуется выполнение более сильного условия, чем сформулированное в п. 2 предположения $\mathcal{G}$.

Предположение $\mathcal{G}1$. *Пусть в п. 2 предположения* $\mathcal{G}$ *выполнено соотношение* $d \ln \|G_t\|/dt \cdot \ln|t|(\ln\ln|t| + \ln\ln\|G_t\|) \to 0$, $t \to \mathcal{T}$, *и при этом* $\|G_t\|$ *является монотонной функцией,* $t \to \mathcal{T}$.

Основным результатом данного раздела является утверждение теоремы 2 о потраекторной оптимальности управления U^*.

Теорема 2. Пусть выполнены условия теоремы 1 и предположение $\mathcal{G}1$. *Если* $\int_{-T}^{T} \|G_t\|^2\, dt \to \infty$, $T \to +\infty$, *то оптимальный в среднем закон управления* U^* *будет также являться решением задачи с потраекторным критерием обобщенного долговременного среднего, т.е.*

$$\limsup_{T\to+\infty} \frac{J_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\, dt} \to \inf_{U\in\mathcal{U}} \quad \textit{с вероятностью единица,} \tag{10}$$

при этом

$$\limsup_{T\to+\infty} \frac{J_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\, dt} = \limsup_{T\to+\infty} \frac{EJ_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\, dt} \qquad \textit{п.н.} \tag{11}$$

Приведем примеры различных классов функций, описывающих динамику нормы матрицы диффузии G_t. Используемый далее знак отношения $f_t \sim g_t$ для двух скалярных неотрицательных функций f_t и g_t означает, что $0 < \lim_{t\to\pm\infty}(f_t/g_t) < \infty$.

Пример 1.

1. Степенное семейство $\|G_t\|^2 \sim |t|^{2\alpha}$, $\alpha \in \mathbb{R}$: при $\alpha \leqslant 0$ имеет место п. 1 предположения $\mathcal{G}$ и для $\alpha > 0$ — п. 2. Так как $d \ln \|G_t\|/dt \sim 1/|t|$, а $\ln\ln\|G_t\| \sim \ln|t|$, то соотношение в предположении $\mathcal{G}1$ возникает при любом числе α. При этом условия теоремы 2 будут выполнены для $\alpha \geqslant -1/2$.

2. Логарифмическое семейство $\|G_t\|^2 \sim \ln^{2\alpha}|t|$, $\beta \in \mathbb{R}$: если $\beta \leqslant 0$, то имеет место п. 1 предположения $\mathcal{G}$ и при $\beta > 0$ — п. 2. В силу того что $d \ln \|G_t\|/dt \sim 1/(|t| \ln|t|)$, а функция $\ln\ln\|G_t\| \sim \ln\ln|t|$, требование предположения $\mathcal{G}1$ выполняется при любом β. Также для каждого $\beta \in \mathbb{R}$ будут выполнены условия теоремы 2.

3. Экспоненциальное семейство $\|G_t\|^2 \sim e^{|t|^\mu}$, $\mu < 1$: при $\mu \leqslant 0$ имеет место п. 1 предположения $\mathcal{G}$ и для $\mu > 0$ — п. 2. Также $d \ln \|G_t\|/dt \sim |t|^{\mu-1}$ и $\ln \ln \|G_t\| \sim |t|^\mu$, т.е. соотношение из предположения $\mathcal{G}1$ следует при любом $0 < \mu < 1$. Очевидно, что условия теоремы 2 выполняются при каждом $\mu < 1$.

4. Заключение

В статье рассмотрена задача стохастического линейного регулятора на бесконечном интервале времени с двусторонним целевым функционалом и переменной матрицей диффузии G_t. В двустороннем квадратичном целевом функционале $J_{2T}(U)$, см. (2), пределы интегрирования имеют противоположный знак и зависят от длины интервала планирования, т.е. $t \in [-T, T]$ в (2), а затем $T \to +\infty$. Показано, что в рамках стандартного условия стабилизируемости детерминированной системы (см. предположение $\mathcal{AB}$) и ограничениях на рост матрицы диффузии, см. предположения $\mathcal{G}$ и $\mathcal{G}1$, известный закон управления U^* в виде линейной обратной связи по состоянию (5)–(7) будет являться оптимальным по критерию обобщенного долговременного среднего (теорема 1) и его потраекторного аналога (теорема 2). Также в статье проведен анализ асимптотического вероятностного поведения X_t^* — оптимальной траектории развития системы, см. уравнение (7). В частности, установлено, что верхняя граница изменений X_t^* в среднем квадратичном может быть определена в зависимости от $\|G_t\|$ (лемма 1). В потраекторной динамике найдена достаточная нормировка, обеспечивающая стремление значений процесса к нулю с вероятностью единица (см. лемму 3) и определяемая через статистическую характеристику (стандартное отклонение) вектора интегральных шумовых воздействий. В качестве направления дальнейших исследований можно выделить изучение задачи трекинга стохастической траектории, обобщая, например, случай модели [7], где эталонная траектория является гауссовским процессом.

ПРИЛОЖЕНИЕ

Доказательство леммы 1. Так как $X_t^* = \Phi(t,0)\chi_t$, где $\chi_t = \int_{-\infty}^t \Phi(0,s)G_s dw_s$, то сначала требуется показать, что существует стохастический интеграл χ_t с бесконечным нижним пределом, а затем дифференцированием проверить, что X_t^* удовлетворяет (7). В силу определения двустороннего винеровского процесса $w_t = w_{-t}^{(2)}$, $t < 0$, где $w_\tau^{(2)}$ — стандартный винеровский процесс, $\tau \geqslant 0$, стохастическое исчисление для интегралов вида χ_t осуществляется по обычным правилам интегрирования по Ито, также см. [6, с. 13–14]. Для $t \geqslant 0$ процесс $X_t^* = \Phi(t,0)X_0^* + \int_0^t \Phi(t,s)G_s dw_s$, где $X_0^* = \chi_0$. Известно, см. [6, теорема 5.1, с. 54], что существование χ_t, $t \in \mathbb{R}$, связано с требованием $\mathrm{E}\|\chi_0\|^2 = \int_{-\infty}^0 \|\Phi(0,s)G_s\|^2\, ds < \infty$, которое выполняется в силу экспоненциальной устойчивости матрицы A_t^* и предположения $\mathcal{G}$. Действительно, $\|\Phi(0,s)\| \leqslant \kappa_0 e^{\kappa s}$, $s \leqslant 0$, и $\limsup_{s\to-\infty} \|G_s\|^2 e^{\gamma s} < \infty$ для любого $\gamma > 0$, тогда, выбирая $\gamma < 2\kappa$, имеем $\mathrm{E}\|\chi_0\|^2 < \infty$. Далее находим, что

$$\text{(П.1)} \qquad \mathrm{E}\|X_t^*\|^2 = \int_{-\infty}^t \mathrm{tr}\{\Phi(t,s)G_s G_s'\Phi'(t,s)\}\, ds \leqslant c \int_{-\infty}^t e^{-2\kappa(t-s)}\|G_s\|^2\, ds,$$

где $\mathrm{tr}(\cdot)$ — след матрицы, здесь и далее в качестве c обозначена некоторая положительная константа, конкретное значение которой несущественно и может меняться от формулы к формуле. Из (П.1) следует, что при ограниченной G_t выражение для $\mathrm{E}\|X_t^*\|^2$ также будет ограничено, $t \in \mathbb{R}$. Если же $\|G_t\| \to +\infty$, $t \to \mathcal{T}$, то при помощи интегрирования по частям, по аналогии с проделанным в [14, лемма 1] для случая $t \to +\infty$, можно показать, что $\limsup_{t\to\mathcal{T}}(\mathrm{E}\|X_t^*\|^2/\|G_t\|^2) < \infty$. Лемма 1 доказана.

Доказательство теоремы 1. Зафиксируем управление $U \in \mathcal{U}$ и определим соответствующий ему процесс по (1). Пусть $x_t = X_t^* - X_t$, $u_t = U_t^* - U_t$, $\bar{x} = X_0^* - X_0$, тогда получается представление

$$J_{2T}(U^*) - J_{2T}(U) = \tag{П.2}$$

$$= 2x'_T\Pi_T X_T^* - 2x'_{-T}\Pi_{-T}X_{-T}^* - \int_{-T}^{T}(x'_tQ_tx_t + u'_tR_tu_t)dt - 2\int_{-T}^{T}x'_t\Pi_tG_tdw_t.$$

Для оценки (П.2) проводится анализ динамики x_t при $t \in [-T, T]$. По построению

$$dx_t = A_tx_tdt + B_tu_tdt. \tag{П.3}$$

Пусть сначала $t \in [0, T]$. Тогда рассмотрение (П.3) с начальным условием $x_0 = \bar{x}$ и предположение $Q_t \geqslant qI$ приводят к решению (П.3) вида $x_T = \bar{\Phi}(T,0)\bar{x} + \int_0^T \bar{\Phi}(T,t)(\bar{k}\sqrt{Q_t}x_t + B_tu_t)dt$, где $\bar{\Phi}(t,s)$ — фундаментальная матрица, соответствующая экспоненциально устойчивой матрице $\bar{A}_t = A_t - \bar{k}\sqrt{Q_t}$ при некоторой константе $\bar{k} > 0$. Оценка приведенного выше соотношения дает

$$\|x_T\|^2 \leqslant \bar{c}e^{-\bar{\kappa}T}\|\bar{x}\|^2 + \bar{c}\int_0^T e^{-\bar{\kappa}(T-s)}(x'_sQ_sx_s + u'_sR_su_s)ds \tag{П.4}$$

с некоторыми константами $\bar{c}, \bar{\kappa} > 0$. Для случая $t \in [-T, 0]$ уравнение (П.3) рассматривается при граничном условии $x_0 = \bar{x}$. В силу $Q_t \geqslant qI$ существует константа $\tilde{k} > 0$, такая что матрица $\tilde{A}_t = A_t + \tilde{k}\sqrt{Q_t}$ является экспоненциально антиустойчивой, т.е. $\|\tilde{\Phi}(s,t)\| \leqslant \tilde{\kappa}e^{-\tilde{\kappa}_1(t-s)}$, $s \leqslant t$, а $\tilde{\kappa}, \tilde{\kappa}_1 > 0$ — константы. Тогда, представив решение (П.3) в виде $x_{-T} = \tilde{\Phi}(-T,0)\bar{x} - \int_{-T}^0 \tilde{\Phi}(-T,s)(\tilde{k}\sqrt{Q_s}x_s + B_su_s)ds$, будем при некоторой константе $\tilde{c} > 0$ иметь оценку

$$\|x_{-T}\|^2 \leqslant \tilde{c}e^{-\tilde{\kappa}_1T}\|\bar{x}\|^2 + \tilde{c}\int_{-T}^0 e^{-\tilde{\kappa}_1(T+s)}(x'_sQ_sx_s + u'_sR_tu_s)ds. \tag{П.5}$$

Тогда ограниченность Π_t, $t \in \mathbb{R}$, в совокупности с элементарным неравенством $2ab \leqslant ca^2 + b^2/c$, которое справедливо при произвольном $c > 0$, и (П.4)–(П.5) приводят к следующей оценке для ожидаемого значения (П.2):

$$\mathrm{E}J_{2T}(U^*) - \mathrm{E}J_{2T}(U) \leqslant c_0e^{-\kappa_1}\|\bar{x}\|^2 + c_1\mathrm{E}\|X_T^*\|^2 + c_2\mathrm{E}\|X_{-T}^*\|^2$$

с некоторыми константами $\kappa_1, c_0, c_1, c_2 > 0$. Применение нормировки $\int_{-T}^{T} \|G_t\|^2$, с учетом результата леммы 1 и условий предположения $\mathcal{G}$, в предельном переходе для $T \to +\infty$, обеспечивает выполнение соотношения

$$\limsup_{T\to+\infty} \frac{\mathrm{E}J_{2T}(U^*)}{\int\limits_{-T}^{T} \|G_t\|^2\,dt} \leqslant \limsup_{T\to+\infty} \frac{\mathrm{E}J_{2T}(U)}{\int\limits_{-T}^{T} \|G_t\|^2\,dt},$$

показывающего, что U^* является решением задачи (3). Следует заметить, что для процессов, определенных при всех $t \in \mathbb{R}$, как в (7), решение соответствующего уравнения представляется в интегральном виде $X_t^* = X_s^* + \int_s^t A_\tau^* X_\tau^* d\tau + + \int_s^t G_\tau\,dw_\tau$ при произвольном $s \in \mathbb{R}$, $s \leqslant t$. Тогда по замечанию [17, замечание 4.3.7, с. 99] известные результаты, в частности справедливость формулы Ито, могут быть распространены на случай таких процессов. Далее, по формуле Ито

(П.6)
$$J_{2T}(U^*) =$$

$$= [(X_{-T}^*)'\Pi_{-T}X_{-T}^*] - [(X_T^*)'\Pi_T X_T^*] + \int_{-T}^{T} \mathrm{tr}(G_t'\Pi_t G_t)\,dt + 2\int_{-T}^{T} (X_t^*)'\Pi_t G_t dw_t.$$

На основании неравенства из леммы 1 и свойства $pI \leqslant \Pi_t \leqslant cI$, $t \in \mathbb{R}$, выписывается двусторонняя оценка для ожидаемого значения (П.6): $\hat{c}_1 \int_{-T}^{T} \|G_t\|^2\,dt \leqslant \leqslant \mathrm{E}J_{2T}(U^*) \leqslant \hat{c}_2 \int_{-T}^{T} \|G_t\|^2\,dt$ при некоторых константах $\hat{c}_1, \hat{c}_2 > 0$, из которой следует (9). Теорема 1 доказана.

Доказательство леммы 2. Для случая $\mathcal{T} = +\infty$ процесс $X_t^* = = \Phi(t,0)X_0^* + \tilde{X}_t^*$, где $\tilde{X}_t^* = \int_0^t \Phi(t,s)G_s dw_s$, $t \geqslant 0$. В [14, лемма 2] было показано, что $\|\tilde{X}_t^*\|^2 \leqslant c_0\|G_t\|^2 \ln t$ п.н. при $t \to +\infty$ и детерминированной константе $c_0 > 0$. Так как X_0^* — случайная величина, а $\|\Phi(t,0\| \leqslant \kappa_0 e^{-\kappa t}$, то из приведенного выше результата для $\|\tilde{X}_t^*\|^2$ сразу получается утверждение доказываемой леммы. При $\mathcal{T} = -\infty$ сначала рассмотрим скалярный процесс z_t с уравнением динамики $dz_t = -\kappa z_t dt + \sigma_t dw_t$, $\kappa > 0$, и коэффициентом диффузии σ_t со свойствами из условия леммы 2. Тогда $z_t = e^{-\kappa t} I_t$, где $I_t = \int_{-\infty}^{t} e^{\kappa s}\sigma_s dw_s$. В стохастическом интеграле I_{-T}, $T \geqslant 0$, можно провести замену времени $\tau = -1/s$ и учесть, что $\tau w_{-1/\tau} = \hat{w}_\tau$, где $\hat{w}_\tau$, $\tau \geqslant 0$, — другой винеровский процесс, см., например, [18, с. 94]. Поэтому

$$I_{-T} = \int_0^{1/T} e^{-\kappa/\tau}\sigma_{-1/\tau}\left(\frac{d\hat{w}_\tau}{\tau} - \frac{\hat{w}_\tau}{\tau^2}d\tau\right).$$

При $T \to +\infty$ для оценки слагаемых в I_{-T} может использоваться локальный закон повторного логарифма [18, следствие 3, с. 93]. Пусть $I_T^{(1)} = = \int_0^{1/T} e^{-\kappa/\tau}\sigma_{-1/\tau}\frac{d\hat{w}_\tau}{\tau}$, тогда $|I_T^{(1)}| \leqslant \hat{c}_1 h_T^{(1)}$ при $h_T^{(1)} = \sqrt{M_T \ln\ln(1/M_T)}$, $M_T =$

$= \int_0^{1/T} e^{-2\kappa/\tau}\sigma^2_{-1/\tau}\frac{d\tau}{\tau^2}$ и некоторой константе $\hat{c}_1 > 0$. Для процесса $I_T^{(2)} = \int_0^{1/T} e^{-\kappa/\tau}\sigma_{-1/\tau}\frac{\hat{w}_\tau}{\tau^2}d\tau$ при $T \to +\infty$ будет иметь место оценка $|I_T^{(2)}| \leqslant \hat{c}_2 h_T^{(2)}$, где $h_T^{(2)} = \int_0^{1/T} \frac{e^{-\kappa/\tau}\sqrt{\tau \ln\ln(1/\tau)}}{\tau^2}|\sigma_{-1/\tau}|d\tau$ и $\hat{c}_2 > 0$ — некоторая константа. При помощи правила Лопиталя нетрудно показать, что

$$\left(h_T^{(1)} + h_T^{(2)}\right) / \sqrt{\left(e^{2\kappa T}\sigma^2_{-T}\ln T\right)} \to c, \quad T \to +\infty.$$

Тогда $\limsup_{T\to+\infty}\{|z_T|/\sqrt{(\sigma^2_{-T}\ln T)}\} < \infty$, и использование этой оценки для каждой из компонент вспомогательного процесса $\hat{X}_{-T} = \int_{-\infty}^{-T} e^{\kappa(T+s)}G_s dw_s$ приводит к тому, что существует константа $\hat{c} > 0$, при которой $\limsup_{T\to+\infty}\{\|\hat{X}_{-T}\|/h_T\} < \hat{c} < \infty$ п.н., если $h_T = \sqrt{\|G_{-T}\|^2 \ln T}$. Далее, для процесса разности $Z_t = X_t^* - \hat{X}_t$ с уравнением динамики $dZ_t = A_t^* Z_t dt + (\kappa I - A_t^*)\hat{X}_t dt$ и решением $Z_t = \int_{-\infty}^t \Phi(t,s)(\kappa I - A_s^*)\hat{X}_s\, ds$ стандартным образом показывается, см., например, [16], что экспоненциальная устойчивость A_t^* и $\dot{h}_t/h_t \to 0$ гарантируют ограниченность отношения $\|Z_t\|/h_t$ при $t \to -\infty$, откуда следует, что и $\limsup_{t\to-\infty}\{\|X_t^*\|/h_t\} < \bar{c} < \infty$ для $h_t = \sqrt{\|G_t\|^2 \ln|t|}$. Лемма 2 доказана.

Доказательство леммы 3. В условиях п. 2 предположения $\mathcal{G}$ использование результата леммы 2 в совокупности с требованием $d\ln\|G_t\|/dt \times \ln|t| \to 0$, $t \to \mathcal{T}$, приводит к тому, что $\lim_{t\to\mathcal{T}}\left\{\|X_t^*\|^2 / \left|\int_0^t \|G_s\|^2\, ds\right|\right\} \leqslant c\lim_{t\to\mathcal{T}}\left\{\|G_t\|^2\ln|t| / \left|\int_0^t \|G_s\|^2\, ds\right|\right\} = 0$ с вероятностью единица. Если матрица диффузии G_t ограничена, то при $\mathcal{T} = +\infty$ вновь используется представление $X_T^* = \Phi(T,0)X_0^* + \tilde{X}_T^*$, где $\tilde{X}_T^* = \int_0^T \Phi(T,s)G_s dw_s$, $T \geqslant 0$, и известный результат [13, теорема 1], согласно которому $\|\tilde{X}_T^*\|^2 / \int_0^T \|G_s\|^2 ds \to 0$ п.н., $T \to +\infty$. Тогда, принимая во внимание наличие убывающей экспоненциальной оценки для $\|\Phi(T,0)\|$, получаем соотношение $\|X_T^*\|^2 / \int_{-T}^T \|G_s\|^2 ds \to 0$, $T \to +\infty$. Для $\mathcal{T} = -\infty$ представление $\|X_{-T}^*\|^2 = \|X_0^*\|^2 - \int_{-T}^0 (X_t^*)'(A_t + A_t')X_t^* dt - \int_{-T}^0 (X_t^*)' G_t dw_t - \int_{-T}^0 (dw_t)' G_t' X_t^* - \int_{-T}^0 \|G_t\|^2\, dt$ позволяет применить при анализе слагаемых результаты [13, лемма 1, лемма 2] с подынтегральной заменой времени $\tau = -t$, и тогда также $\|X_{-T}^*\|^2 / \left|\int_{-T}^0 \|G_s\|^2\, ds\right| \to 0$, $T \to +\infty$. Лемма 3 доказана.

Доказательство теоремы 2. Для оценки (П.2) используются полученные неравенства (П.4) и (П.5). Замена T на t в (П.4), $(-T)$ на t — в (П.5) и последующее интегрирование этих соотношений на $[0,T]$ и $[-T,0]$ приводят к

$$\int_0^T \|x_t\|^2\, dt \leqslant \bar{c}_1\|\bar{x}\|^2 + \bar{c}_1 \int_0^T (x_t' Q_t x_t + u_t' R_t u_t)dt \tag{П.7}$$

и соответственно к

$$\int\limits_{-T}^{0} \|x_t\|^2\,dt \leqslant \tilde{c}_1\|\bar{x}\|^2 + \tilde{c}_1 \int\limits_{-T}^{0} (x_t' Q_t x_t + u_t' R_t u_t)dt \tag{П.8}$$

при некоторых константах $\bar{c}_1, \tilde{c}_1 > 0$. Тогда (П.2) можно оценить как

$$J_{2T}(U^*) \leqslant J_{2T}(U) + c_0\|\bar{x}\|^2 + c_1\|X_T^*\|^2 + c_2\|X_{-T}^*\|^2 -$$
$$- c_3 \int\limits_{-T}^{T} \|x_t\|^2\,dt - 2\int\limits_{-T}^{T} x_t'\Pi_t G_t dw_t,$$

где $c_0, c_1, c_2, c_3 > 0$ — некоторые константы, а затем записать, что

$$J_{2T}(U^*) \leqslant J_{2T}(U) + \mathcal{R}_T^{(0)} + \mathcal{R}_T^{(+)} + \mathcal{R}_T^{(-)}, \tag{П.9}$$

где процессы $\mathcal{R}_T^{(0)} = c_0\|\bar{x}\|^2 + c_1\|X_T^*\|^2 + c_2\|X_{-T}^*\|^2$, $\mathcal{R}_T^{(+)} = -c_3\int_0^T \|x_t\|^2\,dt -$ $-2\int_0^T x_t'\Pi_t G_t dw_t$, $\mathcal{R}_T^{(-)} = -\mathcal{R}_{-T}^{(+)}$. Так как выполнены предположения $\mathcal{G}$, $\mathcal{G}1$ и $\int_{-T}^{T} \|G_t\|^2\,dt \to \infty$, $T \to +\infty$, то с учетом леммы 3 имеем

$$\lim_{T\to+\infty} \left\{ \mathcal{R}_T^{(0)} \Big/ \int\limits_{-T}^{T} \|G_t\|^2\,dt \right\} = 0 \quad \text{п.н.}$$

Далее рассматривается поведение процессов $\mathcal{R}_T^{(+)}$ и $\mathcal{R}_T^{(-)}$ при $T \to +\infty$. Для ограниченной G_t, $t \geqslant 0$, известно, см., например, [12], что $\limsup\limits_{t\to+\infty}\{\mathcal{R}_T^{(+)}/g_T\} \leqslant 0$ п.н. для любой функции $g_T > 0$ и $g_T \to \infty$, $t \to +\infty$. По условию в качестве нормировки g_T можно взять $g_T = \int_{-T}^{T} \|G_t\|^2\,dt$. Если имеет место п. 2 предположения $\mathcal{G}$ и предположение $\mathcal{G}1$, то после использования закона повторного логарифма для стохастических интегралов, см., например, [19], $\mathcal{R}_T^{(+)}$ оценивается в виде $|\mathcal{R}_T^{(+)}| \leqslant L_T$, $T \to +\infty$, где

$$L_T = \hat{c}_1\|G_T\|^2 \sqrt{\int\limits_0^T \|x_t\|^2\,dt \ln\ln\left(\int\limits_0^T \|x_t\|^2\,dt\right)} -$$
$$- \hat{c}_2 \int\limits_0^T \|x_t\|^2 dt + \hat{c}_3\|G_T\|^2 \ln\ln\|G_T\|,$$

а $\hat{c}_1, \hat{c}_2, \hat{c}_3$ — некоторое константы. Применяя аналогичные рассуждения как и при доказательстве в [14, лемма 3], можно определить, что $L_T \leqslant c\|G_T\|^2 \ln\ln\|G_T\|$, и тогда $\limsup_{t\to+\infty}\{\mathcal{R}_T^{(+)}/g_T\} = 0$ п.н. при $g_T =$

$= \|G_T\|^2 \ln\ln\|G_T\|$. Из этого результата и предположения $\mathcal{G}1$ следует, что $\lim_{T\to+\infty}\left\{g_T/\int_{-T}^{T}\|G_t\|^2\,dt\right\} = 0$ п.н. Также необходимо отметить, что результаты по выбору нормировок g_T для процесса $\mathcal{R}_T^{(-)}$ получаются на основе соответствующих соотношений, приведенных выше для $\mathcal{R}_T^{(+)}$, если произвести замену времени $\tau = -t$ в подынтегральных выражениях. Тогда с учетом этих замечаний из (П.9) в пределе при $T \to +\infty$ приходим к неравенству

$$\limsup_{T\to+\infty}\frac{J_{2T}(U^*)}{\int\limits_{-T}^{T}\|G_t\|^2\,dt} \leqslant \limsup_{T\to+\infty}\frac{J_{2T}(U)}{\int\limits_{-T}^{T}\|G_t\|^2\,dt} \qquad \text{с вероятностью единица.}$$

Далее, переходя к (П.6), принимая во внимание (9) и результат леммы 3, очевидно, что для доказательства (11) необходимо исследовать поведение

$$I_T = \int_{-T}^{T}(X_t^*)'\Pi_t G_t\,dw_t = I_T^{(+)} + I_T^{(-)},$$

где $I_T^{(+)} = \int_0^T (X_t^*)'\Pi_t G_t\,dw_t$, $I_T^{(-)} = -I_{-T}^{(+)}$. Точнее, требуется проанализировать $I_T^{(+)}/\Gamma_T$, с $\Gamma_T = \int_0^T \|G_t\|^2\,dt$, при этом случай $I_T^{(-)}/|\Gamma_{-T}|$ рассматривается аналогичным образом путем замены времени. Для ограниченной G_t, $t \geqslant 0$, как было показано в [13], отношение $I_T^{(+)}/\Gamma_T \to 0$ п.н. при $T \to +\infty$. Для $\|G_t\| \to \infty$, $t \to +\infty$, и соотношений в предположении $\mathcal{G}1$ используется закон повторного логарифма для стохастических интегралов, см. [19], когда $\limsup\limits_{T\to+\infty}\left\{|I_T^{(+)}|/\sqrt{\langle I_T^{(+)}\rangle \ln\ln\langle I_T^{(+)}\rangle}\right\} < \infty$ п.н., где $\langle I_T^{(+)}\rangle = \int_0^T \|X_t^*\|^2\|G_t\|^2\|\Pi_t\|^2 dt$. Применение леммы 2 в совокупности с монотонностью $\|G_t\|$ дает оценки $\langle I_T^{(+)}\rangle \leqslant c\|G_T\|^2\int_0^T\|G_t\|^2 dt \ln T$ и $\ln\ln\langle I_T^{(+)}\rangle \leqslant c(\ln\ln T + \ln\ln\|G_T\|)$. Тогда

$$\langle I_T^{(+)}\rangle \ln\ln\langle I_T^{(+)}\rangle \,/\, \Gamma_T^2 \leqslant c\|G_T\|^2\,(\ln\ln T + \ln\ln\|G_T\|)\ln T\,/\,\Gamma_T \to 0, \quad T\to+\infty$$

(здесь равенство нулю получено вследствие предположения $\mathcal{G}1$), поэтому $I_T^{(+)}/\Gamma_T \to 0$ с вероятностью единица. С учетом изложенного

$$I_T/\int_{-T}^{T}\|G_t\|^2\,dt \to 0 \quad \text{п.н.}, \quad T\to+\infty,$$

и имеет место (11). Теорема 2 доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Квакернаак Х., Сиван Р.* Линейные оптимальные системы управления. М.: Наука, 1977.

2. *Mueller M., Cantoni M.* Normalized Coprime Representations for Time-Varying Linear Systems // Proc. 49th IEEE Conf. on Decision and Control. N.Y., 2010. P. 7718–7723.

3. *Tudor C.* Quadratic Control for Linear Stochastic Equations with Pathwise Cost // Stochastic Systems and Optimization. Proc. 6th IFIP WG 7.1 Working Conf. Warsaw, Poland, September 12–16, 1988. Berlin: Springer, 1989. P. 360–369.

4. *Da Prato G., Ichikawa A.* Quadratic Control for Linear Time-Varying Systems // SIAM J. Control Optim. 1990. V. 28. No. 2. P. 359–381.

5. *Makila P.M.* Convoluted Double Trouble // IEEE Contr. Syst. Mag. 2002. V. 22. No. 4. P. 26–31.

6. *Nourdin I.* Selected aspects of fractional Brownian motion. Milan: Springer, 2012.

7. *Altman E., Basar T., Hovakimyan N.* Worst-Case Rate-Based Flow Control with an ARMA Model of the Available Bandwidth // Advances in Dynamic Games and Applications. Boston: Birkhauser, 2000. P. 3–29.

8. *Sun T., Nielsen S.R.K.* Stochastic Optimal Control of a Heave Point Wave Energy Converter Based on a Modified LQG Approach // Ocean Eng. 2018. V. 154. P. 357–366.

9. *Grimble M.J., Johnson M.A.* Optimal control and stochastic estimation: theory and applications. V. 2. N.Y.: John Wiley & Sons, 1986.

10. *Smith P.L., McKenzie C.R.L.* Diffusive Information Accumulation by Minimal Recurrent Neural Models of Decision Making // Neural Comput. 2011. V. 23. No. 8. P. 2000–2031.

11. *Lim S.C., Muniandy S.V.* Self-Similar Gaussian Processes for Modeling Anomalous Diffusion // Phys. Rev. E. 2002. V. 66. No. 2. P. 021114.

12. *Белкина Т.А., Паламарчук Е.С.* О стохастической оптимальности для линейного регулятора с затухающими возмущениями // АиТ. 2013. № 4. С. 110–128.

    Belkina T.A., Palamarchuk E.S. On Stochastic Optimality for a Linear Controller with Attenuating Disturbances // Autom. Remote Control. 2013. V. 74. No. 4. P. 628–641.

13. *Паламарчук Е.С.* Асимптотическое поведение решения линейного стохастического дифференциального уравнения и оптимальность почти наверное для управляемого случайного процесса // Журн. вычислит. математики и мат. физики. 2014. Т. 54. № 1. С. 89–103.

    Palamarchuk E.S. Asymptotic Behavior of the Solution to a Linear Stochastic Differential Equation and Almost Sure Optimality for a Controlled Stochastic Process // Comput. Math. Math. Phys. 2014. V. 54. No. 1. P. 83–96.

14. *Паламарчук Е.С.* Оценка риска в линейных экономических системах при отрицательных временных предпочтениях // Экономика и матем. методы. 2013. Т. 49. № 3. С. 99–116.

15. *Al-Azzawi S., Liu J., Liu X.* Convergence Rate of Synchronization of Systems with Additive Noise // Discrete Contin. Dyn. Syst. Ser. B. 2017. V. 22. No. 2. P. 227–245.

16. *Паламарчук Е.С.* Об обобщении логарифмической верхней функции для решения линейного стохастического дифференциального уравнения с неэкспоненциально устойчивой матрицей // Дифференциальные уравнения. 2018. Т. 54. № 2. С. 195–195.

    Palamarchuk E.S. On the Generalization of Logarithmic Upper Function for Solution of a Linear Stochastic Differential Equation with a Nonexponentially Stable Matrix // Differ. Equat. 2018. V. 54. No. 2. P. 193–200.

17. *Prevot C., Rockner M.* A concise course on stochastic partial differential equations. Berlin: Springer, 2007.
18. *Булинский А.В., Ширяев А.Н.* Теория случайных процессов. М.: Физматлит, 2005.
19. *Wang J.* A Law of the Iterated Logarithm for Stochastic Integrals // Stoch. Proc. Appl. 1993. V. 47. No. 2. P. 215–228.

Статья представлена к публикации членом редколлегии Б.М. Миллером.

Поступила в редакцию 31.05.2019

После доработки 15.07.2019

Принята к публикации 18.07.2019

© 2020 г. М.Х. ПРИЛУЦКИЙ, д-р техн. наук (pril@iani.unn.ru)
(Нижегородский государственный университет)

ПРОГРАММНЫЕ УПРАВЛЕНИЯ ДВУХСТАДИЙНЫМИ СТОХАСТИЧЕСКИМИ ПРОИЗВОДСТВЕННЫМИ СИСТЕМАМИ[1]

Рассматривается проблема программного управления некоторым классом производственных систем, функционирующих в условиях неопределенности. Строится математическая модель, дается постановка оптимизационной задачи программного управления, предлагается алгоритм построения квазиоптимального решения поставленной задачи. Приводятся примеры прикладных задач, формализуемых в рамках построенной математической модели.

Ключевые слова: стохастические производственные системы, двухстадийные системы, программное управление, квазиоптимальное решение.

DOI: 10.31857/S0005231020010067

1. Введение

Рассматривается проблема программного управления некоторым классом производственных систем, функционирующих в условиях неопределенности. В системах для изготовления продуктов используются технологические режимы, в результате применения которых производятся полуфабрикаты. Особенностями рассматриваемых производственных систем является стохастический характер изготовления полуфабрикатов и детерминированный характер производства продуктов из полуфабрикатов. Исходными данными для рассматриваемых производственных систем являются конечные множества используемых технологических режимов, полуфабрикатов и изготавливаемых продуктов производства. Выбор технологического режима определяет вероятности получения того или иного полуфабриката. Известны затраты на использование каждого технологического режима. К началу планируемого периода задан план изготовления продуктов, причем для каждого продукта определен доход, который система получит от его производства. Каждому полуфабрикату соответствует множество продуктов, любой из которых (но только один) может быть изготовлен из этого полуфабриката. При этом изготовление запланированного продукта приносит системе определенный доход. Формально функционирование рассматриваемых производственных систем можно разбить на две стадии. Первая — от применения технологического режима до изготовления полуфабриката. Эта стадия носит стохастический характер. Вторая стадия — от изготовления полуфабриката до

[1] Работа выполнена при финансовой поддержке Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 гг.» в рамках соглашения № 14.578.21.0246 (уникальный идентификатор RFMEFI57817X0246).

производства продукта. Эта стадия носит детерминированный характер. Задачи, рассматриваемые для подобных систем, будем называть двухстадийными, принимая за первую стадию процесс изготовления полуфабрикатов, а за вторую — переработку полуфабрикатов в продукты производства.

Работа является продолжением статей [1, 2], в которых решаются задачи оптимального планирования и оптимального управления. Задача оптимального планирования позволяет находить планы производства, которые являются исходными данными для решения задач оптимального управления. Решение задачи оптимального управления определяет, какие технологические режимы в процессе функционирования системы нужно применять для наилучшего выполнения заданного плана. При этом задача оптимального управления является задачей с обратной связью — выбор очередного технологического режима зависит от того, какие продукты производства уже были изготовлены. Так как использование того или иного технологического режима требует его обеспечения необходимыми материальными и трудовыми ресурсами, то для эффективного функционирования рассматриваемых производственных систем необходимо заранее, до начала планируемого периода, знать, какие технологические режимы будут использоваться в процессе производства, чтобы обеспечить их необходимыми ресурсами. Такая задача в терминах теории управляемых систем [3, 4] носит название задачи поиска программных управлений.

В [1, 2] дается литературный обзор результатов в рассматриваемой области [5–11] и приводятся примеры двухстадийных производственных систем, для которых применимы полученные в работе результаты. Это задачи оптимального планирования и управления процессом переработки газового конденсата, процессом изготовления интегральных схем и процессом производства стали в мартеновских печах [12–17]. Для решения задач программного управления в работе строится математическая модель, в рамках которой исследуются рассматриваемые задачи.

2. Постановка задачи поиска оптимального программного управления двухстадийными стохастическими производственными системами

Как и в [1, 2], пусть I — множество технологических режимов; J — множество полуфабрикатов; K — множество выпускаемых продуктов; $T = \{0, 1, \ldots, T_0\}$ — множество тактов функционирования системы; $P = \|p_{ij}\|$ — матрица вероятностей, где p_{ij} — вероятность того, что, применив технологический режим i, будет получен полуфабрикат j, $\sum_{j \in J} p_{ij} = 1$, $i \in I$, $p_{ij} \geq 0$, $i \in I$, $j \in J$; $K(j)$ — множество продуктов, любой из которых может быть изготовлен из полуфабриката j, $K(j) \subseteq K$, $j \in J$; $\vec{\pi}$ — $|K|$-мерный вектор с целочисленными неотрицательными компонентами — план производства продуктов в планируемом периоде, где π_k — количество k-х продуктов, которые должны быть выпущены в планируемом периоде, $k \in K$; c_i — затраты производственной системы, связанные с использованием i-го технологического режима, $i \in I$; g_k — доход, который получит система от производства одного запланированного k-го продукта, $k \in K$.

Множество состояний системы разобьем на два подмножества — основные и вспомогательные состояния. Основные состояния образуют множество $S = \{\vec{s} \mid s_k \geq 0$ – целые, $s_k \leq T_0$, $k \in K\}$, где s_k определяет количество продуктов k, которые произведены в системе. Вспомогательными состояниями являются всевозможные пары $(\vec{s}, j)$, где $\vec{s} \in S$, j – параметр вспомогательного состояния, $j \in J$. Управлениями в основных состояниях являются элементы множества I – выбор технологического режима. Допустимыми управлениями во вспомогательном состоянии $(\vec{s}, j)$ являются элементы множества $K(j)$ — изготовление из полуфабриката продукта производства. Функционирование системы рассматривается на конечном числе тактов. При этом под одним тактом понимается переход системы из основного состояния в основное — от выбора технологического режима до выпуска продукта производства.

В основном состоянии $\vec{s}$, $\vec{s} \in S$ к системе применяется управление i, $i \in I$ и система с вероятностью p_{ij} переходит во вспомогательное состояние $(\vec{s}, j)$, при этом система несет потери $c_i \geq 0$ — затраты на использование технологического режима. Во вспомогательном состоянии $(\vec{s}, j)$ к системе применяется управление k, $k \in K(j)$, под воздействием которого система переходит в новое основное состояние, отличающееся от состояния $\vec{s}$ лишь в компоненте k, k-я компонента увеличивается на 1 (больше на 1 стало продукта k). При этом переходе система приобретает доход, определяемый функцией $q(\vec{s}, k, \vec{\pi}) = \begin{cases} g_k, & \text{если } s_k < \pi_k, \\ 0, & \text{в противном случае.} \end{cases}$ Здесь $g_k \geq 0$ — доход, который получит система от выпуска запланированного продукта k, $k \in K(j)$.

Множество всех управлений обозначим через $U = V \times W$. В общем случае управление $u \in U$ есть пара функций $v(\vec{s}, t)$ и $w(\vec{s}, j, t)$, определенных соответственно на множествах $S \times T$ и $S \times J \times T$ со значениями из множеств соответственно I и $K(j)$. Содержательно функция $v(\vec{s}, t)$ определяет, какие технологические режимы нужно применять в основных состояниях систем, а функция $w(\vec{s}, j, t)$ определяет, какие продукты нужно выпускать во вспомогательных состояниях. В общем случае введенные функции определяют управления с обратными связями, так как зависят от состояний, в которых система может оказаться в зависимости от тактов функционирования.

Для рассматриваемых задач поиска оптимальных программных управлений (управлений без обратных связей) функция $v(\vec{s}, t)$ определяется целочисленным набором $\vec{x}$, i-я компонента которого x_i задает количество технологических режимов i, которые будут применяться в планируемом периоде, $\vec{x} \in R^{|I|}$, а функция $w(\vec{s}, j, t)$ зависит только от параметра вспомогательного состояния j и рандомизированная, т.е. задается распределением вероятностей, определяемым матрицей $Y = \|y_{jk}\|$, $j \in J$, $k \in K$. Здесь y_{jk} — вероятность изготовления продукта k из полуфабриката j, $j \in J$, $k \in K$. Определенные таким образом вероятности не зависят от состояний и тактов функционирования, $\sum_{k \in K} y_{jk} = 1$, $y_{jk} \geq 0$, $j \in J$, $k \in K$.

Рассмотрим целочисленную случайную величину $\sigma_k = \sigma_k(\vec{x}, Y)$, принимающую значения из множества $\{0, 1, \ldots, T_0\}$ — сколько продуктов k будет выпущено, если в основных состояниях будут применяться технологические режимы из набора $\vec{x}$: x_1 раз первые технологические режимы, x_2 раз вторые

технологические режимы, ..., $x_{|I|}$ раз $|I|$-е технологические режимы, а во вспомогательных состояниях в случае получения полуфабриката j с вероятностью y_{jk} будет выпускаться продукт k, $j \in J$, $k \in K$.

Обозначим через $F(\vec{\pi}, T_0, \vec{x}, Y)$ математическое ожидание полного дохода, который получит система, если известен план производства продуктов $\vec{\pi}$, количество тактов функционирования системы T_0 и к системе будут применяться управления, задаваемые набором $\vec{x}$ (какие технологические режимы и в каких количествах будут использоваться в планируемом периоде) и матрицей Y, задающей распределение вероятностей изготовления тех или иных продуктов. Тогда $F(\vec{\pi}, T_0, \vec{x}, Y) = \sum_{k \in K} g_k E \min(\pi_k, \sigma_k) - \sum_{i \in I} c_i x_i$, где $E \min(\pi_k, \sigma_k)$ — математическое ожидание целочисленной случайной величины $\min(\pi_k, \sigma_k)$. Действительно, если в результате применения программной стратегии, определяемой набором $\vec{x}$ и матрицей Y, будет выпущено продукта k не больше запланированного, то каждый продукт k принесет доход g_k; если будет выпущено продукта k больше запланированного, то доход принесут только выпущенные первые π_k продукты. Отсюда задача нахождения оптимальной программной стратегии для двухстадийных стохастических производственных систем сводится к решению следующей (исходной) задачи математического программирования.

Задача 1. $F(\vec{\pi}, T_0, \vec{x}, Y) = \sum\limits_{k \in K} g_k E \min(\pi_k, \sigma_k) - \sum\limits_{i \in I} c_i x_i \to \max$

при условиях:

$$\sum_{i \in I} x_i = T_0,$$

$$\sum_{k \in K} y_{jk} = 1, \quad j \in J,$$

$$y_{jk} = 0, \text{ если } k \notin K(j), \quad j \in J,$$

$$x_i \geq 0 \text{ – целые}, \quad i \in I,$$

$$y_{jk} \geq 0, \quad j \in J, \quad k \in K.$$

3. Нахождение квазиоптимального программного управления

Из-за существенной сложности функционала $F(\vec{\pi}, T_0, \vec{x}, Y)$, связанной с операцией $E \min(\pi_k, \sigma_k)$, задачу 1 даже при небольших значениях параметров решить не удается. Принципиальные трудности возникают даже при подсчете значения функционала при заданных значениях неизвестных.

Пусть $(\vec{x}^0, Y^0)$ — оптимальное решение исходной задачи 1. Под квазиоптимальным решением задачи 1 будем понимать такое ее допустимое решение $(\vec{x}^*, Y^*)$, что

$$\lim_{T_0 \to \infty} \frac{F(\vec{\pi}, T_0, \vec{x}^0, Y^0) - F(\vec{\pi}, T_0, \vec{x}^*, Y^*)}{F(\vec{\pi}, T_0, \vec{x}^0, Y^0)} = 0.$$

Рассмотрим вспомогательную задачу математического программирования, в которой в отличие от исходной задачи знак математического ожидания внесен под операцию поиска минимума.

Задача 2. $H(\vec{\pi}, T_0, \vec{x}, Y) = \sum\limits_{k\in K} g_k \min(\pi_k, E\sigma_k) - \sum\limits_{i\in I} c_i x_i \to \max$
при сохраненных условиях задачи 1.

Нетрудно показать, что $E\sigma_k = \sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{jk}$, $k \in K$. Действительно, $\sum_{j\in J} p_{ij} y_{jk}$ – вероятность того, что, используя технологический режим i, изготовим продукт k, а x_i – количество i-х технологических режимов, которые будут применяться в планируемом периоде, $i \in I$, $k \in K$.

Преобразуем функционал задачи 2. Учитывая, что

$$\min(\pi_k, E\sigma_k) = -\max(-\pi_k, -E\sigma_k) = \pi_k - \pi_k - \max(-\pi_k, -E\sigma_k) = \\ = \pi_k - \max(0, \pi_k - E\sigma_k) = \pi_k - (\pi_k \otimes E\sigma_k),$$

где $\otimes$ — знак усеченной разности, получим

$$H(\vec{\pi}, T_0, \vec{x}, Y) = \sum_{k\in K} g_k \pi_k - \sum_{k\in K} g_k \left(\pi_k \otimes \sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{ij} \right) - \sum_{i\in I} c_i x_i.$$

Учитывая проделанные преобразования, задачу 2 можно рассматривать как *задачу 3* минимизации функционала

$$\Phi(\vec{\pi}, T_0, \vec{x}, Y) = \sum_{k\in K} g_k \left(\pi_k \otimes \sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{ij} \right) + \sum_{i\in I} c_i x_i$$

при сохраненных условиях задачи 1.

Рассмотрим следующую *задачу 4* математического программирования:

$$\Phi(\vec{\pi}, T_0, \vec{x}, Y) = \sum_{k\in K} g_k \left(\pi_k \otimes \sum_{j\in J} z_{jk} \right) + \sum_{i\in I} c_i x_i \to \min \text{ — при условиях:}$$

$$\sum_{i\in I} x_i = T_0,$$

$$\sum_{k\in K} z_{jk} - \sum_{i\in I} x_i p_{ij} = 0, \quad j \in J,$$

$$z_{jk} = 0, \text{ если } k \notin K(j), \quad k \in K, \quad j \in J,$$

$$x_i \geq 0 \text{ – целые, } i \in I,$$

$$z_{jk} \geq 0, \quad j \in J, \quad k \in K.$$

Покажем, как по решению задачи 4 получить решение задачи 3, а тем самым и задачи 2. Пусть x'_i, z'_{jk}, $i \in I$, $j \in J$, $k \in K$ — решение задачи 4. Тогда решением задачи 3 будут следующие наборы:

$$x_i^0 = x'_i, \quad i \in I,$$

$$y_{jk}^0 = \frac{z'_{jk}}{\sum_{i\in I} x'_i p_{ij}}, \quad \text{если } \sum_{i\in I} x'_i p_{ij} > 0, \tag{1}$$

$$y_{jk}^0 = 0, \quad \text{если } \sum_{i\in I} x'_i p_{ij} = 0, \ j \in J, \ k \in K.$$

Действительно, x_i^0, y_{jk}^0, $i \in I$, $j \in J$, $k \in K$ удовлетворяют ограничениям задачи 3. Пусть существуют такие x_i^*, y_{jk}^*, $i \in I$, $j \in J$, $k \in K$, которые удовлетворяют условиям задачи 3, и выполняется $\Phi(\pi, T_0, \vec{x}^*, Y^*) < \Phi(\pi, T_0, \vec{x}^0, Y^0)$, тогда x_i^*, $z_{jk}^* = y_{jk}^* \sum_{i \in I} x_i^* p_{ij}$, $i \in I$, $j \in J$, $k \in K$, удовлетворяют условиям задачи 4 и на них значение функционала меньше значения функционала задачи на x_i', z_{jk}', $i \in I$, $j \in J$, $k \in K$, что противоречит условиям оптимальности наборов x_i', z_{jk}', $i \in I$, $j \in J$, $k \in K$.

Избавимся в функционале задачи 4 от нелинейности, определяемой усеченной разностью. Рассмотрим *задачу 5* частично-целочисленного линейного программирования.

$$Q(\vec{x}, \vec{v}, \vec{w}, Z) = \sum_{k \in K} g_k v_k + \sum_{i \in I} c_i x_i \to \min \text{ — при условиях:}$$

$$\sum_{i \in I} x_i = T_0,$$

$$\sum_{k \in K} z_{jk} - \sum_{i \in I} x_i p_{ij} = 0, \quad j \in J,$$

$$\sum_{j \in J} z_{jk} + v_k - w_k = \pi_k, \quad k \in K,$$

$$z_{jk} = 0, \quad \text{если } k \notin K(j),\ k \in K,\ j \in J,$$

$$x_i \geq 0 \text{ — целые},\ i \in I,$$

$$z_{jk} \geq 0, \quad v_k \geq 0, \quad w_k \geq 0, \quad j \in J, \quad k \in K.$$

Покажем, что решение задачи 5 определяет решение задачи 4, а тем самым и решение задачи 2. Для этого достаточно показать, что $\pi_k \otimes \sum_{j \in J} z_{jk} = v_k$, если $\pi_k > \sum_{j \in J} z_{jk}$, и $\pi_k \otimes \sum_{j \in J} z_{jk} = 0$, если $\pi_k \leq \sum_{j \in J} z_{jk}$.

Пусть z_{jk}^0, v_k^0, w_k^0, $j \in J$, $k \in K$ оптимальное решение задачи 5. Тогда для любого k, $k \in K$, либо $v_k^0 = 0$, либо $w_k^0 = 0$. Пусть это не так и существует другое оптимальное решение задачи 5 z_{jk}^*, v_k^*, w_k^*, $j \in J$, $k \in K$, для которого найдется такое k, что $v_k^* > 0$ и $w_k^* > 0$. Тогда если $v_k^* > w_k^* > 0$, то при замене $v_k' = v_k^0 - w_k^0$ и $w_k' = 0$ ограничения задачи 5 будут выполнены, а значение функционала уменьшится, что противоречит оптимальности найденного решения задачи 5. Если $w_k^* \geq v_k^* > 0$, то $v_k' = 0$, $w_k' = w_k^0 - v_k^0$ удовлетворяют ограничениям задачи 5 и уменьшают значение функционала.

Таким образом, для решения вспомогательной задачи 2 достаточно решить частично-целочисленную задачу линейного программирования 5, а затем по соотношениям (1) найти оптимальное решение задачи 2.

4. Обоснование квазиоптимальности программного управления

Теорема. Пусть $(\vec{x}^0, Y^0)$ *— оптимальное решение исходной задачи 1, а* $(\vec{x}^*, Y^*)$ *— оптимальное решение вспомогательной задачи 2.*

Тогда $$\lim_{T_0 \to \infty} \frac{F(\vec{\pi}, T_0, \vec{x}^0, Y^0) - F(\vec{\pi}, T_0, \vec{x}^*, Y^*)}{F(\vec{\pi}, T_0, \vec{x}^0, Y^0)} = 0.$$

Из теоремы следует, что при замене оптимального программного управления $(\vec{x}^0, Y^0)$, получающегося при решении исходной задачи 1, на программное управление $(\vec{x}^*, Y^*)$, получающееся при решении вспомогательной задачи 2, математическое ожидание полного дохода уменьшится, однако с ростом T_0 эти потери по отношению к математическому ожиданию полного дохода при оптимальном программном управлении будут стремиться к нулю.

Доказательство теоремы основано на лемме.

Лемма. Для произвольной целочисленной случайной величины σ, $\sigma \in \in \{0, 1, \ldots, n\}$, *и произвольного целого числа* π

$$\min(\pi, E\sigma) - E\min(\pi, \sigma) = \frac{1}{2}(E\,|\pi - \sigma| - |E\sigma - \pi|).$$

Доказательство леммы и теоремы см. в Приложении.

5. Заключение

В статье рассматриваются задачи поиска оптимального программного управления двухстадийными стохастическими производственными системами, первая, стохастическая, стадия которых заключается в изготовлении полуфабриката, а вторая, детерминированная, — в изготовлении из полуфабриката готовой продукции. Решение этой задачи позволяет до начала планируемого периода определить, какие технологические режимы будут использованы в процессе функционирования производственной системы, и тем самым обеспечить эти режимы необходимыми ресурсами. Предлагается процедура решения вспомогательной задачи частично-целочисленного линейного программирования, позволяющая находить квазиоптимальное решение исходной задачи.

Полученные результаты положены в основу программных систем, внедренных в постоянную эксплуатацию при планировании и оперативном управлении процессом производства изделий микроэлектроники ФГУП “ФНПЦ НИИИС им. Ю.Е. Седакова” [12–14], и апробированы при решении задач оптимального планирования и управления для Сургутского завода стабилизации конденсата ООО Сургутгазпром [15–17].

ПРИЛОЖЕНИЕ

Доказательство леммы.

$$\min(\pi, E\sigma) - E\min(\pi, \sigma) = \min(\pi, E\sigma) - \sum_{i=0}^{n} \min(\pi, i)p(\sigma = i) =$$

$$= \min(\pi, E\sigma) - \sum_{i=0}^{\pi} ip(\sigma = i) - \pi \sum_{i=\pi+1}^{n} p(\sigma = i) =$$

$$= \min\left(\pi, \sum_{i=0}^{n} ip(\sigma = i)\right) - \sum_{i=0}^{\pi} ip(\sigma = i) - \pi \sum_{i=\pi+1}^{n} p(\sigma = i) =$$

$$= \min\left(\pi - \sum_{i=0}^{\pi} ip(\sigma = i),\ \sum_{i=\pi+1}^{\pi} ip(\sigma = i)\right) - \pi \sum_{i=\pi+1}^{n} p(\sigma = i) =$$

$$= \min\left(\pi - \pi \sum_{i=\pi+1}^{n} p(\sigma = i) - \sum_{i=0}^{n} ip(\sigma = i),\ \sum_{i=\pi+1}^{\pi} (i - \pi)p(\sigma = i)\right) =$$

$$= \min\left(\sum_{i=0}^{\pi} (\pi - i)p(\sigma = i),\ \sum_{i=\pi+1}^{n} (i - \pi)p(\sigma = i)\right).$$

Обозначим:

$$\alpha(\pi) = \sum_{i=0}^{\pi} (\pi - i)p(\sigma = i) \quad \text{и} \quad \beta(\pi) = \sum_{i=\pi+1}^{n} (i - \pi)p(\sigma = i).$$

Тогда

$$\alpha(\pi) + \beta(\pi) = \sum_{i=0}^{\pi} |\pi - i|\, p(\sigma = i) + \sum_{i=\pi+1}^{n} |\pi - i|\, p(\sigma = i) =$$

$$= \sum_{i=0}^{n} |\pi - i|\, p(\sigma = i) = E\,|\pi - \sigma|\,.$$

С другой стороны

$$\alpha(\pi) = \sum_{i=0}^{\pi} (\pi - i)p(\sigma = i) = \pi - \sum_{i=\pi+1}^{n} (\pi - i)p(\sigma = i) =$$

$$= \pi - E\sigma + E\sigma - \sum_{i=\pi+1}^{n} (\pi - i)p(\sigma = i) =$$

$$= \pi - E\sigma + \sum_{i=0}^{n} ip(\sigma = i) - \sum_{i=\pi+1}^{n} (\pi - i)p(\sigma = i) =$$

$$= \pi - E\sigma + \sum_{i=0}^{n} ip(\sigma = i) - \pi \sum_{i=\pi+1}^{n} p(\sigma = i) - \sum_{i=\pi+1}^{n} ip(\sigma = i) =$$

$$= \pi - E\sigma + \sum_{i=\pi+1}^{n} (i - \pi)p(\sigma = i) = \pi - E\sigma + \beta(\pi).$$

Таким образом, $\alpha(\pi) + \beta(\pi) = E\,|\pi - \sigma|$ и $\alpha(\pi) - \beta(\pi) = \pi - E\sigma$. Отсюда

$$\alpha(\pi) = \frac{1}{2}(E\,|\pi - \sigma| - (E\sigma - \pi)), \quad \beta(\pi) = \frac{1}{2}(E\,|\pi - \sigma| - (\pi - E\sigma)).$$

Тогда

$$\min(\pi, E\sigma) - E\min(\pi, \sigma) = \min(\alpha(\pi), \beta(\pi)) =$$

$$= \frac{1}{2}\min(E\,|\pi-\sigma| - (E\sigma-\pi), E\,|\pi-\sigma| - (\pi-E\sigma)) =$$
$$= \frac{1}{2}(E\,|\pi-\sigma| - \min(E\sigma-\pi, \pi-E\sigma)) =$$
$$= \frac{1}{2}(E\,|\pi-\sigma| - |E\sigma-\pi|).$$

Лемма доказана.

Доказательство теоремы. Из леммы следует, что

$$H(\vec{\pi}, T_0, \vec{x}, Y) - F(\vec{\pi}, T_0, \vec{x}, Y) =$$
$$= \sum_{k\in K} g_k \min(\pi_k, E\sigma_k) - \sum_{i\in I} c_i x_i - \sum_{k\in K} g_k E\min(\pi_k, \sigma_k) + \sum_{i\in I} c_i x_i =$$
$$= \sum_{k\in K} g_k(\min(\pi_k, E\sigma_k) - E(\min(\pi_k, \sigma_k)) =$$
$$= \frac{1}{2}\left(\sum_{k\in K} g_k(E\,|\pi_k-\sigma_k| - |E\sigma_k-\pi_k|)\right).$$

Здесь последнее равенство основано на лемме.

Полученное равенство дает оценку для значения функционала $F(\vec{\pi}, T_0, \vec{x}, Y)$, если известно значение функционала $H(\vec{\pi}, T_0, \vec{x}, Y)$. При этом нужно учитывать, что $F(\vec{\pi}, T_0, \vec{x}, Y) \le H(\vec{\pi}, T_0, \vec{x}, Y)$ и $E\,|\pi-\sigma| \le \sqrt{D(\pi,\sigma)}$, где $D(\pi,\sigma)$ — средний квадрат отклонения (дисперсия) целочисленной случайной величины σ относительно данного целого числа π.

Определим дисперсию и математическое ожидание для данного случая.

$$D(\pi_k, \sigma_k) = \sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{jk}(1 - \sum p_{ij} y_{jk}) + \left(\sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{jk} - \pi_k\right)^2;$$
$$E\sigma_k = \sum_{i\in I} x_i \sum_{j\in J} p_{ij} y_{jk}.$$

Тогда

$$H(\vec{\pi}, T_0, \vec{x}, Y) - \delta(\vec{\pi}, T_0, \vec{x}, Y) \le F(\vec{\pi}, T_0, \vec{x}, Y) \le H(\vec{\pi}, T_0, \vec{x}, Y),$$

где

$$\delta(\vec{\pi}, T_0, \vec{x}, Y) = \frac{1}{2}\sum_{k\in K} g_k\left(\sqrt{D(\pi_k, \sigma_{k)}} - |E\sigma_k - \pi_k|\right).$$

Обозначим:

$$\Delta(F^0, F^*) = \frac{F(\vec{\pi}, T_0, \vec{x}^0, Y^0) - F(\vec{\pi}, T_0, \vec{x}^*, Y^*)}{F(\vec{\pi}, T_0, \vec{x}^0, Y^0)}.$$

Здесь $(\vec{x}^0, Y^0)$ — оптимальное решение исходной задачи 1, а $(\vec{x}^*, Y^*)$ — оптимальное решение задачи 2. Из того, что $F(\vec{\pi}, T_0, \vec{x}^0, Y^0) \le H(\vec{\pi}, T_0, \vec{x}^0, Y^0) \le$ $\le H(\vec{\pi}, T_0, \vec{x}^*, Y^*)$ и $-F(\vec{\pi}, T_0, \vec{x}^*, Y^*) \le -H(\vec{\pi}, T_0, \vec{x}^*, Y^*) + \delta(\vec{\pi}, T_0, \vec{x}^*, Y^*)$, получаем $F(\vec{\pi}, T_0, \vec{x}^0, Y^0) - F(\vec{\pi}, T_0, \vec{x}^*, Y^*) \le \delta(\vec{\pi}, T_0, \vec{x}^*, Y^*)$. Учитывая, что $F(\vec{\pi}, T_0, \vec{x}^0, Y^0) \ge F(\vec{\pi}, T_0, \vec{x}^*, Y^*) \ge H(\vec{\pi}, T_0, \vec{x}^*, Y^*) - \delta(\vec{\pi}, T_0, \vec{x}^*, Y^*)$, получим

$$\Delta(F^0, F^*) = \frac{F(\vec{\pi}, T_0, \vec{x}^0, Y^0) - F(\vec{\pi}, T_0, \vec{x}^*, Y^*)}{F(\vec{\pi}, T_0, \vec{x}^0, Y^0)} \le$$

$$\le \frac{\delta(\vec{\pi}, T_0, \vec{x}^*, Y^*)}{H(\vec{\pi}, T_0, \vec{x}^*, Y^{*0}) - \delta(\vec{\pi}, T_0, \vec{x}^*, Y^*)}.$$

Оценим числитель и знаменатель полученного выражения.

$$\delta(\vec{\pi}, T_0, \vec{x}^*, Y^*) =$$

$$= \frac{1}{2}\sum_{k\in K} g_k \left(\sqrt{\sum_{i\in I} x_i^* \sum_{j\in J} p_{ij} y_{jk}^* \left(1 - \sum_{j\in J} p_{ij} y_{jk}^*\right) + \left(\sum_{i\in I} x_i^* \sum_{j\in J} p_{ij} y_{jk}^* - \pi_k\right)^2} - \right.$$

$$\left. - \left| \sum_{i\in I} x_i^* \sum_{j\in J} p_{ij} y_{jk}^* - \pi_k \right| \right) \le \frac{1}{2}\sum_{i\in I} g_i \sqrt{\sum_{i\in I} x_i^* \sum_{j\in J} p_{ij} y_{jk}^* \left(1 - \sum_{j\in J} p_{ij} y_{jk}^*\right)} \le$$

$$\le \frac{1}{4}\sum_{i\in I} g_i \sqrt{\sum_{i\in I} x_i} \le \frac{|K|}{4} g\sqrt{T_0}.$$

При доказательстве этого неравенства использовалось:

$\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, если $a \ge 0$ и $b \ge 0$;

$\sum_{j\in J} p_{ij} y_{jk} \left(1 - \sum_{j\in J} p_{ij} y_{jk}\right) \le \frac{1}{4}$;

$g = \max_{i\in I} g_i$;

$\sum_{i\in I} x_i = T_0$;

$|K|$ – мощность множества K.

Таким образом, показано, что $\delta(\vec{\pi}, T_0, \vec{x}^*, Y^*) \le \frac{|K|}{4} g\sqrt{T_0} = \alpha(T_0)$. Это неравенство справедливо для любых $\vec{x}, Y, \vec{\pi}, T_0$. Можно показать, что

$$H(\vec{\pi}, T_0, \vec{x}^*, Y^*) \ge \beta(T_0) = g' T_0 \left(\min_{k\in K} \max_{i\in I} \sum_{j|k\in K(j)} p_{ij} \right) \min\left(1, \frac{1}{|I|}\right).$$

Здесь $g' = \min\limits_{i \in I} g_i$. Это неравенство справедливо для любых $\vec{x}, Y, \vec{\pi}, T_0$ таких, что $\sum_{k \in K} \pi_k \geq T_0$. Очевидно, что начиная с некоторого T_0 $\beta(T_0) > \alpha(T_0)$. Тогда

$$\Delta(F^0, F^*) \leq \frac{\alpha(T_0)}{\beta(T_0) - \alpha(T_0)} =$$

$$= \frac{\frac{|K|}{4} g \sqrt{T_0}}{g' T_0 \left(\min\limits_{k \in K} \max\limits_{i \in I} \sum_{j | k \in K(j)} p_{ij} \right) \min\left(1, \frac{1}{|I|}\right) - \frac{|K|}{4} g \sqrt{T_0}},$$

отсюда

$$\lim_{T_0 \to \infty} \Delta(F^0, F^*) = 0.$$

Теорема доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Прилуцкий М.Х.* Оптимальное планирование двухстадийных стохастических производственных систем // АиТ. 2014. № 8. С. 37–47.
   Prilutskii M.Kh. Optimal Planning for Two-Stage Stochastic Industrial Systems // Autom. Remote Control. 2014. V. 75. No. 8. P. 1384–1392.
2. *Прилуцкий М.Х.* Оптимальное управление двухстадийными стохастическими производственными системами // АиТ. 2018. № 5. С. 69–82.
   Prilutskii M.Kh. Optimal Management of Two-Stage Stochastic Production Systems // Autom. Remote Control. 2018. V. 79. No. 5. P. 830–840.
3. *Моисеев Н.Н.* Численные методы в теории оптимальных систем. М.: Наука, 1971.
4. *Красовский Н.Н.* Теория управления движением. М.: Наука, 1968.
5. *Kempf K., Keskinocak P., Uzsoy R.* (ed.) Planning Production and Inventories in the Extended Enterprise // Int. Ser. Oper. Res. & Management Sci. V. 152. N.Y.: Springer, 2010.
6. *Pinedo M.L.* Planning and Scheduling in Manufacturing and Services. N.Y.: Springer-Verlag, 2005.
7. *Weglarz J.* (ed.) Project Scheduling: Recent Models, Algorithms and Applications. N.Y.: Springer, 1999.
8. *Sprecher A.* Resource-constrained project scheduling: Exact methods for the multi-mode Case / Ser. Lect. Notes Econom. Math. Syst. V. 409. Berlin, 1994.
9. *Hapke M., Jaszkiewicz A., Słowiński R.* Fuzzy Multi-Mode Resource-Constrained Project Scheduling with multiple Objectives // Węglarz J. (ed.) Project Scheduling. Int. Ser. Oper. Res. & Management Sci. V. 14. Boston: Springer, 1999.
10. *Armbruster D., Fonteijn J., Wienke M.* Modeling Production Planning and Transient Clearing Functions // Logist. Res. 2012. V. 5. No. 3–4. P. 133–139.
11. *Bügler M., Borrmann A.* Using Swap-Based Search Trees to obtain Solutions for Resource Constrained Project Scheduling Problems // Proc. 85 Annual Meeting Int. Associat. Appl. Math. Mechan. (GAMM). 2014. V. 14. No. 1. P. 809–810.
12. *Прилуцкий М.Х., Власов В.С.* Оптимизационные задачи распределения ресурсов при планировании производства микроэлектронных изделий // Системы управления и информационные технологии. 2009. № 1. С. 38–43.

13. *Прилуцкий М.Х.* Многокритериальные многоиндексные задачи объёмно-календарного планирования // Изв. Акад. наук. Теория и системы управления. 2007. № 1. С. 78–82.

14. *Прилуцкий М.Х., Власов В.С., Кривошеев О.В.* Задачи оптимального планирования как задачи распределения ресурсов в сетевых канонических структурах // Информационные технологии. 2017. Т. 23. № 9. С. 650–657.

15. *Прилуцкий М.Х., Костюков В.Е.* Оптимизационные задачи добычи газа и переработки газового конденсата // Автоматизация в промышленности. 2008. № 6. С. 20–23.

16. *Прилуцкий М.Х., Костюков В.Е.* Потоковые модели для предприятий с непрерывным циклом изготовления продукции // Информационные технологии. 2007. № 10. С. 47–52.

17. *Афраймович Л.Г., Прилуцкий М.Х.* Многоиндексные задачи оптимального планирования производства // АиТ. 2010. № 10. С. 148–155.

Afraimovich L.G., Prilutskii M.Kh. Multiindex Optimal Production Planning Problems // Autom. Remote Control. 2010. V. 71. No. 10. P. 2145–2151.

Статья представлена к публикации членом редколлегии Е.Я. Рубиновичем.

Поступила в редакцию 18.06.2018

После доработки 24.06.2019

Принята к публикации 18.07.2019

Управление в технических системах

© 2020 г. В.Н. БУКОВ, д-р техн. наук (v_bukov@mail.ru)
(ОАО "Бортовые аэронавигационные системы", Москва),
Е.В. ОЗЕРОВ, канд. техн. наук (ozerovevg@yandex.ru)
(ВУНЦ ВВС "Военно-воздушная академия
им. проф. Н.Е. Жуковского и Ю.А. Гагарина", Воронеж),
В.А. ШУРМАН (shurman@niiao.ru)
(Филиал АО "Раменское приборостроительное КБ", Жуковский)

ПАРНЫЙ МОНИТОРИНГ ИЗБЫТОЧНЫХ ТЕХНИЧЕСКИХ СИСТЕМ

Ставится и решается детерминистская задача одновременного контроля технического состояния в реальном времени (мониторинга) как основной (функциональной), так и контролирующей (мониторинговой) частей системы. Предлагается подход, основанный на логическом анализе результатов встроенного контроля пары сопоставимых технических устройств, в общем случае разнородных по изготовлению и инфраструктурной поддержке. Получены структура и правила формирования индикаторной матрицы, позволяющей разделить технические устройства на полностью или частично исправные и неисправные. Приводятся выражения для вероятностей обнаружения неисправностей обеих частей системы и для вероятностей совершения ошибок первого и второго рода. Показаны методические примеры.

Ключевые слова: комплекс оборудования, функциональный модуль, мониторинговый модуль, функциональный узел, булев и не булев формализмы, индикаторная матрица исправности, индикаторное правило логического парного мониторинга.

DOI: 10.31857/S0005231020010079

1. Введение

Возросшие возможности информационного и математического обеспечения процессов управления сложными динамическими системами позволяют принципиально по-новому подойти к удовлетворению постоянно ужесточаемых требований к их отказоустойчивости, в том числе на основе управляемой избыточности [1], которая подразумевает преднамеренную избыточность системы, поддерживаемую специализированными средствами управления и придающую системе такие свойства, как

отказоустойчивость,

повышенная общая производительность/мощность,

существенно увеличенный межрегламентный период,

оперативное изменение различных эксплуатационных характеристик (точность, расходование энергии/ресурса определенных компонентов и др.).

Сказанное в полной мере относится к подвижным объектам [2–4] и технологическим процессам [5] с избыточными комплексами оборудования (КО).

По крайней мере одно из направлений управления избыточностью перспективных КО [1, 6] предполагает выполнение в реальном времени процедуры мониторинга технического состояния [7, 8] компонентов комплекса с целью его реконфигурирования при неправильном функционировании. Термин "мониторинг" как транслитерация англоязычного термина [8, 9], по мнению авторов, точнее передает специфику проверки технического состояния системы в реальном времени, чем широко распространенные термины "контроль" и "диагностирование", по определению [10] относящиеся практически к любым ситуациям.

Среди различных постановок задачи мониторинга можно выделить задачу дихотомического мониторинга, при котором результатом каждый раз является одно из двух суждений: "исправен" или "неисправен"; такой результат используется, например, при формировании индексов готовности компонентов для выбора подходящей конфигурации КО [1, 6].

Разработанный и предлагаемый детерминистский подход относится к мониторингу разнообразных технических устройств с аппаратной избыточностью независимо от их конструктивных особенностей и от характера и обстоятельств возникновения неисправностей. Подход применим при выполнении двух условий:

наличие (или возможность создания) встроенных средств контроля [9] (встроенных средств технического диагностирования [10]),

доступность для управления физических или виртуальных связей между частями технического устройства, несущими функциональную нагрузку, и встроенными в него средствами диагностирования.

2. Состояние проблемы

Диагностированию технического состояния систем (Fault Detection and Isolation – FDI) за последние 20 лет уделено большое внимание [8, 11–16]. Сложившиеся направления исследований могут быть разделены на две группы, связанные с использованием аппаратной или аналитической избыточности. Первая из них подразумевает избыточность конструктивных компонентов (элементов, узлов, подсистем), сопоставительный анализ функционирования которых позволяет при выполнении определенных условий вычислить наличие и место неисправности. Вторая группа направлений предполагает вместо аппаратной избыточности использование математических моделей объектов мониторинга, что создает предпосылки не только для достижения лучших массовых и энергетических характеристик в целом, но и для повышения результативности мониторинга за счет вскрытия внутренних причинно-следственных связей диагностируемых объектов.

Вместе с тем основными средствами диагностики (контроля) технического состояния различных видов оборудования в настоящее время являются встроенные средства технического диагностирования [10] (распространен также [9, 17, 18] термин "встроенные средства контроля" – ВСК), специально вводимые в состав устройств комплекса, и внешние автоматизированные системы контроля (АСК) [19]. Этими средствами обеспечивается контроль (надзор над) функциональных модулей (ФМ) оборудования, выполняющих

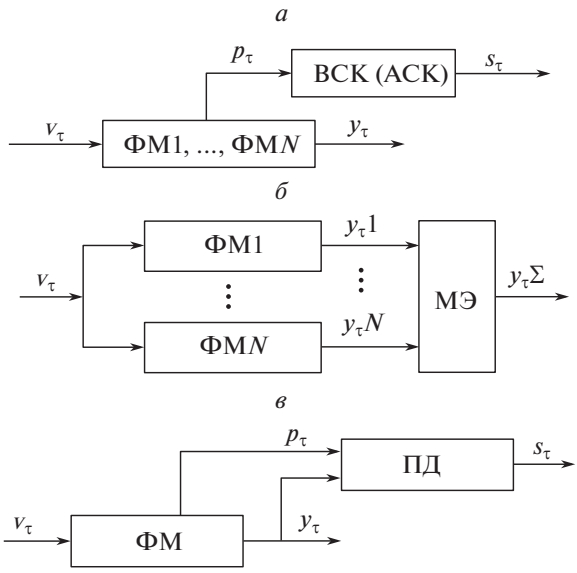


Рис. 1. Схемы контроля ФМ посредством ВСК или АСК: *а* – в обобщенном виде, *б* – мажоритарного контроля, *в* – контроля с использованием правил достоверности.

задачи его прямого предназначения, с целью определения их технического состояния (чаще в терминах состояний “исправен” или “неисправен”).

ВСК контролируют параметры ФМ в соответствии с критериями выполнения требуемых функций [20].

Рисунок 1,*а* иллюстрирует в обобщенном виде такое решение. На рисунке использованы обозначения: τ – текущее время (номер цикла мониторинга), v_τ – входные данные, y_τ – выходные данные, p_τ – контролируемые параметры, s_τ – оценка состояния ФМ.

В общем случае входные v_τ и выходные y_τ данные могут входить в число параметров, контролируемых с помощью ВСК (на схеме рис. 1,*а* и далее это не показано графически во избежание загромождения рисунков излишними деталями непринципиального характера).

Построение ВСК и АСК уровня системы (комплекса) связано с применением различных методов параметрического контроля, основанных обычно на следующих двух подходах.

1. *Мажоритарный контроль.* Неисправный ФМ определяется с помощью мажоритарного элемента (МЭ) путем обработки результатов функционирования нескольких подключенных ФМ. Схема показана на рис. 1,*б*.

Суждение о неисправности ФМ делается на основе значительного (наибольшего или превышающего пороговое значение) отклонения его выхода от выходов большинства других однотипных модулей.

Основные особенности метода включают:

а) предположение о неизменности технического состояния ФМ в пределах цикла τ;

б) предположение о том, что МЭ может быть только исправным;

в) применимо к числу ФМ, превышающему 2;

г) предположение о том, что с учетом правил голосования (равноправное, взвешенное, с дискриминациями и пр.) исправные ФМ внутри каждого цикла τ составляют большинство (доминируют над неисправными);

д) общий поток данных для всех ФМ.

2. *Использование правил достоверности* (ПД). В зависимости от конкретных условий и решений в качестве таких правил могут выступать: сравнение с эталонными моделями, фиксирование нарушений заданных временны́х и/или параметрических интервалов (контроль по допуску на параметр [15]), проверка логических и других соотношений, вычисление инвариантов разных порядков и пр. Такой способ иллюстрируется на рис. 1,*в*, где показана связь ПД с выходом ФМ y_τ, поскольку часто именно она является существенной, что далее иллюстрирует второй пример в разделе 9.

Особенности метода, основанного на использовании ПД:

а) в пределах цикла τ исправность ФМ не изменяется;

б) предполагается, что элемент ПД может быть только исправным, в том числе при наличии эталонной модели, которая может быть только исправной;

в) предполагается, что входные v_τ и выходные y_τ данные в достаточной степени информативны.

Предположение о непременной исправности так называемого заключительного звена (указанных МЭ и ПД) схемы контроля является серьезной проблемой систем диагностирования вообще и мониторинга текущего технического состояния в частности, поскольку оно целиком или частично выпадает из-под диагностирования.

Для решения этой проблемы практикуется, например, многократное мажорирование, при котором результаты мажоритарного контроля нижнего уровня подвергаются мажорированию более высокого уровня. Однако при этом всегда присутствует самый верхний уровень, результаты которого следует принимать “на веру”.

Кроме того, многоуровневому мажоритарному контролю [21] присущи следующие недостатки:

низкая эффективность мажоритарного сравнения сигналов при неоднородной избыточности вычислительных средств;

значительный объем вычислений, связанных с многоуровневым мажоритарным контролем в сочетании со статистической обработкой сигналов трактов;

сложность самого устройства, что вместе с отсутствием у него встроенного самоконтроля снижает соответствующий технический эффект.

Другой известный путь преодоления указанной проблемы заключается в организации самоконтроля самих схем диагностирования. В основном это относится к сложной микропроцессорной технике[1] и сопряжено, как правило, с реализацией тестового контроля, что для мониторинга в реальном времени неприменимо.

[1] URL: https://vikidalka.ru/4-79748.html

Подводя итог краткому обзору, можно отметить, что, в целом, проанализированные подходы обладают серьезными ограничениями, исключающими и в значительной степени затрудняющими или ставящими в зависимость от сильных[2] предположений построение систем мониторинга исправности компонентов КО в реальном времени (в рабочих режимах).

3. Формальные основы логического мониторинга

Сформулируем задачу мониторинга следующим образом. Пусть некоторый функциональный модуль (ФМ) на интервале времени τ решает какую-либо содержательную задачу. Одновременно за его функционированием "наблюдает" мониторинговый модуль (ММ), в решении содержательной задачи участия не принимающий. По выходному сигналу ММ формируется суждение об исправности или неисправности ФМ. При этом возможно неправильное функционирование как ФМ, так и ММ. Кроме того, возможное неправильное функционирование ФМ или ММ не влияет на работоспособность друг друга. Ставится задача получить оценку работоспособности (дихотомическую оценку "исправен" или "неисправен") ФМ и одновременно установить исправность или неисправность ММ.

Схематическое изображение соединения ФМ+ММ, в дальнейшем называемого функциональным узлом (ФУз), показано на рис. 2.

Внешне эта схема похожа на схему рис. 1,*а*, однако принципиальным отличием является допущение возможности неисправного состояния не только ФМ, но и ММ. Жирными стрелками на рис. 2 условно показаны в общем случае многомерные входные v_τ и выходные y_τ каналы данных (сигналы), а также контролируемые параметры p_τ. Все они могут иметь различную физическую природу. Тонкой стрелкой обозначена оценка s_τ технического состояния ФМ, формируемая на выходе ММ и представляющая собой бинарную переменную "исправен" или "неисправен".

Характер рассматриваемых неисправностей ФМ может быть любым как в смысле природы возникновения, так и по проявлению при непременном

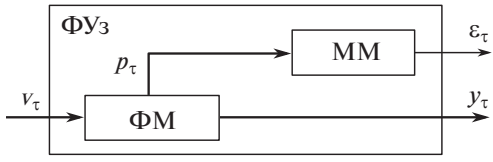


Рис. 2. Схема функционального узла ФМ+ММ.

условии, что эти неисправности ФМ обнаруживаются исправным ММ, представляющим собой соответствующий вариант ВСК.

В известных публикациях [14–17] приняты упрощенные обозначения (логические переменные состояния, булевы переменные): "1" – рассматриваемое

[2] К таковым относится предположение о непременной исправности ВСК или их частей.

устройство исправно, “0” – неисправно. При предположении возможной неисправности как ФМ, так и ММ логика зависимости состояния исправности ФУз от состояния исправности ФМ и/или ММ выражается булевой формулой логического умножения (конъюнкции):

$$
\begin{aligned}
\text{исправный ФМ и исправный ММ:} &\quad 1 \times 1 = 1,\\
\text{отказавший ФМ и исправный ММ:} &\quad 0 \times 1 = 0,\\
\text{исправный ФМ и отказавший ММ:} &\quad 1 \times 0 = 0,\\
\text{отказавший ФМ и отказавший ММ:} &\quad 0 \times 0 = 0.
\end{aligned}
\tag{1}
$$

Здесь исправному состоянию ФУз соответствуют только исправные состояния обоих его модулей.

Отказы ММ, по-видимому, можно в первом приближении подразделить на два вида: простые отказы “залипание на 1” и “залипание на 0” (аналоги “замыкания” и “разрыва” в электрических цепях) и сложный отказ “инверсия значения исправности ФМ”. При этом сложный отказ на каждом отдельном цикле мониторинга, по крайней мере в детерминистской задаче[3], очевидным образом сводится к одному из простых отказов. Если это справедливо, то при отказных состояниях ММ выдаваемый им результат будет неотличим от результата при исправном или неисправном состоянии ФМ (и ФУз).

Подобная ситуация маскирования действительного состояния ФУз и ФМ при отказе ММ имеет место и при использовании не булевой логики состояний с другими обозначениями состояния исправности устройств, например: “1” – при исправном состоянии, “-1” – при неисправном:

$$
\begin{aligned}
\text{исправный ФМ и исправный ММ:} &\quad 1 \times 1 = 1,\\
\text{отказавший ФМ и исправный ММ:} &\quad (-1) \times 1 = (-1),\\
\text{исправный ФМ и отказавший ММ:} &\quad 1 \times (-1) = (-1),\\
\text{отказавший ФМ и отказавший ММ:} &\quad (-1) \times (-1) = 1.
\end{aligned}
\tag{2}
$$

Здесь значение “-1” логического выхода ММ соответствует неправильному функционированию либо ФМ, либо ММ. Одновременное неправильное функционирование ФМ и ММ приводит к значению “1” на выходе ММ, т.е. такое состояние ФУз неразличимо с правильным функционированием обоих модулей.

Возможности указанных формализмов ограничены при их использовании для автономного (самостоятельного) мониторинга ФУз из-за существенной неопределенности возможных результатов.

4. Парный мониторинг на основе логических правил

Для преодоления возникающих неопределенностей предлагается организовать парный мониторинг функциональных узлов ФМ+ММ с использованием любого из формализмов (1) или (2). При этом принимаются не очень существенные и широко распространенные на практике предположения.

[3] В стохастической задаче может возникнуть необходимость различения простых и сложных отказов.

А. Потоки данных через различные ФМ не связаны между собой (функциональная автономность ФМ).

Б. Каждый функциональный узел ФМ+ММ изготавливается на технологической базе и поддерживается инфраструктурными средствами, не зависимыми от базы и средств других ФУз (конструктивная разнородность ФУз).

В. Все ФМ и ММ изготовлены таким образом, что совместимы для образования различных ФУз независимо от технологических и инфраструктурных особенностей (интерфейсная однородность ФУз).

Г. Процесс мониторинга разбит на циклы, внутри которых технические состояния ФМ и ММ неизменны (стационарность неисправностей ФУз).

Принципиально важным следствием предположения Б является практическая невозможность одновременной неисправности двух ФМ и/или двух ММ в двух различных функциональных узлов. Само же такое предположение является распространенным, например, в авиаприборостроении. Так, при создании систем авионики высокой ответственности практикуется разнородное исполнение (наличие не связанных между собой нескольких разработчиков электронной компонентной базы и программного обеспечения) авиационных компонентов. Преследуемая цель: минимизация системных конструктивных и программных ошибок, практически не обнаруживаемых при единственном разработчике систем.

С учетом предположений А и В предлагаемая схема применима к подавляющему большинству технических систем с избыточностью.

Результат оценивания работоспособности ФУз в паре с учетом предположения Г удобно представлять оценочной матрицей (ОМ) размеров 2×2 с бинарными элементами вида

$$S_\tau^{\text{оц}} = \begin{bmatrix} s_\tau^{1-1} & s_\tau^{1-2} \\ s_\tau^{2-1} & s_\tau^{2-2} \end{bmatrix} = \begin{bmatrix} c_{\text{ФМ.1}} \\ c_{\text{ФМ.2}} \end{bmatrix} \times \begin{bmatrix} c_{\text{ММ.1}} & c_{\text{ММ.2}} \end{bmatrix}, \tag{3}$$

где $c_{\text{ФМ}.i}$ — логическое состояние i-го ФМ, $c_{\text{ММ}.j}$ — логическое состояние j-го ММ. Логические состояния модулей $c_{\text{ФМ}.i}$, $c_{\text{ММ}.j}$ и узлов s_τ^{i-j} при различных технических решениях могут соответствовать булеву (1) или не булеву формализму (2). Таким образом, первая строка ОМ – результат оценки исправности ФМ1 двумя ММ, первый столбец ОМ – результат оценки исправности обоих ФМ посредством ММ1 и т.д.

Схематическое решение парного мониторинга с ОМ иллюстрирует рис. 3.

Ключевой особенностью схемы рис. 3 является одновременное или поочередное в пределах одного цикла диагностирование каждым ММ каждого ФМ.

Кроме того, важным обстоятельством является то, что “взаимное проникновение” ФУз в паре происходит исключительно на уровне ММ, осуществляющих контроль ФМ. В то же время ни один из ФМ не вмешивается в работу другого ФМ, как и оба ММ не вмешиваются в работу ни одного из ФМ, в соответствии с принятым, в частности в авионике, принципом разделения.

Функцию заключительного звена в такой схеме играет сочетание аппаратно-программных средств, накапливающих (при мониторинге на одном цикле τ), хранящих и выдающих по запросу значения элементов матри-

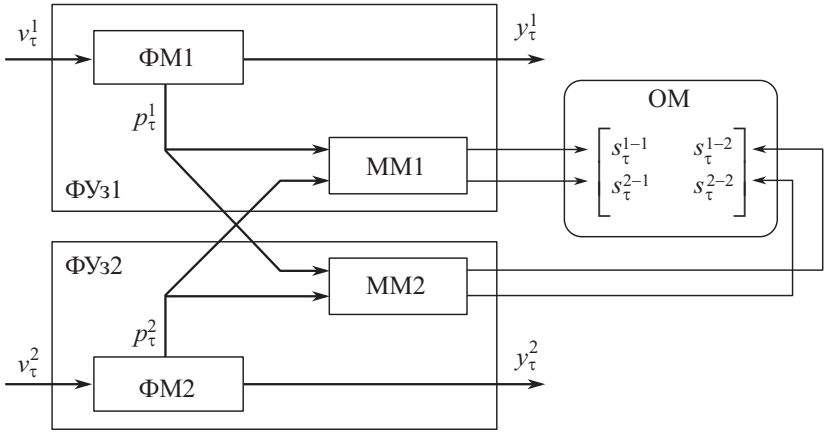


Рис. 3. Схема парного мониторинга функциональных узлов.

цы ОМ $S_\tau^{\text{оц}}$, а также переключающих каналы передачи данных. В зависимости от конкретных условий такое звено может либо принадлежать внешнему модулю, либо дублироваться в каждом из ММ.

Критичность выхода заключительного звена из-под контроля определяется относительной долей аппаратно-программных средств, реализующих это звено. Исследования показывают, что такая доля может быть сведена до весьма малых размеров.

С учетом принятого предположения Б и используемых формул логики (1) и (2) возможны исходы оценивания, т.е. значения логических оценок "исправен" или "неисправен" на выходах ММ, показанные на рис. 4 и 5. На обоих рисунках ОМ (3), относящиеся к разным комбинациям правильно и неправильно функционирующих модулей, выделены различными областями:

A – оба ФУз однозначно исправны;

B – неисправен один из ФМ с указанием, какой именно;

C – неисправен один из ММ с указанием, какой именно;

D – неисправны по одному различному модулю в каждом ФУз;

E – неисправны одновременно ФМ и ММ в одном из ФУз.

При этом в зависимости от используемых формул (1) или (2) области D и E характеризуются различной однозначностью. Если при булевом формализме (1) для ФУз пары, результат мониторинга которой соответствует области D, дается конкретное указание на исправное сочетание модулей ФМ1+ММ2 или ФМ2+ММ1, то при не булевом формализме такого указания нет. Аналогично для пары с результатом в области E булев формализм различает узлы с исправными и неисправными модулями, в то время как не булев формализм этого не делает.

Заключение о полной или частичной исправности и неисправности вместе с дальнейшими действиями определяются следующим образом:

а) оба функциональных узла пары, результат мониторинга которой соответствует области A, могут использоваться по предназначению;

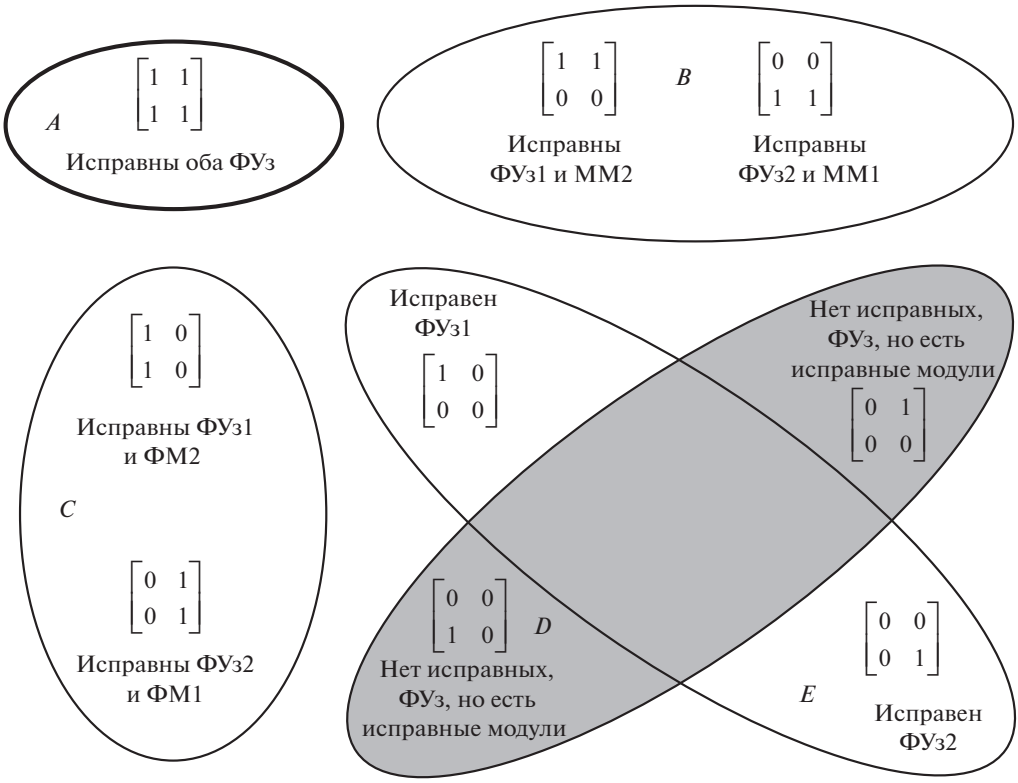


Рис. 4. Исходы парного мониторинга на основе булева формализма (1).

б) кроме того, может использоваться один (конкретный) функциональный узел из пары, результат мониторинга которой соответствует области B или C, а при булевом формализме еще один (конкретный) ФУз из пары, результат мониторинга которой соответствует области E;

в) если модули ФУз конструктивно неразделимы и неисправность любого из модулей объявляется неисправностью узла в целом, то следует отказаться от последующего использования конкретного ФУз пары, результат мониторинга которой соответствует области B или C, при булевом формализме – одного конкретного функционального узла пары в области E, а при не булевом формализме – одного ФУз пары в этой области, но для его выявления следует использовать парный мониторинг в сочетании с другими функциональными узлами;

г) если возможна перекомпоновка функциональных узлов, то с учетом предположения B можно создать дополнительные исправные ФУз, взяв исправный ФМ из неисправного узла пары, попавшей в область C, и объединить его с исправным ММ из неисправного узла пары, попавшей в область B; дополнительные возможности связаны с использованием модулей из ФУз пар в области D.

Однако описанная схема мониторинга обладает недостатком: она не полностью согласуется с практическими ситуациями, когда выходной сигнал

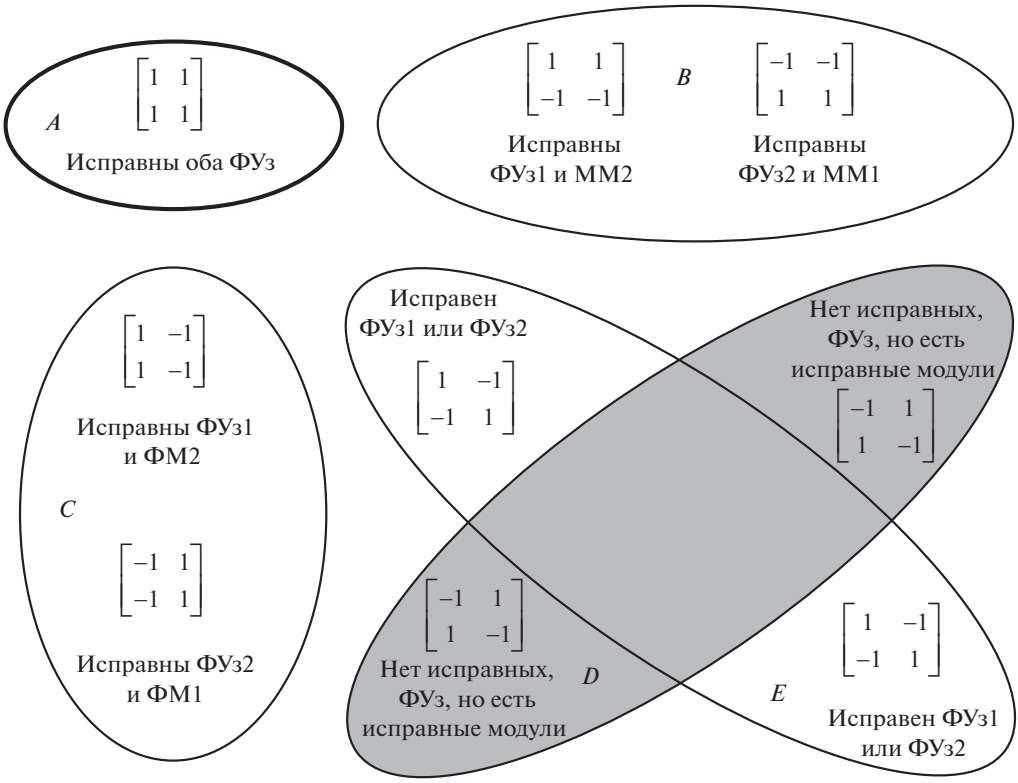


Рис. 5. Исходы парного мониторинга на основе не булева формализма (2).

какого-либо из ММ принимает фиксированное (неизменное) значение “1” или “0” (“константный отказ”).

На основе обобщения и расширения комбинаций, представленных на рис. 4 и 5, сформулируем индикаторное правило логического парного мониторинга, лишенное указанных недостатков.

5. Индикаторное правило логического парного мониторинга

Введем индикаторную матрицу (ИМ) $S_\tau^{\text{инд}}$ размеров 2×2 с бинарными элементами, для которой может применяться любой из формализмов (1) и (2) и в то же время некоторые из элементов которой, названные в статье “странными”, в отличие от (3) могут не соответствовать формулам ни одного из формализмов (1) и (2). Возникновение и использование “странных” элементов поясняются в этом разделе далее.

Схема парного мониторинга с ИМ, который назовем логическим парным мониторингом (ЛПМ), аналогична схеме, показанной на рис. 3, с заменой ОМ $S_\tau^{\text{оц}}$ на ИМ $S_\tau^{\text{инд}}$.

Можно убедиться, что с учетом предположения Б полная группа значений ИМ включает 13 различных матриц.

Используемые далее обозначения соответствуют булеву формализму (1).

Индикаторное правило ЛПМ. Выделение полностью или частично исправной пары функциональных узлов сводится к проверке структуры ИМ

$$S_{\tau}^{\text{инд}} = \begin{bmatrix} s_{\tau}^{1-1} & s_{\tau}^{1-2} \\ s_{\tau}^{2-1} & s_{\tau}^{2-2} \end{bmatrix},$$

где s_{τ}^{i-j} *– выраженный логической переменной результат проверки* i*-го* ФМ *посредством* j*-го* ММ*, уникальный вид которой соответствует* (*является индикатором*) *определенной комбинации исправных и неисправных узлов*:

однозначно исправен ФУз1 *при значениях* ИМ:

$$\underset{\substack{\text{дополнительно} \\ \text{исправен ММ2}}}{\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}}, \underset{\substack{\text{дополнительно} \\ \text{исправен ФМ2}}}{\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix};$$

однозначно исправен ФУз2 *при значениях* ИМ:

$$\underset{\substack{\text{дополнительно} \\ \text{исправен ММ1}}}{\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}}, \underset{\substack{\text{дополнительно} \\ \text{исправен ФМ1}}}{\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix};$$

однозначно нет исправных ФУз *при значениях* ИМ:

$$\underset{\substack{\text{исправны} \\ \text{ФМ2 и ММ1}}}{\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}}, \underset{\substack{\text{исправны} \\ \text{ФМ1 и ММ2}}}{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}, \underset{\substack{\text{исправны} \\ \text{ФМ2 и ММ1}}}{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}, \underset{\substack{\text{исправны} \\ \text{ФМ1 и ММ2}}}{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}};$$

неоднозначно: либо все ФУз *исправны, либо один из них* (*а именно,* ММ *с ложной выдачей* “1”) *неисправен при значении* ИМ:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{4}$$

Указанная неоднозначность и “ошибочное принятие[4] исправного ФМ за исправный” (на выходе неисправного ММ оценка “ФМ исправен”, не зависящая от состояния ФМ, относится к ФМ, который в действительности является исправным) не приводят непосредственно к негативным последствиям в текущем цикле.

[4] Обнаружение неисправности ММ возможно проверкой с тестом, что неприемлемо для мониторинга.

Индикаторные матрицы $S_\tau^{\text{инд}}$ технического состояния модулей при ЛПМ

Неисправность* одного из ФМ	Неисправность** одного из ММ				
	ММ1		ММ2		
	ложная “1”	ложный “0”	ложная “1”	ложный “0”	отсутствует
ФМ1	$\begin{bmatrix}1&0\\1&1\end{bmatrix}$	$\begin{bmatrix}0&0\\0&1\end{bmatrix}$	$\begin{bmatrix}0&1\\1&1\end{bmatrix}$	$\begin{bmatrix}0&0\\1&0\end{bmatrix}$	$\begin{bmatrix}0&0\\1&1\end{bmatrix}$
ФМ2	$\begin{bmatrix}1&1\\1&0\end{bmatrix}$	$\begin{bmatrix}0&1\\0&0\end{bmatrix}$	$\begin{bmatrix}1&1\\0&1\end{bmatrix}$	$\begin{bmatrix}1&0\\0&0\end{bmatrix}$	$\begin{bmatrix}1&1\\0&0\end{bmatrix}$
отсутствует	см. ***	$\begin{bmatrix}0&1\\0&1\end{bmatrix}$	см. ***	$\begin{bmatrix}1&0\\1&0\end{bmatrix}$	$\begin{bmatrix}1&1\\1&1\end{bmatrix}^{***}$

*ненадлежащее выполнение возложенных функций, которое должно обнаруживаться посредством ММ,

**ложная выдача “исправен” или “неисправен”,

***неоднозначно, кроме полной исправности узлов возможна неисправность ММ1 или ММ2 с ложной выдачей “исправен”.

Приведенное правило компактно иллюстрируется таблицей.

Если же в конкретных системах число функциональных узлов превышает два, то должен осуществляться попарный мониторинг всех ФУз, а в случае нечетного их количества какой-то из узлов подвергнется проверке дважды, что не привносит каких-либо методических и технических проблем. Если исправных узлов окажется больше единицы, то выбор предпочтения среди исправных ФУз выходит за рамки собственно мониторинга. Соответствующее решение при выборе конфигурации бортового комплекса описано в [6].

Обоснованность правила подтверждает следующий анализ.

Если ММ исправен, то он будет отражать действительное состояние ФМ, т.е. информировать о его исправности или неисправности. Возможно применение любого из формализмов (1) или (2):

$$(5)\qquad 1 \times \underbrace{1}_{\substack{\text{исправность}\\ \text{ММ}}} = \underbrace{1}_{\substack{\text{истинный}\\ \text{ответ}}}, \qquad 0 \times \underbrace{1}_{\substack{\text{исправность}\\ \text{ММ}}} = \underbrace{0}_{\substack{\text{истинный}\\ \text{ответ}}}$$

$$\text{или} \quad 1 \times 1 = 1, \quad (-1) \times 1 = (-1).$$

Дело обстоит сложнее при неисправном ММ. Пусть его выход не зависит от состояния ФМ и может принимать значение как “1”, так и “0” (или “-1”, в зависимости от формализма). Подробнее для каждого из этих вариантов:

а) неизменная выдача значения “исправен”, т.е. “1” на выходе ММ:

при исправном ФМ получается, что ММ сработал как исправный при любом формализме, и ситуация нераспознаваема, т.е. отличить неисправное состояние функционального узла от исправного невозможно, но и негативных последствий тоже нет:

$$(6)\qquad 1 \times \underbrace{1}_{\substack{\text{фиксированный}\\ \text{отказ ММ}}} = \underbrace{1}_{\substack{\text{истинный}\\ \text{ответ}}},$$

при неисправном ФМ ситуация соответствует только формализму (2):

$$(7)\qquad (-1) \quad \times \quad \underbrace{(-1)}_{\substack{\text{фиксированный}\\ \text{отказ ММ}}} \quad = \underbrace{1}_{\substack{\text{ложный}\\ \text{ответ}}},$$

формализм же (1) неприменим;

б) неизменная выдача значения "неисправен", т.е. "0" или "-1" на выходе ММ:

при исправном ФМ результат соответствует любому из формализмов

$$(8)\quad 1 \quad \times \quad \underbrace{0}_{\substack{\text{фиксированный}\\ \text{отказ ММ}}} \quad = \underbrace{0}_{\substack{\text{истинный}\\ \text{ответ}}} \qquad \text{или} \qquad 1 \quad \times \quad \underbrace{(-1)}_{\substack{\text{фиксированный}\\ \text{отказ ММ}}} \quad = \underbrace{(-1)}_{\substack{\text{истинный}\\ \text{ответ}}}$$

и указывает на наличие неисправности в функциональном узле,

при неисправном ФМ получаем выполнение только формализма (1):

$$(9)\qquad 0 \quad \times \quad \underbrace{0}_{\substack{\text{фиксированный}\\ \text{отказ ММ}}} \quad = \underbrace{0}_{\substack{\text{истинный}\\ \text{ответ}}},$$

формализм (2) неприменим.

Таким образом, при описанном характере неисправностей налицо смешение двух формализмов (1) и (2), что предусмотрено сформулированным индикаторным правилом ЛПМ.

Пусть теперь образована пара функциональных узлов в соответствии с рис. 3, т.е. ФУз1 (ФМ1+ММ1) и ФУз2 (ФМ2+ММ2).

Если оба ФУз исправны, то результат в виде ИМ на выходах ММ совпадает с (3) и содержит все единичные элементы.

Если неисправен только ФМ1 (или ФМ2), то ИМ в соответствии с (5) принимает значение

$$(10)\qquad \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \left(\text{или} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \right).$$

Если неисправен только ММ1 (или ММ2), неизменно выдавая "1", то ИМ в соответствии с (6) не противоречит исправности обоих ФУз

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \left(\text{и} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right),$$

а неизменно выдавая "0", ошибочно приписывает неисправность исправным ФМ1 и ФМ2 одновременно, тем самым в соответствии с (8) раскрывая свое неисправное состояние (противоречит предположению Б) и отрицая исправность своего ФУз

$$(11)\qquad \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \left(\text{или} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \right).$$

При одновременной неисправности ФМ1 и ММ1 (или ФМ2 и ММ2) с неизменной выдачей оценки “1” в соответствии с (4) и (6) получается значение ИМ

$$\begin{bmatrix} \boxed{1} & 0 \\ 1 & 1 \end{bmatrix} \left(\text{или} \begin{bmatrix} 1 & 1 \\ 0 & \boxed{1} \end{bmatrix} \right), \tag{12}$$

указывающее на исправность ФУз2 (или ФУз1). Здесь взятые в рамки значения “1” символизируют появление “странного” значения “1” на выходе ММ1 (или ММ2).

Странность заключается в том, что такое значение не соответствует ни одному из описанных в разделе 3 формализмов (см. комментарий к (7)). И хотя пользователь не может знать о странности одного из элементов ИМ, это не влияет на работоспособность индикаторного правила. Нужное решение формируется безошибочно. В отличие же от (10) в этом случае делается вывод об отсутствии исправных модулей за пределами ФУз2 (ФУз1).

При одновременной неисправности ФМ1 и ММ1 (или ФМ2 и ММ2) с неизменной выдачей оценки “0” в соответствии с (8) и (9) получается значение ИМ

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \left(\text{и} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right), \tag{13}$$

что в силу формализма (1) указывает на исправность только ФУз2 (или ФУз1).

При одновременной неисправности ФМ1 и ММ2 (или ФМ2 и ММ1) с выдачей “1” неисправным ММ получается значение ИМ

$$\begin{bmatrix} 0 & \boxed{1} \\ 1 & 1 \end{bmatrix} \left(\text{или} \begin{bmatrix} 1 & 1 \\ \boxed{1} & 0 \end{bmatrix} \right), \tag{14}$$

со “странным” элементом, о странности которого пользователь не подозревает. Вместе с тем вид ИМ позволяет утверждать об отсутствии исправных ФУз. Выдача “0” неисправным ММ2 (или ММ1) приводит к значению ИМ

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \left(\text{и} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right), \tag{15}$$

что тоже свидетельствует об отсутствии исправных функциональных узлов.

Можно убедиться, что указанный в разделе 3 сложный отказ в виде инверсии оценки состояния ФМ корректно отражается в ИМ (11)–(15).

6. Использование логического парного мониторинга

Использование индикаторного правила ЛПМ позволяет однозначно[5] устанавливать техническое состояние ФМ. Так, например, для подтверждения исправности ФМ1 без акцентирования внимания на других модулях схемы

[5] В исходной детерминистской постановке решаемая задача не предусматривает вероятностные понятия типа ошибок 1-го и 2-го рода. Иное излагается в разделе 8.

ЛПМ (рис. 3) требуется убедиться, что ИМ либо принимает одно из значений

$$\begin{bmatrix}1&1\\0&0\end{bmatrix},\begin{bmatrix}1&0\\1&0\end{bmatrix},\begin{bmatrix}1&1\\0&1\end{bmatrix},\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&1\end{bmatrix},\begin{bmatrix}1&1\\1&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix}\text{ или }\begin{bmatrix}1&1\\1&1\end{bmatrix},$$

либо не принимает ни одного из значений

$$\begin{bmatrix}0&0\\1&1\end{bmatrix},\begin{bmatrix}1&0\\1&1\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix},\begin{bmatrix}0&1\\1&1\end{bmatrix}\text{ или }\begin{bmatrix}0&0\\1&0\end{bmatrix}$$

при обязательном удовлетворении предположения Б.

Другой вариант использования ЛПМ выглядит следующим образом. Если ИМ принимает одно из значений

$$\begin{bmatrix}1&1\\1&1\end{bmatrix},\begin{bmatrix}1&0\\1&0\end{bmatrix}\text{ или }\begin{bmatrix}0&1\\0&1\end{bmatrix},$$

то исправны оба ФМ в паре функциональных узлов. При значениях ИМ

$$\begin{bmatrix}1&1\\0&0\end{bmatrix},\begin{bmatrix}1&1\\0&1\end{bmatrix},\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}1&1\\1&0\end{bmatrix}\text{ или }\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

среди ФМ пары исправен только ФМ1, а при значениях

$$\begin{bmatrix}0&0\\1&1\end{bmatrix},\begin{bmatrix}1&0\\1&1\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix},\begin{bmatrix}0&1\\1&1\end{bmatrix}\text{ или }\begin{bmatrix}0&0\\1&0\end{bmatrix}$$

– только ФМ2. При этом исправность или неисправность ММ не отмечается.

Аналогично можно выбрать комбинации значений ИМ для суждения об исправности только ММ с той лишь особенностью, что если ИМ принимает значение (4), то один из ММ следует подозревать (но не более того) в неисправности.

Для снятия подозрений следует, если это не противоречит позиции разработчика[6], внести управляемую неисправность в один из заведомо исправных ФМ, например в ФМ1. При этом неисправность одного из ММ проявится в виде значения ИМ

$$\begin{bmatrix}1&0\\1&1\end{bmatrix}\quad\text{или}\quad\begin{bmatrix}0&1\\1&1\end{bmatrix},$$

которое в соответствии с таблицей индикаторных матриц укажет на неисправный ММ.

Основная направленность предлагаемой разработки связана с так называемыми избыточными комплексами бортового оборудования, где по разным причинам (вследствие как традиционно практикуемого, так и сознательно

[6] Действие относится к тест-контролю, обычно не применяемому в процедурах мониторинга.

вводимого функционального и структурного резервирования в интересах повышения безопасности, достижения живучести и необслуживаемости в межрегламентные периоды) количество и функциональность компонентов заведомо превышает необходимый минимум.

Поэтому наиболее простая область применения – мониторинг идентичных компонентов, каждый из которых содержит идентичные функциональные и мониторинговые модули. Более отдаленная перспектива – функционально избыточные компоненты, позволяющие решать аналогичные задачи на основе разных физических принципов и технических решений. Для этого, естественно, потребуется развитие подхода.

7. Варианты исполнения функциональных узлов

В зависимости от конструкторских и иных возможностей может реализовываться какая-либо из приводимых далее схем или их комбинация.

1. *Мониторинг ФМ с его дубликатом.* Каждый ММ содержит копию (дубликат) "своего" ФМ идентичной физической природы или построенную на других физических принципах (например, математические модели процессов технического устройства). Рисунок 6,*а* иллюстрирует данный вариант. Здесь СС – схема сравнения (компаратор), подтверждающая "1" или отрицающая "0" совпадение сигналов y_τ на выходе ФМ и $y_\tau^{\text{М}}$ на выходе его физической или математической модели. Следует обратить внимание на то, что

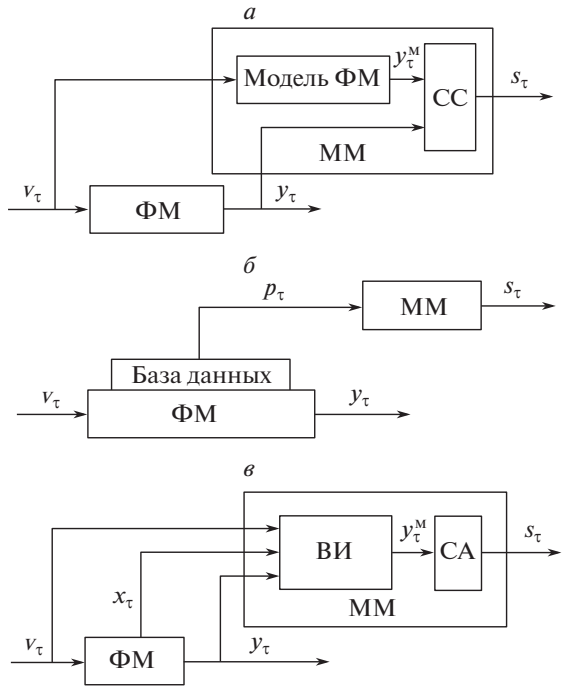


Рис. 6. Схемы функциональных узлов: *а* – с моделью, *б* – с учетом эксплуатационных данных, *в* – с вычислением инвариантов.

предлагаемый подход ЛПМ с натурным дубликатом ФМ в отличие от "почти аналогичного" традиционного дублирования позволяет однозначно выделить в паре исправный ФМ.

2. *Мониторинг ФМ по его эксплуатационным данным.* Схему соответствующего функционального узла поясняет рис. 6,*б*. Подразумевается, что непосредственно с ФМ конструктивно и функционально связан специальный элемент[7] (чип), собирающий и накапливающий данные об условиях его использования и хранения. В число параметров p_τ, хранимых и выдаваемых таким чипом в ММ, входят различные данные о ФМ, включая:

паспортные данные,

результаты испытаний на разных стадиях жизненного цикла,

статистику эксплуатационных показателей и характеристик (оценки достигаемой точности, остаток ресурса, энергетические показатели и пр.),

статистику внешних воздействий во время использования по назначению, при хранении и регламентных работах.

На ММ же возлагается анализ поступающих данных и формирование на основе этого анализа суждения о возможной исправности или неисправности ФМ.

3. *Мониторинг ФМ по специальным соотношениям.* В этом случае ММ по получаемым из ФМ входным v_τ и выходным y_τ данным и промежуточным результатам x_τ определяет некоторые показатели или проверяет характерные соотношения (инварианты), которые при правильном функционировании ФМ должны удовлетворяться, а в случае его неправильного функционирования – нарушаться. Рисунок 6,*в* поясняет соответствующую схему ФУз. Здесь ВИ – вычислитель инвариантов, СА – схема анализа инвариантов. Функционирование парного мониторинга в данном варианте функционального узла поясняет первый из примеров в разделе 9.

Заметим, что правила сертификации изделий авионики по высшей категории А предполагают наличие ММ, встроенного в изделие, который осуществляет на одном и том же потоке входных данных параллельно с ФМ вычисление выходных параметров на альтернативной основе (упрощенно и потому с высокой надежностью) и сравнение своего результата с результатом ФМ, что соответствует рис. 6,*а* и 6,*в*.

8. Эффект использования логического парного мониторинга

Расширим изначальную постановку задачи, добавляя вероятностную составляющую. В первом приближении можно полагать, что все функциональные узлы обладают одинаковыми вероятностными характеристиками исправности, а предположение Б, сделанное в разделе 4, строго выполняется. Тогда эффект, достигаемый при реализации ЛПМ, определяется следующим образом.

Полная группа независимых событий, связанных с техническим состоянием ММ (ВСК) каждого ФУз в каждом отдельном цикле τ мониторинга,

[7] Идея и инициативные проработки принадлежат А.В. Дядищеву.

включает: исправное функционирование, выдачу ложной оценки “исправен” и выдачу ложной оценки “неисправен”. Здесь подразумевается, что ложные значения оценок существуют сами по себе и не связаны с возникновением одноименных истинных оценок.

Введем значения вероятностей возникновения неисправностей ММ (ВСК):

Q_1 – вероятность выдачи ложной оценки “исправен”,

Q_0 – вероятность выдачи ложной оценки “неисправен”.

Тогда при условии $Q_1 + Q_0 \leqslant 1$ вероятность исправного функционирования ВСК имеет значение $P = 1 - Q_1 - Q_0$.

Кроме того, в соответствии с конструктивными решениями и условиями функционирования ММ (ВСК) обнаруживает неисправности ФМ с определенной вероятностью $P_{\text{ВСК}}$, совершая ошибки 1-го рода (ложная тревога) с вероятностью $P_{\text{ВСК}|\,1}$ и 2-го рода (пропуск[8] неисправности) с вероятностью $P_{\text{ВСК}|\,2}$.

При автономном использовании встроенных средств контроля вероятности того, что будет обнаружена возникшая неисправность ФМ или будет совершена ошибка какого-либо рода, определяются формулами:

$$\begin{aligned} P_{\text{обн ФМ}} &= P_{\text{ВСК}}\,(1 - Q_1 - Q_0)\,, \\ P_{\text{обн ФМ}|\,1} &= P_{\text{ВСК}|\,1} + Q_0, \\ P_{\text{обн ФМ}|\,2} &= P_{\text{ВСК}|\,2} + Q_1. \end{aligned}$$

Использование же логического парного мониторинга в отношении контроля технического состояния ФМ исключает последний сомножитель в первой из этих формул и последние слагаемые в остальных, заменяя их на формулы:

$$P_{\text{обн ФМ}} = P_{\text{ВСК}}, \quad P_{\text{обн ФМ}|\,1} = P_{\text{ВСК}|\,1}\,, \quad P_{\text{обн ФМ}|\,2} = P_{\text{ВСК}|\,2}\,.$$

Кроме того, использование ЛПМ позволяет обнаруживать неисправности ММ, совершая ошибки 1-го или 2-го рода, с вероятностями:

$$\begin{gathered} P_{\text{обн ММ}} = P_{\text{ВСК}}, \quad P_{\text{обн ММ}|\,1} = P_{\text{ВСК}|\,1} \;-\text{ для всех значений ИМ}, \\ P_{\text{обн ММ}|\,2} = P_{\text{ВСК}|\,2} \;-\text{ для всех значений ИМ, кроме (4)}, \\ P_{\text{обн ММ}|\,2} = P_{\text{ВСК}|\,2} + Q_1 \;-\text{ для значения ИМ (4)}. \end{gathered}$$

Здесь указанный в разделе 5 случай со значением ИМ, равным (4), составляет исключение, когда ложное подтверждение одним из ММ (ВСК) исправности обоих ФМ оборачивается пропуском соответствующей неисправности этого ММ.

Таким образом, эффект соответствует использованию как бы “абсолютно надежных” ММ (ВСК) в отношении ФМ и “почти абсолютно надежных” ММ (с оговоркой об ИМ (4)) по отношению к самим себе.

[8] Речь идет о неисправностях, которые ВСК должен обнаруживать. Неисправности же “вне ответственности” конкретного ВСК остаются за границами контроля технического состояния соответствующего изделия.

Для более тонкого анализа эффекта использования ЛПМ требуется введение различия вероятностных характеристик модулей разных узлов и, что более существенно, требуется допустить нарушение предположения Б с некоторой вероятностью. Однако это заслуживает особого внимания и не является предметом данной статьи.

9. Примеры

Преобразование векторов. Рассмотрим в качестве примера характерный фрагмент вычислений, выполняемых на борту воздушного судна.

Пусть ФМ – программный модуль преобразования вектора $\overrightarrow{OG}_{\text{н}} = \left[x_{\text{g}} \;\; y_{\text{g}} \;\; z_{\text{g}}\right]^{\text{T}}$ (скорость, орт линии визирования или др.), заданного проекциями на оси нормальной системы координат (СК) $OX_{\text{g}}Y_{\text{g}}Z_{\text{g}}$ (направления осей ориентированы по характерным направлениям местности), в вектор $\overrightarrow{OG}_{\text{св}} = \left[x \;\; y \;\; z\right]^{\text{T}}$, представленный проекциями на оси СК $OXYZ$, связанной с воздушным судном (направления осей ориентированы по характерным направлениям воздушного судна). Для преобразования вектора в ФМ производятся вычисления по формуле

$$\overrightarrow{OG}_{\text{св}} = A^{\text{н}}_{\text{св}}\overrightarrow{OG}_{\text{н}},$$

где

$$A^{\text{н}}_{\text{св}} = \begin{bmatrix} \cos\vartheta\cos\psi & \sin\vartheta & -\cos\vartheta\sin\psi \\ -\cos\gamma\sin\vartheta\cos\psi + \sin\gamma\sin\psi & \cos\gamma\cos\vartheta & \cos\gamma\sin\vartheta\sin\psi + \sin\gamma\cos\psi \\ \sin\gamma\sin\psi\cos\psi + \cos\gamma\sin\psi & -\sin\gamma\cos\vartheta & -\sin\gamma\sin\vartheta\sin\psi + \cos\gamma\cos\psi \end{bmatrix}$$

– ортогональная матрица преобразования координат из нормальной в связанную СК, γ – угол крена, ψ – угол рыскания, ϑ – угол тангажа. Относительную ориентацию этих СК и направления отсчета углов иллюстрирует рис. 7.

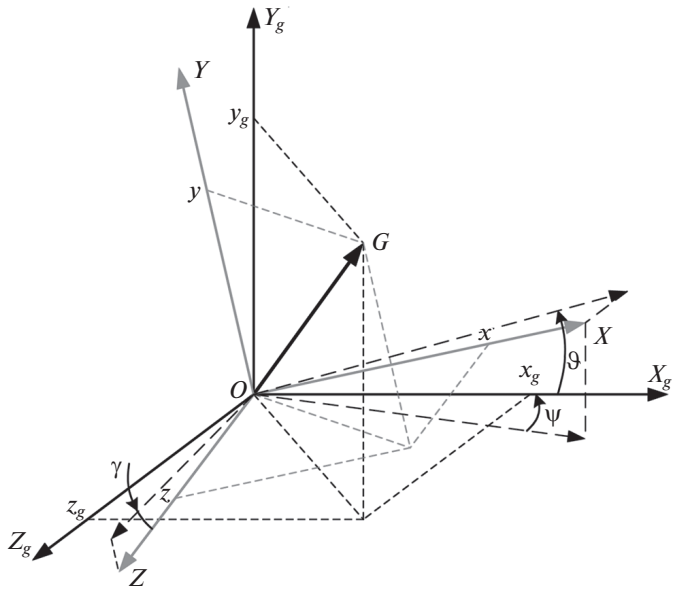


Рис. 7. Системы координат и преобразование проекций вектора.

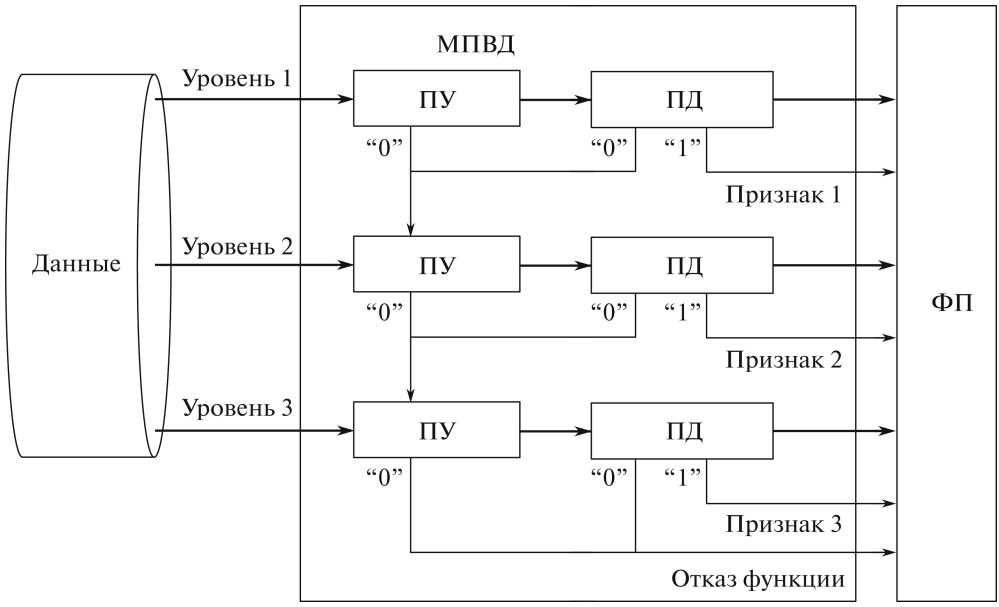


Рис. 8. Модуль проверки входных данных типовой функции авионики.

При этом ММ осуществляет мониторинг ФМ, выполняющего указанное преобразование, путем расчета и сравнения норм векторов (см. рис. 6,*в*):

$$\underbrace{(x_{\text{g}}^2 + y_{\text{g}}^2 + z_{\text{g}}^2)}_{\substack{\text{квадратичная} \\ \text{норма } \overrightarrow{OG}_{\text{н}}}} - \underbrace{(x^2 + y^2 + z^2)}_{\substack{\text{квадратичная} \\ \text{норма } \overrightarrow{OG}_{\text{св}}}} = \Delta.$$

Эти нормы при безошибочных вычислениях должны совпадать, т.е. $\Delta = 0$. Таким образом, в случае совпадения норм векторов до и после преобразования ММ должен выдавать значение логической переменной “1”, в противном случае – “0”.

В результате ЛПМ двух узлов ФМ+ММ соответствующая ИМ $S_\tau^{\text{инд}}$ будет принимать одно из значений, перечисленных в индикаторном правиле, предоставляя тем самым возможность для оценивания работоспособности функциональных узлов в процессе их функционирования.

Проверка входных данных. В авионике принято решаемые в бортовом комплексе прикладные задачи называть функциями. Вычислители, реализующие какую-либо функцию, получают от периферийных систем данные нескольких уровней точности (возможно деление на уровни и по другому признаку).

Поступающие данные подвергаются входной проверке, выполняемой в соответствии с рис. 1,*в* и нацеленной на обнаружение отказов сопрягаемых датчиков информации и каналов связи. На рис. 8 показана упрощенная структура соответствующего ММ – модуля проверки входных данных (МПВД). Укрупненно эта проверка делится на проверку устойчивости (ПУ) приема кодовых слов, поступающих по каналу связи, и применения правил досто-

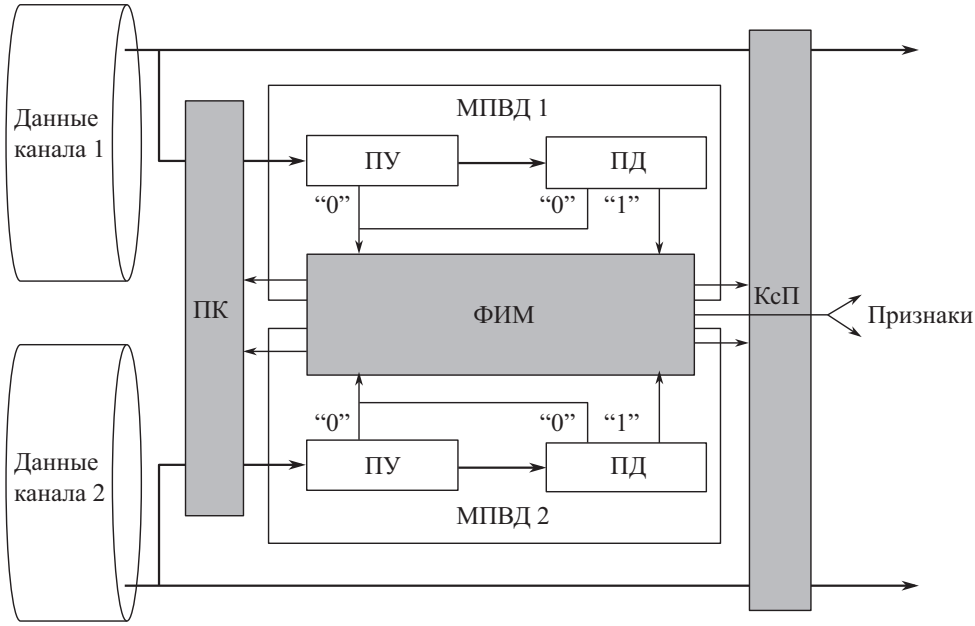


Рис. 9. Организация ЛПМ для проверки данных i-го уровня в двух каналах.

верности (ПД) для получаемой информации. Проверка устойчивости заключается в проверке кодов и анализе временны́х интервалов их поступления, посредством чего обнаруживаются сбои и ошибки передачи данных. Проверка достоверности включает сравнение различных полученных параметров между собой и с заданными границами, позволяющее обнаружить большинство статических и динамических ошибок (независимо от природы их возникновения) источников информации.

Если входные данные уровня 1 отвечают критериям устойчивости и достоверности, то они без каких-либо изменений передаются в следующий модуль, реализующий функцию (функциональное приложение – ФП). Параллельно туда направляется признак исправности данных. Если же прием данных неустойчив или они не удовлетворяют критериям достоверности, то управление передается уровню 2, где проверочные процедуры повторяются, как правило, с другими требованиями. При необходимости цикл повторяется еще раз. Если данные последнего уровня не прошли контроль, то модуль выдает признак отказа функции, не передавая данные для дальнейшей обработки.

Так вкратце выполняются входные проверки в современной авионике. ФМ представляет собой определенный поток данных, а ММ – совокупность процедур параметрического контроля [18]. Проверки же логики работы самого ММ осуществляются периодически по тестовым наборам данных, т.е. вне реального времени и рабочего режима. Вопрос об исправности диагностического теста, который в интересах полноты может быть весьма сложным, остается открытым.

Предлагаемый подход в применении к описанному процессу иллюстрируется на рис. 9, где два ММ (МПВД 1 и МПВД 2) в составе разных вычисли-

тельных каналов, разнородных в указанном выше смысле, участвуют в ЛПМ. Признаки исправности или неисправности данных (в данном примере они в терминах мониторинга играют роль ФМ) и ММ формируются по ИМ $S_{\tau}^{\text{инд}}$ в соответствии с разделами 5 и 6.

Серым цветом на рисунке выделены дополнительные элементы (перекрестный коммутатор – ПК, формирователь индикаторной матрицы – ФИМ, коммутатор с потребителями – КсП), вместе играющие роль заключительного звена. И хотя эти элементы показаны как части схемы за пределами обоих проверочных модулей, при реализации они могут быть разнесены и/или частично дублированы в каждом из модулей.

Подчеркивается, что признаки исправности данных и модулей их проверки формируются одновременно в реальном времени и в рабочем режиме на каждом цикле мониторинга. Наличие перекрестной коммутации на входе и выходе ММ предоставляет дополнительные возможности для реконфигурации системы в целом [6].

Снятие же подозрения с одного из ММ в неисправности, заключающейся в ложной выдаче “1”, при получении значения ИМ (4) обеспечивается повторной проверкой данных с предварительным отключением в элементе ПК входа одного из МПВД (см. раздел 6).

10. Заключение

Показано, что существующие подходы к построению средств контроля технического состояния бортового оборудования обладают особенностями, исключающими, в значительной степени затрудняющими или ставящими в зависимость от сильных предположений построение систем мониторинга исправности компонентов оборудования в реальном времени.

Предложен подход к осуществлению мониторинга путем совмещенного использования самостоятельного и взаимно-перекрестного контроля пары разнородных по производственным платформам избыточных узлов, каждый из которых содержит функциональный и мониторинговый модули, в одинаковой степени допускающие неправильное функционирование. Подход, названный логическим парным мониторингом (ЛПМ), позволяет установить исправность или одну–две неисправности как функциональных, так и мониторинговых модулей одновременно.

Достоинствами предложенного подхода ЛПМ являются:

однозначность определения технического состояния функциональных модулей в условиях возможных неисправностей как самих функциональных модулей, так и модулей, осуществляющих мониторинг их состояния;

определение одиночных и двойных неисправностей в соответствии с таблицей в разделе 5;

использование хорошо развитой технологии разработки и применения встроенных средств технического диагностирования;

относительно незначительное усложнение схематических и программных решений для реализации ЛПМ.

Эффект применения ЛПМ заключается в приведении традиционных встроенных средств контроля в разряд средств как бы “абсолютно надежных” в отношении контроля ФМ и “почти абсолютно надежных” в отношении контроля самих себя.

Приведенные примеры демонстрируют объекты для организации ЛПМ.

СПИСОК ЛИТЕРАТУРЫ

1. *Буков В.Н., Бронников А.М., Агеев А.М., Гамаюнов И.Ф., Озеров Е.В., Шурман В.А.* Концепция управляемой избыточности комплексов бортового оборудования // Науч. чтения по авиации, посвящ. пам. Н.Е. Жуковского: Матер. XVI Всерос. науч.-практ. конф. / Гл. ред. С.П. Халютин (11–12 апр. 2019, Москва). М.: Изд. дом Акад. им. Н.Е. Жуковского, 2019. С. 17–33.
2. *Буков В.Н., Евгенов А.В., Шурман В.А.* Интегрированные комплексы бортового оборудования с управляемой функциональной избыточностью // Актуальные проблемы и перспективные направления развития комплексов авиационного оборудования. Сб. науч. ст. по матер. V Междунар. науч.-практ. конф. “Академические Жуковские чтения” (22-23 нояб. 2017, Воронеж). Воронеж: КВАЛИС, 2018. С. 23–28.
3. *Буков В.Н., Евгенов А.В., Шурман В.А.* Управление функциональной избыточностью перспективных интегрированных комплексов бортового оборудования // Матер. засед. межвед. раб. груп. по подгот. предлож-й, направл. на выявл. перспект. и прорыв. направ. науч.-технич. и инновац. развития авиац. отрасли. М.: Студия Этника, 2018. С. 45–53.
4. *Sollock P.* Reconfigurable Redundancy – The Novel Concept Behind the World’s First Two-Fault-Tolerant Integrated Avionics System // Avionics, Navigation and Instrumentation. P. 243–246. URL: https://www.nasa.gov/centres/johnson/pdf/584731main_Wings-ch4e-pgs242-255.pdf.
5. *Каляев И.А., Мельник Э.В.* Реконфигурируемые информационно-управляющие системы // Матер. пленар. засед. 5-й Рос. мультиконф. по пробл. управл. СПб.: Изд. ЦНИИ Электроприбор, 2012. С. 36–37.
6. *Агеев А.М., Бронников А.М., Буков В.Н., Гамаюнов И.Ф.* Супервизорный метод управления технических систем с избыточностью // Изв. РАН. Теория и системы управления. 2017. № 3. С. 72–82.
7. *Pouliezos A.D., Stavrakakis G.S.* Real time fault monitoring of industrial processes. The Netherlands: Kluwer Acad. Publishers, 1994.
8. DO-297. Integrated modular avionics (IMA) development guidance and certification considerations. Washington: RTCA Inc., 2005.
9. ГОСТ Р 27.605-2013. Надежность в технике. Ремонтопригодность оборудования. Диагностическая проверка.
10. ГОСТ 20911-89. Техническая диагностика. Термины и определения.
11. *Amato F., Cosentino C., Mattei M., Paviglianiti G.* A Direct/Functional Redundancy Scheme for Fault Detection and Isolation on an Aircraft // Aerospace Science and Technology. 2006. No. 10 (4). P. 338–345.
12. *Isermann R., Ball’e P.* Trends in the Application of Model-based Fault Detection and Diagnosis of Technical Processes // Control Eng. Pract. 1997. No. 5 (5). P. 709–719.
13. *Marcos A., Balas G.* A Robust Integrated Controller/Diagnosis Aircraft Application // Int. J. Robust Nonlin. Control. 2005. No. 15. P. 531–551.

14. *Чернодаров А.В.* Контроль, диагностика и идентификация авиационных приборов и измерительно-вычислительных комплексов. М.: Научтехлитиздат, 2017.
15. Диагностирование и прогнозирование технического состояния авиационного оборудования. Уч. пос. для вузов гражд. авиации / Под ред. И.М. Синдеева. М.: Транспорт, 1984.
16. Основы технической диагностики. В 2-х кн. / Под ред. П.П. Пархоменко. М.: Энергия, 1976.
17. Машиностроение: Энц. Т. III-7. Измерения, контроль, испытания и диагностика / Под ред. В.В. Клюева. М.: Машиностроение, 1996.
18. *Долбня Н.А.* Встроенные средства контроля бортовой вычислительной системы под управлением операционной системы реального времени как итеративный агрегированный объект // Вестн. Самарского гос. аэрокосмич. ун-та. 2012. № 5(36). С. 224–228.
19. *Буков В.Н., Базанов А.П., Колодежный А.П., Максименко И.М., Шпилевой Ю.М.* Теоретические основы и средства автоматизированного контроля / Под общ. ред. В.Н. Букова. М.: Изд. ВВИА, 1997.
20. ГОСТ 27.002-2015. Надежность в технике. Основные понятия. Термины и определения.
21. *Авакян А.А, Сучков В.Н., Искандеров Р.Д., Шурман В.А., Копнёнкова М.В., Вовчук Н.Г.* Способ и вычислительная система отказоустойчивой обработки информации критических функций летательных аппаратов. Патент RU 2413975 C2. Бюл. № 7 от 10.03.2011.

Статья представлена к публикации членом редколлегии М.Ф. Караваем.

Поступила в редакцию 17.01.2019

После доработки 04.07.2019

Принята к публикации 18.07.2019

© 2020 г. А.Г. КИРЬЯНОВ, канд. техн. наук (kiryanov@iitp.ru),
А.В. КРОТОВ (krotov@iitp.ru),
Е.М. ХОРОВ, канд. техн. наук (khorov@iitp.ru)
(Институт проблем передачи информации им. А.А. Харкевича РАН, Москва)
И.Ф. АКИЛДИЗ, д-р наук (ian@iitp.ru)
(Институт проблем передачи информации им. А.А. Харкевича РАН, Москва;
Технологический институт Джорджии, Атланта, Джорджия, США)

ПОВЫШЕНИЕ ЭНЕРГОЭФФЕКТИВНОСТИ ПЛОТНЫХ СЕТЕЙ WI-FI С ПРИМЕНЕНИЕМ ОБЛАЧНЫХ ТЕХНОЛОГИЙ[1]

В современном мире одним из лидеров в области беспроводных сетевых технологий, несомненно, является технология Wi-Fi. Рост плотности устройств в сетях Wi-Fi и числа самих сетей привел к высокой интерференции и, как следствие, к снижению производительности сетей Wi-Fi. Одним из эффективных решений для снижения интерференции в сценариях плотного размещения станций является использование облачных систем управления. В статье представлен алгоритм централизованной настройки параметров сети Wi-Fi для такой облачной системы. Алгоритм нацелен на максимизацию энергоэффективности путем решения оптимизационной задачи с ограничениями, в которой необходимо максимизировать разность двух монотонных функций. Валидация и оценка эффективности разработанного алгоритма проводится в среде имитационного моделирования NS-3.

Ключевые слова: энергоэффективность, Wi-Fi, 802.11ax, плотные сети, оптимизация, облачные технологии.

DOI: 10.31857/S0005231020010080

1. Введение

Как было отмечено аналитиками компании Cisco, в 2018 г. объемы данных, передаваемые с использованием технологии Wi-Fi в качестве последней мили, превысили объемы данных, при передаче которых используются только технологии проводной передачи данных [1]. Постоянное увеличение объемов трафика, числа устройств беспроводных сетей и их плотности порождают все новые и новые проблемы, связанные с увеличением пропускной способности сети и обеспечением высокого качества обслуживания различных типов трафика. Сложность в решении данных проблем возникает вследствие большой внутрисетевой и межсетевой интерференции, избежать которую в сетях Wi-Fi довольно сложно. Это во многом обусловлено методом случайного доступа к каналу, который используется в сетях Wi-Fi и допускает возникновение коллизий пакетов, в отличие от сетей LTE, в которых ресурсы беспроводного канала распределяются централизованно, что предотвращает возникновение коллизий.

[1] Исследование выполнено в ИППИ РАН за счет гранта Правительства Российской Федерации (договор No 14.W03.31.0019).

Чтобы повысить производительность плотных сетей Wi-Fi, международный комитет по стандартизации IEEE 802 разрабатывает новое дополнение IEEE 802.11ax [2] к стандарту Wi-Fi. Дополнение IEEE 802.11ax содержит набор методов, которые будут использоваться для уменьшения внутрисетевой и межсетевой интерференции, повышения спектральной эффективности и качества обслуживания пользователей в сценариях с плотным размещением станций: в крупных офисных и жилых зданиях, торговых центрах, аэропортах, стадионах и др.

Кроме того, стоит отметить набирающий популярность тренд использования Wi-Fi сетей — развертывание крупных корпоративных и домашних сетей на базе технологии Wi-Fi, управляемых одним оператором связи. При такой архитектуре сети оказывается полезным наличие единого центра координации, который будет обеспечивать наиболее эффективную совместную работу соседних точек доступа Wi-Fi. Многие производители устройств Wi-Fi, среди которых HP/Aruba Networks, Cisco/Miraki, Huawei, Quantenna Communications и др., уже разработали собственные облачные инфраструктуры для совместного управления множеством точек доступа [3]. Очевидно, что такие облачные системы могут существенно снизить межсетевую интерференцию, так как центр координации имеет исчерпывающее представление об интерференционной картине на каждой из контролируемых точек доступа и теоретически может оптимально распределить канальный ресурс между различными точками доступа. Среди параметров сетей Wi-Fi, значения которых имеет смысл выбирать централизованно, можно отметить номер используемого частотного канала, мощность передатчика, чувствительность приемника, необходимость использования механизма RTS/CTS и даже временное разделение канального ресурса.

Кроме этого, в современных крупных беспроводных сетях число точек доступа Wi-Fi может измеряться сотнями или даже тысячами. В таких сценариях использования немаловажную роль в операционных расходах на поддержание работоспособности беспроводной сети начинают играть расходы на электроэнергию. С учетом этой особенности важной чертой разрабатываемых решений для плотных сетей Wi-Fi является их энергоэффективность.

В данной статье рассматривается задача централизованного управления параметрами работы множества точек доступа Wi-Fi в плотной беспроводной сети с целью повышения пропускной способности сети, достижения низкого энергопотребления и справедливого распределения канального ресурса между различными устройствами. В статье ставится задача глобальной оптимизации и предлагается алгоритм ее решения, который позволяет определить оптимальные значения параметров точек доступа Wi-Fi.

Дальнейшее изложение материала построено следующим образом. В разделе 2 приводится краткий обзор публикаций по данной теме. В разделе 3 формулируется задача глобальной оптимизации, а в разделе 4 приводится описание предложенного алгоритма для ее решения. Исследование эффективности разработанного алгоритма представлено в разделе 5, заключительные выводы приведены в разделе 6.

2. Исследования по теме

В связи со стремительным ростом числа базовых станций и точек доступа в беспроводных сетях расходы на потребление энергии в таких сетях становятся все более существенными. Помимо обеспечения высокой пропускной способности сети, важным аспектом становится энергопотребление сети в целом. Для повышения энергоэффективности сети могут использоваться различные подходы, среди которых аккумулирование энергии от внешних источников, использование более энергоэффективного аппаратного обеспечения, грамотное планирование сети и распределение сетевых ресурсов [4]. В данной статье упор делается на разработку энергоэффективного алгоритма распределения сетевых ресурсов в плотных сетях Wi-Fi с использованием облачных технологий.

Авторы [5] предложили определить энергоэффективность как отношение объема данных, успешно переданных по каналу связи, к количеству затраченной энергии. В [5] авторы рассматривают передатчик, обладающий ограниченным запасом энергии, и сравнивают энергоэффективность различных алгоритмов повторной передачи недоставленных пакетов.

При оптимизации энергопотребления устройств необходимо принимать во внимание не только энергию, потребляемую непосредственно при передаче данных, но и учитывать постоянные энергозатраты, возникающие вне зависимости от передачи данных. В противном случае, как показано в [6], наиболее эффективной стратегией является использование самой низкой скорости передачи, которая достигается при наименьшей мощности передатчика.

В приведенных выше публикациях рассматривалось только одно беспроводное соединение. Очевидно, что определение энергоэффективности должно быть расширено на системы с множеством приемо-передающих устройств. В [7] предложено обобщенное определение энергоэффективности

$$U = \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \frac{r_i}{p_i + p_c}, \tag{1}$$

где U — функция полезности, характеризующая общую энергоэффективность, U_i — функция, характеризующая энергоэффективность соединения i, n — число соединений, r_i — скорость передачи данных по соединению i, p_i — средняя мощность передачи по соединению i, p_c — мощность, потребляемая в режиме ожидания. Основным недостатком предложенной функции полезности является тот факт, что она не отражает суммарный объем потребленной энергии, а представляет собой “сумму энергоэффективностей” отдельных соединений, что имеет менее явный физический смысл.

В [8] предлагается рассмотреть ряд других функций полезности, среди которых произведение энергоэффективностей, а также так называемая глобальная энергоэффекивность (GEE, Global Energy Efficiency). Глобальная энергоэффективность определяется как отношение суммарной скорости передачи данных по всем соединениям к общей потребляемой мощности. В [8] предлагаются быстрые алгоритмы для решения задач максимизации данных функций полезности. Однако для задачи максимизации глобальной энерго-

эффективности оптимальное решение найдено только для случая отсутствия интерференции от соседних передающих устройств.

Задача максимизации глобальной энергоэффективности может быть решена с использованием одного из методов решения задач монотонной оптимизации, заключающегося в последовательном приближении множества решений набором гиперпрямоугольников, имеющих общую вершину в начале координат и задаваемых координатами противоположной вершины [9]. Недостатком данного подхода является слишком медленная сходимость в случае, когда значение одной или нескольких переменных оказывается близким к нулю. Чтобы избежать такого поведения, в данной статье предлагается использовать метод ветвей и границ [10], который лишен отмеченного недостатка. Несмотря на то что метод ветвей и границ успешно использовался для решения задачи максимизации глобальной энергоэффективности в сетях LTE [11], его использование для решения подобной задачи в сетях Wi-Fi оказывается затруднительным. В частности, это обусловлено использованием в сетях Wi-Fi метода множественного доступа с контролем несущей и избеганием коллизий (CSMA/CA, Carrier Sense Multiple Access with Collision Avoidance) и наличием регуляторных ограничений на порог чувствительности приемника, которые необходимо принимать во внимание. Применение метода ветвей и границ для решения задачи максимизации пропускной способности в сетях Wi-Fi описано в [12], где на основе данного метода разработан алгоритм динамической настройки мощности передачи устройств Wi-Fi. Даже при постоянной нагрузке предложенный алгоритм динамически изменяет мощность передачи и позволяет достичь большей пропускной способности сети.

В данной статье обобщается алгоритм, предложенный в [12], для решения задачи максимизации глобальной энергоэффективности в сетях Wi-Fi с учетом требования на справедливое распределение канального ресурса между разными станциями сети.

3. Постановка задачи и ее анализ

Рассмотрим беспроводную сеть, в которой имеется M точек доступа и N пользовательских устройств, каждое из которых ассоциировано (т.е. соединено) с одной точкой доступа. Будем полагать, что данные передаются преимущественно в нисходящем канале, т.е. от точек доступа к пользовательским устройствам. Для обозначения соединения, передатчика или получателя будем использовать один и тот же индекс i. Мощность излучаемого сигнала при передаче данных по соединению i обозначим p_i. При этом затрачиваемая на передачу сигнала мощность равна $\phi_i p_i$, где $\frac{1}{\phi_i}$ — КПД усилителя. Мощность, потребляемую точкой доступа в режиме ожидания, обозначим p_c.

Для простоты будем считать p_c равной для всех точек доступа, тогда суммарное энергопотребление всех точек доступа в режиме ожидания равно Mp_c. В беспроводной среде сигнал передатчика j может быть принят не только получателем j, но и другими устройствами. Пусть $0 \leqslant a_{ij} \leqslant 1$ — коэффициент передачи сигнала между передатчиком j и получателем i, а $0 \leqslant b_{ij} \leqslant 1$ — коэффициент передачи сигнала между передатчиком j и передатчиком i. Будем полагать, что получатели имеют возможность принимать сигнал ненулевой

мощности от соответствующего передатчика, т.е. $\forall i\ a_{ii} > 0$. Кроме этого, естественно положить $\forall i\ b_{ii} = 0$.

Функция полезности (1) имеет следующие недостатки. Во-первых, максимизация такой функции может приводить к несправедливому распределению канального ресурса, например, если $p_c \gg p_i$. В таком случае функция полезности направлена на максимизацию общей пропускной способности, что приведет к несправедливому распределению ресурса. Во-вторых, сумма энергоэффективностей различных соединений не является величиной, отражающей энергоэффективность сети в целом. Исходя из этого, в данной статье предлагается определить энергоэффективность сети в целом как отношение средней пропускной способности к средней потребляемой мощности. В общем случае усредненная пропускная способность может быть определена как

$$U^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}U(r_i)\right), \tag{2}$$

где

$$U(r_i) = \begin{cases} \log(r_i), & \alpha = 1, \\ \dfrac{r_i^{1-\alpha}}{1-\alpha}, & (\alpha \geqslant 0) \wedge (\alpha \neq 1). \end{cases}$$

Например, если $\alpha = 0$, то (2) превращается в арифметическое среднее, если же $\alpha = 1$, то — в геометрическое среднее.

Таким образом, функция полезности $\hat{U}$ может быть определена следующим образом:

$$\hat{U} = \frac{U^{-1}\left(\frac{1}{N}\sum\limits_{i=1}^{N}U(r_i)\right)}{Mp_c + \sum\limits_{i=1}^{N}\phi_i p_i}. \tag{3}$$

В случае $\alpha = 0$ задача становится эквивалентной максимизации глобальной энергоэффективности.

Отметим, что скорость передачи данных r_i представляет собой неубывающую функцию от SINR (signal to interference plus noise ratio) — отношения сигнала к сумме интерференции и шума) γ_i, — которая может быть оценена при помощи существующих моделей ошибок передачи для известных сигнально-кодовых конструкций

$$r_i = f(\gamma_i(\boldsymbol{p})). \tag{4}$$

В свою очередь SINR на получателе i вычисляется так:

$$\gamma_i(\boldsymbol{p}) = \frac{a_{ii}p_i}{n_i + \sum\limits_{\substack{j=1 \\ j\neq i}}^{N} a_{ij}p_j}, \tag{5}$$

где n_i — тепловой шум на получателе i. Таким образом, независимыми переменными в (3) являются мощности передачи p_i.

Рассматриваемая в статье задача заключается в выборе таких мощностей передачи p_i, чтобы максимизировать (3) *с учетом ряда ограничений, возникающих при использовании технологии Wi-Fi:*

$$\begin{gathered} \max_j b_{ij} p_j \leqslant \hat{c}\ \forall i, \text{таких что } p_i > 0; \\ \max_j a_{ij} p_j \leqslant \hat{c}\ \forall i, \text{таких что } p_i > 0; \\ 0 \leqslant p_i \leqslant \hat{p}_i\ \forall i. \end{gathered} \tag{6}$$

Первое ограничение представляет собой принцип прослушивания канала перед началом передачи, которому обязаны следовать все устройства Wi-Fi, так как передача ведется в нелицензируемом спектре. Начать передачу разрешено только в том случае, если принимаемая мощность сигнала не превышает некоторого порогового значения (т.е. устройство не синхронизировалось на прием текущей передачи). Второе условие гарантирует, что получатель также не синхронизован в текущий момент ни на какую чужую передачу. Третье условие отражает ограничение на мощность излучаемого сигнала.

3.1. Анализ функции полезности

Для решения задачи максимизации функции полезности (3) при ограничениях (6) проведем некоторые преобразования. Представим (3) как функцию вектора $\boldsymbol{r}$ скоростей передачи данных. Это можно сделать, решая систему линейных уравнений (5) с учетом известной монотонной зависимости (4).

Прологарифмировав выражение (3), представим его как разность двух неубывающих функций, зависящих от вектора $\boldsymbol{r}$ скоростей передачи данных:

$$\log \hat{U}(\boldsymbol{r}) = V(\boldsymbol{r}) - W(\boldsymbol{r}), \tag{7}$$

где

$$V(\boldsymbol{r}) = \log U^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} U(r_i)\right) \tag{8}$$

и

$$W(\boldsymbol{r}) = \log\left(Mp_c + \sum_{i=1}^{N} \phi_i p_i(\gamma_i(r_i))\right). \tag{9}$$

Очевидно, что (8) является неубывающей функцией $\boldsymbol{r}$. Покажем, что (9) также является неубывающей функцией $\boldsymbol{r}$. В [13, лемма 2] показано, что для двух векторов SINR γ' и γ, таких что $\boldsymbol{\gamma'} \succeq \boldsymbol{\gamma}$ (т.е. каждая компонента вектора γ' не меньше, чем соответствующая компонента вектора γ) следует, что для соответствующих векторов мощностей передачи $\boldsymbol{p'}$ и $\boldsymbol{p}$ верно утверждение $\boldsymbol{p'} \succeq \boldsymbol{p}$. Известно, что если $\boldsymbol{r'} \succeq \boldsymbol{r}$, то $\boldsymbol{\gamma'} \succeq \boldsymbol{\gamma}$. Таким образом, (9) действительно является неубывающей функцией от скорости передачи данных.

Представив (7) как разность двух монотонных функций, можно применить известные оптимизационные методы [10, 11.1.2 DM Functions and DM Constraints]. Согласно [10, теорема 11.1] задача максимизации (7) может быть сведена к задаче максимизации монотонной функции полезности при наличии монотонных ограничений. Для этого необходимо ввести дополнительную переменную $w \in [-\exp(W_{\max}), -\exp(W_{\min})]$ и переформулировать постановку задачи следующим образом:

$$\max_{\boldsymbol{r},w} V(\boldsymbol{r}) - \log(-w) \tag{10}$$

при условии

$$\exp(W(\boldsymbol{r})) + w \leqslant 0. \tag{11}$$

Новая функция полезности (10) является неубывающей функцией от $\boldsymbol{r}$ и w.

Учитывая существующую линейную зависимость между компонентами векторов $\boldsymbol{r}$ и $\boldsymbol{p}$, все ограничения в задаче будут иметь вид

$$g(\boldsymbol{r}, w) \leqslant 0,$$

где $g(\boldsymbol{r}, w)$ — неубывающая функция.

Таким образом, *задача оптимизации энергоэффективности сведена к максимизации монотонной функции* (10) *при монотонных ограничениях* (11) *и* (6) *с учетом зависимостей* (4) *и* (5).

4. Алгоритм

4.1. Статическое решение

Решение изложенной выше оптимизационной задачи максимизации монотонной функции при наличии монотонных ограничений может быть найдено с помощью метода ветвей и границ. Основная идея данного метода состоит в том, чтобы представить рассматриваемую задачу в виде дерева подзадач. В данном случае каждая подзадача описывается областью в пространстве поиска $(\boldsymbol{r}, w)$, представляющей собой гиперпрямоугольник с нижней границей $(\boldsymbol{r}, w)^{(-)}$ и верхней границей $(\boldsymbol{r}, w)^{(+)}$.

Далее дадим краткое описание алгоритма и описание особенностей его применения для решения поставленной задачи. Подробное описание всех шагов алгоритма для оптимизации пропускной способности сети в пространстве $\boldsymbol{r}$ описано авторами в [12].

В качестве входных данных алгоритм принимает параметр ε, задающий требуемую точность решения. Алгоритм оптимизации энергоэффективности состоит из шагов инициализации, ветвления, разбиения и оценивания.

1. *Инициализация.* Создается список подзадач, состоящий из одной подзадачи, описываемой гиперпрямоугольником, для нижней границы $(\boldsymbol{r}, w)^{(-)}$ которого вектор $\boldsymbol{r}$ является нулевым и $w = -\exp(W_{\max})$, а для верхней границы $(\boldsymbol{r}, w)^{(+)}$ $r_i = f\left(\frac{a_{ii}\hat{p}_i}{n_i}\right)$ и $w = -\exp(W_{\min})$.

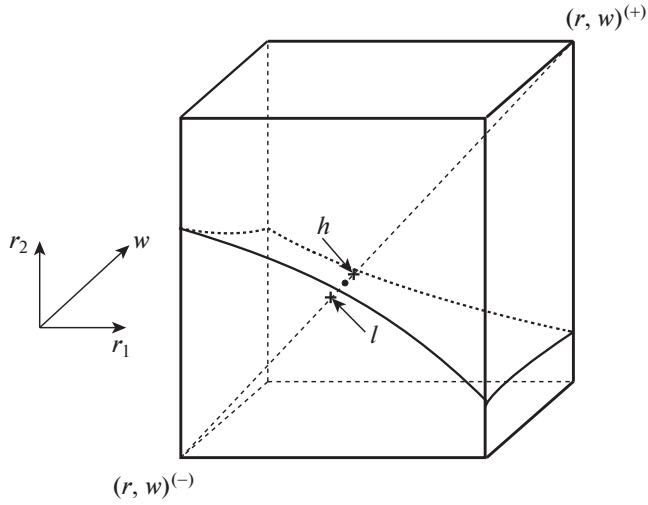


Рис. 1. Шаг оценивания.

2. *Ветвление.* Выбирается одна из подзадач из списка подзадач. В данном алгоритме выбирается последняя добавленная подзадача, т.е. используется алгоритм поиска в глубину аналогично [12].
3. *Разбиение.* Выбирается наиболее длинная сторона гиперпрямоугольника, представляющего выбранную подзадачу, и гиперпрямоугольник разбивается пополам вдоль нее. Заметим, что при этом сравниваются величины, имеющие разные размерности, так как скорости передачи данных (компоненты вектора $\boldsymbol{r}$) измеряются в Мбит/с, а энергопотребление w — в мВт. Так как в этих единицах величины имеют схожий порядок, то для упрощения реализации дополнительная нормализация данных величин перед сравнением не проводится.
4. *Оценивание.* Для каждой из полученных в результате разбиения подзадач методом дихотомии находится приближенное решение внутри рассматриваемого гиперпрямоугольника. Для этого вводятся вспомогательные переменные $\boldsymbol{l}$ и $\boldsymbol{h}$. Изначально $\boldsymbol{l} = (\boldsymbol{r}, w)^{(-)}$ и $\boldsymbol{h} = (\boldsymbol{r}, w)^{(+)}$. Затем вычисляется вектор $\boldsymbol{m} = \frac{1}{2}\boldsymbol{l} + \frac{1}{2}\boldsymbol{h}$ и проверяется, удовлетворяет ли этот вектор условиям (11) и (6). Если условия выполнены, то вектор $\boldsymbol{m}$ заменяет $\boldsymbol{l}$, иначе $\boldsymbol{h}$. Таким образом, происходит приближение к поверхности, ограничивающей область допустимых значений (см. рис. 1). После некоторого фиксированного числа итераций полученный вектор $\boldsymbol{l} = (\boldsymbol{r}_l, w_l)$ является приближенным решением подзадачи и используется для обновления наилучшего найденного решения, если значение функции полезности $V(\boldsymbol{r}_l) - \log(-w_l)$ оказывается больше, чем соответствующее значение для уже найденного решения. Вектор $\boldsymbol{h}$ используется для оценивания верхней границы значения функции полезности с использованием процедуры, описанной в [12]. Если оценка верхней границы оказывается более чем на ε меньше, чем значение функции полезности для наилучшего найденного решения, то подзадача отбрасывается. В противном случае подзадача добавляется в список рассматриваемых подзадач.

5. *Завершение алгоритма.* Когда список рассматриваемых подзадач становится пустым, алгоритм завершается. Если список рассматриваемых подзадач непустой, то алгоритм переходит к следующей итерации, начинающейся с шага 2.

Результатом работы алгоритма является наилучшее найденное к моменту завершения решение. Так как в процессе работы алгоритма не отбрасываются решения, позволяющие улучшить значение функции полезности более чем на ε, то, уменьшая параметр ε, можно получить решение, сколь угодно близкое к оптимальному.

4.2. Динамическое решение

Решение изложенной оптимизационной задачи должно периодически пересчитываться по следующим причинам. Во-первых, канальные условия, сетевой трафик и активность станций меняются со временем. Во-вторых, в ряде случаев невозможно передавать данные одновременно по всем соединениям даже при сниженной мощности передачи. Это приводит к тому, что в оптимальном решении передачи между некоторыми станциями будут запрещены. Чтобы избежать блокировки соединений, необходимо время от времени пересчитывать найденное решение с учетом объема данных, переданного по разным соединениям.

Обозначим через P и R_i суммарную энергию, потраченную с начала эксперимента, и общий объем переданных данных по соединению i с начала эксперимента. Тогда логарифм функции полезности может быть представлен в виде

$$\log \hat{U} = \log U^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} U(R_i)\right) - \log P. \tag{12}$$

В таком случае каждый раз необходимо максимизировать приращение функции полезности, т.е. максимизировать ее производную. Для этого продифференцируем (12), принимая во внимание (2). Для всех $\alpha \geqslant 0$

$$(\log \hat{U})' = \frac{1}{\sum\limits_{i=1}^{N} R_i^{1-\alpha}} \sum_{i=1}^{N} \frac{R_i'}{R_i^{\alpha}} - \frac{P'}{P}. \tag{13}$$

С учетом того что $R_i' = r_i$ и $P' = -w$ необходимо максимизировать функцию

$$(\log \hat{U})' = \frac{1}{\sum\limits_{i=1}^{N} R_i^{1-\alpha}} \sum_{i=1}^{N} \frac{r_i}{R_i^{\alpha}} + \frac{w}{P}. \tag{14}$$

Эта функция также является монотонной функцией от переменных $(\boldsymbol{r}, w)$, поэтому для ее оптимизации используется описанный в подразделе 4.1 алгоритм.

5. Численные результаты

Представим результаты имитационного моделирования по оценке производительности предложенного решения. Рассматривается область пространства размером 100×100 метров, внутри которой равновероятно случайным образом размещаются 10 беспроводных клиентов (получателей). Клиенты обслуживаются точками доступа Wi-Fi (передатчиками), число которых изменяется от 1 до 30. Точки доступа Wi-Fi располагаются внутри данной области, максимизируя минимальное расстояние между любыми двумя точками доступа Wi-Fi и расстояние между точками доступа и границей области.

Для моделирования распространения сигнала используется модель, предложенная в [14]. Согласно этой модели ослабление сигнала на расстоянии d от передатчика рассчитывается следующим образом:

$$d(r) = 40{,}05 + 20\log_{10}(f_c/2{,}4) + 20\log_{10}(\min(r, 10)) + \mathbb{1}_{r>10} \cdot 35\log_{10} 0{,}1r,$$

где $f_c = 5{,}21$ ГГц, $\mathbb{1}_{r>10}$ — функция-индикатор, равная единице, если $r > 10$, и нулю в остальных случаях.

Задача оптимизации, описанная в разделе 3, рассматривается для случая $\alpha = 1$. Как было отмечено, для решения задачи необходимо определить зависимость скорости передачи данных от соотношения SINR сигнал-шум. Данная зависимость для алгоритма выбора скорости передачи данных Minstrel HT была получена с помощью имитационного моделирования в среде NS-3 [15] и может быть аппроксимирована ступенчатой функцией, см. рис. 2.

В данной статье проводится сравнительный анализ эффективности следующих решений.

1. СТД (работа сети согласно стандарту). Моделируется работа сети Wi-Fi по умолчанию без каких-либо изменений или дополнительных настроек. Все станции получают доступ к каналу согласно методу множественного доступа с контролем несущей и избеганием коллизий (CSMA/CA).

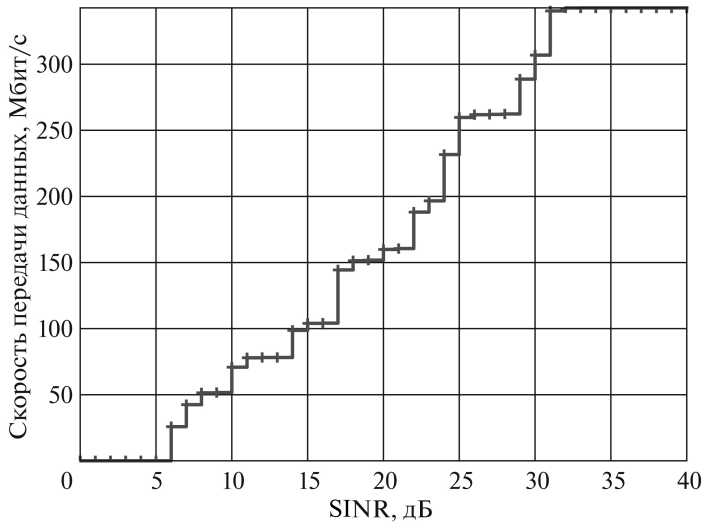


Рис. 2. Зависимость скорости передачи данных от соотношения SINR сигнал-шум.

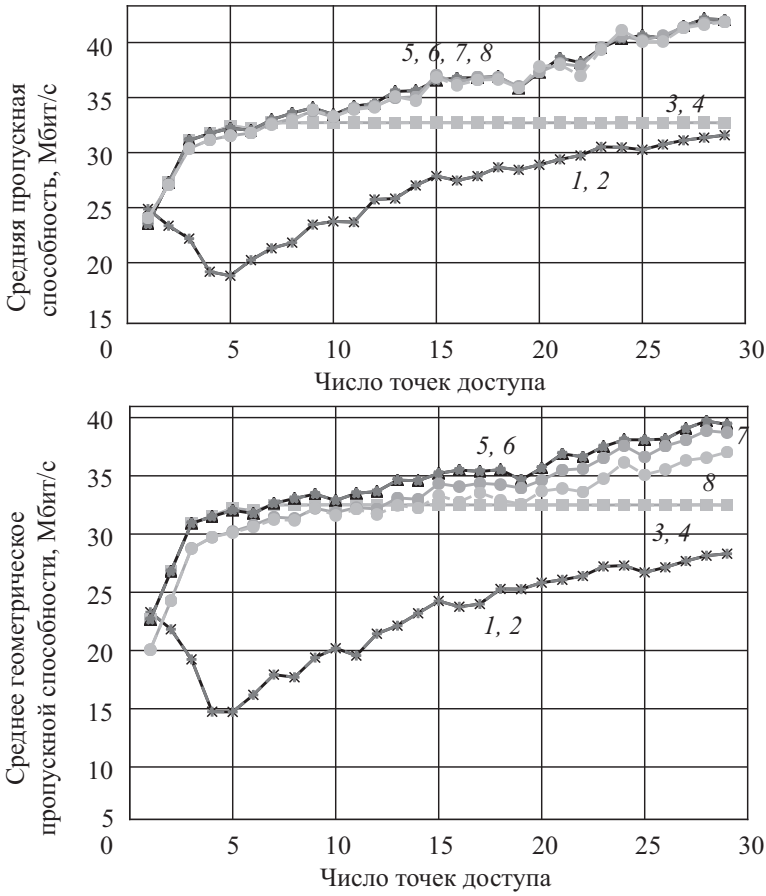


Рис. 3. Пропускная способность сети в зависимости от числа точек доступа: *1* – решение СТД; *2* – решение СТД + ВЫКЛ; *3* – решение РАСП; *4* – решение РАСП + ВЫКЛ; *5* – решение КМ + РАСП; *6* – решение КМ + РАСП + ВЫКЛ; *7* – решение ЭКМ + РАСП; *8* – решение ЭКМ + РАСП + ВЫКЛ.

2. РАСП (передача по расписанию). Передачи различных точек доступа осуществляются согласно расписанию таким образом, чтобы максимизировать среднее геометрическое пропускных способностей различных точек доступа. Для этого решается оптимизационная задача, изложенная в [12].
3. К.М. + РАСП (контроль мощности и передача по расписанию). В дополнение к составлению расписания передач также осуществляется настройка мощности передатчиков, как описано в [12].
4. Э.К.М. + РАСП (энергоэффективный контроль мощности и передача по расписанию) Решение, предложенное в данной статье.

Так как в режиме ожидания точки доступа потребляют значительное количество электроэнергии, то для каждого из решений рассмотрены два сценария: в первом сценарии все точки доступа включены, в то время как во втором сценарии точки доступа, к которым не подключен ни один клиент, выключены (“ВЫКЛ”).

Основные параметры имитационного моделирования

Параметр	Значение
Высота антенны на точке доступа, м	3
Высота антенны на клиенте-получателе, м	1
Максимальная мощность передачи $\hat{p}_i$ $\forall i$, мВт	40
Плотность мощности шума, дБм/Гц	-174
Ширина канала, МГц	80
Мощность шума усилителя, дБ	7
Чувствительность приемника, дБм	-96
Алгоритм выбора скорости передачи	Minstrel HT
Энергопотребление точки доступа в режиме ожидания p_c, мВт	820
КПД усилителя, $1/\phi_i$ $\forall i$	0,1

На рис. 3 показана зависимость средней пропускной способности $\frac{1}{N}\sum\limits_{i=1}^{N} R_i$ и среднего геометрического пропускной способности $\left(\prod\limits_{i=1}^{N} R_i\right)^{1/N}$ от числа обслуживающих точек доступа. При наличии только одной точки доступа результаты всех решений, кроме предложенного в данной статье, практически совпадают. Это объясняется тем, что при присутствии лишь одной точки доступа не имеет значения, как организован доступ к среде: данная точка доступа в любом случае получает все канальные ресурсы. Однако при попытке оптимизировать энергоэффективность, на что нацелено предложенное в данной статье решение, мощность передачи может быть снижена в целях экономии энергии, что приводит к некоторому снижению пропускной способности.

При использовании стандартных настроек работы сети Wi-Fi увеличение числа обслуживающих точек доступа не приносит дополнительной выгоды и даже наоборот может приводить к снижению пропускной способности. Это объясняется появлением скрытых станций и, как следствие, коллизиями пакетов, чего можно избежать при помощи координации работы соседних точек доступа, как предложено в данной статье. При увеличении числа обслуживающих точек доступа пропускная способность растет из-за уменьшения среднего расстояния между передатчиком и приемником, влекущего за собой использование более быстрых сигнально-кодовых конструкций и увеличение скорости передачи данных. Несмотря на то что использование стандартных настроек работы сети Wi-Fi позволяет достичь довольно высокой средней пропускной способности, оно не обеспечивает равномерного распределения канального ресурса между различными клиентами, что выражается в низком значении среднего геометрического пропускной способности.

При использовании расписания передач наблюдается увеличение пропускной способности сети при небольшом числе обслуживающих точек доступа, что объясняется уменьшением среднего расстояния между передатчиком и приемником и, следовательно, использованием более быстрых сигнально-кодовых конструкций. После этого пропускная способность остается постоянной и не изменяется с ростом числа точек доступа, так как в каждый момент времени происходит только одна передача и при этом используется самая быстрая сигнально-кодовая конструкция, а коллизии отсутствуют. Дальней-

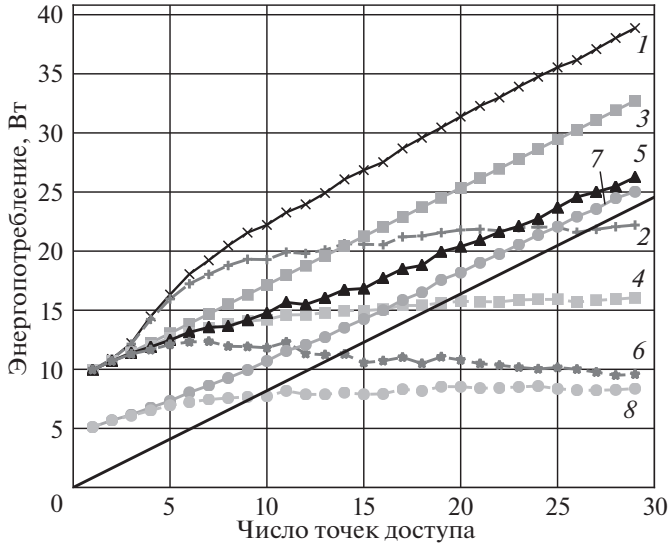


Рис. 4. Энергопотребление сети: *1* — решение СТД; *2* — решение СТД + ВЫКЛ; *3* — решение РАСП; *4* — решение РАСП + ВЫКЛ; *5* — решение КМ + РАСП; *6* — решение КМ + РАСП + ВЫКЛ; *7* — решение ЭКМ + РАСП; *8* — решение ЭКМ + РАСП + ВЫКЛ.

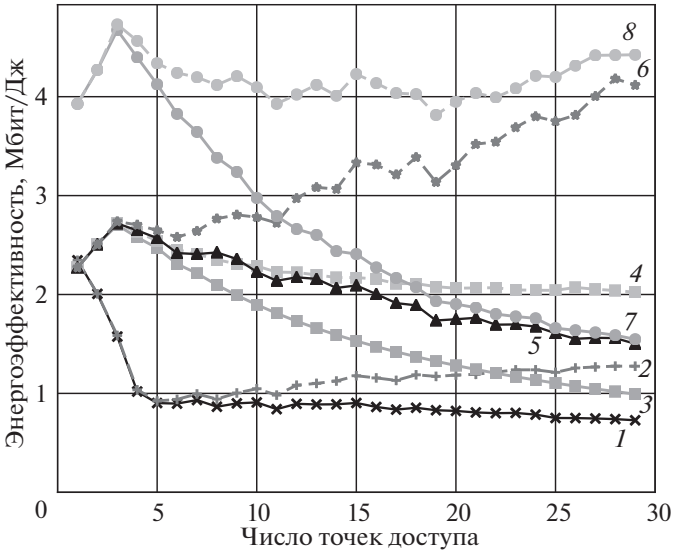
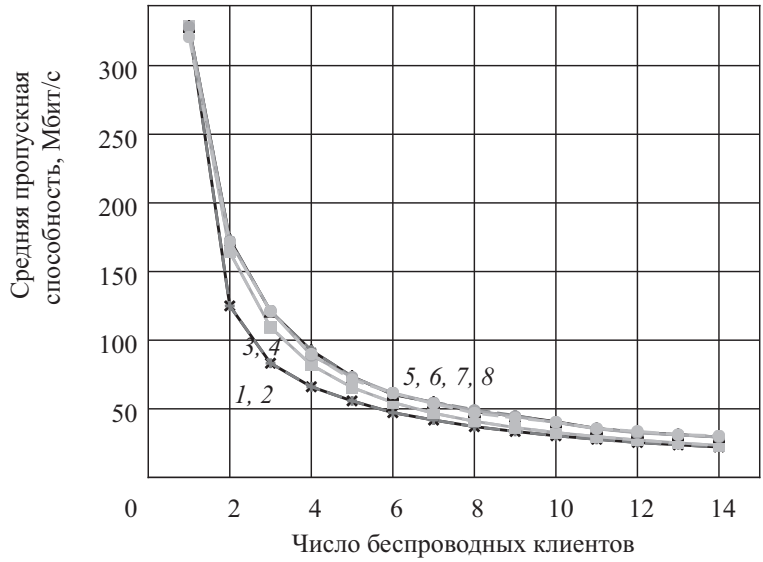


Рис. 5. Энергоэффективность сети: *1* — решение СТД; *2* — решение СТД + ВЫКЛ; *3* — решение РАСП; *4* — решение РАСП + ВЫКЛ; *5* — решение КМ + РАСП; *6* — решение КМ + РАСП + ВЫКЛ; *7* — решение ЭКМ + РАСП; *8* — решение ЭКМ + РАСП + ВЫКЛ.

шее уменьшение среднего расстояния между передатчиком и приемником не приносит дополнительной пользы.

Кривые, соответствующие совместному контролю мощности и расписанию передач, а также энергоэффективному контролю мощности и расписанию пе-



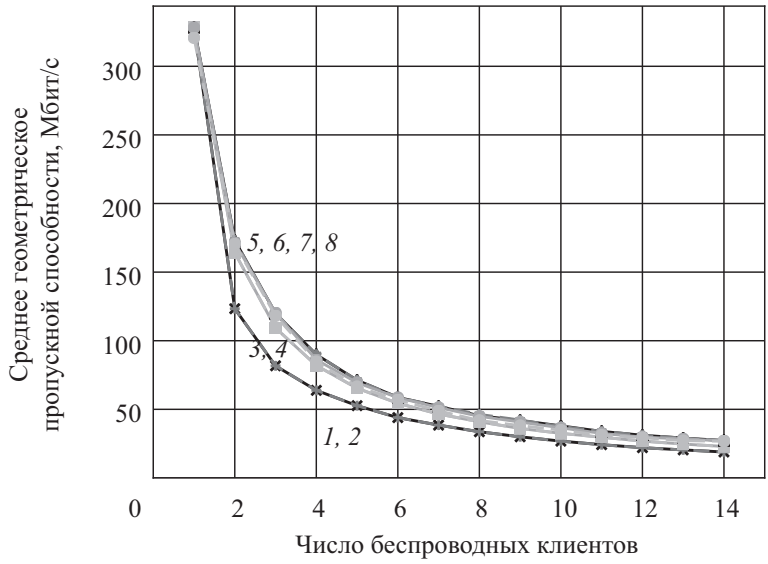


Рис. 6. Пропускная способность сети: *1* — решение СТД; *2* — решение СТД + ВЫКЛ; *3* — решение РАСП; *4* — решение РАСП + ВЫКЛ; *5* — решение КМ + РАСП; *6* — решение КМ + РАСП + ВЫКЛ; *7* — решение ЭКМ + РАСП; *8* — решение ЭКМ + РАСП + ВЫКЛ.

редач, располагаются довольно близко друг другу. Для обоих данных решений динамическая настройка мощности передачи позволяет получить выигрыш в значении среднего геометрического пропускной способности до двух раз. Главное различие данных решений заключается в их энергоэффективности. На рис. 4 и 5 показаны энергопотребление всей сети в ваттах и энергоэффективность, измеренная в Мбит/Дж. Энергоэффективное управление мощностью позволяет существенно снизить энергопотребление, обеспечивая

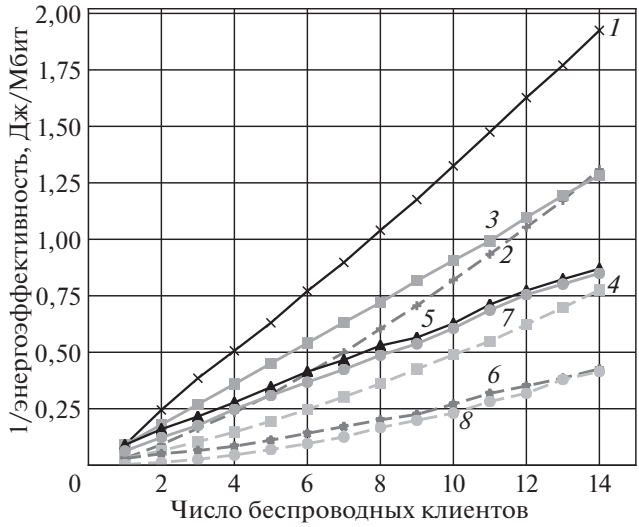


Рис. 7. Величина, обратная энергоэффективности сети: *1* — решение СТД; *2* — решение СТД + ВЫКЛ; *3* — решение РАСП; *4* — решение РАСП + ВЫКЛ; *5* — решение КМ + РАСП; *6* — решение КМ + РАСП + ВЫКЛ; *7* — решение ЭКМ + РАСП; *8* — решение ЭКМ + РАСП + ВЫКЛ.

при этом практически такую же пропускную способность, как и разработанное ранее решение К.М. + РАСП.

На рис. 4 также присутствует сплошная линия, которая показывает минимально возможное энергопотребление сети, обусловленное только компонентой p_c — потребляемой мощностью в режиме ожидания. Результаты применения энергоэффективного контроля мощности и расписания передач оказываются довольно близкими к данному нижнему пределу, который может быть достигнут только в том случае, если сеть не передает данные.

Выключение точек доступа, к которым не подключена ни одна станция (кривые с меткой "ВЫКЛ"), позволяет значительно понизить энергопотребление в сетях с большим числом точек доступа (см. рис. 4) за счет уменьшения слагаемого Mp_c в (3). При этом, как видно из рис. 3, незначительно уменьшается среднее геометрическое пропускной способности, так как слагаемое $\sum_{i=1}^{N} \phi_i p_i$, связанное с мощностью передачи, начинает оказывать более существенное влияние на значение функции полезности, что с точки зрения энергоэффективности делает более выгодным выбор меньшей мощности передачи. Заметим, что в случае выключения неиспользуемых точек доступа применение разработанного решения позволяет значительно повысить энергоэффективность по сравнению с использованием стандартных настроек при всех рассмотренных значениях числа точек доступа, см. рис. 5.

На рис. 6 и 7 показаны зависимости пропускной способности и энергоэффективности сети с 25 точками доступа при изменении числа клиентских устройств. Для всех рассматриваемых решений пропускная способность сети оказывается выше, чем в случае отсутствия централизованного управления.

Для удобства представления результатов на рис. 7 показана величина, обратная энергоэффективности. Как видно из данного рисунка, преимущество использования централизованного управления сетью растет с увеличением числа беспроводных клиентов в сети.

6. Заключение

В данной статье предложен новый алгоритм управления передачей в плотных сетях Wi-Fi с целью повышения энергоэффективности. В основе алгоритма лежит решение оптимизационной задачи по максимизации энергоэффективности с применением метода ветвей и границ. Для оценки эффективности разработанного решения предложенный алгоритм был реализован в среде имитационного моделирования NS-3. Результаты моделирования показали повышение энергоэффективности до 75 % по сравнению с разработанным ранее алгоритмом, при этом обеспечивались схожие среднее геометрическое и среднее арифметическое пропускной способности. В качестве дальнейшего направления исследований планируется оценить эффективность разработанных решений в сценариях с динамически меняющимся трафиком.

СПИСОК ЛИТЕРАТУРЫ

1. *Barnett T., Jain S., Andra U., Khurana T.* Cisco Visual Networking Index (VNI), Complete Forecast Update, 2017–2022 // Americas/EMEAR Cisco Knowledge Network (CKN) Presentation. December, 2018.
2. *Khorov E., Kiryanov A., Lyakhov A., Bianchi G.* A Tutorial on IEEE 802.11ax High Efficiency WLANs // IEEE Communications Surveys & Tutorials. 2019. V. 21. No. 1. P. 197–216. Firstquarter.
3. *Khorov E., Ivanov A., Lyakhov A., Akyildiz I.F.* Cloud Control to Optimize Real-Time Video Transmission in Dense IEEE 802.11aa/ax Networks // Proc. IEEE 15th Int. Conf. on Mobile Ad Hoc and Sensor Systems (MASS). 2018.
4. *Buzzi S., Chih-Lin I., Klein T.E., Poor H.V., Yang C., Zappone A.* A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead // IEEE J-SAC. 2016. V. 34. No. 4. P. 697–709.
5. *Zorzi M., Rao R.R.* Energy-Constrained Error Control for Wireless Channels // IEEE Pers. Comm. Mag. 1997. V. 4. No. 6. P. 27–33.
6. *Li G.Y., Xu Z., Xiong C., Yang C., Zhang S., Chen Y., Xu S.* Energy-Efficient Wireless Communications: Tutorial, Survey, and Open Issues // IEEE Wirel. Commun. 2011. V. 18. No. 6. P. 28–35.
7. *Miao G., Himayat N., Li Y.G., Koc A.T., Talwar S.* Interference-Aware Energy-Efficient Power Optimization // Proc. 2009 IEEE ICC. 2009 P. 1–5.
8. *Venturino L., Zappone A., Risi C., Buzzi S.* Energy-Efficient Scheduling and Power Allocation in Downlink Ofdma Networks with Base Station Coordination // IEEE T. Wirel. Commun. 2015. V. 14. No. 1. P. 1–14.
9. *Zappone A., Jorswieck E.* Energy Efficiency in Wireless Networks via Fractional Programming Theory // Found. Trends Commun. Inform. Theory. 2015. V. 11. No. 3–4. P. 185–396.
10. *Tuy H.* Convex Analysis and Global Optimization. Germany. Springer, 2016.

11. *Zappone A., Bjornson E., Sanguinetti L., Jorswieck E.* Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks // IEEE T. Signal Proces. 2017. V. 65. No. 11. P. 2844–2859.
12. *Кирьянов А.Г., Кротов А.В., Ляхов А.И., Хоров Е.М.* Алгоритм динамического управления мощностью и составления расписания передач в инфраструктурных сетях IEEE 802.11 ax // Информационные процессы. 2019. Т. 19. № 1. С. 16–32.
13. *Стефанюк В.Л., Цетлин М.Л.* О регулировке мощности в коллективе радиостанций // Пробл. передачи информации. 1967. Т. 3. № 4. С. 49–57.
14. *Merlin S.* TGax Simulation Scenarios. [Online].
    https://mentor.ieee.org/802.11/dcn/14/11-14-0980-16-00axsimulation-scenarios.docx
15. The NS-3 Network Simulator. [Online]. http://www.nsnam.org/

Статья представлена к публикации членом редколлегии А.И. Ляховым.

Поступила в редакцию 26.06.2019

После доработки 17.07.2019

Принята к публикации 18.07.2019

© 2020 г. Д.С. ОСИПОВ, канд. техн. наук (d_osipov@iitp.ru)
(Институт проблем передачи информации им. А.А. Харкевича РАН, Москва;
Национальный исследовательский университет
“Высшая школа экономики”, Москва)

ВЕРХНЯЯ ГРАНИЦА ВЕРОЯТНОСТИ ОШИБКИ В СИСТЕМАХ СВЯЗИ, ИСПОЛЬЗУЮЩИХ ОДНОПОЛЬЗОВАТЕЛЬСКИЙ ПРИЕМ НА ОСНОВЕ ПОРЯДКОВЫХ СТАТИСТИК[1]

Рассматривается модель канала, описывающая передачу информации в системах связи, использующих однопользовательский приемник на основе порядковых статистик. Рассматривается передача информации по такому каналу с использованием линейного блокового кода. Целью работы является отыскание верхней границы вероятности ошибки для случая, когда для декодирования используется алгоритм, основанный на полном переборе множества кодовых слов и заданном критерии декодирования.

Ключевые слова: верхняя граница, вероятность ошибки, порядковые статистики, однопользовательский прием, непересекающиеся подканалы, недвоичные линейные коды.

DOI: 10.31857/S0005231020010092

1. Введение

В современных системах связи и управления применяются техники приема и методы теории помехоустойчивого кодирования, позволяющие обеспечить выполнение требований к качеству (т.е. надежности) и скорости связи для весьма широкого круга приложений, в которых возникает необходимость в использовании таких систем. Вместе с тем в ряде случаев традиционные методы приема (например, “жесткий” посимвольный прием или традиционные методы “мягкого” приема, т.е. вычисления оценок надежности для каждого из символов) оказываются неэффективными (например, если передача происходит в условиях воздействия аддитивной помехи, мощность которой существенно выше, чем мощность полезного сигнала, или если параметры канала не известны приемнику или оценки этих параметров существенно отличаются от истинных значений). Для такого рода случаев необходимы специализированные методы приема, устойчивые к искажениям принятого сигнала и ошибкам в определении параметров распределений решающих статистик. Примерами таких методов являются методы приема, основанные на использовании порядковых статистик [1–3]. Настоящая работа является результатом исследования, первые результаты которого были доложены автором в рамках международного симпозиума по проблеме избыточности в информационных

[1] Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150).

системах в Санкт-Петербурге. Ниже будет рассмотрен метод приема, являющийся обобщением метода, предложенного автором в [1]. Для описания систем связи, использующих такой метод, в [4] была предложена модель канала. Эта модель будет описана в разделе 2, где будет описана и схема системы связи, соответствующей такой модели канала, и связь этой модели с моделями каналов, рассмотренными другими авторами. В разделе 3 будут описаны схема кодирования и правило декодирования. В разделе 4 получен вывод верхней границы на вероятность ошибки (на кодовое слово) для случая, в котором для передачи по каналу рассматриваемого типа используется линейный код с известным спектром и для декодирования используется правило декодирования, введенное в разделе 3. Наконец, в разделе 5 приведены результаты имитационного моделирования, свидетельствующие об эффективности используемого подхода и корректности полученных результатов.

2. Описание модели канала

Пусть q и α — натуральные числа такие, что $\alpha \geqslant 2$, $q > \alpha$. Введем следующие обозначения:

$$\mathbb{B}_q^x = \{\mathbf{b} = (b_1, \ldots, b_q)^\intercal : \forall\ i \in \{1:q\}\, b_i \in \{0,1\}\,, w_H(\mathbf{b}) = x\} \tag{1}$$

— множество всех двоичных векторов-столбцов веса x (здесь $w_H(\mathbf{z})$ — вес Хэмминга вектора $\mathbf{z}$).

Кроме того, для любого вектора-столбца $\mathbf{z}$ такого, что $w_H(\mathbf{z}) < \alpha$, определим множества

$$\begin{gathered}\mathbb{S}^1(\mathbf{z},\alpha) = \left\{\mathbf{s} : \mathbf{s} \in \mathbb{B}_q^\alpha, \mathbf{s} \wedge \mathbf{z} = \mathbf{z}\right\},\\ \mathbb{S}^0(\mathbf{z},\alpha) = \left\{\mathbf{s} : \mathbf{s} \in \mathbb{B}_q^\alpha, \mathbf{s} \wedge \mathbf{z} \neq \mathbf{z}\right\}\end{gathered} \tag{2}$$

— множество двоичных векторов-столбцов веса α, покрывающих вектор-столбец $\mathbf{z}$, и множество двоичных векторов-столбцов веса α, не покрывающих вектор-столбец $\mathbf{z}$ (здесь $\wedge$ — символ поэлементной конъюнкции). Рассмотрим векторный канал, входом которого является вектор-столбец $\mathbf{x}$, а выходом вектор-столбец $\mathbf{y}$. Канал задается условиями

$$\begin{gathered}\forall\ \alpha \geqslant 2; \quad q > \alpha; \quad \mathbf{x} \in \mathbb{B}_q^1, \quad \mathbf{y} \in \mathbb{B}_q^\alpha, \quad \frac{1}{2} < p < 1,\\ p(\mathbf{x}) \triangleq \sum_{\mathbf{y} \in \mathbb{S}^1(\mathbf{x},\alpha)} p(\mathbf{y} \mid \mathbf{x}) = p,\end{gathered} \tag{3}$$

т.е. входом канала всегда является двоичный вектор веса 1, а выходом двоичный вектор веса α ($\alpha \geqslant 2$), и каким бы ни был входной вектор, вектор на выходе покрывает его с вероятностью p ($\frac{1}{2} < p < 1$). Кроме того, потребуем выполнения дополнительных условий

$$\begin{gathered}\forall \mathbf{x} \in \mathbb{B}_q^1,\ \mathbf{y}_a \in \mathbb{S}^1(\mathbf{x},\alpha),\ \mathbf{y}_b \in \mathbb{S}^1(\mathbf{x},\alpha),\ a \neq b : p(\mathbf{y}_a \mid \mathbf{x}) = p(\mathbf{y}_b \mid \mathbf{x}) = p_1,\\ \forall \mathbf{x} \in \mathbb{B}_q^1,\ \mathbf{y}_n \in \mathbb{S}^0(\mathbf{x},\alpha),\ \mathbf{y}_l \in \mathbb{S}^0(\mathbf{x},\alpha),\ n \neq l : p(\mathbf{y}_n \mid \mathbf{x}) = p(\mathbf{y}_l \mid \mathbf{x}) = p_0.\end{gathered} \tag{4}$$

Выражения (3) могут быть интерпретированы следующим образом. Представим себе систему связи, использующую канал, состоящий из q непересекающихся подканалов. Будем считать, что однократное использование канала соответствует передаче одного q-ичного символа, причем каждому символу взаимно однозначно ставится в соответствие некоторый вектор веса 1 и длины q (такое отображение рассматривалось во многих работах, в частности, в [5]) и передача ведется по подканалу, соответствующему позиции ненулевого элемента в векторе, т.е. используется позиционная модуляция. Кроме полезного сигнала, на выходы подканалов могут влиять сигналы и аддитивные помехи различного рода (сигналы от других пользователей, преднамеренные помехи, фоновые шумы и т.п.). Приемник измеряет значения некоторого заранее определенного параметра (например, мощности) сигнала на выходе каждого из подканалов (будем называть такие величины "решающими статистиками") и выбирает α номеров подканалов, которым соответствуют "лучшие" (в смысле некоторого критерия) значения решающих статистик (например, α подканалов, у которых мощность сигнала на выходе максимальна). Вероятность p в этом случае интерпретируется как вероятность того, что подканал, по которому передавался полезный сигнал, попал в список на выходе приемника. Условия (4) выполняются в случае, если аддитивные помехи передаются по тем или иным подканалам случайно и равновероятно. Для обеспечения выполнения этого условия достаточно использовать псевдослучайную перестановку на входе при передаче каждого символа (при этом перестановка должна выбираться равновероятно из всего множества возможных перестановок, заново при передаче каждого символа) и обратную перестановку на выходе.

Как уже было сказано, рассмотренная выше модель канала описывает широкий класс реальных систем связи, использующих позиционную модуляцию и передачу по физически разнесенным каналам, в частности многопользовательские каналы с однопользовательским приемом и каналы с аддитивными помехами различного рода. Заметим, что описанная выше модель отличается от дизъюнктивных векторных моделей, описывающих многопользовательские каналы [6] и каналы с аддитивными помехами [7], и модифицированной модели канала, предложенной в [8], так как вес вектора на выходе канала канала описываемого типа фиксирован. С другой стороны, как видно из (3) и (4), исследуемый канал принадлежит к классу дискретных симметричных (в смысле определения [9]) каналов без памяти. Переходные вероятности, характеризующие этот канал, равны

$$p_1 = \frac{p}{\sigma_1}, \quad p_0 = \frac{1-p}{\sigma_0}, \tag{5}$$

где

$$\sigma_1 = |\mathbb{S}^1(\mathbf{x}, \alpha)| = \binom{q-1}{\alpha-1}, \quad \sigma_0 = |\mathbb{S}^0(\mathbf{x}, \alpha)| = \binom{q-1}{\alpha}. \tag{6}$$

Отличие модели канала рассматриваемого типа от обычных моделей дискретных симметричных каналов без памяти состоит в том, что в рассмотренной модели никакой символ на выходе нельзя однозначно отождествить

с каким-либо символом на входе, т.е. не существует понятия “ошибочного” (и, соответственно, “правильного”) приема одиночного символа. Сказанное, в частности, означает, что для рассматриваемого случая невозможно использовать классические границы, такие как [10, 11].

3. Кодирование и декодирование

Опишем теперь схему кодирования и декодирования для канала рассматриваемого типа. Будем считать, что информация кодируется линейным кодом $C(N, K, d)$ над полем $GF(q)$. Кроме того, будем считать, что известен спектр кода, т.е. для любого веса w $(d \leqslant w \leqslant N)$ известно A_w — число слов данного веса в коде C:

$$\forall\, w : d \leqslant w \leqslant N \;\; A_w = |\mathbf{v} \in C : w_H(\mathbf{v}) = w|.$$

Передача t-го символа v_t^m кодового слова $\mathbf{v}^m$ сводится к передаче по каналу соответствующего этому символу двоичного вектора $\mathbf{x}_t^m$. Таким образом, передача кодового слова $\mathbf{v}^m = [v_1^m, \ldots, v_N^m]$ в рассматриваемом случае сводится к передаче соответствующей этому кодовому слову матрицы $X^m = [\mathbf{x}_1^m, \ldots, \mathbf{x}_N^m]$. Поэтому в дальнейшем будем говорить о “передаче кодового слова” и “передаче матрицы, соответствующей кодовому слову”, подразумевая, что эти выражения синонимичны.

Так как выход канала всегда представляет собой вектор веса α, каждой матрице $X^m = [\mathbf{x}_1^m, \ldots, \mathbf{x}_N^m]$ (и, соответственно, каждому кодовому слову $\mathbf{v}^m$) на входе канала на выходе соответствует матрица $Y^j = [\mathbf{y}_1^j, \ldots, \mathbf{y}_N^j]$, такая что $Y^j \in \mathbb{Y}$, где

$$\mathbb{Y} = \left\{ Y : Y = [\mathbf{y}_1, \ldots, \mathbf{y}_N], \forall\, t :\; t = 1, \ldots, N \;\; \mathbf{y}_t \in \mathbb{B}_q^\alpha \right\} \tag{7}$$

— множество всех матриц, которые могут возникнуть на выходе канала. Пусть определена $\Theta\left(Y^j \mid X^l\right)$ — функция достоверности гипотезы о том, что в результате передачи матрицы X^l была принята матрица Y^j. Тогда декодирование сводится к поиску кодового слова $\mathbf{v}^t$, соответствующего матрице X^t такой, что выполняется

$$\forall\, l = 1, \ldots, M \; l \neq t \quad \Theta\Big(Y^j \mid X^t\Big) \geqslant \Theta\left(Y^j \mid X^l\right), \tag{8}$$

поэтому ниже для краткости будем именовать эту функцию функцией декодирования. Декодирование, таким образом, сводится к поиску матрицы, для которой выполняется (8). То, что неравенство в (8) нестрогое, означает, что при определенном выборе функции декодирования максимальное значение функции может достигаться на различных матрицах X^l. В дальнейшем будем считать, что в таких случаях матрица, соответствующая декодированному слову, выбирается случайно из всего множества матриц, для которых выполняется (8).

Так как канал, задаваемый (3), (4), является каналом без памяти и для передачи используется блоковый код, в данной работе ограничимся рассмотрением случая, в котором функция декодирования имеет вид

$$\Theta\left(Y^j|X^l\right) = \prod_{t=1}^{N} \theta\left(\mathbf{y}_t^j, \mathbf{x}_t^l\right), \tag{9}$$

где $X^l = [\mathbf{x}_1^l, \ldots, \mathbf{x}_N^l]$ и $Y^j = [\mathbf{y}_1^j, \ldots, \mathbf{y}_N^j]$. Выберем

$$\theta\left(\mathbf{x}, \mathbf{y}\right) = \begin{cases} \eta\theta & \mathbf{y} \in \mathbb{S}^1\left(\mathbf{x}, \alpha\right), \\ \theta & \mathbf{y} \in \mathbb{S}^0\left(\mathbf{x}, \alpha\right), \end{cases} \tag{10}$$

где $\eta > 1, \theta > 0$.

Такой выбор функции декодирования может быть интерпретирован следующим образом: считается, что гипотеза о том, что вектор-столбец, соответствующий переданному символу, покрывается соответствующим вектором-столбцом в принятой матрице (т.е. подканал, по которому передавался сигнал, попал в список α "лучших"), имеет в η раз более высокую достоверность, чем конкурирующая гипотеза. Заметим, что при выборе параметров функции декодирования в форме $\theta = p_0 = \frac{1-p}{\sigma 0}$, $\eta = \frac{p_1}{p_0} = \frac{p}{1-p}\frac{\sigma_0}{\sigma_1}$ описанный выше декодер эквивалентен декодеру по максимуму правдоподобия. В реальных системах связи вероятность p не может быть точно оценена приемником, поэтому важно отметить, что в настоящей работе будет получена граница для произвольного $\eta > 1$. Ниже будет показано, что при использовании описанного метода приема вероятность ошибки практически не зависит от величины параметра η.

4. Верхняя граница вероятности ошибки декодирования

Целью является получение верхней границы на вероятность ошибочного декодирования для рассматриваемого случая, а именно для случая, в котором для передачи по каналу, заданному (3), (4), используется линейный код с известным спектром и для декодирования используется описанное выше правило. Задача отыскания границ для кодированной передачи с использованием недвоичного кода с известным спектром изучена относительно плохо (по сравнению, например, с двоичным случаем, случаем случайных кодов и кодов с известной композицией) и, как правило, требует исследования свойств конкретного канала, причем зачастую полученные границы определяются не для конкретных кодов, а для ансамблей кодов и справедливы лишь для каналов, удовлетворяющих специфическим дополнительными условиям [12, 13]. Ниже будет приведен вывод верхней границы для канала, заданного (3), (4) и использующего конкретный недвоичный код с известным спектром, а также описанный в настоящем разделе критерий декодирования. Для вывода будут использованы классические подходы, предложенные Галлагером [11] и Фано [14], и граница Думана–Салехи [15], что позволит аналитически решить задачу оптимизации предлагаемой границы для канала рассмотренного типа.

Для того чтобы получить верхнюю границу на вероятность ошибочного декодирования, воспользуемся техникой, восходящей к работам Галлагера [11] и Фано [14]: разобьем все множество матриц, которые могут быть приняты из канала, на два непересекающихся подмножества. Первое подмножество, которое будем условно именовать "плохим" (и обозначать как $\mathbb{Y}_B$), будет включать принятые матрицы, для которых вероятность ошибки декодера велика; второе, условно именуемое "хорошим" (и обозначаемое как $\mathbb{Y}_G$), состоит из всех остальных матриц, которые могут появиться на выходе канала. Пусть передана матрица X^m. Тогда, учитывая, что $\mathbb{Y}_B \cap \mathbb{Y}_G = \emptyset$, можно утверждать, что

$$p\left(err\,|X^m\right) = p\left(err, Y \in \mathbb{Y}_B\,|X^m\right) + p\left(err, Y \in \mathbb{Y}_G\,|X^m\right), \tag{11}$$

где $p(err\big|X^m)$ – условная вероятность ошибки декодера, $p(err, Y \in \mathbb{Y}_B\big|X^m)$ – условная вероятность того, что на выходе канала матрица из "плохого" подмножества и произошла ошибка, $p(err, Y \in \mathbb{Y}_G\big|X^m)$ – условная вероятность того, что на выходе канала матрица из "хорошего" подмножества и произошла ошибка. Предполагается, что вероятность ошибки для матриц из "плохого" подмножества высока, можно записать

$$p\left(err\,|X^m\right) \leqslant p\left(Y \in \mathbb{Y}_b\,|X^m\right) + p\left(err, Y \in \mathbb{Y}_G\,|X^m\right). \tag{12}$$

Оценим первое слагаемое в правой части (12). Рассмотрим матрицы $X^0 = [\mathbf{x}_1^0, \ldots, \mathbf{x}_N^0]$ (соответствующую переданному кодовому слову, которое без ограничения общности будем считать нулевым) и $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ (соответствующую принятой последовательности). Кроме того, рассмотрим множество

$$\mathbb{I}_c\left(Y, X^0\right) \triangleq \left\{t : t \in \{T, \ldots, N\}, \mathbf{y}_t \in \mathbb{S}^1\left(\mathbf{x}_t^0, \alpha\right)\right\} \tag{13}$$

— множество номеров столбцов матрицы Y, которые покрывают соответствующие столбцы матрицы X^0 (здесь $\mathbb{S}^1\left(\mathbf{x}_t^0, \alpha\right)$ — множество векторов-столбцов веса α, покрывающих t-й вектор-столбец матрицы X^0). Введем обозначение

$$i = \left|\mathbb{I}_c\left(Y, X^0\right)\right|. \tag{14}$$

Заметим, что, так как функция декодирования задана в форме (9), (10), вероятность ошибочного декодирования убывает с ростом i. Определим множество $\mathbb{Y}_B$ следующим образом:

$$\mathbb{Y}_B = \left\{Y : Y \in \mathbb{Y}, \left|\mathbb{I}_c\left(Y, X^0\right)\right| < T\right\}, \tag{15}$$

где T — параметр, зависящий от p и удовлетворяющий условиям

$$T \in \mathbb{N}, \qquad 1 \leqslant T \leqslant N - K. \tag{16}$$

При таком определении первое из двух слагаемых в правой части (12) может быть вычислено по формуле

$$p\left(Y \in \mathbb{Y}_b \,|X^m\right) = \sum_{i=0}^{T-1} \binom{N}{i} p^i (1-p)^{N-i}. \tag{17}$$

Для оценки второго слагаемого используется классический подход, который применялся во многих работах (в частности, в [11, 15]). Суть этого подхода в следующем: без ограничения общности будем считать, что по каналу было передано нулевое кодовое слово (или, точнее, матрица, соответствующая нулевому кодовому слову). Разобьем используемый код на подкоды, каждый из которых будет включать нулевое кодовое слово и все слова некоторого веса w. Обозначим множество матриц, соответствующих кодовым словам такого подкода, через $\mathbb{S}_w$, а все множество матриц, соответствующих различным кодовым словам кода C, через $\mathbb{S}_C$. В силу границы объединения верна оценка

$$\begin{gathered} p\left(err, Y \in \mathbb{Y}_G \,|X^m\right) = \\ = P\left(X^l = \arg \max_{t: X^t \in \mathbb{S}_C} \left(\Theta\left(Y \mid X^t\right)\right) \neq X^0, Y \in \mathbb{Y}_G \mid X^0\right) \leqslant \\ \leqslant \sum_{w=d}^{N} P\left(X^l = \arg \max_{t: X^t \in \mathbb{S}_w} \left(\Theta\left(Y \mid X^t\right)\right) \neq X^0, Y \in \mathbb{Y}_G \mid X^0\right). \end{gathered} \tag{18}$$

Для получения аналитического выражения используем границу Думана – Салехи [15]. Для каждого из слагаемых в правой части (18) получим

$$\begin{gathered} P\left(X^l = \arg \max_{t: X^t \in \mathbb{S}_w} \left(\Theta\left(Y \mid X^t\right)\right) \neq X^0, Y \in \mathbb{Y}_G \mid X^0\right) \leqslant \\ \leqslant \left(\sum_{\substack{l \in \mathbb{S}_w \\ l \neq 0}} \sum_{Y \in \mathbb{Y}_G} \left(p_N(Y \,|X^0)\right)^{\frac{1}{\rho}} \left(\psi_N^0\left(Y\right)\right)^{1-\frac{1}{\rho}} \left(\frac{\Theta(Y \,|X^l)}{\Theta(Y \,|X^0)}\right)^s \right)^{\rho}, \end{gathered} \tag{19}$$

где ρ и s — параметры, которые будут выбираться с учетом ограничений

$$1 \geqslant \rho > 0, \quad s > 0 \tag{20}$$

таким образом, чтобы минимизировать оценку (19), а $\psi_N^m(Y)$ — функция перекоса, которая зависит от принятой матрицы Y и (в общем случае) от матрицы X^m, соответствующей переданному слову, и удовлетворяет условиям

$$\forall\, Y \in \mathbb{Y} \quad \psi_N^m(Y) > 0, \quad \sum_{Y \in \mathbb{Y}_G} \psi_N^m(Y) = 1. \tag{21}$$

Параметры ρ и s имеют тот же смысл, что и соответствующие параметры в границе Галлагера [11], и потому также должны выбираться таким образом, чтобы минимизировать правую часть (19). Функция перекоса также должна выбираться таким образом, чтобы минимизировать правую часть (19).

Остальная часть этого раздела будет посвящена именно оптимальному выбору функции перекоса, учитывающему специфику модели канала (3), (4).

В силу требований, предъявляемых к функции $\psi_N^0(Y)$, эта функция должна зависеть только от Y и, возможно, от X^0, поэтому естественно потребовать, чтобы значение этой функции для каждой матрицы Y зависело от числа столбцов в матрице X^0, которые покрывает матрица Y^j. В дальнейшем будем полагать, что функция $\psi_N^0(Y)$ имеет вид

$$\forall\, Y = [\mathbf{y}_1, \ldots, \mathbf{y}_N] : i = \left|\left\{t : \mathbf{y}_t \in \mathbb{S}^1\left(\mathbf{x}_t^0, \alpha\right)\right\}\right| \quad \psi_N^0(Y) = \Psi_i, \tag{22}$$

где Ψ_i — переменные, оптимальные значения которых будут найдены ниже. Введем обозначение

$$\Omega\left(Y, X^0, X^l\right) = \left(\frac{\Theta\left(Y \,|X^l\right)}{\Theta\left(Y^j \,|X^0\right)}\right) = \prod_{j=1}^{N} \omega\left(\mathbf{x}_j^0, \mathbf{x}_j^l, \mathbf{y}_j\right), \tag{23}$$

где $\omega\left(\mathbf{x}_j^0, \mathbf{x}_j^l, \mathbf{y}_j\right)$ имеет вид

$$\omega\left(\mathbf{x}_j^0, \mathbf{x}_j^l, \mathbf{y}_j\right) = \frac{\theta\left(\mathbf{x}_j^l, \mathbf{y}_j\right)}{\theta\left(\mathbf{x}_j^0, \mathbf{y}_j\right)}. \tag{24}$$

Заметим, что в силу (10) каждый из сомножителей $\omega\left(\mathbf{x}_j^0, \mathbf{x}_j^l, \mathbf{y}_j\right)$ в правой части (23) отличен от единицы в том и только том случае, если j-й столбец матрицы Y покрывает соответствующий столбец одной из матриц (X^0 или X^l) и не покрывает соответствующий столбец другой матрицы. Введем следующие обозначения: k — число столбцов матрицы Y таких, что эти столбцы Y покрывают соответствующие столбцы матрицы X^l и не покрывают соответствующие столбцы матрицы X^0; g — число столбцов матрицы Y таких, что эти столбцы Y покрывают соответствующие столбцы матрицы X^0 и не покрывают соответствующие столбцы матрицы X^l.

Тогда функция $\Omega\left(Y, X^0, X^l\right)$ принимает значение

$$\Omega\left(Y, X^0, X^l\right) = \eta^{k-g}. \tag{25}$$

Пусть матрицы X^0 и X^l отличаются в w столбцах. Обозначим переменной h число столбцов таких, что матрицы X^0 и X^l различаются в этих столбцах, а матрица Y покрывает в этих столбцах матрицу X^0. Заметим, что верно неравенство $\max(i + w - n, 0) \leqslant h \leqslant \min(i, w)$. Для каждого набора значений четверки параметров i, h, g и k число матриц Y, для которых эти параметры имеют соответствующие значения, равно

$$\begin{gathered}\mathrm{H}(N, w, i, h, g, k) = \\ = \binom{N-w}{i-h} \sigma_1^{i-h} \sigma_0^{N-w-i+h} \binom{w}{h-g, g, k} \upsilon_{\langle 1,1\rangle}^{h-g} \upsilon_{\langle 1,0\rangle}^{g} \upsilon_{\langle 0,1\rangle}^{k} \upsilon_{\langle 0,0\rangle}^{w-h-k},\end{gathered} \tag{26}$$

где σ_1 и σ_0 задаются (6), а $\upsilon_{\langle 1,1\rangle}$, $\upsilon_{\langle 1,0\rangle}$, $\upsilon_{\langle 0,1\rangle}$ и $\upsilon_{\langle 0,0\rangle}$ задаются выражениями

$$(27)\quad \upsilon_{\langle 1,1\rangle}=\binom{q-2}{\alpha-2},\quad \upsilon_{\langle 1,0\rangle}=\binom{q-2}{\alpha-1},\quad \upsilon_{\langle 0,1\rangle}=\binom{q-2}{\alpha-1},\quad \upsilon_{\langle 0,0\rangle}=\binom{q-2}{\alpha}.$$

Заметим также, что вероятность появления каждой из матриц на выходе канала зависит только от i (количества столбцов X^0, которые покрывает Y) и равна

$$(28)\qquad p(Y\mid X^0)=p_1^i p_0^{N-i}=\left(\frac{p}{\sigma_1}\right)^i\left(\frac{1-p}{\sigma_0}\right)^{N-i}.$$

Заметим, что

а) значения $\Omega\left(Y,X^0,X^l\right)$ пробегают одни и те же значения для любых пар X^0 и X^l, зависят от числа столбцов, в которых X^0 и X^l различаются (это число фиксировано для конкретного подкода, из которого выбираются кодовые слова, соответствующие X^l), и не зависят от конкретной матрицы X^l;

б) для любых X^l число матриц на выходе канала Y, которым для тройки X^0, X^l, Y соответствуют конкретные значения g и k, пробегает одни и те же значения $\mathrm{H}(N,w,i,h,g,k)$ и не зависит от конкретной матрицы X^l.

Таким образом, ни один из сомножителей в правой части (19) не зависит от номеров кодовых слов подкода, по которым ведется суммирование. С учетом полученных выше соотношений (26), (28), (25), (22) выражение (19) можно записать в виде

$$(29)\qquad p\left(err, Y\in\mathbb{Y}_G\,|X^m\right)\leqslant\left(\sum_{\substack{X^l\in\mathbb{S}_w\\ l\neq 0}}\sum_{i=T}^{N}\beta_i\left(s,\rho\right)\Psi_i^{\left(1-\frac{1}{\rho}\right)}\right)^{\rho},$$

где $\beta_i(s,\rho)$ имеет вид

$$(30)\qquad \beta_i(s,\rho)=\sum_{h=h_l}^{h_u}\sum_{g=0}^{h}\sum_{k=0}^{w-h}\left(\mathrm{H}(N,w,i,h,g,k)p_1{}^{\left(\frac{i}{\rho}\right)}\ p_0{}^{\left(\frac{N-i}{\rho}\right)}\eta^{s(k-g)}\right).$$

Подчеркнем, что ни коэффициенты $\beta_i(s,\rho)$, ни значения функции перекоса Ψ_i не зависят от номера кодового слова l (при условии, что все кодовые слова находятся на расстоянии w от переданного кодового слова). Следовательно, выражение (29) может быть записано в следующем виде:

$$(31)\qquad p\left(err, Y\in\mathbb{Y}_G\,|X^m\right)\leqslant(A_w)^{\rho}\left(\sum_{i=T}^{N}\beta_i(s,\rho)\Psi_i^{\left(1-\frac{1}{\rho}\right)}\right)^{\rho}.$$

Отметим, что A_w не зависит от s и ρ (и вообще от каких-либо параметров системы, кроме выбора кода C и величины веса w), а функция вида $z(\xi)=\xi^{\rho}$

является монотонно возрастающей функцией ξ при $0 > \rho \geqslant 1$. Поэтому для того, чтобы минимизировать правую часть (31) при любых фиксированных значениях w, s и ρ, удовлетворяющих исходным условиям ($1 \geqslant \rho > 0$, $s > 0$), достаточно выбрать вектор $\mathbf{\Psi} = [\Psi_0, \Psi_1, \ldots, \Psi_N]$ значений функции $\psi_N^0(Y)$ таким образом, чтобы минимизировать функцию

$$f_0\left(w, \psi_N^0(Y)\right) = \sum_{i=T}^{N} \beta_i(s,\rho)\Psi_i^{\left(1-\frac{1}{\rho}\right)}. \tag{32}$$

С учетом (22) и ограничений (21) эта оптимизационная задача может быть записана в следующем виде

$$\sum_{i=T}^{N} \beta_i(s,\rho)\Psi_i^{\left(1-\frac{1}{\rho}\right)} \underset{\Psi}{\to} \min, \tag{33a}$$

$$\forall\ i = T, \ldots, N \quad \Psi_i > 0 \quad \sum_{Y \in \mathbb{Y}_G} \psi_N^0(Y) = \sum_{i=T}^{N} \gamma_i \Psi_i = 1, \tag{33b}$$

где

$$\gamma_i = \binom{N}{i} \sigma_1^i \sigma_0^{N-i}. \tag{34}$$

Записав условия Каруша – Куна – Таккера для этой задачи, можно показать, что единственным решением является точка

$$\forall\, i = T, \ldots, N \quad \Psi_i = \left(\sum_{i=0}^{N} \gamma_i^{1-\rho}\beta_i^{\rho}\right)^{-1} \left(\frac{\beta_i}{\gamma_i}\right)^{\rho}. \tag{35}$$

Найденное аналитическое выражение (35) для функции перекоса (22) минимизирует правую часть (31) (для канала рассматриваемого типа) при любых фиксированных ρ и s (удовлетворяющих (20)). Подставляя (35) в (31) и учитывая (12) и (17), получим границу:

$$P_e \leqslant \sum_{i=0}^{T-1} \binom{N}{i} p^i (1-p)^{N-i} + \sum_{w=d}^{N} A_w^{\rho} \left(\sum_{i=T}^{N} \left(\beta_i\left(s,\rho,w\right)\right)^{\rho} \gamma_i^{1-\rho}\right). \tag{36}$$

Параметры ρ, s и T находятся численной оптимизацией для каждого конкретного значения p (с учетом ограничений (20), (16)), как это традиционно делается для границы Галлагера и других границ, полученных с использованием техник из [11, 14]. В следующем параграфе будут приведены результаты имитационного моделирования.

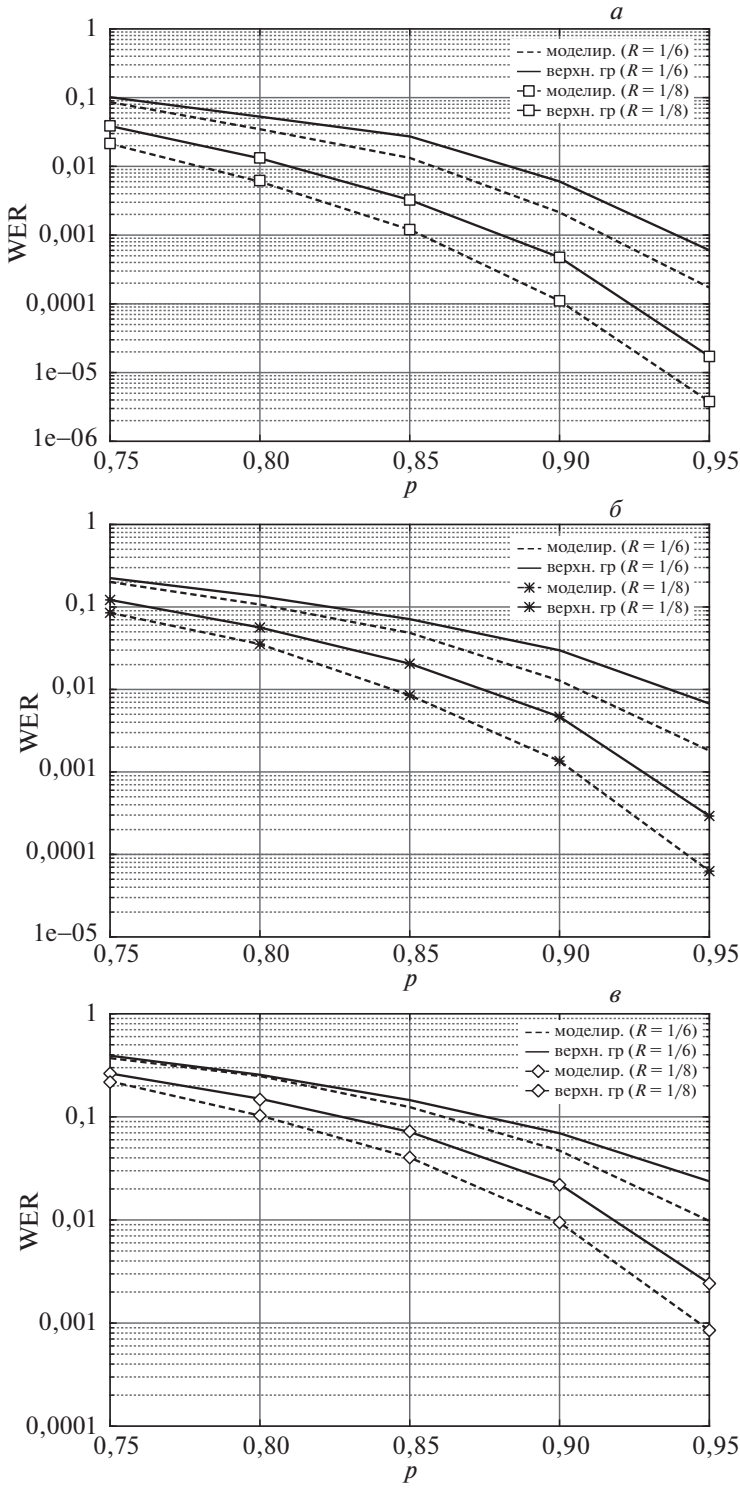


Рис. 1. Зависимость вероятности ошибки WER от вероятности p для $\eta = 12$ и различных α ($а$ – $\alpha = 3$, $б$ – $\alpha = 4$, $в$ – $\alpha = 5$).

5. Моделирование

Для проверки эффективности используемого метода и корректности результатов было проведено имитационное моделирование. При моделировании использовались МДР коды со скоростями $R = 1/6$, $R = 1/7$ и $R = 1/8$ над $GF(2^4)$, полученные из кода Рида – Соломона (15,2,14) выкалыванием проверочных символов (для скоростей $R = 1/6$ и $R = 1/7$) или, напротив, добавлением общей проверки на четность (для скорости $R = 1/8$). В качестве примера на рис. 1 приведено сравнение кривых вероятности ошибки (на кодовое слово) WER в зависимости от вероятности p для скоростей $R = 1/6$ и $R = 1/8$ для различных значений α ($\alpha = 3$, $\alpha = 4$ и $\alpha = 5$).

Как видно их приведенных графиков, предложенное выше выражение действительно позволяет оценить вероятность ошибки сверху, причем точность оценки падает с ростом вероятности p. На рис. 2 приведено сравнение кривых вероятности ошибки (на кодовое слово) WER в зависимости от вероятности p для скорости $R = 1/7$ для $\alpha = 4$ и различных значений параметра η.

Анализируя приведенные графики, можно заметить, что экспериментально полученные кривые вероятности ошибки практически не отличаются для различных величин параметра η. Это подтверждает, что рассматриваемый метод приема (и декодирования) устойчив к изменениям параметра η и, следовательно, не требует информации о распределении решающих статистик и о параметрах таких распределений. Та же особенность присуща и вычисленным с использованием (36) оценкам вероятности.

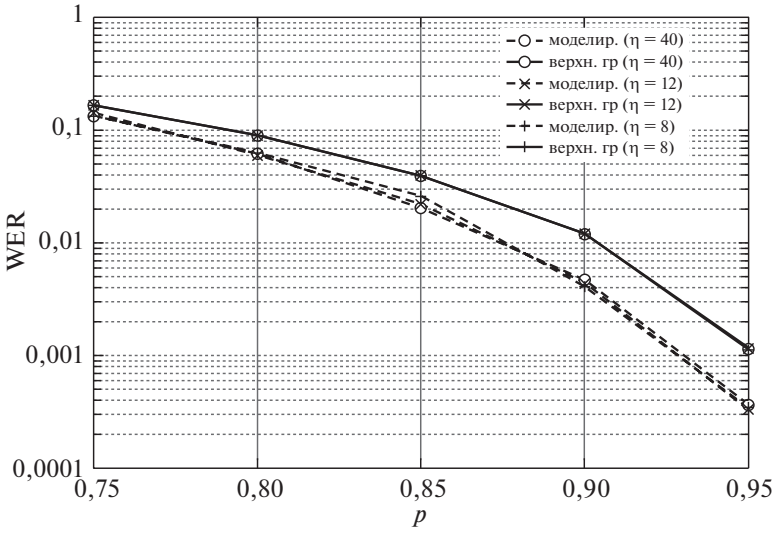


Рис. 2. Зависимость вероятности ошибки WER от вероятности p для $\alpha = 4$, $R = 1/7$.

6. Заключение

В работе рассмотрена модель канала без памяти, описывающего однопользовательский прием на основе порядковых статистик в многопользовательских каналах и каналах с аддитивными помехами. Для передачи по такому

каналу предложено использовать конструкцию Каутса-Синглтона. Предложен метод декодирования, не требующий информации о характеристиках канала и характере аддитивных помех, и получена верхняя граница вероятности ошибки декодирования (на кодовое слово). Состоятельность предложенной верхней границы и устойчивость предлагаемой стратегии приема и декодирования к выбору параметра η подтверждены результатами имитационного моделирования.

СПИСОК ЛИТЕРАТУРЫ

1. *Osipov D.* Reduced-Complexity Robust Detector in a DHA FH OFDMA System under Mixed Interference Multiple Access Communications // MACOM – 7th Int. Work. on Mult. Acc. Comm., Halmstad, Sweden, John Wiley & Sons, 2014. P. 29–34.
2. *Viswanathan R., Gupta S.*, Nonparametric Receiver for FH-MFSK Mobile Radio // IEEE Trans. Commun. 1985. V. 33. No. 2. P. 178–184.
3. *Kreshchuk A., Potapov V.* On applying one-sample goodness-of-fit statistics to coded FSK decoding // REDUNDANCY 2016 – XV Int. Symp. Probl. Redund. Inf. Contr. Syst., St. Petersburg, Russia, IEEE, 2016. P. 66–70.
4. *Osipov D.* An upper bound on the error probability of a communication system with nonparametric detection // REDUNDANCY 2016 – XV Int. Symp. Probl. Redund. Inf. Contr. Syst., St. Petersburg, Russia, IEEE, 2016. P. 100–104.
5. *Kautz W.H., Singleton R.C.* Nonrandom Binary Superimposed Codes // IEEE Transact. Inform. Theory. 1964. No. 4. P. 363–377.
6. *Chang S.C., Wolf J.K.* On the T-User M-Frequency Noiseless Multiple-Access Channels with and without Intensity Information // IEEE Trans. Inform. Theory 1981. V. 27. No. 1. P. 41–48.
7. *Зигангиров К.Ш., Попов С.А., Чепыжов В.В.* Недвоичное сверточное кодирование в канале с преднамеренными помехами // ППИ. 1995. Т. 31. № 2. С. 84–101.
8. *Осипов Д.С., Фролов А.А., Зяблов В.В.* О пропускной способности для пользователя системы множественного доступа в векторном дизъюнктивном канале при наличии ошибок // ППИ. 2013. Т. 49. № 4. С. 13–27.
9. *Cover T.M., Thomas J.A.* Elements of Information Theory. N.Y.: Wiley, 2006.
10. *Herzberg H., Poltyrev G.* Techniques of bounding the probability of decoding error for block coded modulation structures // IEEE Trans. Inform. Theory. 1994. V. 40. No. 3. P. 903–911.
11. *Gallager R.G.* A simple derivation of the coding theorem and some applications // IEEE Trans. Inform. Theory. 1965. V. 11. No. 1. P. 3–18.
12. *Bennatan A., Burshtein D.* On the application of LDPC codes to arbitrary discrete-memoryless channels // IEEE Trans. Inform. Theory. 2004. V. 50. No. 3. P. 417–438.
13. *Erez U., Miller G.* The ML decoding performance of LDPC ensembles over $\mathbb{Z}_q$ // IEEE Trans. Inform. Theory. 2005. V. 51. No. 5. P. 1871–1879.
14. *Fano R.M.* Transmission of Information. M.I.T Press and John Wiley & Sons, 1961.
15. *Duman T.M., Salehi M.* New performance bounds for turbo codes // IEEE Trans. Commun. 1998. V. 46. No. 6. P. 717–723.

Статья представлена к публикации членом редколлегии В.М. Вишневским.

Поступила в редакцию 10.12.2018

После доработки 26.06.2019

Принята к публикации 18.07.2019

Оптимизация, системный анализ и исследование операций

© 2020 г. В.В. ЗЕНКОВ, канд. техн. наук (zenkov-v@yandex.ru)
(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

ПРИМЕНЕНИЕ АППРОКСИМАЦИИ ДИСКРИМИНАНТНОЙ ФУНКЦИИ АНДЕРСОНА И МЕТОДА ОПОРНЫХ ВЕКТОРОВ ДЛЯ РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ КЛАССИФИКАЦИИ

Дискриминантная функция Андерсона имеет ряд свойств, полезных для решения задач классификации и для оценки апостериорных вероятностей классов. В качестве математического аппарата используется один и тот же взвешенный метод наименьших квадратов для аппроксимации дискриминантной функции Андерсона в области нулевых значений как при решении задачи классификации, так и при оценке апостериорных вероятностей классов в заданной точке пространства признаков. В методе опорных векторов задача классификации решается методом квадратичного программирования с количеством ограничений, равным количеству строк обучающей выборки, а для оценки апостериорных вероятностей классов используется дополнительная надстройка – калибратор Платта, преобразующий величину отступа точки от границы в апостериорную вероятность класса, с определением параметров калибратора методом максимального правдоподобия. На нескольких примерах решения задач классификации проведено сравнение эффективности методов по критерию эмпирического риска. Результаты оказались в пользу метода аппроксимации дискриминантной функции Андерсона в области нулевых значений.

Ключевые слова: машинное обучение, классификация, дискриминантная функция Андерсона, метод опорных векторов, SVM, аппроксимация дискриминантной функции Андерсона в области нулевых значений.

DOI: 10.31857/S0005231020010109

1. Введение

Дискриминантную функцию Андерсона (ДФА) здесь получаем из представленного Теодором Андерсоном метода решения байесовой задачи классификации в случае нескольких классов [1] в виде разности средних потерь от отнесения точки в пространстве признаков классов в один из двух конкурирующих классов. ДФА есть функция регрессии. Обучающая выборка с учителем просто преобразуется в выборку регрессионного анализа заменой в выборке номеров (меток) классов на соответствующие разности заданных стоимостей ошибок классификации. В случае двух классов в точках на их границе в пространстве признаков апостериорная вероятность (АпоВ) первого

(для определенности) класса зависит только от определяющих эту ДФА стоимостей двух ошибок классификации. Задача со многими классами сводится к совокупности задач с двумя классами по принципу один против остальных.

Для аппроксимации ДФА по обучающей выборке с учителем предложен эвристический алгоритм [2, 3]. Аппроксимация выполняется в области нулевых значений ДФА, поскольку для классификации важен лишь знак ДФА, а не ее изящные изгибы. Тождественная связь ДФА, АпоВ первого класса и стоимостей ошибок, для которых построена ДФА, лежит в основе методов оценивания АпоВ в заданной точке [4, 5].

Популярная, мощная, широко используемая в машинном обучении разновидность метода опорных векторов (SVM – support vector machine) для случая разделения гиперплоскостью двух перекрывающихся между собой множеств точек также является эвристикой [6]. Это обстоятельство побудило выполнить сравнение на нескольких обучающих выборках двух эвристических методов решения задачи классификации: метода аппроксимации ДФА и метода SVM. Критерием качества классификации является доля неправильно классифицированных точек обучающей выборки – эмпирический риск.

2. Дискриминантная функция Андерсона, ее свойства и связь с апостериорными вероятностями классов

Используя [1], запишем определение ДФА $f_{rs}(x)$, разделяющей классы r и s в d-мерном евклидовом пространстве признаков $x \in R^d$, использующее АпоВ классов и заданные стоимости ошибок классификации, в виде

$$f_{rs}\left(x, C\right) \equiv G_r\left(x\right) - G_s\left(x\right) \equiv \mathrm{M}_{k|x}\left(C_{rk} - C_{sk}\right), \tag{1}$$

где $G_r\left(x\right) = \sum_k C_{rk} p\left(k|x\right)$, $G_s\left(x\right) = \sum_k C_{sk} p\left(k|x\right)$ — средние потери по k в точке x, если точку отнести к классу r или, соответственно, к классу s; C — матрица стоимостей ошибок, C_{ij} — стоимость ошибки, когда точка из класса j ошибочно относится в класс i, $0 \le C_{ij} < \infty$, $C_{ii} = 0$; $p(k|x)$ — АпоВ класса k в точке x, $p\left(k|x\right) = P_k p\left(x|k\right) / p\left(x\right)$, $p\left(x\right) = \sum_k P_k p\left(x|k\right)$, P_k — априорные вероятности классов, $p(x|k)$ — условные распределения признаков классов; k — номера классов от 1 до K, K — количество классов; $M_{k|x}\left(\cdot\right)$ — математическое ожидание по k в точке x. ДФА по определению есть функция регрессии от x. В точке x случайная по k дискретная величина $C_{rk|x} - C_{sk|x} = f_{rs}(x, C) + \varepsilon_{rsk|x}$ имеет распределение АпоВ классов $p(k|x)$, $\sum_k p(k|x) = 1$, среднее $f_{rs}(x, C)$ и случайное отклонение от него $\varepsilon_{rsk|x}$. Дискретная случайная величина $\varepsilon_{rsk|x}$ принимает K значений с вероятностями $p(k|x)$ и имеет нулевое среднее.

Если $f_{rs}(x, C) \le 0$, то точка x относится в класс r и класс s исключается из дальнейшего процесса сравнения, иначе — в класс s и класс r исключается. Так обеспечивается минимум средней стоимости ошибок классификации — байесов критерий решающего правила [1].

В случае двух классов, $K = 2$, имеем $C_{1k|x} - C_{2k|x} = f_{12}(x, C) + \varepsilon_{12k|x}$. Дискретная случайная величина $\varepsilon_{12k|x}$ в точке x принимает два значения: $-C_{21} - f_{12}(x, C)$ с вероятностью $p(1|x)$ и $C_{12} - f_{12}(x, C)$ с вероятностью $1 - p(1|x)$, с нулевым средним и дисперсией $(C_{12}+C_{21})^2 p(1|x)(1-p(1|x))$.

2.1. Свойства ДФА

Перечислим свойства ДФА.

Утверждение 1. Стоимости ошибок классификации можно выбирать при условии, что их сумма равна единице.

Доказательство следует из (1). На результат попарного сравнения $G_r(x)$ и $G_s(x)$ не влияет умножение их на одно и то же положительное число.

Утверждение 2. ДФА есть ограниченная функция регрессии.

Доказательство следует из (1), так как стоимости ошибок классификации ограничены.

Следствие 1. В случае $K = 2$ имеем $-C_{21} < f_{12}(x) < C_{12}$.

Следствие 2. Чтобы аппроксимировать ДФА как функцию регрессии, следует преобразовать обучающую выборку с учителем задачи классификации в выборку задачи регрессионного анализа, заменив номера классов в выборке следующим образом: первый класс на $-C_{21}$, а второй класс — на C_{12}.

Следствие 3. Факт регрессионной зависимости ДФА от признаков позволяет, в частности, выполнять отбор признаков, используемых для решения задачи аппроксимации, по коэффициентам корреляции признаков со столбцом, в котором номера классов заменены стоимостями ошибок классификации. Учитывать при этом необходимо и коэффициенты корреляции признаков между собой [7].

2.2. Связь ДФА с AnoB классов

Утверждение 3. При $K = 2$ для AnoB первого класса и ДФ Андерсона, полученной для заданных C_{12} и C_{21}, имеет место тождество

$$p(1|x) \equiv (C_{12} - f_{12}(x))/(C_{12} + C_{21}). \tag{2}$$

Доказывается по (1) с использованием равенства $p(1|x) + p(2/x) = 1$.

Следствие 4. Задавая стоимости ошибок при условии равенства единице их суммы, тождество (2) *можно представить в виде*

$$p(1|x) \equiv p^* - f_{12}(x, p^*), \tag{3}$$

где скалярный параметр p^* определяет недиагональные элементы матрицы ошибок классификации:

$$C_{12} = p^*, \quad C_{21} = 1 - p^*. \tag{4}$$

Из (3) следует, что в точках на границе классов, если она существует, $f_{12}(x, p^*) = 0$, AnoB первого класса равна p^*, а в точках, относимых в первый класс, где $f_{12}(x, p^*) \leqslant 0$, из (3) следует, что $p(1|x) \geqslant p^*$. Если границы между классами в пространстве признаков классов нет, то p^* задает лишь недиагональные элементы матрицы стоимостей ошибок для ДФА и может находиться в пределах (0, 1), чтобы сумма стоимостей ошибок равнялась единице.

2.3. Условие неразличимости классов

Условие неразличимости классов в задаче с двумя классами для заданного p^* имеет вид

$$\left(\min_x f_{12}(x, p^*) > 0\right) \vee \left(\max_x f_{12}(x, p^*) < 0\right), \tag{5}$$

при выполнении которого все точки надо относить в один соответствующий класс, а параметр p^* есть величина, лишь определяющая стоимости ошибок классификации (4).

2.4. Способы оценивания АпоВ класса

Из тождества (3) вытекают по крайней мере три способа оценивания АпоВ первого класса в задаче с двумя классами.

По одному способу [4] для серии заданных значений параметра p^* строятся аппроксимации ДФА, по которым путем интер- и экстраполяции находится АпоВ класса в заданной точке по одной–двум соседним с точкой аппроксимациям ДФА, или по всем аппроксимациям с использованием аналога ядерных функций. Условия неразличимости классов (5) влияют на выбор предельных значений параметра p^*. Результат естественно зависит от удачного выбора вида аппроксимирующих ДФА функций.

По второму способу [5] величина параметра p^* подбирается итерационно так, чтобы в заданной точке получить нулевое значение аппроксимации ДФА. При этом АпоВ класса будет равна найденному p^*, как это следует из (3). Для аппроксимации ДФА в точке не обязательно использовать аппроксимирующие зависимости сложнее линейных. Но нужно иметь в виду, что в некоторых точках вследствие (5) решение может не существовать.

По третьему способу для произвольно заданного p^*, величина которого из интервала (0, 1) не влияет на результат оценивания АпоВ класса и не влияет так же различимость или неразличимость классов (5), строится аппроксимация ДФА в заданной точке и затем по ней и по p^* вычисляется по (3) оценка АпоВ класса. Для аппроксимации ДФА в точке используется линейная аппроксимирующая функция.

2.5. Аппроксимация ДФА в области нулевых значений

Область нулевых значений не известна. Для аппроксимации ДФА используется прием последовательных приближений к области нулевых значений, использующий взвешенный метод наименьших квадратов [2, 3].

Для аппроксимации ДФА возьмем линейную комбинацию заданных функций от признаков $\lambda'\varphi(x)$, первая компонента – единица, с вектором коэффициентов λ, который находится по обучающей выборке с учителем $\{x_n, k_n\}$, $n = 1 \div N$, k_n — номер класса в строке n, $k_n = \{1, 2\}$, $x_n \subset R^d$ — вектор действительных значений признаков размерности d в строке n. Решается последовательность задач взвешенным методом наименьших квадратов, в которой

на каждом шаге минимизируется по λ_i критерий

$$Q(\lambda_i) = \min_{\lambda_i} \sum_{n=1}^{n=N} \left\{ \left[C_{1k_n} - C_{2k_n} - \lambda_i' \varphi(x_n) \right]^2 \exp\left(-W_i |\lambda_{i-1}' \varphi(x_n)|^k \right) \right\}, \tag{6}$$

где i — номер итерации, $i = 1 \div I$. На первом шаге задача решается без весовой функции, на последующих шагах весовая функция придает больший вес точкам, более близким к нулевой области предыдущей аппроксимации ДФА. I — заданное количество итераций, $I < \infty$. W_i — заданный весовой коэффициент на шаге i, $W_i > 0$, k — заданный показатель степени. Размерность обращаемой матрицы, равная размерности искомого вектора параметров, не зависит от количества строк N в обучающей выборке. Вид весовой функции — не обязательно экспонента. Лучшим значением λ является тот вектор, которому соответствуют меньшие средние по выборке потери (эмпирический риск)

$$R = N^{-1} \left(C_{12} N_2 + C_{21} N_1 \right), \tag{7}$$

где N_1 — количество точек первого класса, ошибочно отнесенных во второй класс, N_2 — количество точек второго класса, ошибочно отнесенных в первый класс.

3. Постановка задачи

Цель работы – сравнение по величине эмпирического риска результатов решения нескольких задач классификации двумя методами: вышеописанного, использующего аппроксимацию ДФА в области нулевых ее значений, и широко известного, мощного и популярного в машинном обучении метода опорных векторов (SVM) в варианте с линейным ядром классификатора.

3.1. Процедуры метода SVM

Для сравнения с методом аппроксимации ДФА использованы три имеющиеся процедуры библиотеки scikit-learn инструментального средства Питон (Python) [8, 9]. Одна из процедур (LinearSVC) исключительно ориентирована на функцию ядра линейного типа, другие (SVC, NuSVC) позволяют задавать вид функции ядра, являясь в этом смысле более универсальными. ДФА используется в линейном относительно искомых коэффициентов виде, поэтому и процедуры SVM использованы в линейном варианте. Процедура, реализующая метод аппроксимации ДФА в области нулевых значений, также написана на Питоне.

Проблема регуляризации не затрагивалась, сравнивалось качество разделения двух наборов точек гиперплоскостью в пространстве признаков, а также в пространстве с координатами – произведениями и степенями признаков не выше второго порядка.

3.2. Исходные данные

Исходные данные – обучающие выборки с учителем – взяты в основном из примеров, ранее использованных в работах автора [2–5]. Чтобы иметь возможность проверить некоторые из приведенных результатов, в качестве одного из примеров использованы данные из репозитория UCI [10], описания которых приведены в [11]. Доступны в интернете и конкурсные данные ТМШ [12]. Данные из репозитория и конкурсные данные являются реальными. Остальные были сгенерированы в качестве модельных примеров с заданными нормальными законами условных распределений двух признаков и с заданными априорными вероятностями двух классов.

Данные из репозитория были проверены. Из них были удалены строки с отсутствующими величинами некоторых из 9 признаков.

Данные из репозитория и из конкурсной задачи представлялись в двух вариантах: с полным набором признаков и с подмножеством признаков, отобранных по коэффициентам корреляции признаков с искомой величиной — оценками ДФА, полученными путем замены номеров классов в выборке стоимостями ошибок классификации, в данном случае значениями 0,5 и −0,5. При отборе учитывались и корреляции признаков между собой. Если уменьшение количества признаков до 5 из 9 в задаче из репозитория (пример № 8, таблица) практического смысла не имело (цель — получить еще один вариант данных для сравнения методов классификации), то в конкурсной задаче выбор 3 из 216 исходных признаков, пример № 2, имело практический смысл — удобство использования полученной аппроксимации ДФА и снижение переобучения. В конкурсной задаче со всеми 216 признаками, пример № 6, не имело смысла решать полиномиальный вариант из-за гигантского количества членов. Метод аппроксимации ДФА попросту не смог бы решить такую задачу в лоб, в то время как SVM задачи, в которых количество признаков или функций от признаков больше количества строк обучающей выборки, решать может.

В задаче из репозитория отбор 5 функций от признаков по корреляции выполнялся из 54 членов полинома второго порядка, полученного по 9 исходным признакам (пример № 8). Из наиболее коррелированных с номером класса членов полинома были отброшены те, которые имели корреляцию между собой более 0,8. Осталось 5 из 54 членов. Корреляция 0,8 задавалась исключительно из соображения получить не слишком много членов.

4. Решение задачи

Коэффициенты весовой функции при обращении к процедуре аппроксимации ДФА (6) изменялись в итерационном процессе по правилу

$$W_i = w \times i, \quad i = 0 \div I, \tag{8}$$

где шаг изменения w подбирался вручную из нескольких значений, чтобы получить поменьше величину эмпирического риска (7) в итерационном процессе, см. рис. 1–5. Автоматический перебор значений шага w для (8) из некоторого диапазона для поиска наилучшего значения не выполнялся, потому что итерационные процессы в несколько десятков шагов занимали мало времени,

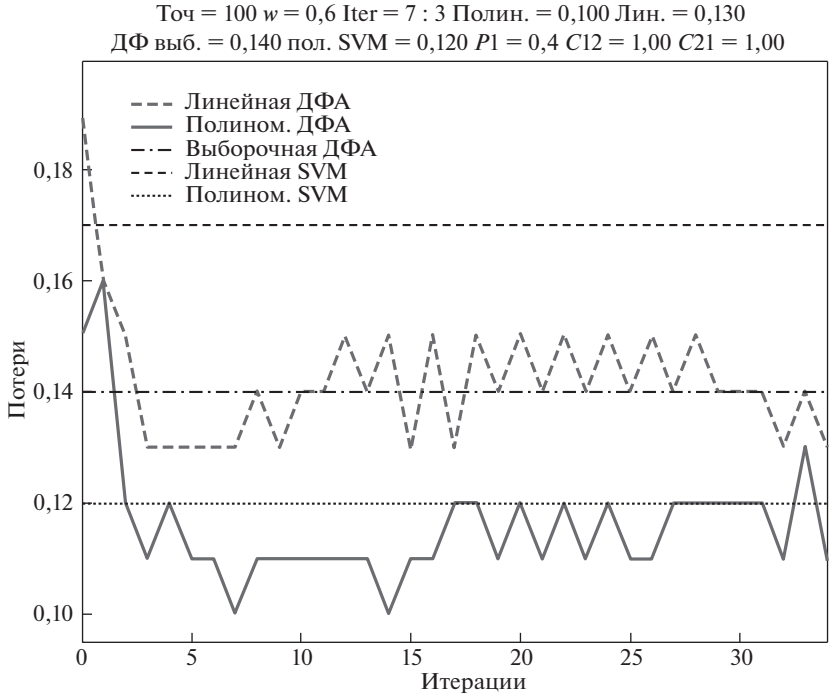


Рис. 1. Пример № 1.

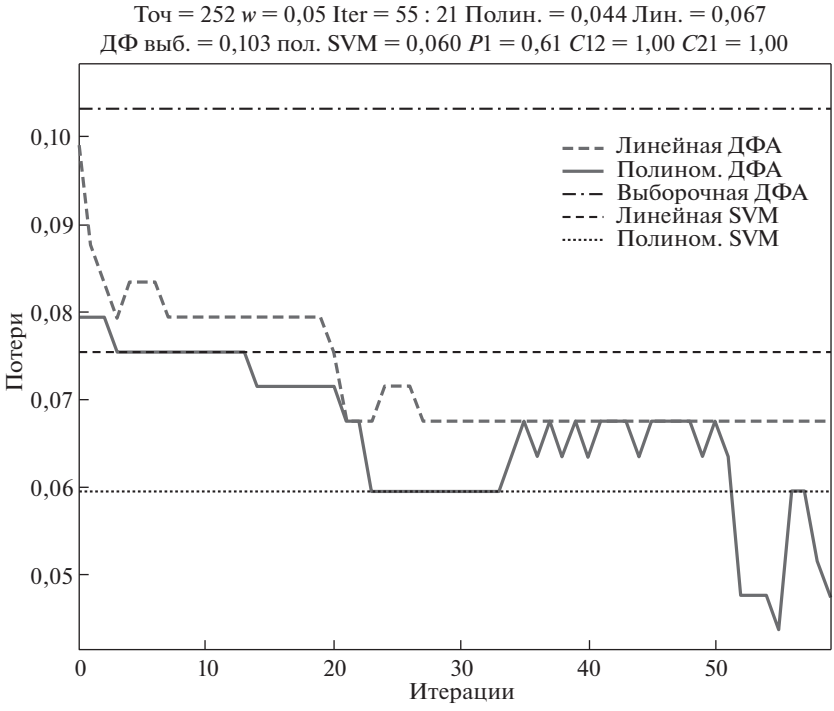


Рис. 2. Пример № 2.

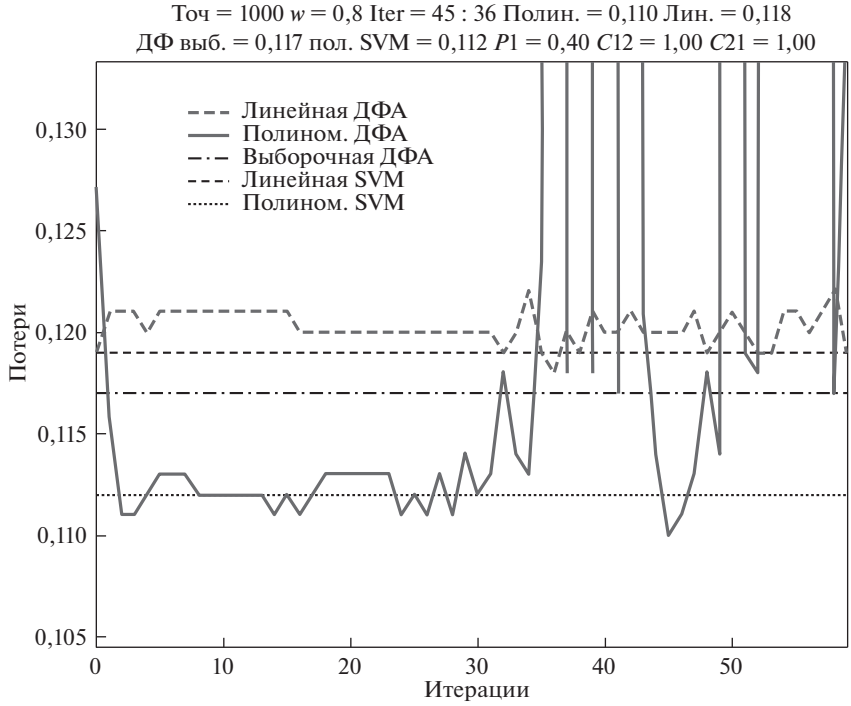


Рис. 3. Пример № 3.

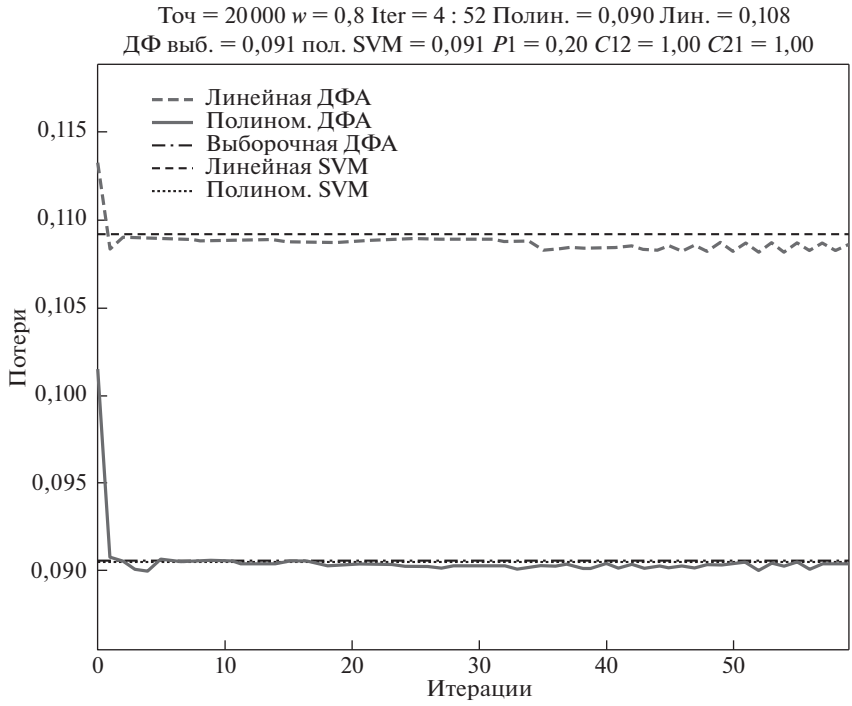


Рис. 4. Пример № 4.

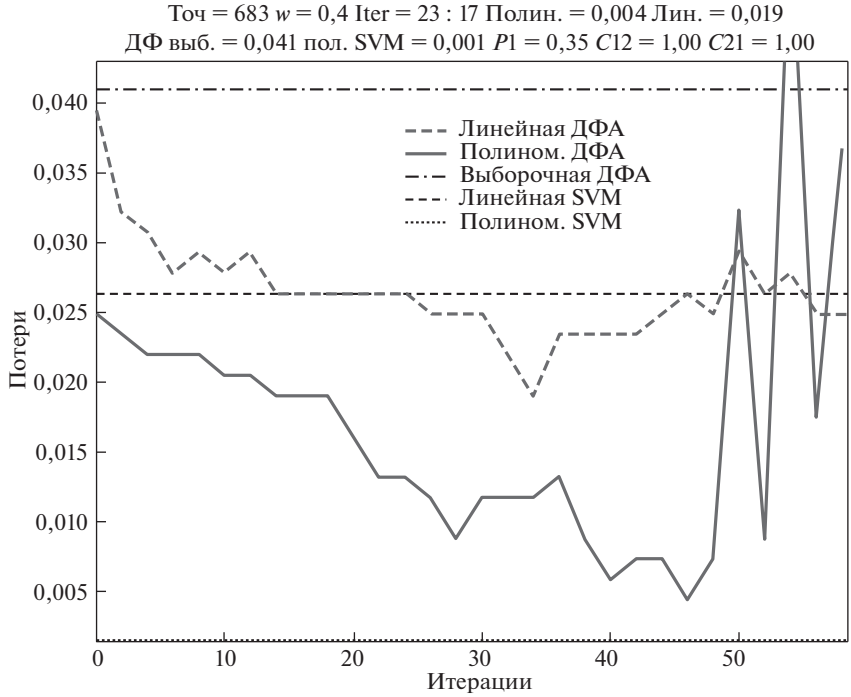


Рис. 5. Пример № 7.

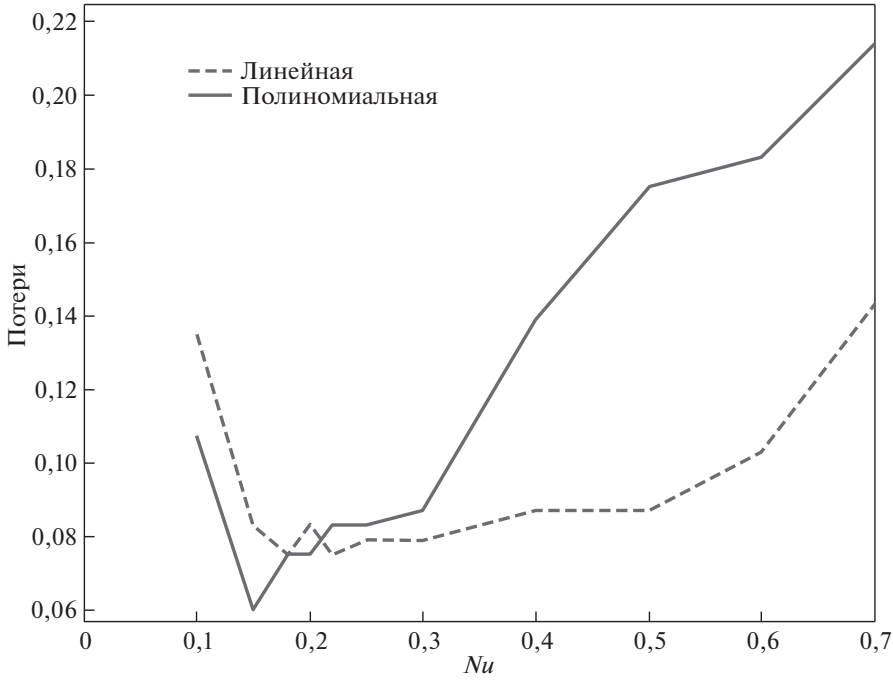


Рис. 6. Зависимость потерь от параметра Nu.

ручной выбор не был обременителен и визуальный контроль итерационного процесса облегчал подбор w и I. Показатель степени k в (6) равен единице. Лишь в одном случае (пример № 7, полином), когда аппроксимация ДФА не была лучше SVM, пытались долго и безуспешно искать и лучшее значение w и k в (6).

Количество шагов итераций I выбирается достаточно большим, но не слишком, потому что график изменения потерь от итераций с некоторого момента становится хаотическим из-за уменьшения части выборки, выделяемой весовой функцией с ростом W_i по правилу (8) и влияющей на результат,

Характер изменения по итерациям весового коэффициента в виде арифметической прогрессии в (8) хорошо зарекомендовал себя при решении автором предыдущих задач.

Процедуры метода SVM также настраивались в каждом примере параметрами: C в SVC и LinearSVC; Nu в NuSVC. И если параметр C (коэффициент штрафной функции) изменял эмпирический риск (7) сравнительно в малых пределах, то параметр Nu изменял (7) в широких, см. рис. 6. Параметр Nu — верхняя граница ошибок обучения. Он должен находиться в интервале (0, 1]. По умолчанию равен 0,5.

Остальные параметры процедур в примерах не изменялись. Для LinearSVC они были установлены отличными от значений по умолчанию на уровнях: dual=False. Параметр dual определяет функцию потерь. Значение ‘hinge‘ – это стандартная потеря SVM (используемая, например, классом SVC для решения задач классификации), в то время как значение ‘squared_hinge‘ – это использование квадрата потерь. Параметр penalty, задающий норму штрафа, установлен в значение ‘l1‘. Норма ‘l2‘ – это стандарт, используемый в SVC. Параметр max_iter = 100000 (максимальное количество итераций, которое можно выполнять). Для SVC и NuSVC установлены: kernel = ’linear’ – определяет тип ядра, который используется в алгоритме.

Проблема регуляризации не затрагивалась, сравнивалось качество разделения двух наборов точек гиперплоскостью в пространстве признаков, а также в пространстве с координатами – произведениями и степенями признаков не выше второго порядка.

Результаты решения задач разными методами представлены в таблице.

Графики итерационных процессов получения аппроксимаций ДФА, на которых горизонтальными линиями изображены решения задач процедурами SVM, представлены на рис. 1–4, соответствующих примерам № 1–4. Рисунок 5 соответствует примеру № 7. Рисунок для примера № 6 не представлен, так как графики вырождаются в горизонтальные линии. Рисунок для примера № 8 не представлен для экономии места.

Рисунок 6 показывает зависимости потерь (7) от параметра Nu, полученных при ручном поиске лучших значений для процедуры NuSVC в случаях линейной и полиномиальной дискриминантных функций для условий примера № 2. Из рисунка видно, что полагаться на установленное по умолчанию значение параметра $Nu = 0{,}5$ не стоит, если стремиться к получению более точного решения задачи классификации методом SVM, реализованным в про-

Таблица

№	Кол-во точек	Кол-во призн., вид модели		ДФА				LinearSVC		SVC		NuSVC		Выб, ДФ
				R	t	I	w	R	t	R	t	R	t	R
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	100	2	Линейн.	**0,130**	+	35	0,6	0,180	0,02	**0,170**	0,000	**0,170**	0,000	
		5	Полином.	**0,100**	0,32	35	0,6	0,140	0,00	**0,120**	0,000	0,150	0,000	0,140
2	252	3	Линейн.	**0,067**	+	60	0,05	**0,075**	0,00	**0,075**	0,016	**0,075**	0,000	
		9	Полином.	**0,044**	1,06	60	0,05	**0,060**	0,05	0,067	0,781	**0,060**	0,078	0,103
3	1000	2	Линейн.	**0,118**	+	60	0,8	**0,119**	0,00	**0,119**	0,047	**0,119**	0,047	
		5	Полином.	**0,110**	3,72	60	0,8	0,114	0,06	**0,112**	0,047	0,116	0,062	0,117
4	20000	2	Линейн.	**0,1082**	+	60	0,8	0,1092	0,11	**0,1091**	22,57	0,1099	22,25	
		5	Полином.	**0,0900**	59,1	60	0,8	0,0907	0,72	**0,0905**	28,93	0,0924	37,28	0,0906
5	50000	2	Линейн.	**0,1405**	+	60	0.8	0,14108	0,25	0,1411	98,00	**0,1409**	152	
		5	Полином.	**0,1205**	142	60	0,8	**0,12120**	1,72	0,1213	89,14	0,1214	217	0,1213
6	252	216	Линейн.	**0**	7,5	60	0,8	**0**	0,91	**0**	0,016	**0**	0,03	
7	683	9	Линейн.	**0,019**	+	30	0.4	0,0278	0,03	**0,0264**	0,109	0,0278	0,016	
		54	Полином.	0,0044	1,91	30	0,4	**0,0015**	27,9	**0,0015**	26,78	0,0029	0,031	0,041
8	683	5	Линейн.	**0,023**	+	30	0,4	0,0322	0,03	**0,0293**	0,258	**0,0293**	0,016	
		20	Полином.	**0,019**	1,31	30	0,4	0,0307	0,23	**0,0264**	433,0	0,0278	0,047	0,042

цедуре NuSVC. Аналогичное замечание справедливо и в отношении реализаций SVM в процедурах LinearSVC и SVC, хотя в них изменения параметра C не столь существенно сказываются на результате.

Пояснения к таблице. Столбцы: 1 – номер примера; 2 – количество точек выборки; 3 – количество признаков и количество использованных членов в полиноме второго порядка; 4 – вариант дискриминантной функции; 5 – эмпирический риск по итерациям в двух видах аппроксимаций ДФА; 6 – суммарное время на выполнение заданных итераций по двум видам аппроксимации ДФА; 7 – количество заданных итераций в каждом виде аппроксимаций ДФА; 8 – коэффициент весовой функции; 9 – эмпирический риск метода SVM LinearSVC; 10 – время решения SVM LinearSVM в секундах; 11–14 – аналогично 9 и 10, но для SVC и NuSVC; 15 – эмпирический риск в предположении, что обучающая выборка подчиняется двум нормальным условным законам распределения признаков и параметры их получены по выборке (для примеров 1, 3–5 предположение справедливо по построению. Дискриминантная функция для нормально распределенных признаков двух классов является полиномом второго порядка).

В столбцах 5, 9, 11 и 13 полужирным шрифтом отмечены лучшие эмпирические риски сравниваемых методов, причем в столбцах 9, 11 и 13 отмечены и лучшие риски среди реализаций метода SVM. В столбце 5 отмечены лучшие результаты для метода аппроксимации ДФА по сравнению с лучшим результатом среди реализаций SVM.

В одном из 15 примеров, в примере № 7, рис. 5, в полиномиальном варианте все реализации метода SVM дали лучший результат, чем метод аппроксимации ДФА. Причем в отличие от других примеров в примере № 7 выполнялся тщательный поиск лучшего решения не только в широком диапазоне коэффициента w в (8), но испытывались и другие весовые функции с разными степенями k в весовой функции.

Работа выполнена на ноутбуке SMARTBOOK 116C Prestigio, CPU x64, 144GHz 144GHz. Оперативная память 2 ГБ. Дисковая память 32 ГБ. Операционная система Windows 10 домашняя 32-разрядная.

4.1. Способ проверить некоторые результаты

Чтобы можно было проверить результаты сравнения методов на общедоступных наборах данных [10, 12], приводим две аппроксимации ДФА. Для сравнения с ними решений, получаемых методом SVM, можно использовать доступные в разных инструментальных средствах реализации метода SVM подобно взятым из Питона [8, 9].

Линейная модель для примера № 7, рак легких, репозиторий [10, 11]:

$$\begin{aligned} f_{12}(x) = {} & -0{,}45946108 + 0{,}02133231x1 + 0{,}01406552x2 + \\ & + 0{,}02060665x3 + 0{,}01399266x4 - 0{,}00209362x5 + 0{,}0264621x6 + \\ & + 0{,}01320144x7 + 0{,}01350233x8 + 0{,}02748957x9, \end{aligned} \tag{9}$$

где переменные $x1$–$x9$ обозначают признаки, перечисленные в списке атрибутов под номерами 2–10 в [11]. Если $f_{12}(x) < 0$, то точку следует отнести в класс с меткой 2 — доброкачественный (benign), иначе — в класс с мет-

кой 4 — злокачественный (malignant). Количество ошибочно классифицированных точек — 13 из 683 точек обучающей выборки. Из 699 строк выборки [10] были предварительно исключены 16 строк с неполными данными.

Полиномиальная модель для примера № 2, диагностика заболевания, конкурсная задача [12]:

$$f_{12}(x) = -0{,}81685233 - 0{,}54083596x1 - 0{,}34128288x2 - 0{,}1861543x3 + \\ + 0{,}01519183x1^{**}2 + 0{,}30754489x1x2 + 0{,}42900778x1x3 + \\ + 0{,}0584966x2^{**}2 - 0{,}06766405x2x3 + 0{,}01869679x3^{**}2, \tag{10}$$

где $x1$, $x2$, $x3$ — соответственно элементы в столбцах 22, 104 и 115 обучающей выборки. В первом столбце 0 — метка здорового пациента, 1 — больного. Три признака отобраны из 216 по коэффициентам корреляции с первым столбцом и коэффициентам корреляции между собой.

Если $f_{12}(x) < 0$, то пациент здоров, иначе – болен некоторой болезнью.

Из 252 строк обучающей выборки ошибочно классифицированы 11 строк.

5. Заключение

1. Метод аппроксимации дискриминантной функции Андерсона в области нулевых значений (ДФА) по обучающей выборке с учителем, используемый для решения задач классификации с целью минимизации эмпирического риска, является эвристическим методом. Метод опорных векторов (SVM) также является эвристическим методом. Эвристика этих методов побуждает выполнять их сравнение между собой на модельных примерах и на реальных обучающих выборках. Метод опорных векторов был представлен тремя реализациями, имеющимися в инструментальном средстве Питон. Метод ДФА был написан автором на том же языке. Имеется и вариант метода, написанный автором на МАТЛАБе.

2. Из 15 примеров лишь в одном эмпирические риски, полученные всеми тремя реализациями SVM, оказались меньше, чем эмпирический риск, полученный методом аппроксимации ДФА. В одном примере все реализации SVM и ДФА дали одинаковый, нулевой, результат. В остальных примерах метод аппроксимации ДФА был лучше всех трех реализаций SVM.

3. При сравнении выполнялась ручная настройка параметров методов: w в ДФА, C в SVC и LinearSVC, Nu в NuSVC. Для тех обучающих выборок, которые можно скопировать из интернета, приведены дискриминантные функции, полученные методом аппроксимации ДФА. Для них можно попытаться найти лучшие параметры настроек реализаций методов SVM, чтобы сравнить с результатами работы.

4. Примеры решений задач эвристическими методами не могут дать исчерпывающего ответа на вопрос о том, какой из методов лучше. Так, по показателю эмпирического риска в некоторых случаях один метод оказывается лучше другого. По используемому математическому аппарату (взвешенному методу наименьших квадратов) ДФА проще SVM, использущего квадратичное программирование с количеством ограничений, равным количеству строк в обучающей выборке. Но SVM может решать задачу, когда количество признаков больше количества строк в выборке. В ДФА в таких случаях отбирает-

ся меньшее количество признаков по коэффициентам корреляции. SVM для оценок апостериорных вероятностей классов в точках пространства признаков должен использовать специальные надстройки типа калибратора Платта с оценкой параметров методом максимального правдоподобия, а в ДФА для оценки апостериорных вероятностей классов используется тот же взвешенный метод наименьших квадратов, который используется для аппроксимации ДФА в окрестности нулевых значений.

СПИСОК ЛИТЕРАТУРЫ

1. *Anderson T.W.* An Introduction to Multivariate Statistical Analysis. Third edition. John Wiley & Sons, 2003. 721 p.
2. *Зенков В.В.* Аппроксимация дискриминантных функций в окрестности нулевых значений // Изв. АН СССР. Техн. кибернетика. 1973. № 2. С. 152–156.
3. *Зенков В.В.* Использование взвешенного метода наименьших квадратов при аппроксимации дискриминантной функции цилиндрической поверхностью в задачах классификации // АиТ. 2017. № 9. С. 145–158.

   Zenkov V.V. Using Weighted Least Squares to Approximate the Discriminant Function with a Cylindrical Surface in Classification Problems // Autom. Remote Control. 2017. V. 78. No. 9. P. 1662–1673.
4. *Зенков В.В.* Оценка апостериорной вероятности класса по серии дискриминантных функций Андерсона // АиТ. 2019. № 3. С. 68–71.

   Zenkov V.V. Evaluation of the Posterior Probability of a Class with a Series of Anderson Discriminant Functions // Autom. Remote Control. 2019. V. 80. No. 3. P. 447–458.
5. *Зенков В.В.* Оценка вероятности принадлежности точки классу по аппроксимации одной дискриминантной функции // АиТ. 2018. № 9. С. 46–58.

   Zenkov V.V. Estimating the Probability of a Class at a Point by the Approximation of one Discriminant Function // Autom. Remote Control. 2018. V. 79. No. 9. P. 1580–1590.
6. *Воронцов К.В.* Математические методы обучения по прецедентам (теория обучения машин). http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf
7. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ, 3-е изд. : Пер. с англ. М.: Изд. дом "Вильямс", 2007. 912 с.
8. Scikit learn. 1.4. Support Vector Machines.

   https://scikit-learn.org/stable/modules/svm
9. Sklearn.svm.LinearSVC.

   https://scikit-learn.org/0.20/modules/generated/sklearn.svm.LinearSVC.html
10. UC Irvine Machine Learning Repository. http://mlr.cs.umass.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data
11. UC Irvine Machine Learning Repository. http://mlr.cs.umass.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.names
12. Данные для задания на ТМШ 2014.

    http://www.machinelearning.ru /wiki/images/e/e1/School-VI-2014-task-3.rar

Статья представлена к публикации членом редколлегии В.И. Васильевым.

Поступила в редакцию 05.04.2019

После доработки 03.06.2019

Принята к публикации 18.07.2019

© 2020 г. С.И. УВАРОВ, канд. техн. наук (uvarov53@gmail.com)
(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

УСОВЕРШЕНСТВОВАННЫЙ ГЕНЕРАТОР 3-КНФ ФОРМУЛ

Рассматривается задача выполнимости (SAT-problem) булевых формул, заданных в конъюнктивной нормальной форме с ограничением, что каждый дизъюнкт содержит по три литерала переменных (3-КНФ). В эмпирических исследованиях широко используется генерация случайных формул с фиксированной длиной дизъюнкта. Феноменом этого метода является многократно подтвержденная линейная зависимость числа дизъюнктов формулы от числа булевых переменных в точке «фазового перехода» – от статуса выполнимых к статусу невыполнимых (когда доля невыполнимых формул становится превалирующей). Предложен и исследован метод генерации случайных формул, имеющий меньший коэффициент (3,49) пропорциональности между числом дизъюнктов и числом переменных в точке «фазового перехода» (для известного метода генерации этот коэффициент равен 4,23).

Ключевые слова: задача выполнимости (SAT-problem), конъюнктивная нормальная форма (КНФ), дизъюнкт, литерал, булевы переменные.

DOI: 10.31857/S0005231020010110

1. Введение

В работе исследуется генерация сложных примеров для задачи выполнимости логических формул (SAT-problem). Задача выполнимости является опорной для обширного класса NP-complete. Многие практически важные задачи несложными преобразованиями могут быть приведены к задаче выполнимости. Таковыми, например, являются задачи верификации аппаратуры и программного обеспечения, планирования, составления расписаний, комбинаторного анализа [1]. Важнейшим применением задачи выполнимости является автоматическое доказательство теорем – основа искусственного интеллекта.

Задача выполнимости относится к труднорешаемым задачам, при этом вопрос о возможности существования решения, алгоритмическая сложность которого ограничена некоторым полиномом от длины записи логической формулы, остается открытым.

Многие из встречающихся частных примеров в действительности оказываются достаточно простыми. Как правило, достаточно простыми являются и взятые из практики (индустриальные) примеры.

Поиск путей построения сложных примеров важен как для понимания природы сложности задачи, так и для построения тестовых примеров (benchmarks) с целью экспериментальной оценки алгоритмов [2]. Сложные

примеры активно способствуют выявлению недостатков в разработанных алгоритмах и тем самым указывают пути их дальнейшего совершенствования.

Считается, что относительно сложные примеры удается строить с помощью случайных чисел. Генерируемые при помощи случайных чисел формулы широко изучаются, поскольку это достаточно естественная модель, которая проливает свет на фундаментальные структурные свойства задачи выполнимости. Такие формулы хорошо отражают специфику задач в системах логических доказательств [3].

Первое известное значимое исследование было выполнено на формулах со случайной длиной дизъюнктов. Было продемонстрировано [4], что при некоторых параметрах генератора таких формул задача выполнимости для них может быть решена в среднем за $O(MN^2)$ шагов. Последующие исследования [2–6] показали, что семейство формул, со случайной длиной дизъюнктов, должно рассматриваться как простое. Простота анализа формул со случайной длиной дизъюнктов объясняется тем, что они часто содержат пустые дизъюнкты, дизъюнкты из единственного литерала (юниты) и тривиальные дизъюнкты. В настоящее время эмпирические исследования смещены к методам построения формул, в которых генерация таких простых дизъюнктов исключена [7, 8]. Подробное изложение результатов по рассматриваемой тематике содержится в [1].

Подавляющее большинство исследователей используют вариации генератора с фиксированной длиной дизъюнкта. Для 3-КНФ длина дизъюнкта равна трем. Литералы переменных, включенные в дизъюнкт, выбираются с использованием последовательности случайных чисел. Такой генератор хорош тем, что если подать на его вход нужную последовательность чисел, он способен породить любую формулу. Он очень удобен для теоретических исследований: его простота позволила аналитически доказать, что (в пределе) при увеличении количества переменных при $R \leqslant 3{,}52$ (R – отношение числа дизъюнктов к числу переменных) генератор порождает пренебрежимо малое число невыполнимых формул, а при $4{,}51 \leqslant R$ алгоритм порождает пренебрежимо малое число выполнимых формул [9, 10].

Известна интересная гипотеза [5] о том, что нетривиальные формулы, являющиеся невыполнимыми при меньшем числе ограничений (дизъюнктов), как правило являются более сложными для доказательства невыполнимости – анализа. Поэтому для построения сложных примеров представляет интерес поиск способов генерации формул, значительный процент которых является невыполнимыми при $R \leqslant 3{,}52$.

2. Генераторы случайных формул

Будем использовать следующие обозначения. $\mathbf{X}$ – множество из N булевых переменных X_i, $\mathbf{X} = \{X_1, \ldots, X_N\}$. Литералами переменной X_i назовем термы x_i^0 и x_i^1, где $x_i^0 = X_i$ и $x_i^1 = \neg X_i$.

Рассматриваются формулы $F(M) = C_1^3 \wedge C_2^3 \wedge \ldots \wedge C_M^3$, представленные в конъюнктивной нормальной форме 3-КНФ, являющиеся конъюнкцией набора из M трехлитеральных дизъюнктов C_j^3. Исследуется их выполнимость, т.е.

возможность отыскания набора значений булевых переменных, на которых формула принимает значение «истина».

Трехлитеральным дизъюнктом $C_j^3(x_k^p, x_\ell^q, x_m^r)$ является дизъюнкция трех литералов переменных X_ι из множества $\mathbf{X}$. Пример дизъюнкта с конкретными значениями верхних индексов литералов: $C_j^3(x_k^0, x_\ell^1, x_m^0) = x_k^0 \vee x_\ell^1 \vee x_m^0 = X_k \vee \neg X_\ell \vee X_m$.

Большое число публикаций, связанных с задачей выполнимости булевых формул, заданных в 3-КНФ, представляют результаты эмпирических исследований. Основным инструментом таких исследований является генератор случайных формул. Обычно для заданного числа N переменных желательно строить набор из V невыполнимых формул F_v $(v = 1, \ldots, V)$.

Формула, не имеющая дизъюнктов, выполнима. Искомая невыполнимая формула строится так, что, пока она остается выполнимой, к ранее построенным дизъюнктам добавляется новый. При получении невыполнимого набора дизъюнктов процесс генерации формулы заканчивается. После того, как формула стала невыполнимой, добавление новых дизъюнктов не может сделать ее выполнимой.

Формулу F_v удобно продуцировать на основе потока дизъюнктов. Принадлежащие такому потоку дизъюнкты строятся на основе одной последовательности $\mathbf{S}_v$ случайных натуральных чисел ξ_ν, предположительно равномерно распределенных в диапазоне от 1 до $2N(N-1)(N-2)$.

Разные формулы строятся на основе отличающихся друг от друга последовательностей случайных чисел.

Первый генератор с независимым выбором литералов (далее НВЛ-генератор).

Для построения очередного дизъюнкта C_{j+1} формулы F_v (в предположении, что при построении предшествующих дизъюнктов использованы ω чисел из $\mathbf{S}_v$) выбираются три последовательных случайных числа $\xi_{\omega+1}$, $\xi_{\omega+2}$, $\xi_{\omega+3}$. В произвольным образом упорядоченном множестве $\mathbf{A}$, состоящем из $2N$ литералов от N, переменных выбираем элементы, имеющие номера $\xi_{\omega+1} \pmod{2N}$, $\xi_{\omega+2} \pmod{2N}$ и $\xi_{\omega+3} \pmod{2N}$. Пусть этими элементами являются литералы x_k^p, x_ℓ^q и x_m^r, из них строим очередной дизъюнкт $C_{j+1}(x_k^p, x_\ell^q, x_m^r)$.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

При использовании такого генератора с небольшой вероятностью (порядка $1/N^2$) продуцируется дизъюнкт, содержащий более одного литерала одной переменной. В этом случае дизъюнкт будет тавтологией, если верхние индексы литералов, представляющих одну переменную, различны, либо будет дизъюнктом, в котором число различных литералов меньше трех. Такая ситуация уменьшает сложность анализа формулы, поэтому следует предпочесть следующий генератор.

Второй генератор формул без простых дизъюнктов (далее БПД-генератор) исключает возможность построения дизъюнкта, содержащего более одного литерала одной переменной.

Для построения очередного дизъюнкта C^3_{j+1} формулы F_v из $\mathbf{S}_v$ выбираются три последовательных случайных числа $\xi_{\omega+1}$, $\xi_{\omega+2}$, $\xi_{\omega+3}$. Затем в упорядоченном множестве $\mathbf{A}$ выбирается элемент x^p_k с номером $\xi_{\omega+1} \pmod{2N}$. Этот элемент полагается первым литералом дизъюнкта. Второй литерал назначается из множества $\mathbf{A} - \{x^0_k, x^1_k\}$, содержащего $2N-2$ элементов. Выбирается элемент x^q_ℓ с номером $\xi_{\omega+2} \pmod{2N-2}$. Третий литерал дизъюнкта выбирается из множества $\mathbf{A} - \{x^0_k, x^1_k, x^0_\ell, x^1_\ell\}$, содержащего $2N-4$ элементов. Это элемент с номером $\xi_{\omega+2} \pmod{2N-4}$. Выбранный элемент x^r_m становится третьим литералом дизъюнкта $C^3_{j+1}(x^p_k, x^q_\ell, x^r_m)$.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

Поскольку при больших N вероятность того, что дизъюнкт формулы, построенной первым генератором, содержит литералы одной переменной, невелика (порядка $1/N^2$), то результаты эмпирических исследований с этими генераторами близки.

Отметим, что существенная разница в частоте использования различных переменных обычно является одним их факторов упрощения анализа формул и часто используется алгоритмами решения задачи выполнимости (SAT-solver). С целью получения более сложных для анализа формул в настоящей публикации предлагается третий генератор, особенностью которого является выравнивание частоты использования в формуле литералов различных переменных.

Третий генератор, выравнивающий нагрузку на литералы (далее ВНЛ-генератор), использует статистическую информацию о количестве использования литералов переменных в построенных дизъюнктах.

Для построения очередного дизъюнкта C^3_{j+1} формулы F_v выбираем литерал, который минимальное число T раз встречается в ранее построенных дизъюнктах. Пусть это литерал x^t_f переменной X_f. Его полагаем первым литералом генерируемого дизъюнкта.

В отличие от первых двух, третий генератор при задании первого литерала очередного дизъюнкта не использует элементы последовательности $\mathbf{S}_v$ случайных чисел.

Для выбора второго и третьего литералов дизъюнкта используются числа $\xi_{\eta+1}$ и $\xi_{\eta+2}$ из $\mathbf{S}_v$ (η чисел использованы при построении предшествующих дизъюнктов формулы F_v).

Для генерации второго литерала дизъюнкта строится множество $\mathbf{B}$ литералов (переменных, отличных от переменной X_f), входящих в построенные дизъюнкты формулы не более чем $T+2$ раз. Если множество оказалось пустым, повторяем построение множества $\mathbf{B}$ с новым увеличенным на единицу значением T ($T = T+1$). Если множество не пусто, фиксируем число его элементов B и порядок элементов в множестве $\mathbf{B}$. Вторым литералом назначаем элемент x^q_ℓ, имеющий номер $\xi_{\eta+1} \pmod{B}$ в множестве $\mathbf{B}$.

При построении третьего литерала дизъюнкта снова строится множество $\mathbf{B}$ литералов (переменных, отличных от переменных X_f и X_ℓ), входящих в построенные дизъюнкты формулы не более чем $T+2$ раз. Если множество оказалось пустым, повторяем построение множества $\mathbf{B}$ с новым увеличенным

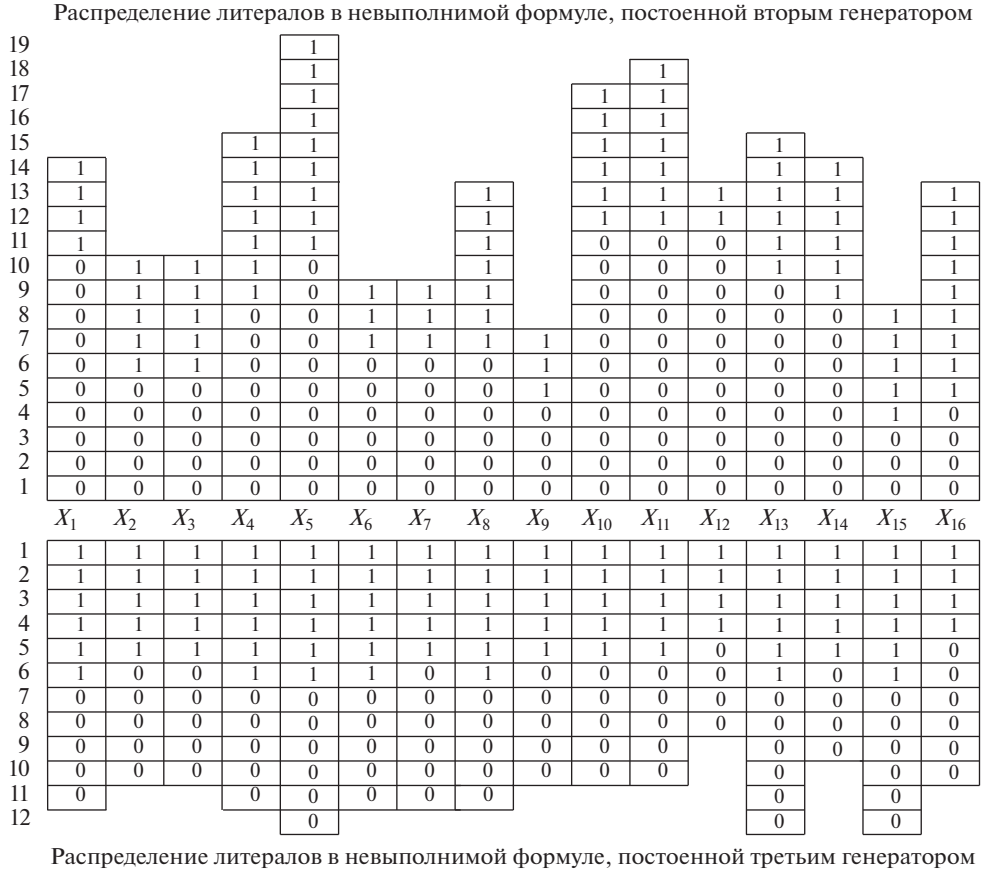


Рис. 1. Иллюстрация неравномерности использования литералов в 3-КНФ формулах, построенных БПД- и ВНЛ-генераторами.

на единицу значением T $(T = T + 1)$. Если множество не пусто, фиксируем число его элементов B и порядок элементов в множестве $\mathbf{B}$. Третьим литералом назначаем элемент x^r_m, имеющий номер $\xi_{\eta+2} \pmod B$ в множестве $\mathbf{B}$. Построение дизъюнкта $C^3_{j+1}(x^t_f, x^q_\ell, x^r_m)$ завершено.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

На рис. 1 представлены две гистограммы, иллюстрирующие количественное присутствие литералов 16 переменных в двух невыполнимых формулах, построенных ВНЛ-генератором (внизу) и БПД-генератором (вверху). Литералы, представляющие переменные без инверсии, обозначены нулем по значению верхнего индекса, а литералы, представляющие инвертированные переменные, обозначены единицей. В формуле от 16 переменных, построенной третьим генератором, число использованных литералов каждого типа отличается не более чем на два. В формуле от 16 переменных, построенной вторым генератором, 15-я переменная без инверсии используется трижды, а, например, 10-я переменная – одиннадцать раз.

Отметим, что все три описанных генератора могут породить формулу, дизъюнкты которой имеют одинаковые наборы литералов. Вероятность того, что в формуле из $M = kN$ дизъюнктов два будут содержать одинаковые наборы литералов, будет порядка $1/N^2$. Присутствие в формуле таких дизъюнктов не приводит к существенному упрощению ее анализа. Поэтому усложнение генераторов, направленное на исключение возможности генерации формул с дизъюнктами, содержащими идентичные наборы литералов, считается нецелесообразным.

3. Эмпирическое исследование генератора

Замечательным свойством второго БПД-генератора является соблюдаемая с высокой точностью линейная зависимость $\mathfrak{M}_2$ от N.

Если генерировать формулы содержащие по $\lfloor \mathfrak{M}(N) \rfloor$ дизъюнктов от N переменных, математическое ожидание вероятности выполнимости формул будет больше либо равно 0,5. Если генерировать формулы содержащие по $\lceil \mathfrak{M}(N) \rceil$ дизъюнктов от N переменных, математическим ожиданием вероятности невыполнимости будет величина большая, либо равная 0,5.

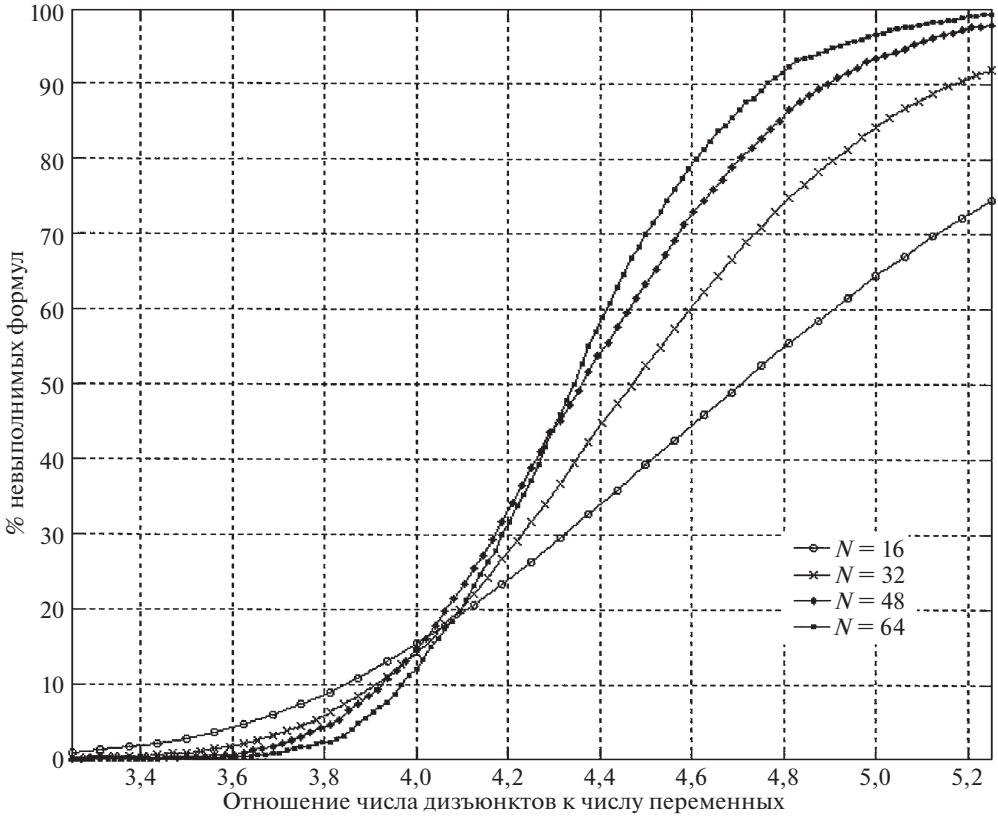


Рис. 2. Иллюстрация статистики перехода от статуса выполнимости к статусу невыполнимости для 3-КНФ формул, порожденных БПД-генератором.

Этот удивительный результат [6] до сих пор не нашел удовлетворительного объяснения.

Феномен линейной зависимости от N точки «фазового перехода» 3-КНФ формул (порожденных БПД-генератором) от статуса выполнимых к статусу невыполнимых был многократно подтвержден последующими публикациями, например [7, 8], и его статистическая достоверность не вызывает сомнения.

Для большей наглядности график процентной доли формул, ставших невыполнимыми после включения в их состав M дизъюнктов, строят относительно величины $R = M/N$, являющейся отношением числа дизъюнктов к числу переменных. На рис. 2 совместно представлены такие зависимости для формул с числом переменных 16, 32, 48, 64, построенных БПД-генератором. Результированы данные по 2000, 4000, 16000 и 32000 формулам для $N = 64$, $N = 48$, $N = 32$ и $N = 16$. В выбранных координатах на графике «фазовый переход» становится все более крутым с увеличением числа переменных.

Проведенные исследования [6–8] показали колоколообразную зависимость медианной сложности анализа формул от числа дизъюнктов. При этом максимальная медианная сложность имела место в непосредственной близости от точки «фазового перехода».

Медианная сложность в определенном смысле является усредненной сложностью анализа формул, содержащих фиксированное число дизъюнктов. В одной из пионерских работ [5], ориентированных на построение сложных для анализа формул, отмечено, что самыми сложными, как правило, оказываются формулы, являющиеся невыполнимыми при минимальном числе дизъюнктов.

Этот эмпирический результат объяснен [5] тем, что задача выполнимости, как и большинство других NP-comlete задач, может рассматриваться как поиск решения, удовлетворяющего определенным ограничениям. Интуитивно ясно, что если ограничений мало, то решение найти легко, так как при этом обычно имеет место множество возможных решений. Аналогично, если ограничений слишком много, достаточно интеллектуальный алгоритм обычно способен быстро отбрасывать большинство тупиковых ветвлений в дереве поиска. Таким образом, разумно ожидать, что самые сложные задачи – это задачи, которые, с одной стороны, не перегружены ограничениями, с другой стороны, имеющиеся ограничения оставляют возможность лишь для небольшого числа решений [6]. Эмпирическими исследованиями подтверждено, что это действительно имеет место [6–8].

Приведенные рассуждения являются мотивацией для поиска алгоритмов построения 3-КНФ формул, являющихся невыполнимыми при меньшем числе дизъюнктов, чем в невыполнимых формулах, построенных ставшими классическими НВЛ- и БПД-генераторами.

На рис. 3 представлены результаты исследований, для формул с числом переменных 16, 32, 48, 64, построенных ВНЛ-генератором. Для $N = 64$, $N = 48$, $N = 32$ и $N = 16$ построено 2000, 4000, 16000 и 32000 формул соответственно.

Как и для второго БПД-генератора, «фазовый переход» на графике рис. 3 становится все более крутым с увеличением числа переменных. Заметно,

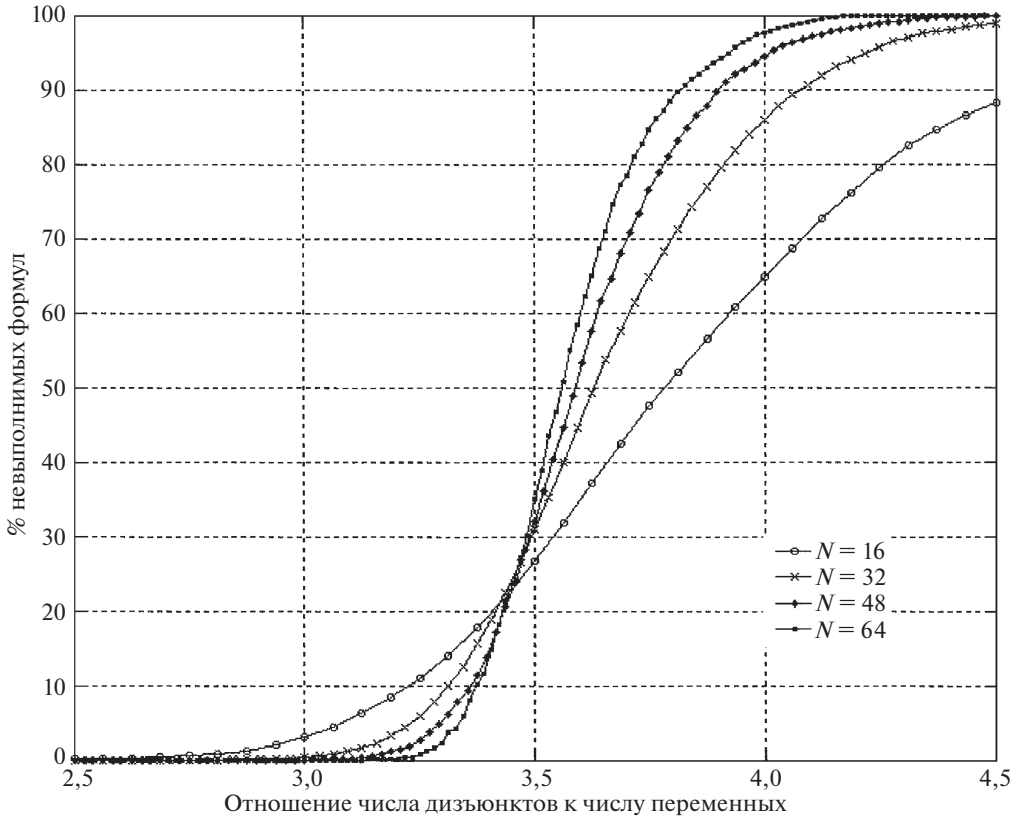


Рис. 3. Иллюстрация статистики перехода от статуса выполнимости к статусу невыполнимости для 3-КНФ формул, порожденных ВНЛ-генератором.

что при одинаковом числе переменных для формул, порожденных третьим ВНЛ-генератором, «фазовый переход» более крутой, чем для формул, порожденных БПД-генератором (см. рис. 2). Важно, что при заданном числе переменных «фазовый переход» для формул, построенных ВНЛ-генератором,

Таблица 1. Связь эмпирических вероятностей невыполнимости формулы со значениями $\lfloor \mathfrak{M}(N) \rfloor$ и $\lceil \mathfrak{M}(N) \rceil$

Второй генератор	Третий генератор
$\lfloor \mathfrak{M}_2(3) \rfloor = 19$ (47,835) $\lceil \mathfrak{M}_2(3) \rceil = 20$ (53,056)	$\lfloor \mathfrak{M}_3(3) \rfloor = 14$ (44,160) $\lceil \mathfrak{M}_3(3) \rceil = 15$ (51,900)
$\lfloor \mathfrak{M}_2(16) \rfloor = 75$ (48,919) $\lceil \mathfrak{M}_2(16) \rceil = 76$ (52,366)	$\lfloor \mathfrak{M}_3(16) \rfloor = 60$ (47,503) $\lceil \mathfrak{M}_3(16) \rceil = 61$ (52,031)
$\lfloor \mathfrak{M}_2(32) \rfloor = 143$ (48,756) $\lceil \mathfrak{M}_2(32) \rceil = 144$ (52,419)	$\lfloor \mathfrak{M}_3(32) \rfloor = 116$ (47,503) $\lceil \mathfrak{M}_3(32) \rceil = 117$ (53,656)
$\lfloor \mathfrak{M}_2(48) \rfloor = 209$ (49,000) $\lceil \mathfrak{M}_2(48) \rceil = 210$ (51,500)	$\lfloor \mathfrak{M}_3(48) \rfloor = 172$ (48,825) $\lceil \mathfrak{M}_3(48) \rceil = 173$ (53,350)
$\lfloor \mathfrak{M}_2(64) \rfloor = 278$ (49,850) $\lceil \mathfrak{M}_2(64) \rceil = 279$ (52,700)	$\lfloor \mathfrak{M}_3(64) \rfloor = 227$ (46,800) $\lceil \mathfrak{M}_3(64) \rceil = 228$ (50,700)

происходит при существенно меньшем соотношении числа дизъюнктов к числу переменных, чем для формул, построенных БПД-генератором.

Процент невыполнимых формул построенных БПД- и ВНЛ-генераторами при заданном числе переменных и заданном числе дизъюнктов, представлен в табл. 1.

При $N = 3$ вероятность того, что содержащая M дизъюнктов формула, порожденная БПД-генератором, будет невыполнимой, может быть вычислена аналитически. Если переменных всего три, то для получения невыполнимой формулы необходимо и достаточно, чтобы среди дизъюнктов формулы присутствовали восемь различных дизъюнктов, каждый из которых содержит литералы различных переменных. Если общее число дизъюнктов в формуле меньше восьми, формула выполнима.

Вероятность того, что формула, содержащая восемь случайных дизъюнктов, будет невыполнимой, равна $7!/8^7 = 0{,}0024$. При увеличении числа дизъюнктов в формуле вероятность того, что формула невыполнима, будет возрастать, стремясь к единице.

Как указано в [6], вероятность генерирования невыполнимой формулы, составленной из M дизъюнктов, соответствует вероятности того, что за M обращений к генератору случайных чисел с диапазоном (1:8) будут сгенерированы все восемь возможных чисел (т.е. в полученной последовательности будет представлено каждое из восьми чисел).

Обозначим через P_M^k вероятность того, что среди M дизъюнктов k будут различными. Вероятности P_M^k $(1 \leqslant k \leqslant 8)$ связаны рекуррентными соотношениями:

$$P_M^k = (k/8)P_{M-1}^k + ((9-k)/8)P_{M-1}^{k-1}$$

при начальных условиях $P_0^0 = 1, P_0^j = 0, j = \overline{1,8}$ и граничных условиях $P_M^0 = 0$.

Вычисления по этим формулам дают $P_{19}^8 = 0{,}478348$, а $P_{20}^8 = 0{,}530558$. Таким образом, для трех переменных точка «фазового перехода» находится между 19 и 20 дизъюнктами.

В ячейках табл. 1 в круглых скобках приведены процентные доли невыполнимых формул.

Аппроксимация методом наименьших квадратов данных, представленных в табл. 1, подтверждает хорошо известную для БПД-генератора линейную зависимость числа $\mathfrak{M}_2$ от числа переменных N.

$$\mathfrak{M}_2(N) = 4{,}23N + 7{,}18.$$

Аппроксимация представленных в табл. 1 эмпирических данных выявляет линейную зависимость $\mathfrak{M}_3$ от числа переменных N и для ВНЛ-генератора.

$$\mathfrak{M}_3(N) = 3{,}49N + 4{,}46.$$

Обе эти зависимости представлены на рис. 4. Полученные зависимости позволяют предположить, что величины 4,23 и 3,49, возможно, являются

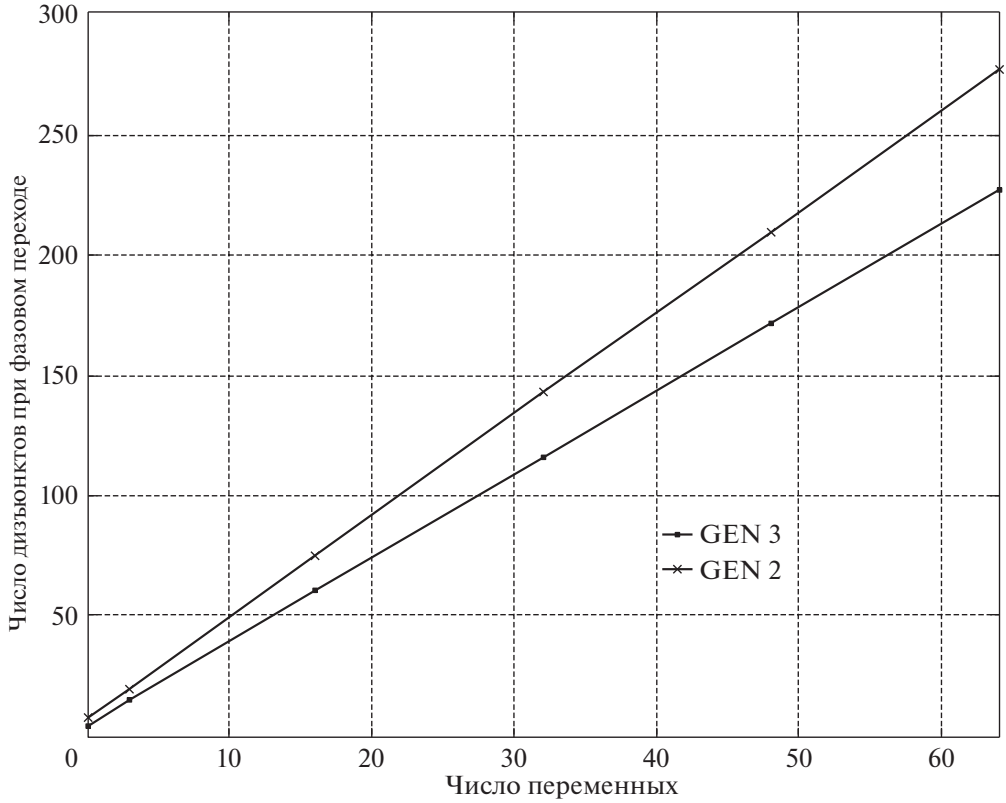


Рис. 4. Иллюстрация эмпирической линейной зависимости $\mathfrak{M}_2$ и $\mathfrak{M}_3$ от N.

асимптотами (при $N \to \infty$) для отношения числа дизъюнктов к числу переменных в точке «фазового перехода» для БПД- и ВНЛ-генераторов соответственно.

В табл. 2 представлены экспериментальные данные для зависимостей $\mathfrak{M}_2(N) = \alpha_2(N)N + 7{,}18$ и $\mathfrak{M}_3(N) = \alpha_3(N)N + 4{,}46$ при числах переменных, превышающих 64. Приведенные результаты позволяют оценить стабильность коэффициентов пропорциональности $\alpha_2(N)$ и $\alpha_3(N)$ при увеличении числа булевых переменных. Каждое из представленных значений получено на основе анализа 200 формул.

Значение коэффициента пропорциональности с вероятностью 0,99 находится в доверительном интервале $(\alpha_i(N) - \delta_i(N), \alpha_i(N) + \Delta_i(N))$, $i = 2, 3$.

Таблица 2. Зависимость коэффициентов пропорциональности от N

N	80	96	112	128	160	192	224	256	288	320
$\alpha_2(N)$	4,23	4,20	4,24	4,22	4,23	4,23	4,21	4,23	4,23	4,23
$\delta_2(N)$	0,04	0,04	0,05	0,04	0,03	0,04	0,02	0,02	0,02	0,02
$\Delta_2(N)$	0,09	0,05	0,03	0,06	0,05	0,04	0,03	0,03	0,03	0,03
$\alpha_3(N)$	3,52	3,49	3,51	3,50	3,51	3,50	3,50	3,49		
$\delta_3(N)$	0,03	0,03	0,02	0,01	0,01	0,02	0,01	0,01		
$\Delta_3(N)$	0,04	0,02	0,01	0,01	0,02	0,01	0,02	0,01		

Невыполнимые формулы полученные ВНЛ-генератором, в среднем существенно сложнее для анализа, чем невыполнимые формулы, полученные БПД-генератором. Анализ формул от 256 переменных, построенных ВНЛ-генератором, длится в среднем в 20 раз дольше, чем для формул, построенных БПД-генератором.

4. Заключение

Предложен и исследован ВНЛ-генератор (выравнивающий нагрузку на литералы) случайных формул 3-КНФ. Этот генератор продуцирует формулы, являющиеся невыполнимыми, при меньшем числе дизъюнктов, чем у широко используемого в настоящее время БПД-генератора (формул без простых дизъюнктов), для которого асимптотическое значение R в точке «фазового перехода» оценивается как 4,23.

Для предложенного в настоящей работе ВНЛ-генератора асимптотическое значение R (при $N \to \infty$) в точке «фазового перехода» оценивается как 3,49.

Характерная для НВЛ,- БПД- и ВНЛ-генераторов линейная зависимость числа дизъюнктов от числа переменных в точке «фазового перехода» удобна при построении тестовых примеров для испытания алгоритмов решения задачи выполнимости (SAT-problem). По заданному числу переменных легко вычисляется количество дизъюнктов, при котором больше половины из генерируемых случайных 3-КНФ формул будут выполнимыми, а при добавлении по одному дизъюнкту в каждую формулу больше половины формул станут невыполнимыми.

Проведенные исследования свидетельствуют в пользу выдвинутой в литературе [5–8] гипотезы о том, что анализ невыполнимых формул, имеющих меньшее число дизъюнктов, как правило, сложнее, чем анализ формул, содержащих больше дизъюнктов. Представляет интерес разработка способов построения формул, невыполнимых при еще меньшем числе дизъюнктов.

Представляет также интерес поиск зависимости количества дизъюнктов, минимально необходимого для невыполнимости нетривиальной 3-КНФ формулы, от числа булевых переменных.

СПИСОК ЛИТЕРАТУРЫ

1. *Biere A., Heule M., Maaren H., Walsh T.* Handbook of Satisfiability // IOS Press, 2009. P. 1–966.
2. *Cook S., Mitchell D.* Finding hard instances of the satisfiability problem: a survey // DIMACS Ser. Discret Math. Theoret. Comput. Sci., Amer. Math. Sos. 1997. V. 35. P. 1–17.
3. *Beame P., Karp R., Pitassi T., Saks M.* On the Complexity of Unsatisfiability Proofs for Random k-CNF Formulas // 30thSTOC, Dallas, TX. May 1998. P. 561–571.
4. *Goldberg A.* On the complexity of the satisfiability problem // Appl. Math. Comput. J. Amer. Math. Sos. 1997. V. 35. No. 1. P. 1–17.
5. *Mitchell D., Selman B., Levesque H.* Hard and easy distribution of SAT problem // Proc. Tenth National Conf. Artific. Intelligence (AAAI-92), San Jose, CA. 1997. P. 459–465.

6. *Crawford J., Auton I.* Experimental Results on the Crossover Point in Satisfiability Problems // Proc. AAAI-93, Washington, DC. 1993. P. 21–27.
7. *Gomes C., Kautz H., Sabharwal A., Selman B.* Satisfiability Solvers / Handbook Knowlege Represent., Elsevier B.V. 2008. P. 88–134.
8. *Heule M.* Minimal unsatisfiable cores of random formulas // Proc. SAT competition. 2013. P. 105.
9. *MohammadTghi Hajighayi, Gregory Sorkin.* The Satisfiability Threshold of Random 3-SAT Is at Least 3.52 // 2003/10/13. arXiv preprint math /0310193.
10. *Kaporis Alexis C., Kirousis Lefteris M., Laias Efthimios G.* The Probabilistic Analysis of a Greedy Satisfiability Algorithm // Random Struct. & Algorithms. 2006. V-28(40). P. 444–480.

Статья представлена к публикации членом редколлегии А.А. Лазаревым.

Поступила в редакцию 06.10.2017

После доработки 15.04.2019

Принята к публикации 18.07.2019

СОДЕРЖАНИЕ

CONTENTS

Linear Systems

Nonlinear Systems

Stochastic Systems

Control in Technical Systems

Optimization, System Analysis, and Operations Research