

ISSN 0005-2310

А Т АВТОМАТИКА И ЭЛЕМЕХАНИКА



7

2023



Российская Академия Наук

АТ АВТОМАТИКА И ТЕЛЕМЕХАНИКА

Журнал основан в 1936 году

Выходит 12 раз в год

7

ИЮЛЬ

Москва

2023

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)

Главный редактор:

Галяев А.А.

Заместители главного редактора:

Соболевский А.Н., Рубинович Е.Я., Хлебников М.В.

Ответственный секретарь:

Родионов И.В.

Редакционный совет:

Васильев С.Н., Желтов С.Ю., Каляев И.А., Кулешов А.П., Куржанский А.Б.,
Мартынюк А.А. (Украина), Пешехонов В.Г., Попков Ю.С.,
Федосов Е.А., Черноусько Ф.Л.

Редакционная коллегия:

Алескеров Ф.Т., Бахтадзе Н.Н., Бобцов А.А., Виноградов Д.В., Вишневский В.М.,
Воронцов К.В., Глумов В.М., Граничин О.Н., Губко М.В., Каравай М.Ф.,
Кибзун А.И., Краснова С.А., Красносельский А.М., Крищенко А.П.,
Кузнецов Н.В., Кузнецов О.П., Кушнер А.Г., Лазарев А.А., Ляхов А.И.,
Маликов А.И., Матасов А.И., Меерков С.М. (США), Миллер Б.М.,
Михальский А.И., Мунасыпов Р.А., Назин А.В., Немировский А.С. (США),
Новиков Д.А., Олейников А.Я., Пакшин П.В., Пальчунов Д.Е.,
Поляков А.Е. (Франция), Рапопорт Л.Б., Рублев И.В., Степанов О.А.,
Уткин В.И. (США), Фрадков А.Л., Хрусталеv М.М., Цыбаков А.Б. (Франция),
Чеботарев П.Ю., Щербаков П.С.

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 «Тематическая редакция»

Стохастические системы

© 2023 г. А.М. ГОРЦЕВ, д-р техн. наук (dekanat@fpmk.tsu.ru),
Л.А. НЕЖЕЛЬСКАЯ, д-р физ.-мат. наук (ludne@mail.ru)
(Национальный исследовательский Томский государственный университет)

АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ ОДНОЛИНЕЙНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ВХОДЯЩИМ МАР-ПОТОКОМ СОБЫТИЙ

Рассматривается однолинейная система массового обслуживания с входящим МАР-поток запросов (МАР-поток — Markovian Arrival Process) с двумя состояниями. Выводятся явные выражения для стационарного распределения вероятностей состояний и явные выражения для числовых характеристик системы: вероятности простоя обслуживающего прибора, математического ожидания числа запросов в системе, математического ожидания длины очереди. Численные результаты представлены в таблицах и в построенных на их основе графических зависимостях указанных характеристик, показанных на рисунках. Изучается рекуррентный МАР-поток с двумя состояниями как частный случай коррелированного МАР-потока запросов.

Ключевые слова: МАР-поток запросов, однолинейная система массового обслуживания (СМО), стационарное распределение вероятностей состояний системы, числовые характеристики.

DOI: 10.31857/S0005231023070012, EDN: FCERPV

1. Введение

Математические модели систем и сетей массового обслуживания (СМО, СеМО) достаточно адекватно описывают поведение реальных физических, технических, экономических и других объектов и систем и в связи с этим получили широкое распространение в научной среде. Одними из основных элементов СМО и СеМО являются случайные входящие потоки запросов. Практически на протяжении всего прошлого века исследования в области СМО и СеМО основывались на предположении о некоррелированном характере входящих потоков запросов, т.е. в качестве последних рассматривались простейшие потоки (стационарные пуассоновские потоки запросов). Однако начиная с конца 20-го века модель стационарного пуассоновского потока в связи с интенсивным развитием телекоммуникационных сетей и систем, беспроводных и мобильных сетей связи перестала быть адекватной реальным информационным потокам запросов в таких системах и сетях.

Быстрая смена цифровых технологий обеспечила проникновение цифровых сетей во все сферы человеческой деятельности. Все это было бы невозможным без использования и развития методов и алгоритмов математического моделирования сетевых технологий. Начиная с конца 20-го века в современной теории очередей начались интенсивные исследования нового направления — системы массового обслуживания с коррелированными потоками (системы с дважды стохастическими потоками). Стимулом к появлению дважды стохастических потоков событий — новой математической модели, наиболее адекватно учитывающей коррелированный характер реальных информационных потоков — послужили практические потребности исследований современных телекоммуникационных сетей, в которых разнородные информационные потоки являются существенно нестационарными и коррелированными.

Дважды стохастические потоки характеризуются двумя стохастиками: запросы в потоке наступают в случайные моменты времени (первая стохастика), интенсивность (сопровождающий процесс) потока является случайным процессом (вторая стохастика). Дважды стохастические потоки делятся на два типа: первый тип — потоки, сопровождающий процесс (интенсивность) которых есть непрерывный случайных процесс [1, 2]; второй тип — потоки, сопровождающий процесс (интенсивность) которых есть кусочно-постоянный случайный процесс с конечным (произвольным) числом состояний. Впервые результаты исследований потоков второго типа были опубликованы практически одновременно в 1979 г. в работах [3–5]. В [3, 4] указанные потоки получили название МС (Markov Chain)-потоки, в [5] — MVP (Markov Versatile Processes)-потоки. В [6, 7] описанные выше потоки названы МАР (Markovian Arrival Process)-потоками. Основным свойством введенных потоков является их коррелированность. Подчеркнем, что МАР (МС)-потоки являются наиболее подходящей математической моделью коррелированных потоков запросов в реальных телекоммуникационных системах и сетях [8].

В монографии [8], в своем роде единственной в мировой литературе, приведено систематизированное изложение СМО и СеМО с коррелированными потоками. В [8] подчеркивается, что аналитическое исследование СМО и СеМО с коррелированными потоками — достаточно затруднительный процесс, тем более нахождение характеристик СМО и СеМО в явном виде представляет собой сложную задачу, порой неразрешимую.

В настоящей статье проводится аналитическое исследование однолинейной СМО с ожиданием, входящим классическим МАР-потоком запросов с двумя состояниями [6, 7] и экспоненциальным обслуживанием.

Для стационарного режима функционирования СМО выводятся явные аналитические формулы для вероятности простоя обслуживающего прибора, средней длины очереди и среднего числа запросов в системе.

Отметим, что исследования, связанные с анализом СМО и СеМО с входящими МАР-потоками запросов, проводятся с 90-х гг. прошлого века до на-

стоящего времени. В частности, авторами настоящей статьи решены задачи по оценке состояний и параметров МАР-потока запросов в условиях его полной наблюдаемости, а также в условиях его неполной наблюдаемости (при наличии мертвого времени). По этому поводу здесь приводятся некоторые ссылки на публикации авторов [9–14].

Кроме того, отличие рассматриваемой системы от систем, функционирующих в синхронной случайной среде, состоит в том, что в синхронной среде рассматриваются синхронные потоки, у которых состояние управляющего процесса (сопровождающего процесса) изменяется в случайные моменты времени, являющиеся моментами наступления событий. Таким образом, синхронная случайная среда всегда предполагает отличную от нуля вероятность смены состояний управляющего процесса в момент наступления событий синхронного потока, в противоположность МАР-потоку: вероятность появления события потока в момент изменения состояния управляющего процесса может равняться нулю (если вероятность всегда равна единице, то имеет место синхронный поток). Таким образом, рассматриваемая в настоящей статье математическая модель случайной среды представляет собой обобщение математической модели синхронной случайной среды, в чем и состоит новизна проведенного исследования.

Эволюцию от простейшего потока к современным математическим моделям информационных потоков в телекоммуникационных системах и сетях — к моделям коррелированных потоков, в частности к МАР-потоку — можно проследить в упомянутой монографии [8], где, кроме того, приводится обширная библиография по исследуемой области моделей СМО и СеМО. Из последних работ по данной тематике отметим статью [15]. Подчеркнем, что общим для работ, в которых исследуются СМО и СеМО с входящим МАР-потоком запросов, является проводимый в них численный анализ систем и сетей обслуживания. В настоящей статье продолжаются исследования, начатые в [16].

2. Математическая модель системы. Постановка задачи

Изучается однолинейная СМО с ожиданием. На вход обслуживающего прибора поступает МАР-поток событий (запросов, сообщений и т.д.), сопровождающий процесс $\lambda(t)$ которого есть кусочно-постоянный случайный процесс с двумя состояниями — S_1 и S_2 . Будем говорить, что если $\lambda(t) = \lambda_i$, то имеет место i -е состояние (S_i) процесса $\lambda(t)$ (потока), $i = 1, 2$, $\lambda_1 > \lambda_2 > 0$. Длительность пребывания процесса $\lambda(t)$ в состоянии S_i есть случайная величина с экспоненциальной функцией распределения $F_i^*(t) = 1 - \exp\{-\lambda_i t\}$, $t \geq 0$, $i = 1, 2$.

В момент окончания i -го состояния потока (процесса $\lambda(t)$) возможны следующие мгновенные изменения состояния системы: 1) наступает событие потока, и процесс $\lambda(t)$ переходит из состояния S_i в состояние S_j ; совместная вероятность описанной ситуации есть $P_1(\lambda_j | \lambda_i)$, $i, j = 1, 2$; 2) не наступает

событие потока, и процесс $\lambda(t)$ переходит из состояния S_i в состояние S_j ; совместная вероятность этой ситуации есть $P_0(\lambda_j|\lambda_i)$, $i, j = 1, 2$; $i \neq j$. При этом $P_0(\lambda_j|\lambda_i) + P_1(\lambda_j|\lambda_i) + P_1(\lambda_i|\lambda_i) = 1$, $i, j = 1, 2$, $i \neq j$. Здесь первично наступление события в состоянии S_i , а затем переход из состояния потока S_i в состояние S_j с вероятностью $P_1(\lambda_j|\lambda_i)$, либо первично ненаступление события в состоянии S_i , а затем переход из состояния потока S_i в состояние S_j с вероятностью $P_0(\lambda_j|\lambda_i)$.

Рассматривается стационарный режим функционирования СМО. В сделанных предположениях $\lambda(t)$ — сопровождающий стационарный кусочно-постоянный транзитивный марковский процесс с двумя состояниями S_1 и S_2 . Если процесс $\lambda(t)$ находится в состоянии S_i , то длительность обслуживания сообщения на обслуживающем приборе распределена по экспоненциальному закону $F^{(i)}(\tau) = 1 - \exp\{-\mu_i\tau\}$, $\tau \geq 0$, с интенсивностью μ_i ($\mu_i > 0$), $i = 1, 2$.

Замечание 1. Сопровождающий случайный процесс $\lambda(t)$ для МАР-потока не совпадает с интенсивностью потока, так как в состоянии S_1 значение интенсивности потока есть $\lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, в состоянии S_2 соответственно $\lambda_2[1 - P_0(\lambda_1|\lambda_2)]$. Тогда средняя интенсивность МАР-потока равна [17]

$$\begin{aligned} \lambda &= \lambda_1[1 - P_0(\lambda_2|\lambda_1)]\pi_1 + \lambda_2[1 - P_0(\lambda_1|\lambda_2)]\pi_2, \\ (1) \quad \pi_1 &= \frac{\lambda_2[1 - P_1(\lambda_2|\lambda_2)]}{\lambda_1[1 - P_1(\lambda_1|\lambda_1)] + \lambda_2[1 - P_1(\lambda_2|\lambda_2)]}, \\ \pi_2 &= \frac{\lambda_1[1 - P_1(\lambda_1|\lambda_1)]}{\lambda_1[1 - P_1(\lambda_1|\lambda_1)] + \lambda_2[1 - P_1(\lambda_2|\lambda_2)]}, \end{aligned}$$

где π_1, π_2 — априорные вероятности состояний S_1, S_2 процесса $\lambda(t)$ (потока) в стационарном режиме.

Пусть: $\tau_k = t_{k+1} - t_k$, $k = 1, 2, \dots$, — значение длительности k -го интервала между моментами наступления запросов потока t_k и t_{k+1} ($\tau_k \geq 0$). Так как рассматривается стационарный режим, то плотность вероятности значений длительности k -го интервала равна $p(\tau_k) = p(\tau)$, $\tau \geq 0$, для любого $k \geq 1$. Тогда момент времени t_k без потери общности полагается равным нулю, т.е. момент наступления запроса есть $\tau = 0$. В [11] получено явное выражение для плотности вероятности $p(\tau)$:

$$\begin{aligned} p(\tau) &= \gamma z_1 e^{-z_1\tau} + (1 - \gamma) z_2 e^{-z_2\tau}, \quad \tau \geq 0, \\ \gamma &= \{z_2 - \lambda_1\pi_1(0)[1 - P_0(\lambda_2|\lambda_1)] - \lambda_2\pi_2(0)[1 - P_0(\lambda_1|\lambda_2)]\} (z_2 - z_1)^{-1}, \\ (2) \quad z_{1,2} &= \left[(\lambda_1 + \lambda_2) \mp \sqrt{(\lambda_1 - \lambda_2)^2 + 4\lambda_1\lambda_2 P_0(\lambda_1|\lambda_2) P_0(\lambda_2|\lambda_1)} \right] / 2, \\ \pi_1(0) &= \frac{P_1(\lambda_1|\lambda_2) + P_1(\lambda_1|\lambda_1)P_0(\lambda_1|\lambda_2)}{P_1(\lambda_1|\lambda_2) + P_1(\lambda_2|\lambda_1) + P_1(\lambda_1|\lambda_1)P_0(\lambda_1|\lambda_2) + P_1(\lambda_2|\lambda_2)P_0(\lambda_2|\lambda_1)}, \\ \pi_2(0) &= 1 - \pi_1(0). \end{aligned}$$

В (2) $\pi_i(0)$ — стационарная вероятность того, что процесс $\lambda(\tau)$ в момент времени $\tau = 0$ (в момент наступления запроса МАР-потока) находится в состоянии S_i , $i = 1, 2$; z_1, z_2 — корни характеристического уравнения $z^2 - (\lambda_1 + \lambda_2)z + \lambda_1\lambda_2[1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)] = 0$, при этом из (2) следует, что $0 < z_1 < z_2$, γ — величина, зависящая от параметров потока.

Рассмотрим два смежных интервала (t_k, t_{k+1}) , (t_{k+1}, t_{k+2}) , значения длительностей которых есть $\tau_k = t_{k+1} - t_k$, $\tau_{k+1} = t_{k+2} - t_{k+1}$ соответственно; их местоположение на временной оси в силу стационарности потока произвольно. Тогда, полагая $k = 1$, будем рассматривать два интервала (t_1, t_2) , (t_2, t_3) со значениями длительностей $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$; $\tau_1 \geq 0$, $\tau_2 \geq 0$. При этом $\tau_1 = 0$ соответствует моменту t_1 наступления запроса потока; $\tau_2 = 0$ соответствует моменту t_2 наступления следующего запроса потока. Совместная плотность вероятности при этом есть [11, 13]

$$(3) \quad p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2) + \gamma(1 - \gamma) \times \\ \times \frac{P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1)}{1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)} \times \\ \times (z_1e^{-z_1\tau_1} - z_2e^{-z_2\tau_1})(z_1e^{-z_1\tau_2} - z_2e^{-z_2\tau_2}), \quad \tau_1 \geq 0, \quad \tau_2 \geq 0,$$

где $z_1, z_2, p(\tau_k)$ определены в (2) для $\tau = \tau_k$, $k = 1, 2$.

Из (3) следует, что в общем случае МАР-поток является коррелированным потоком; только в частных случаях он переходит в рекуррентный либо вырождается в простейший.

Частный случай 1: $P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1) = 0$ — рекуррентный МАР-поток запросов с двумя состояниями. При этом $p(\tau)$ выражается формулой (2), где $\gamma = [z_2 - \lambda_1P_1(\lambda_1|\lambda_1) - \lambda_2P_1(\lambda_2|\lambda_2)](z_2 - z_1)^{-1}$.

Из (3) получаем $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2)$. Так как моменты наступления запросов в потоке t_1, \dots, t_k порождают вложенную цепь Маркова $\{\lambda(t_k)\}$, то можно показать, что для произвольного k , $k \geq 2$, $p(\tau_1, \dots, \tau_k) = p(\tau_1) \dots p(\tau_k)$.

Произведение $\gamma(1 - \gamma)$ из (3) представимо в виде

$$(4) \quad \gamma(1 - \gamma) = \frac{z_1z_2}{(z_2 - z_1)^2} \{ \lambda_1[1 - P_0(\lambda_2|\lambda_1)] - \lambda_2[1 - P_0(\lambda_1|\lambda_2)] \} \times \\ \times \{ \pi_1(0)\lambda_1[1 - P_1(\lambda_1|\lambda_1)] - \pi_2(0)\lambda_2[1 - P_1(\lambda_2|\lambda_2)] \} \times \\ \times \{ \lambda_1\lambda_2[1 - P_0(\lambda_2|\lambda_1)][1 - P_1(\lambda_2|\lambda_2)] + \lambda_1\lambda_2[1 - P_0(\lambda_1|\lambda_2)][1 - P_1(\lambda_1|\lambda_1)] \}^{-1}.$$

Из (4) вытекают частные случаи 2 и 3.

Частный случай 2: $\lambda_1[1 - P_0(\lambda_2|\lambda_1)] - \lambda_2[1 - P_0(\lambda_1|\lambda_2)] = 0$ — простейший поток с параметром z_1 . Из (2) находим $z_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\gamma = 1$; $p(\tau) = z_1e^{-z_1\tau}$, $\tau \geq 0$.

Частный случай 3: $\pi_1(0)\lambda_1[1 - P_1(\lambda_1|\lambda_1)] - \pi_2(0)\lambda_2[1 - P_1(\lambda_2|\lambda_2)] = 0$ — простейший поток с параметром z_1 . Из (2) находим $z_1 = \lambda_2[P_1(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_2)P_0(\lambda_2|\lambda_1)]$, $\gamma = 1$; $p(\tau) = z_1e^{-z_1\tau}$, $\tau \geq 0$.

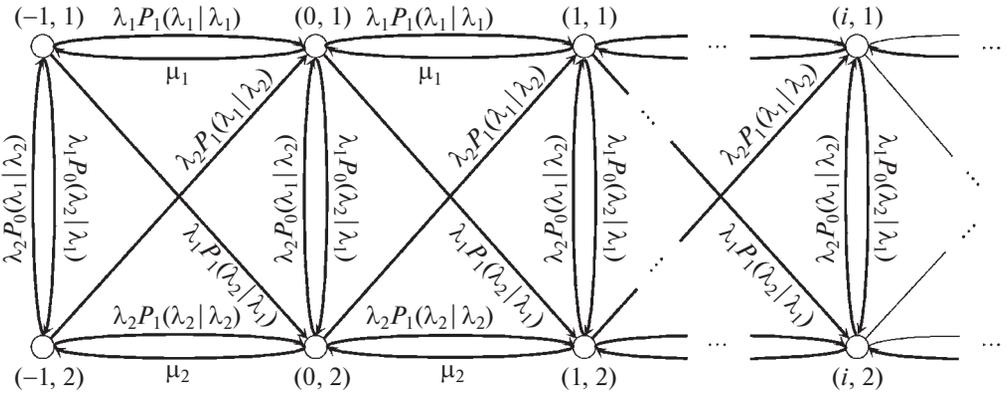


Рис. 1. Стохастический граф переходов процесса $\lambda(t)$ из состояния в состояние.

Задача анализа изучаемой СМО заключается в нахождении явного аналитического вида числовых характеристик системы: а) вероятности простоя обслуживающего прибора, б) среднего числа запросов в очереди, в) среднего числа запросов в системе.

Пусть $i(t)$ число запросов в очереди в произвольный момент времени t ($i(t) = 0, 1, \dots$). Так как входящий МАР-поток является коррелированным, то случайный процесс $i(t)$ не является марковским. Для того чтобы построить марковский процесс, необходимо учесть состояние входящего МАР-потока. Последнее обеспечивается введением дополнительной переменной $j(t)$ — состояния входящего МАР-потока (состояния сопровождающего процесса $\lambda(t)$ в произвольный момент времени t), $j(t) = 1, 2$. Если $j(t) = 1$, то $\lambda(t) = \lambda_1$; если $j(t) = 2$, то $\lambda(t) = \lambda_2$, что обеспечивает марковость двумерного процесса $(i(t), j(t))$.

Замечание 2. В силу того, что интенсивность обслуживания обслуживающим прибором в состоянии S_j есть μ_j ($\mu_j > 0$), $j = 1, 2$, то компонента $j(t)$ двумерного марковского процесса $(i(t), j(t))$ должна быть наблюдаемой так же, как наблюдаема компонента $i(t)$. Тогда сопровождающий процесс $\lambda(t)$, который в общем случае является ненаблюдаемым, нужно рассматривать в виде наблюдаемого процесса, управляющего сменой состояний МАР-потока запросов.

Обозначим состояние системы (в силу того, что рассматривается стационарный режим ее функционирования) как (i, j) , $i = 0, 1, \dots$, $j = 1, 2$. Здесь возможны еще два состояния $(-1, 1)$ и $(-1, 2)$, при которых запросы в системе отсутствуют (длина очереди равна нулю и обслуживающий прибор простаивает).

Сделанные выше предпосылки дают возможность представления математической модели изучаемой СМО в виде связанного стохастического графа [18], представленного на рис. 1. Здесь вершинам графа соответствуют состояния СМО; каждой дуге графа поставлены в соответствие инфинитезимальные

характеристики (интенсивности переходов из состояния в состояние), причем петли в каждом состоянии опущены; каждая вершина графа (каждое состояние) достижима и возвратна.

3. Вывод числовых характеристик системы

Обозначим через $P(i, 1)$, $P(i, 2)$ стационарные (финальные) вероятности состояний системы ($i = -1, 0, \dots$). Для сечений стохастического графа $G_{i1} = \{(i - 1, 1; i, 1), (i, 1; i - 1, 1), (i, 1; i + 1, 1), (i + 1, 1; i, 1), (i, 1; i, 2), (i, 2; i, 1), (i - 1, 2; i, 1), (i, 1; i + 1, 2)\}$, $G_{i2} = \{(i - 1, 2; i, 2), (i, 2; i - 1, 2), (i, 2; i + 1, 2), (i + 1, 2; i, 2), (i, 2; i, 1), (i, 1; i, 2), (i - 1, 1; i, 2), (i, 2; i + 1, 1)\}$, $i = 0, 1, \dots$, справедлива бесконечная система разностных уравнений с постоянными коэффициентами:

$$(5) \quad \begin{aligned} & \mu_1 P(i + 1, 1) - (\lambda_1 + \mu_1) P(i, 1) + \lambda_1 P_1(\lambda_1 | \lambda_1) P(i - 1, 1) + \\ & \quad + \lambda_2 P_0(\lambda_1 | \lambda_2) P(i, 2) + \lambda_2 P_1(\lambda_1 | \lambda_2) P(i - 1, 2) = 0, \\ & \mu_2 P(i + 1, 2) - (\lambda_2 + \mu_2) P(i, 2) + \lambda_2 P_1(\lambda_2 | \lambda_2) P(i - 1, 2) + \\ & \quad + \lambda_1 P_0(\lambda_2 | \lambda_1) P(i, 1) + \lambda_1 P_1(\lambda_2 | \lambda_1) P(i - 1, 1) = 0, \quad i = 0, 1, \dots \end{aligned}$$

Решение системы (5) ищется в виде $P(i, 1) = \xi^i$, $P(i, 2) = C \xi^i$ ($i = 0, 1, \dots$). При этом характеристическое уравнение для (5) выпишется в виде

$$(6) \quad \begin{aligned} & (\xi - 1) \left\{ \mu_1 \mu_2 \xi^3 - [\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)] \xi^2 + \right. \\ & \quad + [\lambda_1 \lambda_2 + \lambda_1 \mu_2 P_1(\lambda_1 | \lambda_1) + \lambda_2 \mu_1 P_1(\lambda_2 | \lambda_2) - \lambda_1 \lambda_2 P_0(\lambda_1 | \lambda_2) P_0(\lambda_2 | \lambda_1)] \xi - \\ & \quad \left. - \lambda_1 \lambda_2 [P(\lambda_1 | \lambda_1) P_1(\lambda_2 | \lambda_2) - P_1(\lambda_1 | \lambda_2) P_1(\lambda_2 | \lambda_1)] \right\} = 0. \end{aligned}$$

Рассмотрим условия существования стационарного режима функционирования СМО (условия существования вероятностей $P(i, 1)$, $P(i, 2)$, $i = -1, 0, \dots$). Математическое ожидание случайной величины τ — длительности интервала между соседними событиями в МАР-потоке запросов — определится в виде

$$(7) \quad E(\tau) = \int_0^{\infty} \tau p(\tau) d\tau,$$

где плотность $p(\tau)$ задана в (2). Подставляя ее в (7), находим $E(\tau) = [\gamma z_2 + (1 - \gamma) z_1] / z_1 z_2$. Тогда среднее число запросов во входящем коррелированном МАР-потоке в единицу времени запишется в виде $\lambda = 1/E(\tau) = \lambda_1 [1 - P_0(\lambda_2 | \lambda_1)] \pi_1 + \lambda_2 [1 - P_0(\lambda_1 | \lambda_2)] \pi_2$, что совпадает с (1). С другой стороны, математическое ожидание числа обслуженных в единицу времени запросов есть $\mu = \mu_1 \pi_1 + \mu_2 \pi_2$.

Рассмотрим ситуацию, когда $\lambda = \mu$, или $(\mu_1 - \lambda_1[1 - P_0(\lambda_2|\lambda_1)])\pi_1 + (\mu_2 - \lambda_2[1 - P_0(\lambda_1|\lambda_2)])\pi_2 = 0$. Из последнего соотношения следует, что равенство нулю возможно только в случае, когда $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$. Подставляя эти выражения для μ_1 и μ_2 в (6), находим характеристическое уравнение для рассматриваемой ситуации:

$$(8) \quad \begin{aligned} & \lambda_1\lambda_2(\xi - 1)^2 \left\{ [1 - P_0(\lambda_1|\lambda_2)][1 - P_0(\lambda_2|\lambda_1)]\xi^2 - \right. \\ & \quad \left. - [2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)]\xi + \right. \\ & \quad \left. + [P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1)] \right\} = 0. \end{aligned}$$

Так как характеристическое уравнение (8) имеет кратные корни, то общее решение системы (5), в которой $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$, выразится в виде

$$(9) \quad \begin{aligned} P(i, 1) &= D_1\xi_1^i + D_2i\xi_2^i + D_3\xi_3^i + D_4\xi_4^i, \\ P(i, 2) &= B_1D_1\xi_1^i + B_2D_2i\xi_2^i + B_3D_3\xi_3^i + B_4D_4\xi_4^i, \quad i = 0, 1, \dots \end{aligned}$$

В (9) $P_s(i, 1) = D_s\xi_s^i$; $P_s(i, 2) = B_sD_s\xi_s^i$ — частные решения системы (5); B_s, D_s — константы, определяемые из граничных условий, $s = \overline{1, 4}$, $\xi_1 = \xi_2 = 1$,

$$(10) \quad \begin{aligned} \xi_{3,4} &= \left\{ [2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)] \mp \right. \\ & \left. \mp \left([2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)]^2 - 4[1 - P_0(\lambda_1|\lambda_2)][1 - P_0(\lambda_2|\lambda_1)]b \right)^{\frac{1}{2}} \right\} \times \\ & \quad \times \left\{ 2[1 - P_0(\lambda_1|\lambda_2)][1 - P_0(\lambda_2|\lambda_1)] \right\}^{-1}, \\ b &= P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1). \end{aligned}$$

Здесь возможны три случая: $b > 0$, $b < 0$, $b = 0$.

Рассмотрим случай $b > 0$. Тогда из (10) следует $0 < \xi_3 < 1 < \xi_4$. Поскольку $P(i, 1), P(i, 2)$ — вероятности, то для них должно выполняться условие нормировки

$$\sum_{i=-1}^{\infty} P(i, 1) + \sum_{i=-1}^{\infty} P(i, 2) = 1.$$

Необходимым условием выполнения последнего равенства является выполнение предельных соотношений: $\lim_{i \rightarrow \infty} P(i, 1) = 0$, $\lim_{i \rightarrow \infty} P(i, 2) = 0$ при $i \rightarrow \infty$.

В противном случае ряды $\sum_{i=-1}^{\infty} P(i, 1)$, $\sum_{i=-1}^{\infty} P(i, 2)$ будут расходящимися.

С учетом сказанного общее решение (9), в котором $D_1 = D_2 = D_4 = 0$, принимает вид

$$(11) \quad P(i, 1) = D_3\xi_3^i, \quad P(i, 2) = B_3D_3\xi_3^i, \quad i = 0, 1, \dots$$

Найдем константу B_3 . Подставим (11) в первое уравнение системы (5), в котором $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$. После достаточно трудоемких преобразований находим $B_3 < 0$. Тогда из (11) следует, что $D_3 < 0$. Неравенство $D_3 < 0$ приводит к противоречию: $P(i, 1) < 0$, $i \geq 0$; $P(i, 2) > 0$, $i \geq 0$. Если положить $D_3 = 0$, то $P(i, 1) = P(i, 2) = 0$, $i \geq 0$, т.е. противоречие устраняется. Отсюда следует, что при $\lambda = \mu$ финальное распределение $P(i, 1)$, $P(i, 2)$, $i \geq 0$, не существует и тем более не существует при $\lambda > \mu$.

Изучим ситуацию $\lambda < \mu$. С учетом (6) общее решение системы (5) принимает вид

$$(12) \quad \begin{aligned} P(i, 1) &= A_1\xi_1^i + A_2\xi_2^i + A_3\xi_3^i + A_4\xi_4^i, \\ P(i, 2) &= C_1A_1\xi_1^i + C_2A_2\xi_2^i + C_3A_3\xi_3^i + C_4A_4\xi_4^i, \quad i = 0, 1, \dots, \end{aligned}$$

где $P_s(i, 1) = A_s\xi_s^i$; $P_s(i, 2) = C_sA_s\xi_s^i$ — частные решения системы (5); C_s, A_s — константы, находящиеся из граничных условий, $s = \overline{1, 4}$; $\xi_4 = 1$; ξ_1, ξ_2, ξ_3 — корни кубического уравнения из (6), являющиеся вещественными и положительными: $0 < \xi_1 < \xi_2 < 1 < \xi_3$. При этом необходимо выполнение предельных соотношений $\lim P(i, 1) = \lim P(i, 2) = 0$ при $i \rightarrow \infty$. Откуда следует, что $A_3 = A_4 = 0$. Тогда общее решение (12) выразится в виде

$$(13) \quad \begin{aligned} P(i, 1) &= A_1\xi_1^i + A_2\xi_2^i, \\ P(i, 2) &= C_1A_1\xi_1^i + C_2A_2\xi_2^i, \quad i = 0, 1, \dots. \end{aligned}$$

Подставив частное решение $P_s(i, 1) = A_s\xi_s^i$, $P_s(i, 2) = C_sA_s\xi_s^i$, $i = 0, 1, \dots$, в первое уравнение системы (5) сначала для $s = 1$, затем для $s = 2$, получаем константу C_s в виде

$$(14) \quad C_s = -\frac{\mu_1\xi_s^2 - (\lambda_1 + \mu_1)\xi_s + \lambda_1P_1(\lambda_1|\lambda_1)}{\lambda_2[P_0(\lambda_1|\lambda_2)\xi_s + P_1(\lambda_1|\lambda_2)]}, \quad s = 1, 2.$$

Для нахождения величин A_i , $i = 1, 2$, и вероятностей $P(-1, 1)$, $P(-1, 2)$ привлекаются граничные уравнения и условие нормировки. Сечения стохастического графа

$$\begin{aligned} G_{-1,1} &= \{(-1, 1; 0, 1), (0, 1; -1, 1), (-1, 1; 0, 2), (-1, 1; -1, 2), (-1, 2; -1, 1)\}, \\ G_{-1,2} &= \{(-1, 2; 0, 2), (0, 2; -1, 2), (-1, 2; 0, 1), (-1, 2; -1, 1), (-1, 1; -1, 2)\}, \\ G &= \{(i, 1; i + 1, 2), (i, 1; i, 2), (i, 2; i + 1, 1), (i, 2; i, 1), \quad i = -1, 0, 1, \dots\} \end{aligned}$$

соответственно определяют граничные уравнения:

$$(15) \quad \begin{aligned} \mu_1P(0, 1) - \lambda_1P(-1, 1) + \lambda_2P_0(\lambda_1|\lambda_2)P(-1, 2) &= 0, \\ \mu_2P(0, 2) - \lambda_2P(-1, 2) + \lambda_1P_0(\lambda_2|\lambda_1)P(-1, 1) &= 0, \\ \lambda_1[1 - P_1(\lambda_1|\lambda_1)] \sum_{i=-1}^{\infty} P(i, 1) - \lambda_2[1 - P_1(\lambda_2|\lambda_2)] \sum_{i=-1}^{\infty} P(i, 2) &= 0. \end{aligned}$$

Присоединяя к (15) условие нормировки

$$P(-1, 1) + P(-1, 2) + \sum_{i=0}^{\infty} [P(i, 1) + P(i, 2)] = 1,$$

с учетом (13) получаем систему уравнений для нахождения неизвестных A_i , $i = 1, 2$, $P(-1, 1)$, $P(-1, 2)$. Решая (15), находим

$$(16) \quad P(-1, 1) = a_{11}A_1 + a_{12}A_2, \quad P(-1, 2) = a_{21}A_1 + a_{22}A_2,$$

$$A_1 =$$

$$= (1 - \xi_1) \frac{\pi_1 [C_2 + a_{22}(1 - \xi_2)] - \pi_2 [1 + a_{12}(1 - \xi_2)]}{[1 + a_{11}(1 - \xi_1)][C_2 + a_{22}(1 - \xi_2)] - [1 + a_{12}(1 - \xi_2)][C_1 + a_{21}(1 - \xi_1)]},$$

$$A_2 =$$

$$= -(1 - \xi_2) \frac{\pi_1 [C_1 + a_{21}(1 - \xi_1)] - \pi_2 [1 + a_{11}(1 - \xi_1)]}{[1 + a_{11}(1 - \xi_1)][C_2 + a_{22}(1 - \xi_2)] - [1 + a_{12}(1 - \xi_2)][C_1 + a_{21}(1 - \xi_1)]},$$

$$a_{11} = \frac{\mu_1 + \mu_2 P_0(\lambda_1 | \lambda_2) C_1}{\lambda_1 [1 - P_0(\lambda_1 | \lambda_2) P_0(\lambda_2 | \lambda_1)]}, \quad a_{12} = \frac{\mu_1 + \mu_2 P_0(\lambda_1 | \lambda_2) C_2}{\lambda_1 [1 - P_0(\lambda_1 | \lambda_2) P_0(\lambda_2 | \lambda_1)]},$$

$$a_{21} = \frac{\mu_2 C_1 + \mu_1 P_0(\lambda_2 | \lambda_1)}{\lambda_2 [1 - P_0(\lambda_1 | \lambda_2) P_0(\lambda_2 | \lambda_1)]}, \quad a_{22} = \frac{\mu_2 C_2 + \mu_1 P_0(\lambda_2 | \lambda_1)}{\lambda_2 [1 - P_0(\lambda_1 | \lambda_2) P_0(\lambda_2 | \lambda_1)]}.$$

Величины C_1 , C_2 определены в (14); вероятности π_1 , π_2 — в (1); ξ_1 , ξ_2 — корни кубического уравнения в (6) ($0 < \xi_1 < \xi_2 < 1$).

Формулы (13), (16) позволяют найти явные выражения для числовых характеристик системы: $P(-1)$ — вероятность простоя обслуживающего прибора; $E(I)$ — средняя длина очереди; $E(I + 1)$ — среднее число запросов в системе, где I — случайная величина длины очереди в СМО.

$$P(-1) = (a_{11} + a_{21})A_1 + (a_{12} + a_{22})A_2,$$

$$(17) \quad E(I) = A_1(1 + C_1) \frac{\xi_1}{(1 - \xi_1)^2} + A_2(1 + C_2) \frac{\xi_2}{(1 - \xi_2)^2},$$

$$E(I + 1) = \frac{A_1(1 + C_1)}{(1 - \xi_1)^2} + \frac{A_2(1 + C_2)}{(1 - \xi_2)^2},$$

где C_1 , C_2 определены в (14); A_1 , A_2 , a_{11} , a_{21} , a_{12} , a_{22} — в (16); ξ_1 , ξ_2 — корни кубического уравнения в (6) ($0 < \xi_1 < \xi_2 < 1$).

Приведенные здесь и ниже исходные данные для расчета числовых характеристик (17) выбраны таким образом, чтобы показать, насколько поведение последних соответствует физическим представлениям о процессе обслуживания в изучаемой СМО.

Таблица 1. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b > 0$

$P_1(\lambda_1 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,780	0,718	0,680	0,651	0,627	0,616
1/6	0,787	0,728	0,693	0,667	0,645	0,636
1/8	0,790	0,734	0,700	0,675	0,654	0,645
1/10	0,792	0,737	0,704	0,679	0,659	0,651
1/12	0,794	0,739	0,706	0,682	0,663	0,655
1/13	0,794	0,739	0,707	0,684	0,664	0,656

Таблица 2. Зависимость средней длины очереди $E(I)$ от λ_1 для $b > 0$

$P_1(\lambda_1 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,052	0,097	0,145	0,196	0,249	0,276
1/6	0,047	0,085	0,125	0,167	0,209	0,231
1/8	0,045	0,080	0,116	0,153	0,191	0,210
1/10	0,043	0,076	0,110	0,145	0,180	0,198
1/12	0,042	0,074	0,106	0,140	0,173	0,190
1/13	0,042	0,073	0,105	0,138	0,171	0,187

Таблица 3. Зависимость среднего числа запросов в системе $E(I + 1)$ от λ_1 для $b > 0$

$P_1(\lambda_1 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,272	0,379	0,465	0,545	0,622	0,659
1/6	0,260	0,357	0,432	0,500	0,564	0,595
1/8	0,254	0,346	0,416	0,478	0,537	0,565
1/10	0,251	0,340	0,406	0,466	0,521	0,547
1/12	0,249	0,335	0,400	0,457	0,510	0,536
1/13	0,248	0,334	0,398	0,454	0,506	0,531

В табл. 1–3 приведены зависимости характеристик $P(-1)$, $E(I)$, $E(I + 1)$ от параметра λ_1 ($\lambda_1 = 2, 4, \dots, 10, 11$) при фиксированных значениях параметров $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$; $P_1(\lambda_1|\lambda_1) = P_1(\lambda_2|\lambda_1) = P_0(\lambda_2|\lambda_1) = P_1(\lambda_2|\lambda_2) = \frac{1}{3}$ для $b > 0$ и $P_1(\lambda_1|\lambda_2) = \frac{1}{4}$ ($P_0(\lambda_1|\lambda_2) = \frac{5}{12}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{6}$ ($P_0(\lambda_1|\lambda_2) = \frac{1}{2}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{8}$ ($P_0(\lambda_1|\lambda_2) = \frac{13}{24}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{10}$ ($P_0(\lambda_1|\lambda_2) = \frac{17}{30}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{12}$ ($P_0(\lambda_1|\lambda_2) = \frac{7}{12}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{13}$ ($P_0(\lambda_1|\lambda_2) = \frac{23}{39}$), вычисленные по формулам (17).

Поведение отмеченных числовых характеристик (17) в зависимости от параметра λ_1 при $b > 0$ соответствует физическим представлениям о процессе обслуживания в изучаемой однолинейной СМО с входящим коррелированным МАР-потокком запросов.

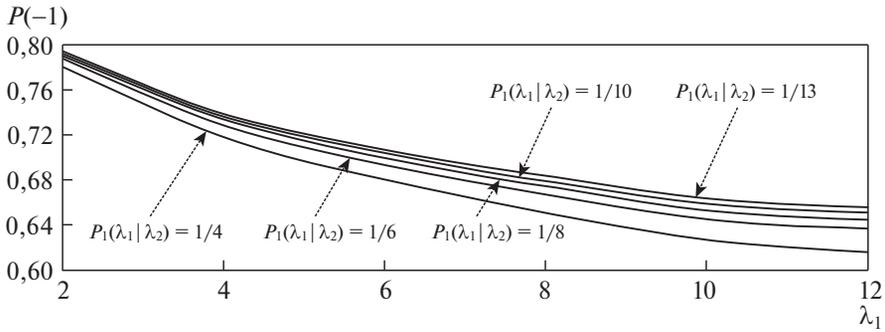


Рис. 2. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b > 0$.

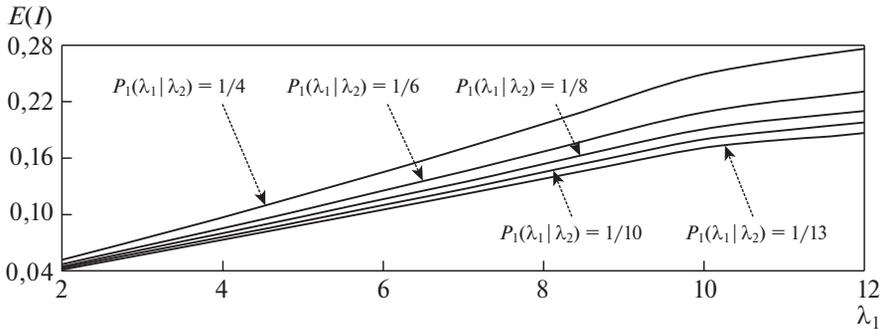


Рис. 3. Зависимость средней длины очереди $E(I)$ от λ_1 для $b > 0$.

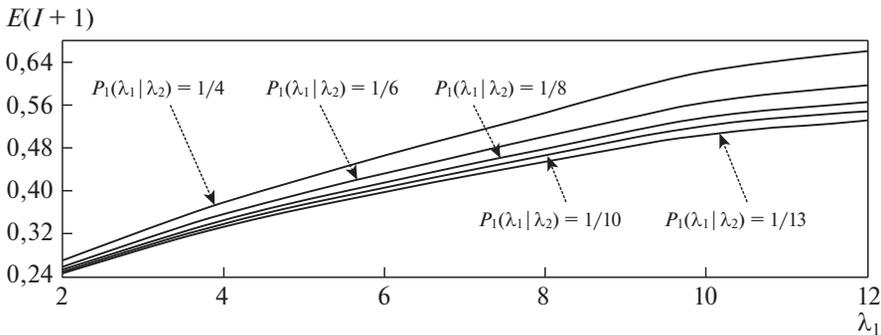


Рис. 4. Зависимость среднего числа запросов в системе $E(I+1)$ от λ_1 для $b > 0$.

На рис. 2–4 приведены графики числовых характеристик (17), построенные для численных значений табл. 1–3 соответственно.

Рассмотрим случай $b < 0$. В первую очередь изучим вопрос о существовании стационарного режима, т.е. ситуацию, когда $\lambda = \mu$. Тогда из (10) следует, что $\xi_3 < 0$, $\xi_4 > 1$ и, аналогично случаю $b > 0$, общее решение системы выписывается в виде (11). Так как $\xi_3 < 0$, то этот факт влечет за собой отрицательность вероятности $P(i, 1)$ для $i = 1, 3, \dots$, что противоречит ее определению; чтобы снять это противоречие, нужно положить $D_3 = 0$: $P(i, 1) = P(i, 2) = 0$,

Таблица 4. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b < 0$

$P_1(\lambda_2 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,796	0,729	0,687	0,654	0,627	0,615
1/6	0,815	0,748	0,705	0,672	0,645	0,633
1/8	0,824	0,757	0,714	0,681	0,654	0,642
1/10	0,830	0,762	0,719	0,686	0,659	0,647
1/12	0,833	0,765	0,722	0,690	0,662	0,650
1/13	0,834	0,767	0,724	0,691	0,664	0,652

Таблица 5. Зависимость средней длины очереди $E(I)$ от λ_1 для $b < 0$

$P_1(\lambda_2 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,044	0,091	0,143	0,200	0,260	0,291
1/6	0,035	0,077	0,124	0,177	0,232	0,261
1/8	0,031	0,071	0,116	0,166	0,220	0,248
1/10	0,029	0,067	0,112	0,161	0,213	0,241
1/12	0,027	0,065	0,109	0,157	0,209	0,236
1/13	0,027	0,064	0,108	0,156	0,207	0,234

Таблица 6. Зависимость среднего числа запросов в системе $E(I + 1)$ от λ_1 для $b < 0$

$P_1(\lambda_2 \lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0,249	0,362	0,457	0,546	0,633	0,675
1/6	0,219	0,329	0,419	0,504	0,587	0,528
1/8	0,206	0,314	0,402	0,485	0,566	0,606
1/10	0,199	0,305	0,392	0,474	0,554	0,594
1/12	0,194	0,300	0,386	0,467	0,547	0,586
1/13	0,193	0,298	0,384	0,465	0,544	0,583

$i \geq 0$. Последнее означает, что для случая $b < 0$ финальное распределение $P(i, 1)$, $P(i, 2)$, $i \geq 0$, не существует при $\lambda = \mu$ и тем более не существует при $\lambda > \mu$.

Изучим ситуацию $\lambda < \mu$. С учетом (6) общее решение системы (5) выразится в виде (12). Для рассматриваемого случая $b < 0$ имеет место $\xi_4 = 1$; ξ_1, ξ_2, ξ_3 — корни кубического уравнения из (6), являющиеся вещественными: $\xi_1 < 0$, $0 < \xi_2 < 1 < \xi_3$. Отсюда следует, что константы A_1, A_3, A_4 в (12) полагаются равными нулю, при этом общее решение (12) выпишется в виде

$$(18) \quad P(i, 1) = A_2 \xi_2^i, \quad P(i, 2) = C_2 A_2 \xi_2^i, \quad i = 0, 1, \dots$$

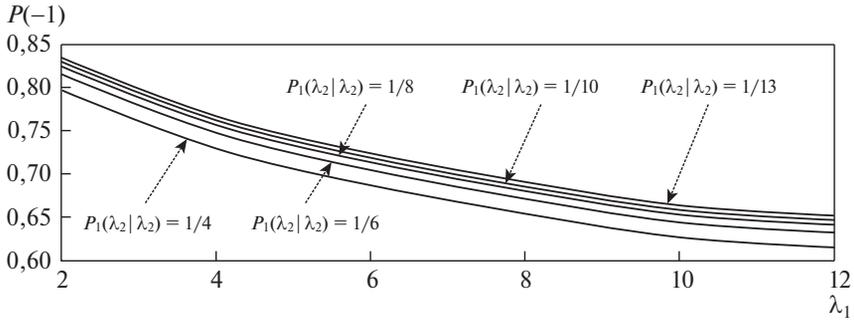


Рис. 5. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b < 0$.

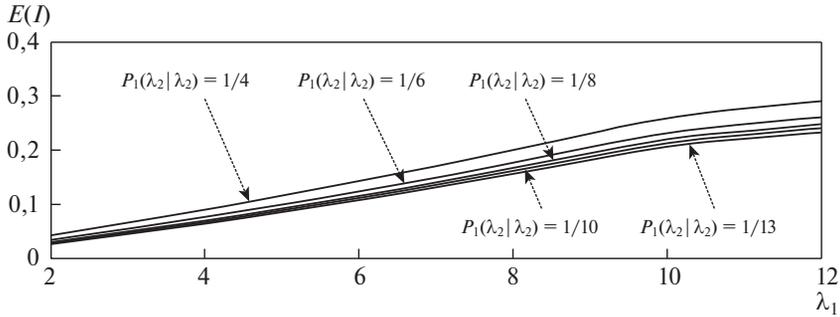


Рис. 6. Зависимость средней длины очереди $E(I)$ от λ_1 для $b < 0$.

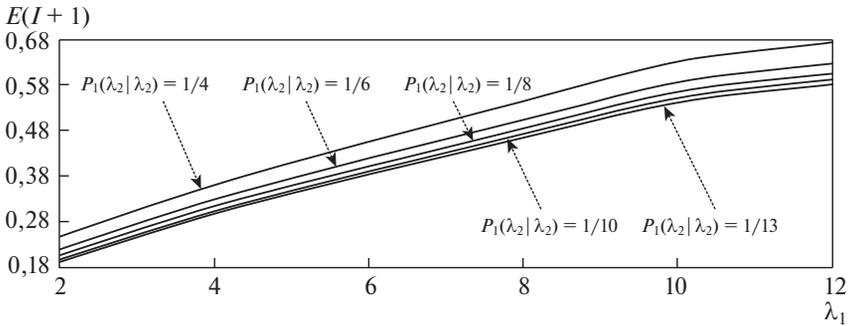


Рис. 7. Зависимость среднего числа запросов в системе $E(I + 1)$ от λ_1 для $b < 0$.

В (18) константа C_2 определена формулой (14) для $s = 2$. Для определения константы A_2 и вероятностей $P(-1, 1)$, $P(-1, 2)$ нужно привлечь уравнения (15) и условие нормировки. Откуда находим

$$\begin{aligned}
 P(-1, 1) &= a_{12}A_2; & P(-1, 2) &= a_{22}A_2; \\
 (19) \quad A_2 &= \frac{1 - \xi_2}{1 + C_2 + (a_{12} + a_{22})(1 - \xi_2)},
 \end{aligned}$$

C_2 определена в (14) для $s = 2$; a_{12} , a_{22} определены в (16); ξ_2 — корень кубического уравнения в (6) ($0 < \xi_2 < 1$).

Формулы (18), (19) позволяют определить характеристики системы:

$$(20) \quad \begin{aligned} P(-1) &= (a_{12} + a_{22})A_2; \\ E(I) &= A_2\xi_2 \frac{1 + C_2}{(1 - \xi_2)^2}, \quad E(I + 1) = \frac{(1 + C_2)A_2}{(1 - \xi_2)^2}, \end{aligned}$$

где C_2 определена в (14) для $s = 2$; a_{12} , a_{22} — в (16); A_2 — в (19); ξ_2 — корень кубического уравнения в (6) ($0 < \xi_2 < 1$).

В табл. 4–6 приведены зависимости характеристик $P(-1)$, $E(I)$, $E(I + 1)$ от параметра λ_1 ($\lambda_1 = 2, 4, \dots, 10, 11$) при фиксированных значениях параметров $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$; $P_1(\lambda_1|\lambda_1) = P_1(\lambda_2|\lambda_1) = P_0(\lambda_2|\lambda_1) = P_1(\lambda_1|\lambda_2) = \frac{1}{3}$ для $b < 0$ и $P_1(\lambda_2|\lambda_2) = \frac{1}{4}$ ($P_0(\lambda_1|\lambda_2) = \frac{5}{12}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{6}$ ($P_0(\lambda_1|\lambda_2) = \frac{1}{2}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{8}$ ($P_0(\lambda_1|\lambda_2) = \frac{13}{24}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{10}$ ($P_0(\lambda_1|\lambda_2) = \frac{17}{30}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{12}$ ($P_0(\lambda_1|\lambda_2) = \frac{7}{12}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{13}$ ($P_0(\lambda_1|\lambda_2) = \frac{23}{39}$), вычисленные по формулам (20).

На рис. 5–7 приведены графики числовых характеристик (20), построенные для численных значений табл. 4–6 соответственно.

Поведение отмеченных числовых характеристик (20), как и в случае $b > 0$, в зависимости от параметра λ_1 при $b < 0$ соответствует физическим представлениям о процессе обслуживания в изучаемой однолинейной СМО с входящим коррелированным МАР-потокм запросов.

4. Частный случай. Рекуррентный МАР-поток запросов

Для изучаемого частного случая имеем $b = 0$, что влечет за собой, как следует из (3), рекуррентность МАР-потока запросов. Сначала рассмотрим условия существования стационарных вероятностей $P(i, 1)$, $P(i, 2)$, $i \geq 0$. По-прежнему рассмотрим ситуацию, когда $\lambda = \mu$. Тогда характеристическое уравнение (8) примет вид

$$(21) \quad \begin{aligned} \lambda_1\lambda_2(\xi - 1)^2\xi \left\{ [1 - P_0(\lambda_2|\lambda_1)][1 - P_0(\lambda_1|\lambda_2)]\xi - \right. \\ \left. - [2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)] \right\} = 0, \end{aligned}$$

при этом общее решение системы (5) выпишется в виде (9). Корни характеристического уравнения (21) есть

$$(22) \quad \xi_1 = \xi_2 = 1, \quad \xi_3 = 0, \quad \xi_4 = \frac{1}{1 - P_0(\lambda_1|\lambda_2)} + \frac{1}{1 - P_0(\lambda_2|\lambda_1)} > 1.$$

С учетом (22) в общем решении (9) нужно положить $D_1 = D_2 = D_4 = 0$ и тогда получаем $P(i, 1) = P(i, 2) = 0$, $i \geq 0$. Последнее означает, что для случая $b = 0$ и при $\lambda = \mu$ финальное распределение $P(i, 1)$, $P(i, 2)$, $i \geq 0$, не существует и тем более не существует при $\lambda > \mu$.

Таблица 7. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b = 0$

$P_1(\lambda_1 \lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0,831	0,781	0,750	0,726	0,707	0,698
1/6	0,888	0,854	0,833	0,818	0,805	0,799
1/8	0,916	0,891	0,875	0,863	0,854	0,850
1/10	0,933	0,913	0,900	0,891	0,883	0,880
1/12	0,944	0,927	0,917	0,909	0,902	0,900
1/13	0,949	0,933	0,923	0,916	0,910	0,907

Таблица 8. Зависимость средней длины очереди $E(I)$ от λ_1 для $b = 0$

$P_1(\lambda_1 \lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0,030	0,056	0,083	0,111	0,139	0,152
1/6	0,013	0,023	0,033	0,043	0,053	0,058
1/8	0,007	0,012	0,018	0,023	0,028	0,030
1/10	0,004	0,008	0,011	0,014	0,017	0,019
1/12	0,003	0,005	0,008	0,010	0,0129	0,013
1/13	0,002	0,005	0,006	0,008	0,010	0,011

Таблица 9. Зависимость среднего числа запросов в системе $E(I + 1)$ от λ_1 для $b = 0$

$P_1(\lambda_1 \lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0,199	0,275	0,333	0,385	0,432	0,454
1/6	0,124	0,169	0,200	0,226	0,248	0,259
1/8	0,091	0,122	0,143	0,160	0,174	0,181
1/10	0,071	0,095	0,111	0,124	0,134	0,139
1/12	0,059	0,078	0,091	0,101	0,109	0,113
1/13	0,054	0,072	0,083	0,092	0,100	0,103

Изучим ситуацию $\lambda < \mu$. Характеристическое уравнение (6) для случая $b = 0$ выпишется в виде

$$(23) \quad \xi(\xi - 1) \left\{ \mu_1 \mu_2 \xi^2 - [\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)] \xi + \right. \\ \left. + [\lambda_1 \lambda_2 + \lambda_1 \mu_2 P_1(\lambda_1|\lambda_1) + \lambda_2 \mu_1 P_1(\lambda_2|\lambda_2) - \lambda_1 \lambda_2 P_0(\lambda_1|\lambda_2) P_0(\lambda_2|\lambda_1)] \right\} = 0.$$

Корни характеристического уравнения (23) есть $\xi_3 = 0$, $\xi_4 = 1$,

$$(24) \quad \xi_{1,2} = \left\{ (\lambda_1 \mu_2 + \lambda_2 \mu_1 + \mu_1 \mu_2) \mp \left[(\lambda_1 \mu_2 + \lambda_2 \mu_1 + \mu_1 \mu_2)^2 - \right. \right. \\ \left. \left. - 4 \mu_1 \mu_2 (\lambda_1 \lambda_2 + \lambda_1 \mu_2 P_1(\lambda_1|\lambda_1) + \lambda_2 \mu_1 P_1(\lambda_2|\lambda_2) - \right. \right. \\ \left. \left. - \lambda_1 \lambda_2 P_0(\lambda_1|\lambda_2) P_0(\lambda_2|\lambda_1)) \right]^{\frac{1}{2}} \right\} / 2 \mu_1 \mu_2,$$

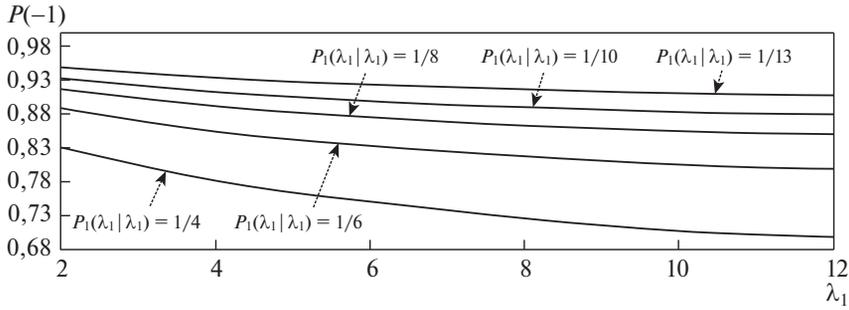


Рис. 8. Зависимость вероятности простоя $P(-1)$ от λ_1 для $b = 0$.

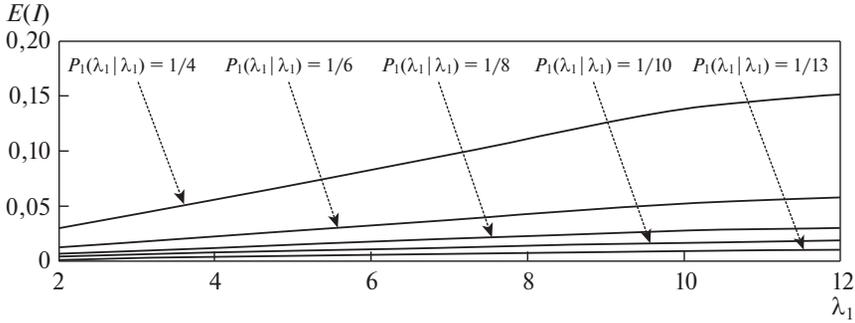


Рис. 9. Зависимость средней длины очереди $E(I)$ от λ_1 для $b = 0$.

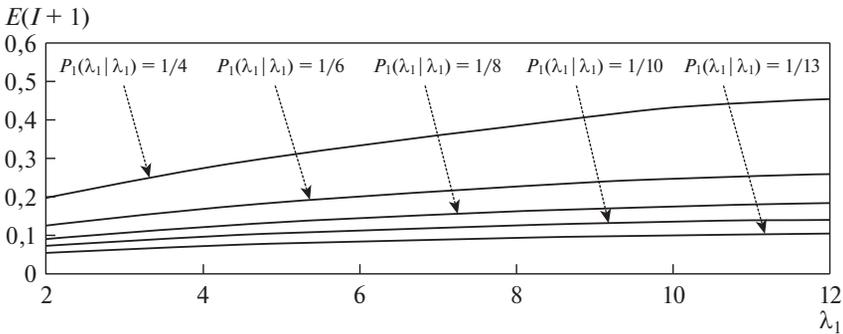


Рис. 10. Зависимость среднего числа запросов в системе $E(I+1)$ от λ_1 для $b = 0$.

$0 < \xi_1 < 1 < \xi_2$. С учетом (23) и (24) общее решение (12) системы (5) примет вид

$$(25) \quad P(i, 1) = A_1 \xi_1^i, \quad P(i, 2) = C_1 A_1 \xi_1^i, \quad i = 0, 1, \dots$$

В (25) константа C_1 определена в (14) для $s = 1$. Для определения константы A_1 и вероятностей $P(-1, 1)$, $P(-1, 2)$ воспользуемся уравнениями (15) и условием нормировки. Тогда находим

$$(26) \quad \begin{aligned} P(-1, 1) &= a_{11} A_1; & P(-1, 2) &= a_{21} A_1; \\ A_1 &= \frac{1 - \xi_1}{1 + C_1 + (a_{11} + a_{21})(1 - \xi_1)}, \end{aligned}$$

C_1 определена в (14) для $s = 1$; a_{21} , a_{11} определены в (16); ξ_1 — в (24).

Формулы (25), (26) позволяют найти характеристики системы:

$$(27) \quad P(-1) = (a_{21} + a_{11})A_1; \\ E(I) = A_1 \xi_1 \frac{1 + C_1}{(1 - \xi_1)^2}, \quad E(I + 1) = \frac{(1 + C_1)A_1}{(1 - \xi_1)^2},$$

где C_1 определена в (14) для $s = 1$; a_{21} , a_{11} определены в (16); A_1 — в (26); ξ_1 — в (24).

В табл. 7–9 приведены зависимости характеристик $P(-1)$, $E(I)$, $E(I+1)$ от параметра λ_1 ($\lambda_1 = 2, 4, \dots, 10, 11$) при фиксированных значениях параметров $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$ для $b = 0$ и ($P_1(\lambda_i|\lambda_j) = P_1(\lambda_j|\lambda_i) = \frac{1}{4}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{1}{2}$); ($P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{6}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{2}{3}$); ($P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{8}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{3}{4}$); ($P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{10}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{4}{5}$); ($P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{12}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{5}{6}$); ($P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{13}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{11}{13}$); $i, j = 1, 2$, $i \neq j$, вычисленные по формулам (27).

Поведение отмеченных числовых характеристик, как и в случаях $b > 0$, $b < 0$, в зависимости от параметра λ_1 при $b = 0$ соответствует физическим представлениям о процессе обслуживания в изучаемой однолинейной СМО с входящим рекуррентным МАР-потокком запросов.

На рис. 8–10 приведены графики числовых характеристик (27), построенные для численных значений табл. 7–9 соответственно.

5. Заключение

Поставленные во втором разделе статьи задачи анализа однолинейной СМО с входящим коррелированным МАР-потокком запросов с двумя состояниями решены в полном объеме.

Подытожим полученные результаты и приведем итоговые формулы.

Случай $b > 0$. Стационарные вероятности $P(i, 1)$, $P(i, 2)$ определяются в виде $P(i, 1) = A_1 \xi_1^i + A_2 \xi_2^i$, $P(i, 2) = C_1 A_1 \xi_1^i + C_2 A_2 \xi_2^i$, $i = 0, 1, \dots$, где константы C_s , $s = 1, 2$, определены формулой (14); ξ_1 , ξ_2 ($0 < \xi_1 < \xi_2 < 1$) — корни кубического уравнения (6); вероятности $P(-1, 1)$, $P(-1, 2)$ и константы A_1 , A_2 определены формулами (16); числовые характеристики $P(-1)$, $E(I)$, $E(I + 1)$ — формулами (17).

Случай $b < 0$. Стационарные вероятности $P(i, 1)$, $P(i, 2)$ определяются в виде $P(i, 1) = A_2 \xi_2^i$, $P(i, 2) = C_2 A_2 \xi_2^i$, $i = 0, 1, \dots$, где константа C_2 определена формулой (14) для $s = 2$; ξ_2 ($0 < \xi_2 < 1$) — корень кубического уравнения (6); вероятности $P(-1, 1)$, $P(-1, 2)$ и константа A_2 определены формулами (19); числовые характеристики $P(-1)$, $E(I)$, $E(I + 1)$ — формулами (20).

Случай $b = 0$. Стационарные вероятности $P(i, 1)$, $P(i, 2)$ определяются в виде $P(i, 1) = A_1 \xi_1^i$, $P(i, 2) = C_1 A_1 \xi_1^i$, $i = 0, 1, \dots$, где константа C_1 определена формулой (14) для $s = 1$; ξ_1 ($0 < \xi_1 < 1$) — корень (24) характеристического уравнения (23); вероятности $P(-1, 1)$, $P(-1, 2)$ и константа A_1 определены

формулами (26); числовые характеристики $P(-1)$, $E(I)$, $E(I + 1)$ — формулами (27).

При выводе формул (17), (20), (27) привлечены метод введения дополнительной переменной и метод диаграмм интенсивностей переходов (метод сечений стохастического графа) [8]. Случай $b = 0$ приводит к вырождению входящего коррелированного МАР-потока запросов в рекуррентный.

Аналитические формулы (17), (20), (27) позволяют без привлечения численных методов находить значения числовых характеристик при заданных параметрах МАР-потока запросов. Приведенные в статье графические зависимости числовых характеристик соответствуют физическим представлениям о процессе обслуживания в изучаемой СМО.

СПИСОК ЛИТЕРАТУРЫ

1. *Cox D.R.* The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // Proceedings of the Cambridge Philosophical Society. 1955. V. 51. No. 3. P. 433–441.
2. *Kingman J.F.C.* On doubly stochastic Poisson processes // Proceedings of the Cambridge Philosophical Society. 1964. V. 60. No. 4. P. 923–930.
3. *Башарин Г.П., Кокотушкин В.А., Наумов В.А.* О методе эквивалентных замен расчета фрагментов сетей связи. Ч. 1 // Изв. АН СССР. Техн. кибернетика. 1979. № 6. С. 92–99.
4. *Башарин Г.П., Кокотушкин В.А., Наумов В.А.* О методе эквивалентных замен расчета фрагментов сетей связи. Ч. 2 // Изв. АН СССР. Техн. кибернетика. 1980. № 1. С. 55–61.
5. *Neuts M.F.* A versatile Markov point process // J. Appl. Probab. 1979. V. 16. No. 4. P. 764–779.
6. *Lucantoni D.M.* New results on the single server queue with a batch Markovian arrival process // Communications in Statistics Stochastic Models. 1991. V. 7. No. 1. P. 1–46.
7. *Lucantoni D.M., Neuts M.F.* Some steady-state distributions for the MAP/SM/1 queue // Communications in Statistics Stochastic Models. 1994. V. 10. No. 3. P. 575–598.
8. *Вишневецкий В.М., Дудин А.Н., Клименок В.И.* Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: Техносфера, 2018. 564 с.
9. *Горцев А.М., Соловьев А.А.* Вероятность ошибки при оценивании состояний потока физических событий // Изв. вузов. Физика. 2016. Т. 59. № 5. С. 54–60.
10. *Горцев А.М., Нежелская Л.А., Соловьев А.А.* Оптимальная оценка состояний МАР-потока событий в условиях непродлевающегося мертвого времени // АиТ. 2012. № 8. С. 49–63.
Gortsev A.M., Nezhel'skaya L.A., Solov'ev A.A. Optimal state estimation in MAP event flows with unextendable dead time // Autom. Remote Control. 2012. V. 73. No. 8. P. 1316–1326.
11. *Горцев А.М., Соловьев А.А.* Совместная плотность вероятностей длительности интервалов МАР-потока событий и условия его рекуррентности // Вест. Томск.

гос. ун-та. Управление, вычислительная техника и информатика. 2012. № 3(20). С. 32–41.

12. Горцев А.М., Соловьев А.А. Сравнение МП- и ММ-оценок длительности непродлевающегося мертвого времени в МАР-потоке событий // Вест. Том. гос. ун-та. Управление, вычислительная техника и информатика. 2015. № 4(33). С. 13–22.
13. Горцев А.М., Соловьев А.А. Совместная плотность вероятностей длительности интервалов потока физических событий при непродлевающемся мертвом времени // Изв. вузов. Физика. 2014. Т. 57. № 7. С. 103–111.
14. Горцев А.М., Соловьев А.А. Оценка максимального правдоподобия длительности непродлевающегося мертвого времени в потоке физических событий // Изв. вузов. Физика. 2015. Т. 58. № 11. С. 141–149.
15. Бинь Сунь, Дудин С.А., Дудина О.С., Дудин А.Н. Модель обслуживания мобильных пользователей в соте с адаптивной модуляцией, учитывающая влияние случайной среды // АиТ. 2021. № 5. С. 86–105.
Bin Sun, Dudin S.A., Dudina O.S., Dudin A.N. A Customer Service Model in an Adaptive-Modulation Mobile Communication Cell with Allowance for Random Environment // Autom. Remote Control. 2021. V. 82. No. 5. P. 812–826.
16. Горцев А.М., Нежелская Л.А. Аналитическое исследование однолинейной СМО с входящим асинхронным потоком событий // АиТ. 2022. № 8. С. 65–80.
17. Нежелская Л.А. Оценка состояний дважды стохастических потоков событий. Томск: Изд-во Том. гос. ун-та, 2020. 210 с.
18. Медведев Г.А. Анализ стохастических графов, описывающих поведение шаговых систем автоматического поиска // Автоматика и вычислительная техника. 1968. № 4. С. 27–30.

Статья представлена к публикации членом редколлегии В.М. Вишневым.

Поступила в редакцию 27.02.2023

После доработки 25.04.2023

Принята к публикации 11.05.2023

Стохастические системы

© 2023 г. В.Л. ХАЦКЕВИЧ, д-р техн. наук (vlkhats@mail.ru)
(Военный учебно-научный центр Военно-воздушных сил
«Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина»,
Воронеж)

О НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ПРОЦЕССАХ С НЕЧЕТКИМИ СОСТОЯНИЯМИ

Изучены непрерывные случайные процессы с нечеткими состояниями. Установлены свойства их числовых характеристик (ожиданий и корреляционных функций), соответствующие свойствам характеристик числовых случайных процессов. Полученные результаты опираются на свойства нечетко случайных величин. Рассмотрены приложения к задаче преобразования случайного сигнала с нечеткими состояниями линейной динамической системой.

Ключевые слова: непрерывные случайные процессы, нечеткие состояния, нечеткие ожидания, корреляционные функции.

DOI: 10.31857/S0005231023070024, EDN: FCKKGGK

1. Введение

Нечеткое моделирование активно используется при решении различных прикладных задач, когда исходные данные неполные или слабо формализованные [1–3]. Отметим, в частности, недавние работы по развитию нечеткого подхода в теории управления [4, 5].

С другой стороны, при исследовании динамических процессов в условиях ограниченной исходной информации один из возможных подходов заключается в трактовке их параметров как реализации некоторых случайных процессов [6].

В данной работе сочетаются эти подходы, точнее исследуются непрерывные случайные процессы с нечеткими состояниями. А именно, время считается непрерывным и множество возможных нечетких состояний считается несчетным (непрерывным). При этом сечение непрерывного нечеткого случайного процесса в любой момент времени представляет собой нечетко случайную величину. В данном исследовании используются известные результаты по теории нечетко случайных величин [7–9], а также дифференцируемости и интегрируемости нечеткозначных функций ([10, 11] и [12, 13] соответственно).

Установленные в данной статье свойства являются модификациями известных [6, гл. 1] результатов о математических ожиданиях и корреляционных функциях стандартных непрерывных случайных процессов. Ранее они, по-видимому, не отмечались. Подчеркнем, что в данном случае важную роль играют нечеткие ожидания (неслучайные нечеткозначные функции времени), отражающие тренды нечетко случайных процессов.

В качестве приложения рассмотрена задача о преобразовании нечетко случайного сигнала линейной динамической системой, описываемой линейным дифференциальным уравнением с постоянными коэффициентами. В ней по характеристикам входного сигнала (нечеткому ожиданию и корреляционной функции) определяются аналогичные характеристики выходного сигнала. Эти результаты являются развитием известных для вещественных стационарных случайных процессов (см., например, [6, гл. 7]).

Отметим отличие излагаемых здесь подхода и результатов от работ, посвященных случайным процессам с дискретными нечеткими состояниями и непрерывным временем (см., например, [14, 15]), где обсуждаются свойства вероятностей дискретных нечетких состояний в зависимости от времени.

Ниже под нечетким числом \tilde{z} , заданным на универсальном пространстве R вещественных чисел, будем понимать совокупность упорядоченных пар $(x, \mu_{\tilde{z}}(x))$, где функция принадлежности $\mu_{\tilde{z}} : R \rightarrow [0, 1]$ определяет степень принадлежности $\forall x \in R$ множеству \tilde{z} [1, гл. 5].

Будем использовать интервальное представление нечетких чисел. А именно, каждому нечеткому числу поставим в соответствие совокупность его α -интервалов.

Как известно, множество α -уровня нечеткого числа \tilde{z} с функцией принадлежности $\mu_{\tilde{z}}(x)$ определяется соотношением

$$Z_{\alpha} = \{x \mid \mu_{\tilde{z}}(x) \geq \alpha\} \quad (\alpha \in (0, 1]), \quad Z_0 = cl\{x \mid \mu_{\tilde{z}}(x) > 0\},$$

где cl — обозначает замыкание множества.

Будем считать, что все α -уровни нечеткого числа — замкнутые и ограниченные интервалы вещественной оси. Обозначим левую границу интервала вещественной оси. Обозначим левую границу интервала Z_{α} через $z^{-}(\alpha)$, а правую через $z^{+}(\alpha)$. Таким образом, $Z_{\alpha} = [z^{-}(\alpha), z^{+}(\alpha)]$. Иногда $z^{-}(\alpha)$ и $z^{+}(\alpha)$ называют соответственно левым и правым α -индексами (индексами) нечеткого числа. Ниже предполагается, что они измеримы и ограничены на $[0, 1]$. Совокупность таких нечетких чисел будем обозначать через J .

Под суммой нечетких чисел понимается нечеткое число, индексы которого являются суммами соответствующих индексов слагаемых. Умножение нечеткого числа на положительное число означает умножение индексов на это число. Умножение на отрицательное вещественное число означает умножение индексов на это число и перемену их местами. Равенство нечетких чисел понимается как равенство всех соответствующих α -индексов (при $\forall \alpha \in [0, 1]$).

Множество J можно метризовать различными способами. В частности, для нечетких чисел $\tilde{z}_1, \tilde{z}_2 \in J$ определяют метрику следующим равенством (см., например, [16])

$$(1) \quad d(\tilde{z}_1, \tilde{z}_2) = \left(\int_0^1 ((z_1^-(\alpha) - z_2^-(\alpha))^2 + (z_1^+(\alpha) - z_2^+(\alpha))^2) d\alpha \right)^{1/2},$$

где $z_i^-(\alpha)$ и $z_i^+(\alpha)$ являются соответственно левым и правым индексами \tilde{z}_i ($i = 1, 2$).

2. Нечеткое ожидание, ожидание и ковариация нечетко случайных величин

Пусть (Ω, Σ, P) — вероятностное пространство, где Ω — множество элементарных событий, Σ — σ -алгебра, состоящая из подмножеств множества Ω , P — вероятностная мера.

Рассмотрим отображение $\tilde{X} : \Omega \rightarrow J$. Его интервалы α -уровня $X_\alpha(\omega)$ при фиксированном $\omega \in \Omega$ определяются формулами

$$X_\alpha(\omega) = \{r \in R : \mu_{\tilde{X}(\omega)}(r) \geq \alpha\} \quad \alpha \in (0, 1], \quad X_0(\omega) = cl \{r \in R : \mu_{\tilde{X}(\omega)}(r) > 0\},$$

где $\mu_{\tilde{X}(\omega)}(r)$ — функция принадлежности нечеткого числа $\tilde{X}(\omega)$. Интервал $X_\alpha(\omega)$ представим в виде $X_\alpha(\omega) = [X^-(\omega, \alpha), X^+(\omega, \alpha)]$. Его границы $X^-(\omega, \alpha)$, $X^+(\omega, \alpha)$ называют соответственно левым и правым индексами отображения \tilde{X} .

Отображение $\tilde{X} : \Omega \rightarrow J$ называют нечетко случайной величиной, кратко Н.С.В. (см., например, [7, 8]), если вещественнозначные функции $X^\pm(\omega, \alpha)$ измеримы для всех $\alpha \in [0, 1]$. В этом случае при любом $\alpha \in [0, 1]$ индексы являются вещественными случайными величинами.

В дальнейшем будем рассматривать класс \mathcal{X} нечетко случайных величин, для которых индексы $X^-(\omega, \alpha)$ и $X^+(\omega, \alpha)$ являются квадратично суммируемыми на $\Omega \times [0, 1]$ функциями.

Отметим, что Н.С.В. можно трактовать как случайный элемент в метрическом пространстве J с метрикой (1).

Для Н.С.В. $\tilde{X}(\omega)$ положим

$$(2) \quad x^-(\alpha) = \int_{\Omega} X^-(\omega, \alpha) dP, \quad x^+(\alpha) = \int_{\Omega} X^+(\omega, \alpha) dP.$$

Нечеткое число с α -индексами, определяемыми в (2), называют (см., например, [9]) нечетким ожиданием Н.С.В. \tilde{X} . Будем обозначать его как $M(\tilde{X})$, а его индексы — $[M(\tilde{X})]_{\alpha}^{\pm}$.

На множестве \mathcal{X} нечетко случайных величин рассмотрим метрику, задаваемую для Н.С.В. \tilde{X}_1, \tilde{X}_2 равенством

$$(3) \quad \rho(\tilde{X}_1, \tilde{X}_2) = \left(\int_0^1 \int_{\Omega} ((X_1^-(\omega, \alpha) - X_2^-(\omega, \alpha))^2 + (X_1^+(\omega, \alpha) - X_2^+(\omega, \alpha))^2) dP d\alpha \right)^{1/2}.$$

Определенное в (2) нечеткое ожидание обладает свойствами, аналогичными свойствам математических ожиданий вещественных случайных величин. А именно, имеет место

Утверждение 1 (см., например, [17, 18]). *Нечеткое ожидание, определяемое посредством (2), обладает следующими свойствами:*

1. Если $\tilde{X}(\omega) = \tilde{X}$ при почти всех $\omega \in \Omega$, то $M(\tilde{X}) = \tilde{X}$ (идемпотентность).

2. Нечеткое ожидание $M: \mathcal{X} \rightarrow J$ аддитивно, т.е. $M(\tilde{X}_1 + \tilde{X}_2) = M(\tilde{X}_1) + M(\tilde{X}_2)$ ($\forall \tilde{X}_1, \tilde{X}_2 \in \mathcal{X}$).

3. Нечеткое ожидание $M: \mathcal{X} \rightarrow J$ однородно, т.е. $M(C\tilde{X}) = CM(\tilde{X})$ ($\forall \tilde{X} \in \mathcal{X}, \forall C \in R$).

4. Нечеткое ожидание $M: \mathcal{X} \rightarrow J$ непрерывно как отображение из \mathcal{X} с метрикой (3) в J с метрикой (1).

5. Справедливо следующее экстремальное свойство нечеткого ожидания $M(\tilde{X})$ для заданной Н.С.В. $\tilde{X} \in \mathcal{X}$:

$$\begin{aligned} & \rho^2(\tilde{X}, M(\tilde{X})) = \\ & = \int_0^1 \int_{\Omega} \left((X^-(\omega, \alpha) - [M(\tilde{X})]_{\alpha}^-)^2 + (X^+(\omega, \alpha) - [M(\tilde{X})]_{\alpha}^+)^2 \right) dP d\alpha \leq \\ & \leq \int_0^1 \int_{\Omega} ((X^-(\omega, \alpha) - V_{\alpha}^-)^2 + (X^+(\omega, \alpha) - V_{\alpha}^+)^2) dP d\alpha = \rho^2(\tilde{X}, \tilde{V}) \quad (\forall \tilde{V} \in J), \end{aligned}$$

где $X^{\pm}(\omega, \alpha)$ и V_{α}^{\pm} — индексы Н.С.В. $\tilde{X}(\omega)$ и нечеткого числа \tilde{V} соответственно.

Рассмотрим вопрос о дефазификации нечеткого ожидания.

Как известно, среднее нечеткого числа \tilde{z} при интервальном подходе определяется равенством [19]

$$(4) \quad z_{cp} = \frac{1}{2} \int_0^1 (z^-(\alpha) + z^+(\alpha)) d\alpha,$$

где $z^{\pm}(\alpha)$ — индексы нечеткого числа \tilde{z} .

В связи с (4) ожидание $m(\tilde{X})$ Н.С.В. $\tilde{X} \in \mathcal{X}$ определяют как усредняющий функционал посредством формулы

$$(5) \quad m(\tilde{X}) = \frac{1}{2} \int_0^1 \left([M(\tilde{X})]^- (\alpha) + [M(\tilde{X})]^+ (\alpha) \right) d\alpha,$$

где $M^\pm(\alpha)$ — индексы нечеткого ожидания $M(\tilde{X})$, задаваемые в (2). По существу, это один из способов дефазификации нечеткого ожидания.

Из определения (5) и утверждения 1 вытекает

Утверждение 2 (см., например, [17, 18]). *Ожидание Н.С.В. (5) обладает следующими свойствами:*

1. Если $\tilde{X}(\omega) = \tilde{X}$ при почти всех $\omega \in \Omega$, то $m(\tilde{X}) = X_{\text{ср}}$ (идемпотентность).

2. Ожидание $m : \mathcal{X} \rightarrow R$ аддитивно, т.е. $m(\tilde{X}_1 + \tilde{X}_2) = m(\tilde{X}_1) + m(\tilde{X}_2)$ ($\forall \tilde{X}_1, \tilde{X}_2 \in \mathcal{X}$).

3. Ожидание $m : \mathcal{X} \rightarrow R$ однородно, т.е. $m(C\tilde{X}) = Cm(\tilde{X})$ ($\forall \tilde{X} \in \mathcal{X}, \forall C \in R$).

4. Ожидание $m : \mathcal{X} \rightarrow R$ непрерывно.

5. Справедливо следующее экстремальное свойство ожидания $m(\tilde{X})$ для $\tilde{X} \in \mathcal{X}$:

$$\rho(\tilde{X}, m(\tilde{X})) \leq \rho(\tilde{X}, a) \quad (\forall a \in R),$$

где для вещественного числа a полагаем $a_\alpha^- = a_\alpha^+ = a$.

Для Н.С.В. \tilde{X} и \tilde{Y} ковариацию определяют формулой [20]

$$(6) \quad cov(\tilde{X}, \tilde{Y}) = \frac{1}{2} \int_0^1 (cov(X_\alpha^-, Y_\alpha^-) + cov(X_\alpha^+, Y_\alpha^+)) d\alpha,$$

а дисперсию $D(\tilde{X}) = cov(\tilde{X}, \tilde{X})$. В (6) ковариации вещественных случайных величин X_α^\pm и Y_α^\pm определены стандартной [21, гл. I] формулой

$$cov(X_\alpha^\pm, Y_\alpha^\pm) = E(X_\alpha^\pm - E(X_\alpha^\pm))(Y_\alpha^\pm - E(Y_\alpha^\pm)).$$

Здесь и ниже символ E обозначает математическое ожидание скалярной случайной величины, т.е. для случайной величины $\xi(\omega)$ полагаем $E\xi = \int_\Omega \xi(\omega) dP$.

Имеет место

Утверждение 3 [20]. *Справедливы следующие свойства ковариации (6) Н.С.В.:*

1. *Симметричность:* $cov(\tilde{X}, \tilde{Y}) = cov(\tilde{Y}, \tilde{X}), \forall \tilde{X}, \tilde{Y} \in \mathcal{X}$.

2. *Аддитивность*: $\text{cov}(\tilde{X} + \tilde{Y}, \tilde{Z}) = \text{cov}(\tilde{X}, \tilde{Y}) + \text{cov}(\tilde{Y}, \tilde{Z})$, $\forall \tilde{X}, \tilde{Y}, \tilde{Z} \in \mathcal{X}$.
3. *Положительная однородность*: $\text{cov}(C_1\tilde{X}, C_2\tilde{Y}) = C_1C_2\text{cov}(\tilde{X}, \tilde{Y})$, $\forall \tilde{X}, \tilde{Y} \in \mathcal{X}, \forall C_1, C_2 \in \mathbb{R} : C_1C_2 > 0$.
4. $D(C\tilde{X}) = C^2D(\tilde{X})$, $\forall \tilde{X} \in \mathcal{X}, \forall C \in \mathbb{R}$.
5. $D(\tilde{X} + \tilde{Y}) = D(\tilde{X}) + D(\tilde{Y}) + 2\text{cov}(\tilde{X}, \tilde{Y})$, $\forall \tilde{X}, \tilde{Y} \in \mathcal{X}$.

3. Непрерывные случайные процессы с нечеткими состояниями

Пусть $[t_0, T]$ расширенный отрезок числовой оси. Ниже, как и в разделе 2, (Ω, Σ, P) — вероятностное пространство, \mathcal{X} — совокупность Н.С.В. с квадратично суммируемыми на $\Omega \times [0, 1]$ индексами.

Непрерывным случайным процессом с нечеткими состояниями или нечетко случайным процессом (Н.С.П.) будем называть отображение $\tilde{X} : [t_0, T] \rightarrow \mathcal{X}$, т.е. функцию $\tilde{X}(\omega, t)$, значениями которой при $\forall t \in [t_0, T]$ являются Н.С.В. Обозначим α -индексы Н.С.П. $\tilde{X}(\omega, t)$ через $X_\alpha^\pm(\omega, t)$.

Далее будем рассматривать Н.С.П., для которых функции $X_\alpha^\pm(\omega, t)$ квадратично суммируемы по совокупности переменных на $\Omega \times [0, 1] \times [t_0, T]$.

Определим нечеткое ожидание $M(\tilde{X}(\omega, t))$ Н.С.П. $\tilde{X}(\omega, t)$ при каждом $t \in [t_0, T]$ как нечеткое ожидание, соответствующей Н.С.В., т.е. нечеткозначную функцию, имеющую индексы $\left[M(\tilde{X}(\omega, t)) \right]_\alpha^\pm = \int_\Omega X_\alpha^\pm(\omega, t) dP$.

Из определения нечеткого ожидания Н.С.П. и свойств нечетких ожиданий Н.С.В. вытекает:

Утверждение 4. Справедливы следующие свойства:

1. *Нечеткое ожидание от неслучайной нечеткозначной функции $\tilde{z} : [t_0, T] \rightarrow J$ равно самой этой функции $M(\tilde{z}(t)) = \tilde{z}(t)$.*
2. *Неслучайную скалярную функцию $\varphi : [t_0, T] \rightarrow \mathbb{R}$ можно выносить за знак нечеткого ожидания*

$$M(\varphi(t)\tilde{X}(t)) = \varphi(t)M(\tilde{X}(t)),$$

где $\tilde{X}(t)$ — нечетко-случайный процесс.

3. *Нечеткое ожидание суммы двух Н.С.П. равно сумме нечетких ожиданий слагаемых*

$$M(\tilde{X}(t) + \tilde{Y}(t)) = M(\tilde{X}(t)) + M(\tilde{Y}(t)).$$

4. *Справедливо следующее экстремальное свойство нечеткого ожидания $M(\tilde{X}(t))$ для Н.С.П. $\tilde{X}(t)$:*

$$\int_{t_0}^T \rho^2(\tilde{X}(t), M(\tilde{X}(t))) dt \leq \int_{t_0}^T \rho^2(\tilde{X}(t), \tilde{V}(t)) dt.$$

Здесь $\tilde{V} : [t_0, T] \rightarrow J$ — произвольная неслучайная нечеткозначная функция с квадратично суммируемыми на $[t_0, T]$ индексами, а метрика ρ задана в (3).

В соответствии с (5) определим ожидание Н.С.П. $\tilde{X}(t)$ формулой

$$m(\tilde{X}(t)) = \frac{1}{2} \int_0^1 \left([M(\tilde{X}(t))]_{\alpha}^{-} + [M(\tilde{X}(t))]_{\alpha}^{+} \right) d\alpha.$$

Заметим, что для $m(\tilde{X}(t))$ при $\forall t \in [t_0, T]$ справедливо утверждение 2.

В данной работе используется понятие дифференцируемости скалярного случайного процесса в среднем квадратичном [21, гл. II]. А именно, скалярный случайный процесс $\xi(t)$, $t \in R$ называют дифференцируемым в среднем квадратичном в точке t , если существует случайная величина $\xi'(t)$, такая что математическое ожидание

$$E \left| \frac{\xi(t+h) - \xi(t)}{h} - \xi'(t) \right|^2 \rightarrow 0 \quad \text{при } h \rightarrow 0.$$

Ниже будем еще использовать понятие производной от нечеткозначных функций (см., например, [10, 11]).

Н.С.П. $\tilde{X}(\omega, t)$ с α -интервалами $[X_{\alpha}^{-}(\omega, t), X_{\alpha}^{+}(\omega, t)]$ назовем дифференцируемым в точке t (по Сеиккала), если его α -индексы дифференцируемы по t как скалярные случайные процессы в среднем квадратичном, а $\frac{\partial}{\partial t} X_{\alpha}^{-}(\omega, t)$ и $\frac{\partial}{\partial t} X_{\alpha}^{+}(\omega, t)$ являются соответственно нижним и верхним α -индексами некоторой Н.С.В., называемой производной (сравни с [11] для нечеткозначной функции). В этом случае производную по t Н.С.П. $\tilde{X}(t)$ будем обозначать $\tilde{X}'(t) = \frac{\partial}{\partial t} \tilde{X}(\omega, t)$.

Теорема 1. Пусть Н.С.П. $\tilde{X}(t)$ дифференцируем при $t \in (t_0, T)$ и для $\forall \alpha \in [0, 1]$ найдутся суммируемые на Ω функции $\varphi_{\alpha}^{\pm}(\omega)$ такие, что $|\frac{\partial}{\partial t} X_{\alpha}^{\pm}(\omega, t)| \leq \varphi_{\alpha}^{\pm}(\omega)$ ($\forall t \in (t_0, T)$, $\omega \in \Omega$). Тогда нечеткое ожидание производной Н.С.П. равно производной от нечеткого ожидания:

$$(7) \quad M\tilde{X}'(t) = \left(M\tilde{X}(t) \right)'$$

Действительно, по определению производной и нечеткого ожидания можем записать $\left[M(\tilde{X}'(t)) \right]_{\alpha}^{\pm} = \int_{\Omega} \frac{\partial}{\partial t} X_{\alpha}^{\pm}(\omega, t) dP$ ($\forall \alpha \in [0, 1]$).

В условиях теоремы 1 согласно теореме о дифференцировании по параметру под знаком интеграла Лебега можно вынести производную в правой части последнего равенства за знак интеграла. Тогда, используя интервальный признак равенства нечетких чисел, получим (7).

Ниже будем использовать понятие интеграла от нечеткозначной функции (см., например, [12, 13]).

Интегралом по промежутку $[t_0, T]$ от Н.С.П. $\tilde{X}(\omega, t)$ назовем Н.С.В. $\tilde{Y}(\omega) = \int_{t_0}^T \tilde{X}(\omega, t) dt$ с α -индексами

$$Y_{\alpha}^{\pm}(\omega) = \int_{t_0}^T X_{\alpha}^{\pm}(\omega, t) dt,$$

где $X_{\alpha}^{\pm}(t)$ — соответствующие α -индексы Н.С.П. $\tilde{X}(t)$, а интегралы от случайных процессов $X_{\alpha}^{\pm}(t)$ понимаются в среднем квадратичном [21, гл. II]. Если такой интеграл существует, то Н.С.П. назовем интегрируемым на $[t_0, T]$.

Теорема 2. Пусть $\tilde{X}(\omega, t)$ — интегрируемый на $[t_0, T]$ Н.С.П. Тогда

$$(8) \quad M \left(\int_{t_0}^T \tilde{X}(\omega, t) dt \right) = \int_{t_0}^T M \left(\tilde{X}(\omega, t) \right) dt.$$

Действительно, по определению нечеткого ожидания левая часть (8) имеет α -индексы $\left[M \left(\int_{t_0}^T \tilde{X}(\omega, t) dt \right) \right]_{\alpha}^{\pm} = \int_{\Omega} \left[\int_{t_0}^T \tilde{X}(\omega, t) dt \right]_{\alpha}^{\pm} dP = \int_{\Omega} \left(\int_{t_0}^T \tilde{X}_{\alpha}^{\pm}(\omega, t) dt \right) dP$. Поскольку случайные величины $X_{\alpha}^{\pm}(\omega, t)$ квадратично суммируемы на $\Omega \times [t_0, T]$, то в последнем выражении можно поменять порядок интегрирования на основании теоремы Фубини. Тогда

$$\begin{aligned} \left[M \left(\int_{t_0}^T \tilde{X}(\omega, t) dt \right) \right]_{\alpha}^{\pm} &= \int_{t_0}^T \left(\int_{\Omega} X_{\alpha}^{\pm}(\omega, t) dP \right) dt = \\ &= \int_{t_0}^T \left[M(\tilde{X}(\omega, t)) \right]_{\alpha}^{\pm} dt = \left[\int_{t_0}^T M(\tilde{X}(\omega, t)) dt \right]_{\alpha}^{\pm}. \end{aligned}$$

Здесь используется определение нечеткого ожидания. Тогда формула (8) следует из равенства соответствующих индексов при $\forall [0, 1]$.

Пример 1. Пусть случайные числовые процессы $\xi_i(\omega, t)$ ($i = 1, 2, 3$; $\omega \in \Omega$, $t \in [t_0, T]$) квадратично суммируемы на $\Omega \times [t_0, T]$ и таковы, что $\xi_1(\omega, t) < \xi_2(\omega, t) < \xi_3(\omega, t)$ при всех $\omega \in \Omega$, $t \in [t_0, T]$.

Рассмотрим Н.С.П. $\tilde{X}(t)$, для которого при каждом $\omega \in \Omega$, $t \in [t_0, T]$ нечеткое число $\tilde{X}(\omega, t)$ имеет треугольный вид $(\xi_1(\omega, t), \xi_2(\omega, t), \xi_3(\omega, t))$. То есть функция принадлежности $\tilde{X}(\omega, t)$ при любом $\omega \in \Omega$, $t \in [t_0, T]$ описывается формулой

$$\mu_{\omega, t}(x) = \begin{cases} \frac{x - \xi_1(\omega, t)}{\xi_2(\omega, t) - \xi_1(\omega, t)}, & \text{если } x \in [\xi_1(\omega, t), \xi_2(\omega, t)]; \\ \frac{x - \xi_3(\omega, t)}{\xi_2(\omega, t) - \xi_3(\omega, t)}, & \text{если } x \in [\xi_2(\omega, t), \xi_3(\omega, t)]; \\ 0 & \text{в противном случае.} \end{cases}$$

В этом случае

$$X_{\alpha}^{-}(\omega, t) = (1 - \alpha)\xi_1(\omega, t) + \alpha\xi_2(\omega, t), \quad X_{\alpha}^{+}(\omega, t) = (1 - \alpha)\xi_3(\omega, t) + \alpha\xi_2(\omega, t).$$

Тогда нечеткое ожидание $M(\tilde{X}(t))$ определяется формулами для α -индексов

$$\begin{aligned} [M(\tilde{X})]_{\alpha}^{-}(t) &= (1 - \alpha) \int_{\Omega} \xi_1(\omega, t) dP + \alpha \int_{\Omega} \xi_2(\omega, t) dP = \\ &= (1 - \alpha)E\xi_1(t) + \alpha E\xi_2(t) \quad (\forall \alpha \in [0, 1]) \end{aligned}$$

и

$$\begin{aligned} [M(\tilde{X})]_{\alpha}^{+}(t) &= (1 - \alpha) \int_{\Omega} \xi_3(\omega, t) dP + \alpha \int_{\Omega} \xi_2(\omega, t) dP = \\ &= (1 - \alpha)E\xi_3(t) + \alpha E\xi_2(t) \quad (\forall \alpha \in [0, 1]), \end{aligned}$$

где E обозначает математическое ожидание вещественной случайной величины.

Пример 2. В условиях примера 1 дополнительно предположим, что случайные процессы $\xi_i(\omega, t)$ при $i = 1, 2, 3$ дифференцируемы по t в среднем квадратическом, причем для производных выполнено соотношение $\frac{\partial}{\partial t}\xi_1(\omega, t) \leq \frac{\partial}{\partial t}\xi_2(\omega, t) \leq \frac{\partial}{\partial t}\xi_3(\omega, t)$ при всех $\omega \in \Omega$, $t \in [t_0, T]$. Тогда производная $\tilde{X}'(t)$ Н.С.П. $\tilde{X}(t)$ имеет треугольный вид $\left(\frac{\partial}{\partial t}\xi_1(\omega, t), \frac{\partial}{\partial t}\xi_2(\omega, t), \frac{\partial}{\partial t}\xi_3(\omega, t) \right)$.

В частности, пусть заданы случайные величины $\xi_1(\omega) < \xi_2(\omega) < \xi_3(\omega)$. Рассмотрим треугольный нечеткий процесс $(e^t\xi_1(\omega), e^t\xi_2(\omega), e^t\xi_3(\omega))$. Его производная существует и совпадает с исходным Н.С.П.

Пример 3. Пусть в условиях примера 1 случайные процессы $\xi_i(\omega, t)$ дополнительно интегрируемы по t на промежутке $[t_0, T]$. Тогда Н.С.П. $\int_{t_0}^T \tilde{X}(t) dt$ имеет треугольный вид $\left(\frac{\partial}{\partial t}\xi_1(\omega, t)dt, \frac{\partial}{\partial t}\xi_2(\omega, t)dt, \frac{\partial}{\partial t}\xi_3(\omega, t)dt \right)$.

Ниже рассмотрим понятие ковариационной функции Н.С.П. и ее свойства. Ковариационной функцией Н.С.П. $\tilde{X}(t)$ назовем величину

$$(9) \quad K_{\tilde{X}}(t, s) = cov(\tilde{X}(t), \tilde{X}(s)) = \frac{1}{2} \int_0^1 \left(K_{X_{\alpha}^{-}}(t, s) + K_{X_{\alpha}^{+}}(t, s) \right) d\alpha.$$

Здесь $K_{X_{\alpha}^{-}}(t, s)$ и $K_{X_{\alpha}^{+}}(t, s)$ — ковариационные функции случайных процессов $X_{\alpha}^{-}(\omega, t)$ и $X_{\alpha}^{+}(\omega, t)$ соответственно, равные

$$(10) \quad K_{X_{\alpha}^{\pm}}(t, s) = E \left[(X_{\alpha}^{\pm}(\omega, t) - EX_{\alpha}^{\pm}(\omega, t))(X_{\alpha}^{\pm}(\omega, s) - EX_{\alpha}^{\pm}(\omega, s)) \right].$$

Дисперсия Н.С.П. $\tilde{X}(t)$ определяется равенством $D_{\tilde{X}}(t) = K_{\tilde{X}}(t, t)$.

Отметим, что (9) является модификацией на нечеткий случай стандартного определения ковариационной функции скалярных случайных процессов (см., например, [21, гл. II]).

Согласно (9), (10) и свойствам ковариации Н.С.В. (утверждение 3) имеет место

Утверждение 5. Ковариационная функция Н.С.П. обладает следующими свойствами:

1. *Симметричность. Для непрерывного Н.С.П. $\tilde{X}(t)$ при $\forall t_1, t_2 \in [t_0, T]$ имеет место равенство*

$$2. K_{\tilde{X}}(t_1, t_2) = K_{\tilde{X}}(t_2, t_1).$$

3. *Если $\tilde{X}(t)$ — непрерывный Н.С.П. и $\varphi(t)$ — неслучайная числовая функция, то для Н.С.П. $\tilde{Y}(t) = \varphi(t)\tilde{X}(t)$ ковариационная функция $K_{\tilde{Y}}(t_1, t_2)$ имеет вид $K_{\tilde{Y}}(t_1, t_2) = \varphi(t_1)\varphi(t_2)K_{\tilde{X}}(t_1, t_2)$.*

4. *Если $\tilde{Y}(t) = \tilde{X}(t) + \varphi(t)$, то $K_{\tilde{Y}}(t_1, t_2) = K_{\tilde{X}}(t_1, t_2)$.*

5. *Справедливо соотношение $|K_{\tilde{X}}(t_1, t_2)| \leq \sqrt{D_{\tilde{X}}(t_1)D_{\tilde{X}}(t_2)}$.*

Взаимосвязь между корреляционными функциями дифференцируемого Н.С.П. и его производной характеризует

Теорема 3. Пусть определены и непрерывны по совокупности переменных t, s, α вторые производные $\frac{\partial^2 K_{X_\alpha^-}(t, s)}{\partial t \partial s}$ и $\frac{\partial^2 K_{X_\alpha^+}(t, s)}{\partial t \partial s}$ ковариационных функций (10) для Н.С.П. $\tilde{X}(t)$. Тогда ковариационная функция $K'_{\tilde{X}}(t, s)$ производной $\tilde{X}'(t)$ Н.С.П. $\tilde{X}(t)$ задается формулой

$$(11) \quad K'_{\tilde{X}}(t, s) = \frac{\partial^2 K_{\tilde{X}}(t, s)}{\partial t \partial s}.$$

Доказательство. По определению (9)

$$K_{\tilde{X}'}(t, s) = \frac{1}{2} \int_0^1 \left(K_{(\tilde{X}_\alpha^-)'}(t, s) + K_{(\tilde{X}_\alpha^+)'}(t, s) \right) d\alpha.$$

При этом по известному свойству (скалярных) случайных процессов

$$K_{(\tilde{X}_\alpha^-)'}(t, s) = \frac{\partial^2 K_{X_\alpha^-}(t, s)}{\partial t \partial s}, \quad K_{(\tilde{X}_\alpha^+)'}(t, s) = \frac{\partial^2 K_{X_\alpha^+}(t, s)}{\partial t \partial s}.$$

Тогда $K_{\tilde{X}'}(t, s) = \frac{1}{2} \int_0^1 \frac{\partial^2}{\partial t \partial s} \left(K_{X_\alpha^-}(t, s) + K_{X_\alpha^+}(t, s) \right) d\alpha$. Вынося здесь вторую смешанную производную за знак интеграла, получим формулу (11).

Отметим, что законность последней операции обеспечивается непрерывностью $\frac{\partial^2 K_{X_\alpha^-}(t, s)}{\partial t \partial s}$ и $\frac{\partial^2 K_{X_\alpha^+}(t, s)}{\partial t \partial s}$ по совокупности переменных t, s, α .

Рассмотрим интеграл с переменным верхним пределом $\tilde{Y}(t) = \int_{t_0}^t \tilde{X}(\omega, s) ds$

от Н.С.П. $\tilde{X}(t)$. Имеет место

Теорема 4. Пусть ковариационные функции $K_{X_\alpha^\pm}(t, s)$ от α -индексов $X_\alpha^\pm(t)$ Н.С.П. $\tilde{X}(t)$ суммируемы по совокупности переменных t, s, α . Тогда ковариационная функция интеграла $\tilde{Y}(t)$ равна

$$(12) \quad K_{\tilde{Y}}(t, s) = \int_{t_0}^t \int_{t_0}^s K_{\tilde{X}}(\tau_1, \tau_2) d\tau_1 d\tau_2.$$

Доказательство. По определению (9) $K_{\tilde{Y}}(t, s) = \frac{1}{2} \int_0^1 \left(K_{Y_\alpha^-}(t, s) + K_{Y_\alpha^+}(t, s) \right) d\alpha$. При этом на основании определения нечеткого интеграла $Y_\alpha^\pm(t) = \int_{t_0}^t X_\alpha^\pm(\tau) d\tau$. Тогда в силу известного свойства интеграла от скалярного случайного процесса получим

$$K_{Y_\alpha^-}(t, s) = \int_{t_0}^t \int_{t_0}^s K_{X_\alpha^-}(\tau_1, \tau_2) d\tau_1 d\tau_2, \quad K_{Y_\alpha^+}(t, s) = \int_{t_0}^t \int_{t_0}^s K_{X_\alpha^+}(\tau_1, \tau_2) d\tau_1 d\tau_2.$$

Следовательно, $K_{\tilde{Y}}(t, s) = \frac{1}{2} \int_0^1 \left(\int_{t_0}^t \int_{t_0}^s (K_{X_\alpha^-}(\tau_1, \tau_2) + K_{X_\alpha^+}(\tau_1, \tau_2)) d\tau_1 d\tau_2 \right) d\alpha$. Меняя в правой части порядок интегрирования на основании теоремы Фубини и используя (9), установим (12).

Отметим, что законность последней операции обеспечивается суммируемостью $K_{X_\alpha^-}(\tau_1, \tau_2)$ и $K_{X_\alpha^+}(\tau_1, \tau_2)$ по совокупности переменных τ_1, τ_2, α .

4. Задача о преобразовании нечетко-случайного сигнала линейной динамической системой

Рассмотрим ситуацию, когда на вход некоторого устройства A (см. рис. 1) поступает непрерывный случайный сигнал $y(t)$, а на выходе наблюдается непрерывный случайный сигнал $z(t)$.

Устройство A называют линейной динамической системой, если связь между входным и выходным сигналами описывается дифференциальным уравнением n -го порядка с постоянными коэффициентами.

В литературе рассматривается задача (см., например, [6, гл. 7]) об установлении взаимосвязей между числовыми характеристиками (математическими ожиданиями и соответственно ковариационными функциями) входного и выходного случайных сигналов. Эта задача в предположении стационарности рассматриваемых случайных сигналов решается с использованием частотной характеристики системы, прямого и обратного преобразования Фурье, теоремы Винера–Хинчина. При этом случайный процесс называют стационарным

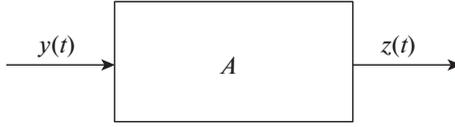


Рис. 1.

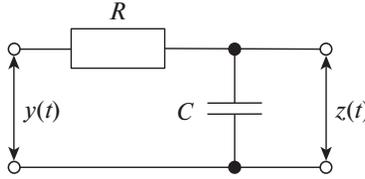


Рис. 2.

(в широком смысле), если его математическое ожидание не зависит от времени, а ковариационная функция зависит лишь от разности аргументов.

Рассмотрим аналогичную задачу, когда входной и выходной сигналы являются непрерывными случайными сигналами с нечеткими состояниями (кратко нечетко случайными сигналами, или Н.С.С.). При этом не предполагается стационарности в каком-либо смысле. В отличие от известных методик, ниже используется метод функции Грина. Проведем рассмотрение на примерах, которые показывают суть предлагаемого подхода.

Пример 4. Рассмотрим интегрирующую RC-цепочку (рис. 2), описываемую дифференциальным уравнением

$$z'(t) + \beta z(t) = \beta y(t), \quad \beta = \frac{1}{RC} > 0,$$

где R — сопротивление, а C — емкость.

Пусть на вход поступает непрерывный Н.С.С. $\tilde{Y}(t)$. Определим связь между нечеткими ожиданиями (а также ожиданиями) выходного и входного сигналов данной системы. По условию выходной Н.С.С. $\tilde{X}(t)$ удовлетворяет нечеткому случайному дифференциальному уравнению

$$(13) \quad \tilde{X}'(t) + \beta \tilde{X}(t) = \beta \tilde{Y}(t).$$

Отметим, что нечеткие дифференциальные уравнения рассмотрены, например, в [22–24].

Возьмем нечеткое ожидание от обеих частей равенства (13). С учетом свойства аддитивности и однородности ожидания Н.С.П. (утверждение 4), а также теоремы 1 получим

$$(M\tilde{X})'(t) + \beta M\tilde{X}(t) = \beta M\tilde{Y}(t).$$

Согласно определению равенства нечетких чисел и операций между ними, а также определению производной для α -индексов и в силу положительности

$\beta > 0$ это уравнение эквивалентно совокупности соотношений для индексов

$$(14) \quad \frac{\partial}{\partial t}(M\tilde{X})_{\alpha}^{\pm}(t) + \beta(M\tilde{X})_{\alpha}^{\pm}(t) = \beta(M\tilde{Y})_{\alpha}^{\pm}(t) \quad (\forall \alpha \in [0, 1]).$$

Предположим дополнительно, что функции $(M\tilde{Y})_{\alpha}^{\pm}(t)$ ограничены по t на всей числовой оси. Тогда ограниченное на всей оси решение уравнения (14), на основании [25, гл. II], существует, единственно, асимптотически устойчиво по Ляпунову и имеет вид

$$(15) \quad \begin{aligned} (M\tilde{X})_{\alpha}^{\pm}(t) &= \beta \int_{-\infty}^{\infty} G_1(t-s)(M\tilde{Y})_{\alpha}^{\pm}(s) ds = \\ &= \beta \int_{-\infty}^t e^{-\beta(t-s)}(M\tilde{Y})_{\alpha}^{\pm}(s) ds \quad (\forall \alpha \in [0, 1]). \end{aligned}$$

Здесь $G_1 = \begin{cases} e^{-\beta t} & \text{при } t > 0; \\ 0 & \text{при } t < 0 \end{cases}$ — функция Грина задачи об ограниченных решениях (скалярного) уравнения (14).

Формула (15) характеризует связь между нечеткими ожиданиями входного и выходного Н.С.С. системы, описываемой уравнением (13).

С учетом (15), (5) устанавливается формула, связывающая ожидания входного и выходного Н.С.С. рассматриваемой системы (13)

$$m(\tilde{X}(t)) = \beta \int_{-\infty}^t e^{-\beta(t-s)} m(\tilde{Y}(s)) ds.$$

Заметим, что согласно определению равенства нечетких чисел, а также по определению производной Н.С.П. из (13) для α -индексов $\tilde{X}(t)$ следует

$$(16) \quad (X_{\alpha}^{\pm})'(t) + \beta X_{\alpha}^{\pm}(t) = \beta Y_{\alpha}^{\pm}(t) \quad (\forall \alpha \in [0, 1]).$$

Рассматривая (16) как уравнение в гильбертовом пространстве \mathcal{H} случайных величин с конечным вторым моментом. Предположим дополнительно, что функции $Y_{\alpha}^{\pm}(t)$ ограничены по t в пространстве \mathcal{H} на всей числовой оси. Тогда, согласно [25, гл. II], получим, что ограниченное на всей числовой оси решение уравнения (16) существует, единственно, асимптотически устойчиво по Ляпунову и задается формулой

$$(17) \quad X_{\alpha}^{\pm}(t) = \beta \int_{-\infty}^t e^{-\beta(t-s)} Y_{\alpha}^{\pm}(s) ds \quad (\forall \alpha \in [0, 1]).$$

Отметим, что в формуле (17) слева выражения $X_{\alpha}^{\pm}(t)$ определяют α -индексы нечеткого числа $\tilde{X}(t)$. В частности, выражение $X_{\alpha}^{+}(t)$ монотонно невозрастающее, а выражение $X_{\alpha}^{-}(t)$ монотонно неубывающее по α , поскольку таковы α -индексы $Y_{\alpha}^{+}(s)$ и $Y_{\alpha}^{-}(s)$ [1, гл. 5] и в силу свойства монотонности интеграла.

Вычислим ковариационную функцию выходного ограниченного на всей оси Н.С.П. \tilde{X} системы, описываемой уравнением (13). Согласно (10), (17) ковариационная функция $K_{X_\alpha^\pm}(t, s)$ имеет вид

$$K_{X_\alpha^\pm}(t, s) = \beta^2 E \left[\left(\int_{-\infty}^t e^{-\beta(t-\tau_1)} Y_\alpha^\pm(\omega, \tau_1) d\tau_1 - E \int_{-\infty}^t e^{-\beta(t-\tau_1)} Y_\alpha^\pm(\omega, \tau_1) d\tau_1 \right) \times \right. \\ \left. \times \left(\int_{-\infty}^s e^{-\beta(s-\tau_2)} Y_\alpha^\pm(\omega, \tau_2) d\tau_2 - E \int_{-\infty}^s e^{-\beta(s-\tau_2)} Y_\alpha^\pm(\omega, \tau_2) d\tau_2 \right) \right].$$

Меняя местами операции математического ожидания E и интегрирования во внутренних скобках, получим

$$(18) \quad K_{X_\alpha^\pm}(t, s) = \beta^2 E \left[\left(\int_{-\infty}^t e^{-\beta(t-\tau_1)} (Y_\alpha^\pm(\omega, \tau_1) - E(Y_\alpha^\pm(\omega, \tau_1))) d\tau_1 \right) \times \right. \\ \left. \times \left(\int_{-\infty}^s e^{-\beta(s-\tau_2)} (Y_\alpha^\pm(\omega, \tau_2) - E(Y_\alpha^\pm(\omega, \tau_2))) d\tau_2 \right) \right] = \\ = \beta^2 \int_{-\infty}^t \int_{-\infty}^s e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} K_{Y_\alpha^\pm}(\tau_1, \tau_2) d\tau_1 d\tau_2.$$

В силу (18) и (9) ковариационные функции выходного и входного Н.С.С. системы, определяемой (13), связаны соотношением

$$K_{\tilde{X}}(t, s) = \frac{\beta^2}{2} \int_{-\infty}^t \int_{-\infty}^s e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} \int_0^1 (K_{Y_\alpha^-}(\tau_1, \tau_2) + K_{Y_\alpha^+}(\tau_1, \tau_2)) d\alpha d\tau_1 d\tau_2 = \\ = \beta^2 \int_{-\infty}^t \int_{-\infty}^s e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} K_{\tilde{Y}}(\tau_1, \tau_2) d\tau_1 d\tau_2.$$

Отметим, что сходимость рассматриваемых несобственных интегралов и законность приведенных рассуждений обеспечиваются экспоненциальными оценками соответствующих подынтегральных выражений.

Пример 5. На вход линейной динамической системы, описываемой дифференциальным уравнением

$$(19) \quad z''(t) + a_1 z'(t) + a_2 z(t) = y(t),$$

поступает Н.С.С. $\tilde{Y}(t)$. Укажем характеристики выходного Н.С.С. $\tilde{X}(t)$.

Согласно (19) имеем нечеткое дифференциальное уравнение

$$(20) \quad \tilde{X}''(t) + a_1 \tilde{X}'(t) + a_2 \tilde{X}(t) = \tilde{Y}(t).$$

Аналогично примеру 4 ожидание выходного сигнала $M(\tilde{X})(t)$ удовлетворяет дифференциальному уравнению

$$(M\tilde{X})''(t) + a_1(M\tilde{X})'(t) + a_2M(\tilde{X})(t) = M(\tilde{Y})(t).$$

Будем считать коэффициенты $a_1, a_2 > 0$. Тогда для α -индексов ожидания получим

$$(21) \quad \left[(M\tilde{X})_{\alpha}^{\pm} \right]''(t) + a_1 \left[(M\tilde{X})_{\alpha}^{\pm} \right]'(t) + a_2 (M\tilde{X})_{\alpha}^{\pm}(t) = M(\tilde{Y})_{\alpha}^{\pm}(t) \quad (\forall \alpha \in [0, 1]).$$

Пусть функции $(M\tilde{Y})_{\alpha}^{\pm}(t)$ ограничены по t на всей числовой оси. Предположим дополнительно, что корни характеристического уравнения $\lambda^2 + a_1\lambda + a_2 = 0$, соответствующего (19), вещественные и отрицательные $\lambda_1 < \lambda_2 < 0$. Тогда ограниченное на всей числовой оси решение уравнения (21) существует, единственно, асимптотически устойчиво по Ляпунову и имеет вид

$$(22) \quad (M\tilde{X})_{\alpha}^{\pm}(t) = \int_{-\infty}^t G_2(t-s)(M\tilde{Y})_{\alpha}^{\pm}(s) ds \quad (\forall \alpha \in [0, 1]).$$

Здесь G_2 — функция Грина задачи об ограниченных решениях скалярного уравнения (19) (см., например, [26, гл. 2, § 8]). В сделанных предположениях она имеет вид

$$G_2(t) = \begin{cases} (e^{\lambda_2 t} - e^{\lambda_1 t})(\lambda_2 - \lambda_1)^{-1} & \text{при } t > 0; \\ 0 & \text{при } t < 0. \end{cases}$$

Формула (22) характеризует связь нечетких ожиданий входного и выходного Н.С.С. системы, описываемых уравнением (20).

В предположении ограниченности на всей числовой оси в пространстве \mathcal{H} α -индексов $Y_{\alpha}^{\pm}(t)$, аналогично рассуждениям примера 4 устанавливается следующая формула:

$$(23) \quad \begin{aligned} K_{\tilde{X}}(t, s) &= \frac{1}{2} \int_{-\infty}^t \int_{-\infty}^s G_2(t-\tau_1)G_2(s-\tau_2) \times \\ &\times \int_0^1 \left(K_{\tilde{Y}_{\alpha}^{-}}(\tau_1, \tau_2) + K_{\tilde{Y}_{\alpha}^{+}}(\tau_1, \tau_2) \right) d\alpha d\tau_1 d\tau_2 = \\ &= \int_{-\infty}^t \int_{-\infty}^s G_2(t-\tau_1)G_2(s-\tau_2)K_{\tilde{Y}}(\tau_1, \tau_2) d\tau_1 d\tau_2. \end{aligned}$$

Формула (23) характеризует связь корреляционных функций входного и выходного Н.С.С. системы, описываемых уравнением (20).

Заметим, что в условиях примеров 4, 5 выходные процессы асимптотически устойчивы по Ляпунову в том смысле, что их α -индексы асимптотически устойчивы по Ляпунову (см., например, [25, гл. II]). Только такие процессы физически реализуемы. В литературе рассматриваются и другие подходы к устойчивости решений нечетких дифференциальных уравнений (см., например, [3, гл. 8]).

5. Заключение

Отправным пунктом данного исследования является статья [20], в которой обсуждаются ковариации Н.С.В. Благодаря ей, появилось понятие ковариационной функции Н.С.П., изученное в настоящей статье.

Свойства нечетких ожиданий Н.С.П. (утверждение 4) и ковариационных функций (утверждение 5) естественным образом вытекают из соответствующих свойств нечетких ожиданий и ковариаций для нечетко-случайных величин (утверждения 1–3).

Существенное содержание и научную новизну данной работы составляют теоремы 1–4, связанные с характеристиками дифференцируемых и интегрируемых Н.С.П. При их доказательстве существенно использовались определения и свойства дифференцируемости и интегрируемости нечеткозначных функций. Теоремы 1–4 являются развитием известных (см., например, [6]) утверждений для стандартных непрерывных случайных процессов. Отметим еще экстремальное свойство нечетких ожиданий Н.С.П. (утверждение 4), которое автору представляется новым.

Примеры 1–3 носят иллюстративный характер. Примеры 4–5 показывают возможность применения развитой теории к прикладным задачам, в частности к задаче преобразования Н.С.С. линейной динамической системой. Результаты раздела 4 допускают развитие на случай периодических и почти периодических Н.С.С., в том числе для задачи о спектральных разложениях Н.С.С.

СПИСОК ЛИТЕРАТУРЫ

1. *Аверкин А.Н.* Нечеткие множества в моделях управления и искусственного интеллекта. М.: Наука. 1986. 312 с.
2. *Buckley J.J., Eslami E., Feuring T.* Fuzzy mathematics in economic and engineering. Heidelberg, New-York: Physica-Verl. 2002. 282 p.
3. *Пегат А.* Нечеткое моделирование и управление. М.: БИНОМ. Лаборатория знаний. 2015. 798 с.
4. *Мочалов И.А., Хрисат М.С., Шихаб Еддин М.Я.* Нечеткие дифференциальные уравнения в задачах управления. Часть II // Информационные технологии. 2015. Т. 21. № 4. С. 243–250.

5. Деменков Н.П., Микрин Е.А., Мочалов И.А. Нечеткое оптимальное управление линейными системами. Часть 1. Позиционное управление // Информационные технологии. 2019. Т. 25. № 5. С. 259–270.
6. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и их инженерные приложения. М.: Кнорус. 2016. 439 с.
7. Puri M.L., Ralesku D.A. Fuzzy random variables // Journal of Mathematical Analysis and Applications. 1986. V. 114. P. 409–422.
8. Nguyen H.T., Wu B. Fundamentals of Statistics with Fuzzy Data. Springer-Verlag Berlin Heidelberg. 2006. 204 p.
9. Шведов А.С. Оценивание средних и ковариаций нечетко-случайных величин // Прикладная эконометрика. 2016. Т. 42. С. 121–138.
10. Puri M.L., Ralescu D.A. Differential of fuzzy functions // J. Math. Anal. Appl. 1983. V. 91. P. 552–558.
11. Seikkala S. On the fuzzy initial value problem // Fuzzy Sets Syst. 1987. V. 24. No. 3. P. 319–330.
12. Aumann R.J. Integrals of set-valued functions // J. Math. Anal. Appl. 1987. No. 3. P. 1–12.
13. Hukuhara M. Integration des applications mesurables dont la valeur est un compact convexe // Func. Ekvacioj. 1967. No. 11. P. 205–223.
14. Деменков Н.П., Микрин Е.А., Мочалов И.А. Марковские и полумарковские процессы с нечеткими состояниями. Ч. 1. // Информационные технологии. 2020. Т. 26. № 6. С. 323–334.
15. Деменков Н.П., Микрин Е.А., Мочалов И.А. Марковские и полумарковские процессы с нечеткими состояниями. Ч. 2. // Информационные технологии. 2020. Т. 26. № 7. С. 387–393.
16. Diamond P., Kloeden P. Metric Spaces of Fuzzy Sets // Fuzzy Sets Syst. 1990. V. 35. No. 2. P. 241–249.
17. Хацкевич В.Л. О некоторых свойствах нечетких ожиданий и нелинейных нечетких ожиданий нечетко-случайных величин // Известия вузов. Математика. 2022. № 11. С. 97–109.
18. Хацкевич В.Л. Экстремальные свойства средних нечетко-случайных величин // Итоги науки и техники. Сер. Соврем. мат. и ее прил. Темат. обз. 2022. Т. 204. С. 160–169.
19. Dubois D., Prade H. The mean value of fuzzy number // Fuzzy Sets and Syst. 1987. V. 24. No. 3. P. 279–300.
20. Feng Y., Hu L., Shu H. The variance and covariance of fuzzy random variables // Fuzzy Syst. 2001. V. 120. No. 3. P. 487–497.
21. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика. М.: Наука. 1985. 322 с.
22. Kaleva O. Fuzzy differential equations // Fuzzy Sets and Syst. 1987. V. 24. No. 3. P. 301–317.
23. Jong Yeoul Park, Han H. Existence and uniqueness theorem for a solution of fuzzy differential equations // Int. J. Math. Mathem. Sci. 1999. No. 22(2). P. 271–280.
24. Ahmad L., Farooq M., Abdullah S. Solving nth order fuzzy differential equation by fuzzy Laplace transform // Ind. J. Pure Appl. Math. 2014. No. 2. P. 1–20.

25. *Далецкий Ю.Л., Крейн М.Г.* Устойчивость решений дифференциальных уравнений в банаховом пространстве. М.: Наука. 1970. 533 с.
26. *Красносельский М.А., Бурд В.Ш., Колесов Ю.С.* Нелинейные почти периодические колебания. М.: Наука. 1970. 351 с.

Статья представлена к публикации членом редколлегии А.И. Кибзуном.

Поступила в редакцию 03.08.2022

После доработки 12.01.2023

Принята к публикации 30.03.2023

Управление в технических системах

© 2023 г. П.А. АКИМОВ, канд. физ.-мат. наук (akmpavel@rambler.ru),
А.И. МАТАСОВ, д-р физ.-мат. наук (alexander.matasov@gmail.com)
(Московский государственный университет им. М.В. Ломоносова)

МЕТОД ГАРАНТИРУЮЩЕГО ОЦЕНИВАНИЯ ПРИ КАЛИБРОВКЕ БЛОКА ГИРОСКОПОВ

Статья посвящена применению метода гарантирующего оценивания к задаче калибровки блока гироскопов. Построены математические модели, описывающие кинематику блока гироскопов на калибровочном стенде, исследованы границы их применимости и возникающие погрешности. Для проблем гарантирующего оценивания разработана методика численного решения, основанная на сведениях к задачам l_1 -аппроксимации.

Ключевые слова: гарантирующий подход к оцениванию, выпуклая оптимизация, блок гироскопов, калибровка, Python.

DOI: 10.31857/S0005231023070036, EDN: FCNJCE

1. Введение

В данной работе предложена методика калибровки блока датчиков угловой скорости (ДУС), гироскопов, основанная на методе гарантирующего оценивания [1–5]. Цель калибровки состоит в оценке параметрических погрешностей (сдвигов нулей, масштабных коэффициентов и углов перекаса) в блоке, состоящем из трех датчиков угловой скорости. Для определения этих параметров производится серия измерений на стенде, позволяющем с высокой точностью управлять скоростью вращения и ориентацией блока. При калибровке решается задача выбора режимов вращения и обработки полученных сигналов.

Во многих традиционных методах калибровки используются динамические модели оценивания неизвестных параметров по измерениям датчиков инерциальной навигационной системы (ИНС) — ньютонометров (акселерометров) и гироскопов [6–8]. Тем не менее, как указано в [9], определение погрешностей ДУС без привлечения информации от ньютонометров в ряде случаев является более предпочтительной. Такая ситуация возникает, во-первых, когда точность ньютонометров не позволяет использовать их сигналы при стендовых испытаниях и, во-вторых, при калибровке лазерных гироскопов на виброподвесах. Основным инструментом решения проблем оценивания служат фильтр Калмана или метод наименьших квадратов [10]. В силу наличия

большого количества неизвестных параметров в задаче калибровки (погрешности датчиков, погрешности стенда), нелинейно влияющих на результаты измерений, существенны два вопроса: как построить математическую модель стендовых испытаний, учитывающую все эти факторы, и как минимизировать влияние погрешностей на результат оценки. Вариант ответа на эти вопросы представлен в данной статье, которая является продолжением исследования по применению гарантирующего подхода в инерциальной навигации, описанного, например, в [11], где рассматривается калибровка блока ньютонометров. Однако в отличие от указанной статьи здесь этот метод применяется к другому классу систем, характеризующемуся большим количеством неизвестных параметров и нелинейными эффектами, что требует построения других моделей и приводит к иным, структурно более сложным проблемам оценивания.

Метод гарантирующего оценивания дает возможность решить проблемы оценивания неизвестных параметров при “наихудших” реализациях ошибок измерений, максимально снижающих точность оценки. При этом вместо вероятностных гипотез используются характерные максимальные значения погрешностей стенда. Получаемое данным методом решение задает направления вращения блока гироскопов на стенде, т.е. явным образом описывают оптимальные режимы калибровки.

Структура статьи следующая. В первом разделе строится математическая модель стендовых испытаний блока гироскопов, осуществляется переход от основных кинематических уравнений к нескольким вариантам линейной модели, возникающим в результате осреднения сигналов, исследуются границы применимости линейных моделей. Во втором и третьем разделах формулируются проблемы гарантирующего оценивания и соответствующие им вспомогательные вариационные задачи l_1 -аппроксимации, которые могут быть решены численно. Четвертый раздел посвящен определению оптимальных планов эксперимента, применению алгоритмов оценивания в модельных задачах и анализу точности полученных решений.

2. Математическая модель

2.1. Основные предположения и соотношения кинематики блока гироскопов

Рассмотрим математическую модель, описывающую основные кинематические соотношения блока гироскопов на стенде при калибровке [12]. Пусть эксперимент проводится в точке M , жестко связанной с Землей и находящейся на ее поверхности. Введем следующие обозначения:

$Mz = Mz_1z_2z_3$ — приборная система координат, жестко связанная с блоком гироскопов;

$Mx = Mx_1x_2x_3$ — система координат, жестко связанная с неподвижным относительно Земли основанием стенда;

$D(t)$ — ортогональная матрица ориентации Mz относительно Mx . По определению матрицы ориентации для любого вектора l его координаты в системах отсчета Mz, Mx связаны соотношением $l_z = D(t)l_x$, а строки матрицы ориентации состоят из координат базисных векторов Mz в системе Mx ;

$\Omega(t)$ — вектор угловой скорости вращения блока гироскопов относительно стенда;

$\omega(t)$ — вектор абсолютной угловой скорости вращения блока гироскопов;

u_x — координаты угловой скорости вращения Земли в системе Mx , u — абсолютная величина угловой скорости вращения Земли.

В проекциях на Mx абсолютная угловая скорость блока гироскопов описывается соотношением

$$\omega_x(t) = \Omega_x(t) + u_x.$$

В проекциях на оси трехгранника Mz данное равенство примет вид

$$(1) \quad \omega_z(t) = D(t) (\Omega_x(t) + u_x).$$

В начальный момент времени матрица ориентации D известна с некоторой точностью, эту оценку $D(0)$ обозначим как D_{init} :

$$(2) \quad D(0) = D_{\text{init}}(I_3 + \hat{\beta}), \quad \beta = (\beta_1, \beta_2, \beta_3)^T, \quad \hat{\beta} = \begin{pmatrix} 0 & \beta_3 & -\beta_2 \\ -\beta_3 & 0 & \beta_1 \\ \beta_2 & -\beta_1 & 0 \end{pmatrix},$$

$$|\beta_i| \leq \beta_{\text{max}}, \quad i = 1, 2, 3.$$

Погрешности начальной выставки блока гироскопов — углы малого поворота β_i — неизвестны, но ограничены по величине значением β_{max} . Обозначение I_n здесь и далее используется для единичной матрицы размерности $n \times n$, а символ $\hat{\beta}$ обозначает кососимметрическую матрицу, построенную по вектору β согласно вышеуказанному правилу.

Для блока гироскопов выходными сигналами (измерениями) являются показания каждого из ДУС — компоненты вектора $\omega_z(t)$. Введем следующую модель измерений:

$$(3) \quad \zeta(t) = \omega_z(t) + \Gamma\omega_z(t) + \nu_0 + \delta\nu(t).$$

Здесь $\zeta(t) \in \mathbf{R}^3$ — значения измерений, $\Gamma \in \mathbf{R}^{3 \times 3}$ — неизвестная матрица, описывающая ошибки масштабных коэффициентов датчиков и ошибки в знании ориентации осей чувствительности, $\nu_0 \in \mathbf{R}^3$ — неизвестные смещения нулей в показаниях датчиков, $\delta\nu(t) \in \mathbf{R}^3$ — непараметрические флуктуирующие ошибки измерения. Матрицу Γ без ограничения общности будем считать симметричной [5].

Задача калибровки состоит в определении значений Γ , ν_0 по совокупности доступных измерений $\zeta(t)$. При этом на стенде можно управлять угловой

скоростью вращения $\Omega_x(t)$, и именно формирование оптимальной стратегии движения блока на стенде является одной из целей математической задачи калибровки. Представим угловую скорость $\Omega_x(t)$ в виде

$$\Omega_x(t) = s'(t)w,$$

где $s'(t) \in \mathbf{R}$ — величина угловой скорости вращения, $w \in \mathbf{R}^3$ — единичный вектор направления угловой скорости вращения Mz в проекциях на Mx . Далее будем рассматривать частный случай движения блока гироскопов на стенде, состоящий из нескольких однотипных экспериментов. В рамках каждого из них блок вращается вокруг фиксированной оси с задаваемой угловой скоростью, затем блок переводится в новое положение, задается новое направление его вращения и эксперимент повторяется. Направление оси вращения и величина угловой скорости известны с некоторыми погрешностями, т.е. считаются известными функция $s(t) \geq 0$ и вектор y ($\|y\|_2 = 1$), которые связаны с истинными значениями следующими выражениями:

$$(4) \quad s'(t) = s(t) + \varepsilon(t), \quad w = (I_3 + \hat{\alpha})y, \quad \Omega_x(t) = (s(t) + \varepsilon(t))(I_3 + \hat{\alpha})y.$$

Углы малого поворота $\alpha \in \mathbf{R}^3$ и соответствующая им кососимметрическая матрица $\hat{\alpha}$ определяют неизвестные погрешности в задании вектора вращения блока гироскопов, а скалярная функция $\varepsilon(t)$ — погрешность в задании величины угловой скорости. Как и в случае с углами β , известными являются максимальные допустимые значения для α : $|\alpha_i| \leq \alpha_{\max}$, $i = 1, 2, 3$.

На каждом этапе испытаний погрешности α , β неизменны, но для разных экспериментов они разные, поскольку предполагается, что задание программной, целевой, угловой скорости и ориентации блока происходит при каждом эксперименте независимо. Таким образом формируется несколько серий измерений $\zeta(t)$, соответствующих разным режимам вращения и разным реализациям погрешностей.

2.2. Линеаризация уравнений и осреднение сигнала

Подставив (1), (4) в (3), получим выражение для измерений

$$(5) \quad \zeta(t) = (I_3 + \Gamma) D(t) (s'(t)w + u_x) + \nu_0 + \delta\nu(t).$$

Помимо сигнала $\zeta(t)$, от времени зависят две составляющие угловой скорости, $D(t)u_x$ и $s'(t)D(t)w$, и неопределенные погрешности $\delta\nu(t)$, $\varepsilon(t)$. В этом разделе рассматривается математическая модель, соответствующая режиму вращения блока гироскопов с постоянной по направлению угловой скоростью в течение периода времени длительностью T . Процедура калибровки будет состоять из последовательности таких режимов вращения вокруг различных направлений.

Далее построим “осредненный” по времени аналог уравнения измерений, для чего необходимо учесть эффект осреднения при вращении Mz относительно Mx . Под осреднением понимается процедура расчета средних арифметических значений от серии измерений на длительном интервале времени T (несколько десятков минут), в течение которого система делает не один полный оборот.

Воспользуемся тем фактом, что $D(t)$ — матрица перехода от Mx к Mz , а Mz жестко связана с блоком гироскопов, вращающимся относительно Mx с угловой скоростью Ω_x . Переход от Mx к Mz можно представить состоящим из трех этапов.

1. Переход от Mx к Mx_{fix} — неподвижному относительно основания стенда трехграннику, орт $e_{\text{fix}3}$ которого по направлению совпадает с w . Соответствующую матрицу перехода обозначим как $D_{\text{fix}} = (d_{\text{fix}1}; d_{\text{fix}2}; d_{\text{fix}3})$, ее третья строка равна $d_{\text{fix}3} = w^T$, а первая и вторая ортогональны ей и между собой и могут быть выбраны любым подходящим образом.

2. Вращение относительно оси $Mx_{\text{fix}3} = Mz_{\text{cir}3}$ со скоростью $s'(t)$, переводящее Mx_{fix} в трехгранник Mz_{cir} , жестко связанный с блоком гироскопов. Зависящий от времени угол вращения в плоскости $Mx_{\text{fix}1}x_{\text{fix}2}$ обозначим через $\psi(t)$. В силу принятых обозначений имеем $\frac{d\psi(t)}{dt} = s'(t)$, а матрица перехода от Mx_{fix} к Mz принимает вид

$$D_{\text{cir}}(t) = \begin{pmatrix} \cos \psi(t) & -\sin \psi(t) & 0 \\ \sin \psi(t) & \cos \psi(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

3. Переход от Mz_{cir} к Mz , задающийся не известной точно постоянной ортогональной матрицей, которую для дальнейших выкладок удобно представить в виде $D' = (d'_1, d'_2, d'_3)$.

Таким образом, матрицу $D(t)$ можно представить в виде произведения фиксированных и зависящих от времени матриц перехода

$$D(t) = D' D_{\text{cir}}(t) D_{\text{fix}}.$$

Получим выражение для $s'(t)D(t)w$, воспользовавшись тем, что по определению первые две строки матрицы D_{fix} ортогональны w :

$$\begin{aligned} s'(t)D(t)w &= s'(t)D' D_{\text{cir}}(t) D_{\text{fix}} w = \\ (6) \quad &= s'(t)D' \begin{pmatrix} \cos \psi(t) & -\sin \psi(t) & 0 \\ \sin \psi(t) & \cos \psi(t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = s'(t)d'_3. \end{aligned}$$

По определению матрицы перехода D' столбец d'_3 состоит из проекций ортов приборного трехгранника на ось $Mz_{\text{cir}3}$. Так как направление $Mz_{\text{cir}3}$

не меняется во времени и вращение происходит вокруг этой оси, то проекции ортов приборного трехгранника на данное направление также остаются постоянными. Поэтому d'_3 можно определить на основании априорной информации (2):

$$(7) \quad d'_3 = D(t)w = D(0)w = D_{\text{init}}(I_3 + \hat{\beta})w,$$

т.е. при вращении вокруг фиксированной оси направление вектора $D(t)\Omega_x(t)$ остается постоянным и в результате его осреднения получаем вектор $s'D_{\text{init}}(I_3 + \hat{\beta})w$, где s' — среднее значение $s'(t)$.

В силу рассматриваемого режима движения результат осреднения вектора $D(t)u_x$ будет иметь специальную структуру, описанную в лемме 1.

Лемма 1. Пусть величина угловой скорости вращающейся системы координат описывается функцией $s'(t) = s + \varepsilon(t)$, “программная” угловая скорость вращения s постоянна и $s > \varepsilon_{\text{max}}$, а вращение происходит вокруг фиксированного направления w . Тогда при осреднении сигнала во времени проекции вектора вращения Земли на оси Mz описываются равенством

$$\bar{u}_z = D_{\text{init}}(I_3 + \hat{\beta})ww^T u_x + u^\perp.$$

При этом неизвестный вектор u^\perp ортогонален вектору $D_{\text{init}}(I_3 + \hat{\beta})w$ и по модулю его компоненты оцениваются сверху как

$$|u_i^\perp| \leq u \left(\frac{4}{T(s - \varepsilon_{\text{max}})} + C \frac{\varepsilon_{\text{max}}}{s} \right) \stackrel{\text{def}}{=} u_{\text{max}},$$

где C — ограниченная величина, зависящая от начального и конечного значений угла поворота на стенде.

Получим явный вид уравнения измерений, перейдя к осредненным величинам в уравнении (5) и используя выражения (4), (7), а также результаты леммы 1:

$$(8) \quad \begin{aligned} \tilde{\zeta} &= (I_3 + \Gamma) \left(s'D_{\text{init}}(I_3 + \hat{\beta})w + D_{\text{init}}(I_3 + \hat{\beta})ww^T u_x + u^\perp \right) + \nu_0 + \delta\tilde{\nu} = \\ &= (I_3 + \Gamma) D_{\text{init}}(I_3 + \hat{\beta}) \left((s + \varepsilon)(I_3 + \hat{\alpha})y + (I_3 + \hat{\alpha})yy^T(I_3 + \hat{\alpha})^T u_x \right) + \\ &\quad + (I_3 + \Gamma) u^\perp + \nu_0 + \delta\tilde{\nu}. \end{aligned}$$

Здесь ε — результат осреднения шума $\varepsilon(t)$ — и погрешность $\delta\tilde{\nu}$, являющаяся результатом осреднения во времени шума $\delta\nu(t)$, будем считать ограниченными по абсолютной величине

$$|\varepsilon| \leq \varepsilon_{\text{max}}, \quad |\delta\tilde{\nu}_j| \leq \nu_{\text{max}}, \quad j = 1, 2, 3,$$

где константа ν_{max} считается известной, характеризует априорное знание о величине погрешности гироскопа.

Таблица 1. Порядки величин для параметров модели

Величина	Порядок величины
α, α_{\max}	$1' \approx 2,9 \cdot 10^{-4}$
β, β_{\max}	$5' \approx 1,5 \cdot 10^{-3}$
$\varepsilon(t)$	$5 \cdot 10^{-6} \text{ 1/c}$
$\varepsilon, \varepsilon_{\max}$	$1 \cdot 10^{-8} \text{ 1/c}$
s	$17,5 \text{ 1/c}$
T	$600\text{--}1200 \text{ c}$
Γ_{ii}	$5 \cdot 10^{-3} (5 \cdot 10^{-5})$
$\Gamma_{ij}, i \neq j$	$5 \cdot 10^{-3} (5 \cdot 10^{-5})$
ν_0	$2,4 \cdot 10^{-7} \text{ 1/c} (5 \cdot 10^{-8} \text{ 1/c})$
ν_{\max}	$1,2 \cdot 10^{-8} \text{ 1/c}$
u	$7,292115 \cdot 10^{-5} \text{ 1/c}$

В зависимости от масштабов переменных $\alpha, \beta, \Gamma, \varepsilon, \nu$ можно по-разному упрощать модель измерений, пренебрегая теми или иными группами слагаемых. В табл. 1, см. ниже, зафиксируем характерные масштабы параметров модели, соответствующие типовым точностям стендов и гироскопов, а также требования к точности оценки параметров Γ, ν_0 (указаны в скобках).

Важно заметить, что в рассматриваемом случае величины $\nu_{\max}, \varepsilon_{\max}$ значительно меньше, чем характерные амплитуды $\delta\nu(t), \varepsilon(t)$, что соответствует случаю, когда исходный сигнал был осреднен. Предлагаемый в статье подход не является применимым исключительно при вышеуказанных масштабах переменных, и в зависимости от реальной задачи могут быть применены иные способы упрощения исходной модели измерений.

При выбранных значениях в выражении для измерений слагаемые делятся на несколько групп: не являющиеся малыми, такие, например, как $sy, \tilde{\zeta}$; линейно зависящие от малых параметров $\alpha, \beta, \varepsilon, u$; пренебрежимо малые, второго и третьего порядка малости, гарантированно не превосходящие ν_{\max} ; нелинейные слагаемые, которыми нельзя пренебречь в силу того, что они зависят от $s\Gamma D_{\text{init}}\hat{\alpha}, s\Gamma D_{\text{init}}\hat{\beta}, sD_{\text{init}}\hat{\alpha}\hat{\beta}, \Gamma u^\perp, u^\perp$ и могут превышать ν_{\max} .

После исключения малых слагаемых уравнение (8) примет вид

$$\begin{aligned} \tilde{\zeta} = & D_{\text{init}} \left(s(I_3 + \hat{\alpha} + \hat{\beta})y + s\hat{\beta}\hat{\alpha}y + \varepsilon y + (I_3 + \hat{\alpha} + \hat{\beta})yy^T u_x - yy^T \hat{\alpha}u_x \right) + \\ & + u^\perp + \Gamma u^\perp + \Gamma D_{\text{init}}(sy + yy^T u_x) + \Gamma D_{\text{init}}s(\hat{\alpha} + \hat{\beta})y + \nu_0 + \delta\tilde{\nu}. \end{aligned}$$

Для того, чтобы слагаемыми $sy^T \hat{\beta}\hat{\alpha}y, \Gamma D_{\text{init}}s(\hat{\alpha} + \hat{\beta})y$ также можно было пренебречь, достаточно, чтобы они по величине были сопоставимы с неустраиваемой помехой $\delta\tilde{\nu}$, т.е. $\|sy^T \hat{\beta}\hat{\alpha}y\|_\infty \leq \nu_{\max}, \|\Gamma D_{\text{init}}s(\hat{\alpha} + \hat{\beta})y\|_\infty \leq \nu_{\max}$. Поскольку масштаб компонент матрицы Γ известен (см. табл. 1), $|\Gamma_{ij}| \leq \Gamma_{\max}$, введем

дополнительное ограничение на величину угловой скорости s :

$$(9) \quad s \leq \frac{\nu_{\max}}{2 \max\{\alpha_{\max}\beta_{\max}, \Gamma_{\max}(\alpha_{\max} + \beta_{\max})\}} \stackrel{\text{def}}{=} s_{\max}.$$

Иными словами, вращение блока гироскопов на стенде с меньшей угловой скоростью позволяет снизить влияние погрешностей α, β на результат оценивания.

Если вращение будет слишком медленным, то может оказаться, что ошибка осреднения — слагаемое Γu^\perp — превосходит требуемую точность оценки, откуда получаем второе ограничение на параметр s : $\|\Gamma u^\perp\|_\infty \leq \nu_{\max}$, или с учетом априорного знания масштабов компонент матрицы Γ_{\max} и леммы 1

$$\|\Gamma u^\perp\|_\infty \leq 3\Gamma_{\max} u \left(\frac{4}{(s - \varepsilon_{\max})T} + C \frac{\varepsilon_{\max}}{s} \right) \leq \nu_{\max}.$$

В силу того, что слагаемое $\Gamma_{\max} C u \varepsilon_{\max}/s$ мало по сравнению с ν_{\max} , ограничение на s принимает вид

$$(10) \quad 3\Gamma_{\max} u \frac{4}{(s - \varepsilon_{\max})T} \leq \nu_{\max}, \quad \text{или} \quad s \geq \frac{12\Gamma_{\max} u}{\nu_{\max} T} + \varepsilon_{\max} \stackrel{\text{def}}{=} s_{\min}.$$

Скомпенсировать влияние u^\perp также можно, радикально увеличив T , однако, как будет показано далее, это достигается иным способом — при помощи скаляризации.

Помимо вращения со скоростью sy возможен еще один режим стендовых испытаний, при котором блок ДУС неподвижен относительно основания стенда и гироскопы измеряют угловую скорость вращения Земли. Как нетрудно проверить, выражение для измерений в этом случае может быть получено из (5) подстановкой $s' = 0$, $s = 0$, $\varepsilon(t) = 0$, $D(t) = D_{\text{init}}(I_3 + \hat{\beta})$ и переходом к осредненным сигналам $\tilde{\zeta}$:

$$\begin{aligned} \tilde{\zeta} &= (I_3 + \Gamma) D_{\text{init}}(I_3 + \hat{\beta}) u_x + \nu_0 + \delta\tilde{\nu} = \\ &= D_{\text{init}} u_x + \Gamma D_{\text{init}} u_x + D_{\text{init}} \hat{\beta} u_x + \nu_0 + \delta\tilde{\nu} \end{aligned}$$

с точностью до слагаемых второго порядка малости.

Подводя итог данного раздела, сформулируем модель осредненного сигнала гироскопов с учетом введенных выше ограничений (9), (10).

$$(11) \quad \begin{aligned} \tilde{\zeta} &= D_{\text{init}} \left(s(I_3 + \hat{\alpha} + \hat{\beta})y + \varepsilon y + (I_3 + \hat{\alpha} + \hat{\beta})y y^T u_x - y y^T \hat{\alpha} u_x \right) + \\ &+ \Gamma D_{\text{init}} (s y + y y^T u_x) + u^\perp + \nu_0 + \delta\tilde{\nu}, \quad s \in \{0\} \cup [s_{\min}, s_{\max}]. \end{aligned}$$

2.3. Модели измерений и скаляризация

В уравнении (11) входной информацией являются слагаемые $\tilde{\zeta}$, $D_{\text{init}}(s y + y y^T u_x)$, а “полезным сигналом” — слагаемые $\Gamma D_{\text{init}}(s y + y y^T u_x) + \nu_0$. Погрешности измерений состоят из вектора $\delta\tilde{\nu}$, возникающего при осреднении

флуктуационного шума $\delta\tilde{\nu}(t)$, и неизвестных систематических погрешностей, обусловленных неточностью стенда и меняющихся при каждом новом стендовом испытании. После переноса известных слагаемых в левую часть уравнения (11) получим следующую линейную модель измерений:

$$(12) \quad z(s, y) = \Gamma D_{\text{init}}(sy + yy^T u_x) + \nu_0 + r + \delta\nu',$$

в которой сами “измерения” $z = z(s, y)$ и их погрешности $r = r(s, y, \alpha, \beta, \varepsilon)$, $\delta\nu'$ определены следующим образом:

$$(13) \quad \begin{aligned} z &= \tilde{\zeta} - D_{\text{init}}(sy + yy^T u_x), \\ r &= D_{\text{init}} \left(s(\hat{\alpha} + \hat{\beta})y + \varepsilon y + (\hat{\alpha} + \hat{\beta})yy^T u_x - yy^T \hat{\alpha} u_x \right), \end{aligned}$$

$$(14) \quad \delta\nu' = \delta\tilde{\nu} + u^\perp, \quad |\delta\nu'_j| \leq \nu_{\text{max}} + u_{\text{max}} \stackrel{\text{def}}{=} \nu'_{\text{max}}, \quad j = 1, 2, 3.$$

Для определенности назовем эту модель трехмерной (поскольку $z(s, y) \in \mathbf{R}^3$) или моделью-1.

В модели (12)–(14) составляющая $\delta\tilde{\nu}$ вследствие осреднения не имеет внятного спектра и поэтому исчерпывающе описывается неравенством $|\delta\tilde{\nu}_j| \leq \nu_{\text{max}}$. Составляющая u^\perp не является столь же произвольной (см. лемму 1). Поэтому принятое в (14) ограничение $|\delta\nu'_j| \leq \nu'_{\text{max}}$ является загрубленным.

Воспользуемся методом скаляризации [11], который состоит в том, что исходные трехмерные уравнения измерений скалярно домножаются на известный вектор, в данном случае — на $\tilde{y} = D_{\text{init}}y$. После такого шага все слагаемые в r , представляющие собой произведение кососимметричной матрицы на вектор y , сокращаются. Скалярное произведение $y^T D_{\text{init}}^T u^\perp$ также приближенно можно считать равным нулю в силу леммы 1:

$$(15) \quad 0 = w^T (I_3 - \hat{\beta}) D_{\text{init}}^T u^\perp = y^T (I_3 - \hat{\alpha}) (I_3 - \hat{\beta}) D_{\text{init}}^T u^\perp \approx y^T D_{\text{init}}^T u^\perp$$

— с точностью до малых слагаемых, имеющих порядок $(\alpha_{\text{max}} + \beta_{\text{max}})u \times \left(\frac{4}{T(s - \varepsilon_{\text{max}})} + C \frac{\varepsilon_{\text{max}}}{s} \right)$. Данная точность приближения является достаточной, поскольку

$$(\alpha_{\text{max}} + \beta_{\text{max}})u \left(\frac{4}{T(s - \varepsilon_{\text{max}})} + C \frac{\varepsilon_{\text{max}}}{s} \right) \leq \nu_{\text{max}}$$

при рассматриваемых масштабах параметров α_{max} , β_{max} , s , T , ε_{max} .

Следовательно, появляется возможность перейти к одномерной (скалярной) модели измерений с меньшими по масштабу погрешностями:

$$(16) \quad z_{\text{scal}} = z_{\text{scal}}(s, y) = \tilde{y}^T \Gamma D_{\text{init}}(sy + yy^T u_x) + \tilde{y}^T \nu_0 + r_{\text{scal}} + \tilde{y}^T \delta\tilde{\nu},$$

где

$$(17) \quad z_{\text{scal}} = \tilde{y}^T \tilde{\zeta} - s - y^T u_x, \quad r_{\text{scal}} = \varepsilon - y^T \hat{\alpha} u_x.$$

Назовем эту модель измерений скалярной или моделью-2. В ней учтена специфика слагаемого, содержащего величину u^\perp , почти исчезающего при скаляризации (см. (15)). Поэтому модель для помехи измерений z_{scal} в этом случае более адекватна, чем для трехмерной модели-1, этим и объясняется ее лучшая точность.

3. Формулировки задач гарантирующего оценивания

В соответствии с изложенными в [5, 11] идеями сформулируем задачу оценки неизвестных матрицы Γ и вектора ν_0 в виде проблемы гарантирующего оценивания. Вектор неизвестных оцениваемых параметров в задаче состоит из компонент погрешностей Γ, ν_0 :

$$\gamma = (\Gamma_{11}, \Gamma_{21}, \Gamma_{31}, \Gamma_{12}, \Gamma_{22}, \Gamma_{32}, \Gamma_{13}, \Gamma_{23}, \Gamma_{33})^T, \quad q = (\gamma^T, \nu_{01}, \nu_{02}, \nu_{03})^T \in \mathbf{R}^{12}.$$

Необходимо оценить скалярную величину $a^T q$, где вектор $a \in \mathbf{R}^{12}$ задан. К примеру, при оценивании компоненты Γ_{11} имеем $a = (1, 0, \dots, 0)^T$ и т.д.

Оценка ищется в виде линейного функционала от измерений

$$l(\Phi) = \int_{(y,s) \in S} \Phi_0^T(y, s) z(y, s) dy ds + \sum_{k=1}^K \Phi^T(k) z(y(k), s(k)),$$

где интеграл берется по множеству

$$S = \{y \in \mathbf{R}^3 : \|y\|_2 = 1\} \times \{0\} \cup [s_{\min}, s_{\max}]$$

(далее для краткости множество, по которому берется интеграл, не будет указываться), $\Phi_0(\cdot) : S \rightarrow \mathbf{R}^3$, $\Phi^{(k)} \in \mathbf{R}^3$, а набор значений $\{y(k), s(k)\}_{k=1}^K$ задает совокупность изолированных положений и величин вектора угловой скорости вращения в рамках эксперимента. Каждому элементу (y, s) соответствуют измерения $z(y, s)$, полученные в результате обработки сигналов гироскопов при вращении с соответствующей угловой скоростью, а структура измерений может задаваться моделями-1,-2, описанными в предыдущем разделе. Сначала сформулируем задачу гарантирующего оценивания для трехмерной модели-1, а затем по аналогии — задачу для скалярной модели.

С помощью формального обозначения, использующего дельта-функцию Дирака, $l(\Phi)$ можно записать в более компактном виде:

$$l(\Phi) = \int_{y,s} \Phi^T(y, s) z(y, s) dy ds,$$

$$\Phi(y, s) = \Phi_0(y, s) + \sum_{k=1}^K \Phi^T(k) \delta(y - y(k), s - s(k)).$$

Обозначим множество всех функций $\Phi(\cdot)$ вышеуказанного вида через \mathcal{F} .

Будем рассматривать проблему гарантирующего оценивания скалярного параметра $a^T q$, которая состоит в определении оценщика Φ , доставляющего минимум гарантированной ошибке оценки [11]

$$(18) \quad I(\Phi) \rightarrow \inf_{\Phi \in \mathcal{F}},$$

где функционал $I(\Phi)$ представляет собой точную верхнюю грань ошибки $|l(\Phi) - a^T q|$:

$$(19) \quad I(\Phi) = \sup_{(q, \alpha, \beta, \varepsilon, \delta \nu') \in \mathcal{B}'} |l(\Phi) - a^T q|.$$

Точная верхняя грань ищется среди множества \mathcal{B}' — множества всех допустимых значений неизвестных параметров $(q, \alpha, \beta, \varepsilon, \delta \nu')$:

$$(20) \quad q \in \mathbf{R}^{12}, \quad |\alpha_j| \leq \alpha_{\max}, \quad |\beta_j| \leq \beta_{\max}, \quad |\delta \nu'_j| \leq \nu'_{\max}, \quad j = 1, 2, 3, \quad |\varepsilon| \leq \varepsilon_{\max}.$$

Решение проблемы (18)–(20) позволит определить оптимальный план проведения эксперимента. На практике возникают задачи оценки каждой компоненты вектора q , для этих целей необходимо решить 12 задач вида (20), в каждой из которых от нуля отлична только одна компонента a .

Для задачи гарантирующего оценивания может быть сформулирована эквивалентная проблема l_1 -аппроксимации, которая, в свою очередь, может быть численно решена.

Утверждение 1. Функция $\Phi(y, s) \in \mathcal{F}$ является решением проблемы гарантирующего оценивания (18)–(20) для модели-1 (12)–(13) тогда и только тогда, когда она является решением оптимизационной задачи

$$(21) \quad \int \left(\nu'_{\max} \|\Phi\|_1 + \alpha_{\max} \|C_\alpha \Phi\|_1 + \beta_{\max} \|C_\beta \Phi\|_1 + \varepsilon_{\max} |y^T D_{\text{init}}^T \Phi| \right) dy ds \rightarrow \inf_{\Phi \in \mathcal{F}}$$

при ограничениях

$$(22) \quad \left(\begin{array}{c} \int v(y, s) \otimes \Phi dy ds \\ \int \Phi dy ds \end{array} \right) = a,$$

где введены следующие обозначения:

$$(23) \quad \begin{aligned} v &= D_{\text{init}}(sy + yy^T u_x), \\ C_\alpha &= (s\hat{y} + y^T u_x \hat{y} - \hat{u}_x y y^T) D_{\text{init}}^T, \\ C_\beta &= (s\hat{y} + y^T u_x \hat{y}) D_{\text{init}}^T. \end{aligned}$$

Здесь символ \otimes соответствует произведению Кронекера, вектор v и матрицы C_α, C_β являются функциями s, y . Доказательство утверждения 1 приведено в Приложении.

Похожим образом получим формулировки проблем гарантирующего оценивания для модели-2, определив одномерный оценитель χ :

$$\chi(y, s) = \chi_0(y, s) + \sum_{k=1}^K \chi(k) \delta(y - y(k), s - s(k)),$$

при помощи которого оценка неизвестного скалярного параметра $a^T q$ ищется в виде

$$l(\chi) = \int_{y,s} \chi(y, s) z_{\text{scal}}(y, s) dy ds.$$

Обозначим множество всех функций $\chi(y, s)$ описанной выше структуры как \mathcal{X} .

Утверждение 2. Функция $\chi(y, s) \in \mathcal{X}$ является решением проблемы гарантирующего оценивания (18)–(20) для скалярной модели-2 (16)–(17) тогда и только тогда, когда она является решением оптимизационной задачи

$$(24) \quad \int (\nu_{\max} \|D_{\text{init}} y\|_1 + \alpha_{\max} \|\hat{u}_x y\|_1 + \varepsilon_{\max}) |\chi(y, s)| dy ds \rightarrow \inf_{\chi \in \mathcal{X}}$$

при ограничениях

$$(25) \quad \int \chi(y, s) \begin{pmatrix} v(y, s) \otimes D_{\text{init}} y \\ D_{\text{init}} y \end{pmatrix} dy ds = a.$$

Доказательство утверждения 2 аналогично доказательству утверждения 1.

4. Дискретные оптимизационные задачи

В вариационной задаче l_1 -аппроксимации (21)–(22) в качестве искомой переменной выступает вектор-функция $\Phi(y, s)$, аргумент которой принимает континуальное множество значений, что затрудняет поиск численного решения. Для схожих проблем гарантирующего оценивания доказано, что оптимальный оценитель Φ отличен от нуля на конечном множестве точек [5, 11]. В данной работе не будем получать аналитическое решение, вместо этого рассмотрим дискретный аналог вышеуказанных проблем и будем решать их численно. Пусть функция $\Phi(y, s)$ отлична от нуля в конечном наборе точек $\{(y(k), s(k))\}_{k=1}^K$, соответствующих различным значениям вектора угловой скорости блока на стенде, и принимает в них значения $\Phi(k)$. Это дискретное множество положений может быть задано с помощью введения сферических

координат для определения вектора y и выбора “сетки значений” для широты и долготы с заданным шагом.

В результате перехода в (21)–(22) от интегралов к конечным суммам получим следующую оптимизационную задачу:

$$(26) \quad \sum_{k=1}^K \left(\nu'_{\max} \|\Phi(k)\|_1 + \alpha_{\max} \|C_{\alpha}(k)\Phi(k)\|_1 + \beta_{\max} \|C_{\beta}(k)\Phi(k)\|_1 + \varepsilon_{\max} |y^T(k)D_{\text{init}}^T\Phi(k)| \right) \rightarrow \inf_{\Phi(1), \dots, \Phi(K)}$$

при ограничениях

$$(27) \quad \begin{pmatrix} \sum_{k=1}^K v(k) \otimes \Phi(k) \\ \sum_{k=1}^K \Phi(k) \end{pmatrix} = a,$$

где векторы $v(k)$ и матрицы $C_{\alpha}(k)$, $C_{\beta}(k)$ зависят от известных аргументов $y(k)$, $s(k)$:

$$\begin{aligned} v(k) &= D_{\text{init}}(s(k)y(k) + y(k)y^T(k)u_x), \\ C_{\alpha}(k) &= (s(k)\hat{y}(k) + y^T(k)u_x\hat{y}(k) - \hat{u}_x y(k)y^T(k)) D_{\text{init}}^T, \\ C_{\beta}(k) &= (s(k)\hat{y}(k) + y^T(k)u_x\hat{y}(k)) D_{\text{init}}^T. \end{aligned}$$

Важно подчеркнуть, что в данной задаче “дискретная сетка” значений $\{(y(k), s(k))\}_{k=1}^K$ считается априори заданной и определению подлежат только значения $\Phi(k)$. Домножив функционал (26) на константу ν'^{-1}_{\max} и введя обозначения

$$(28) \quad \begin{aligned} \frac{\alpha_{\max}}{\nu'_{\max}} C_{\alpha}(k)\Phi(k) &= x_{\alpha}(k), \\ \frac{\beta_{\max}}{\nu'_{\max}} C_{\beta}(k)\Phi(k) &= x_{\beta}(k), \\ \frac{\varepsilon_{\max}}{\nu'_{\max}} y^T(k)D_{\text{init}}^T\Phi(k) &= x_{\varepsilon}(k), \end{aligned}$$

$$x = \left(\Phi^T(1), \dots, \Phi^T(K), x_{\alpha}^T(1), \dots, x_{\alpha}^T(K), x_{\beta}^T(1), \dots, x_{\beta}^T(K), x_{\varepsilon}(1), \dots, x_{\varepsilon}(K) \right)^T \in \mathbf{R}^{10K},$$

запишем задачу (26)–(27) в сокращенной форме, соответствующей классической проблеме l_1 -аппроксимации

$$(29) \quad \|x\|_1 \rightarrow \inf_{x \in \mathbf{R}^{10K}}$$

при линейных ограничениях $A_{eq}x = a_{eq}$.

Матрица и вектор из уравнения ограничений могут быть представлены в блочном виде:

$$A_{eq} = \begin{pmatrix} A_\alpha & I_{3K} & 0_{3K \times 3K} & 0_{K \times K} \\ A_\beta & 0_{3K \times 3K} & I_{3K} & 0_{K \times K} \\ A_\varepsilon & 0_{K \times 3K} & 0_{K \times 3K} & I_K \\ A_\Phi & 0_{12 \times 3K} & 0_{12 \times 3K} & 0_{12 \times K} \end{pmatrix} \in \mathbf{R}^{(7K+12) \times 10K},$$

$$a_{eq} = \begin{pmatrix} 0_{7K \times 1} \\ a \end{pmatrix} \in \mathbf{R}^{7K+12},$$

где в силу введенных обозначений (27), (28)

$$A_\alpha = \frac{\alpha_{\max}}{\nu_{\max}} \begin{pmatrix} C_\alpha(1) & 0_{3 \times 3} & \dots & 0_{3 \times 3} \\ 0_{3 \times 3} & C_\alpha(2) & \dots & 0_{3 \times 3} \\ \dots & \dots & \ddots & \dots \\ 0_{3 \times 3} & \dots & 0_{3 \times 3} & C_\alpha(K) \end{pmatrix},$$

$$A_\beta = \frac{\beta_{\max}}{\nu_{\max}} \begin{pmatrix} C_\beta(1) & 0_{3 \times 3} & \dots & 0_{3 \times 3} \\ 0_{3 \times 3} & C_\beta(2) & \dots & 0_{3 \times 3} \\ \dots & \dots & \ddots & \dots \\ 0_{3 \times 3} & \dots & 0_{3 \times 3} & C_\beta(K) \end{pmatrix},$$

$$A_\varepsilon = \frac{\varepsilon_{\max}}{\nu_{\max}} \begin{pmatrix} y^\top(1)D_{\text{init}}^\top & 0_{1 \times 3} & \dots & 0_{1 \times 3} \\ 0_{1 \times 3} & y^\top(2)D_{\text{init}}^\top & \dots & 0_{1 \times 3} \\ \dots & \dots & \ddots & \dots \\ 0_{1 \times 3} & \dots & 0_{1 \times 3} & y^\top(K)D_{\text{init}}^\top \end{pmatrix},$$

$$A_\Phi = \begin{pmatrix} v_1(1) & v_1(2)I_3 & \dots & v_1(K)I_3 \\ v_2(1)I_3 & v_2(2)I_3 & \dots & v_2(K)I_3 \\ v_3(1)I_3 & v_3(2)I_3 & \dots & v_3(K)I_3 \\ I_3 & I_3 & \dots & I_3 \end{pmatrix}.$$

Аналогичным образом может быть сформулирован дискретный аналог скаляризованной модели (24)–(25), в котором ищется минимум суммы модулей компонент неизвестного вектора при линейных ограничениях-равенствах:

$$(30) \quad \sum_{k=1}^K (\nu_{\max} \|D_{\text{init}} y(k)\|_1 + \alpha_{\max} \|\hat{u}_x y(k)\|_1 + \varepsilon_{\max}) |\chi(k)| \rightarrow \inf_{\chi(1), \dots, \chi(K)}$$

при ограничениях

$$(31) \quad \sum_{k=1}^K \chi(k) \begin{pmatrix} v(k) \otimes D_{\text{init}} y(k) \\ D_{\text{init}} y(k) \end{pmatrix} = a.$$

В матричном виде эта задача записывается следующим образом:

$$(32) \quad \|x_\chi\|_1 \rightarrow \inf_{x_\chi \in \mathbf{R}^K}$$

при ограничениях $A_\chi x_\chi = a$.

Здесь введены обозначения

$$x_{\chi k} = \rho_k \chi(k), \quad \rho_k = \nu_{\max} \|D_{\text{init}} y(k)\|_1 + \alpha_{\max} \|\hat{u}_x y(k)\|_1 + \varepsilon_{\max}, \quad k = 1, \dots, K;$$

$$A_\chi = \begin{pmatrix} \rho_1^{-1} v_1(1) D_{\text{init}} y(1) & \rho_2^{-1} v_1(2) D_{\text{init}} y(2) & \dots & \rho_K^{-1} v_1(K) D_{\text{init}} y(K) \\ \rho_1^{-1} v_2(1) D_{\text{init}} y(1) & \rho_2^{-1} v_2(2) D_{\text{init}} y(2) & \dots & \rho_K^{-1} v_2(K) D_{\text{init}} y(K) \\ \rho_1^{-1} v_3(1) D_{\text{init}} y(1) & \rho_2^{-1} v_3(2) D_{\text{init}} y(2) & \dots & \rho_K^{-1} v_3(K) D_{\text{init}} y(K) \\ \rho_1^{-1} D_{\text{init}} y(1) & \rho_2^{-1} D_{\text{init}} y(2) & \dots & \rho_K^{-1} D_{\text{init}} y(K) \end{pmatrix} \in \mathbf{R}^{12 \times K}.$$

Выпуклые оптимизационные задачи вышеуказанного вида могут быть решены при помощи различных численных методов, таких как метод внутренней точки [13], ADMM [14], метод вариационно-взвешенных квадратических приближений [5]. В отличие от задачи для трехмерной модели (29) проблемы (32) характеризуются меньшей размерностью неизвестного вектора и матриц ограничений (в 7–10 раз меньше переменных и ограничений), что делает их более подходящими для решения в случае больших K .

Таким образом, решение проблем l_1 -аппроксимации (29), или (32) позволяет найти оптимальные значения оценщиков $\Phi(k), \chi(k)$, которым соответствуют целевые значения угловой скорости $s(k)y(k)$ вращения блока гироскопов на стенде. Как правило, ненулевым значениям $\Phi(k)$ или $\chi(k)$ соответствует небольшое количество положений угловой скорости вращения (что является распространенным свойством решений гарантирующего оценивания, см. обоснование, например, в книге [5]), обозначим это подмножество индексов через \mathcal{K} , $\mathcal{K} \subset \{1, \dots, K\}$.

Алгоритм гарантирующего оценивания для погрешностей блока ДУС представляет собой серию шагов, на каждом из которых для каждого $k \in \mathcal{K}$:

- 1) производится вращение блока гироскопов с угловой скоростью $s(k)y(k)$ и формируется массив показаний гироскопов $\zeta(t) \in \mathbf{R}^3$;
- 2) проводится осреднение сигнала $\zeta(t)$ на отрезке времени T , в течение которого происходило вращение с фиксированной скоростью: $\tilde{\zeta} = \left(\sum_{t=0}^T \zeta(t) \right) / (T + 1)$;
- 3) в соответствии с (13)–(17) формируются измерения для линейных моделей оценивания $z(y(k), s(k))$, $z_{\text{scal}}(y(k), s(k))$.

После чего строится оценка неизвестного параметра $a^T q$ в виде

$$\sum_{k \in \mathcal{K}} \Phi^T(k) z(y(k), s(k)) \quad \text{либо} \quad \sum_{k \in \mathcal{K}} \chi(k) z_{\text{scal}}(y(k), s(k)).$$

5. Численные примеры

Рассмотрим несколько численных примеров, иллюстрирующих применение предложенных методов гарантирующего оценивания. Практическая реализация алгоритма состоит из нескольких этапов: решения проблем гарантирующего оценивания; моделирования сигнала — измерений $\zeta(t)$, при заданных “истинных” значениях погрешностей и неизвестных параметров Γ, ν_0 ; и построения оценок Γ, ν_0 , которые затем сравниваются с “истинными” значениями. Соответствующие программы реализованы на языке Python, а для решения проблем l_1 -аппроксимации (29), (32) используются готовые процедуры из пакета `cvxpy`¹.

Характерные значения погрешностей и параметров модели выбирались согласно табл. 1. Границы допустимых значений величины угловой скорости определены формулами (9), (10): $s_{\min} = 1,25^\circ/\text{с}$, $s_{\max} = 3,28^\circ/\text{с}$. В модельном примере будем считать, что модуль угловой скорости принимает значение $s = 0$ (отсутствие вращения) и два значения из отрезка $[s_{\min}; s_{\max}]$: $s_1 = 1,5^\circ/\text{с}$ и $s_2 = 2^\circ/\text{с}$, векторы $y(k)$ расположены на единичной сфере равномерно, а также $D_{\text{init}} = I_3$. Опишем полученные решения для каждой из групп неизвестных параметров Γ, ν_0 .

Для полной и скаляризованной моделей оптимальные оценщики для диагональных компонент Γ_{ii} имеют вид

$$\begin{aligned}\Phi_{ii}(y, s) &= \chi_0 e_i \delta(y - e_i, s - s_2) - \chi_0 e_i \delta(y + e_i, s - s_2), \\ \chi_{ii}(y, s) &= \chi_0 \delta(y - e_i, s - s_2) + \chi_0 \delta(y + e_i, s - s_2),\end{aligned}$$

где e_i — единичный вектор с i -й компонентой, равной 1, χ_0 — некая величина, численно определяемая в решении оптимизационной задачи.

Иными словами, чтобы оценить, к примеру, компоненту Γ_{11} — масштабный коэффициент ошибки первого гироскопа — необходимо провести две серии измерений, вращая блок вдоль оси чувствительности этого гироскопа с максимальной угловой скоростью $s = s_2$ сначала в одну сторону ($y(1) = (1, 0, 0)^T$), затем в другую ($y(2) = (-1, 0, 0)^T$).

Для внедиагональных элементов $\Gamma_{12} = \Gamma_{21}$ оптимальные оценщики описываются формулами:

$$\begin{aligned}\Phi_{12}(y, s) &= \begin{pmatrix} \Phi_1 \\ \Phi_1 \\ 0 \end{pmatrix} \delta(y - e(\pi/4), s - s_2) + \begin{pmatrix} \Phi_2 \\ -\Phi_1 \\ 0 \end{pmatrix} \delta(y - e(3\pi/4), s - s_2) + \\ &+ \begin{pmatrix} -\Phi_2 \\ -\Phi_2 \\ 0 \end{pmatrix} \delta(y - e(5\pi/4), s - s_2) + \begin{pmatrix} -\Phi_1 \\ \Phi_2 \\ 0 \end{pmatrix} \delta(y - e(7\pi/4), s - s_2), \\ \chi_{12}(y, s) &= \chi_0 \delta(y - e(\pi/4), s - s_2) + \chi_0 \delta(y - e(3\pi/4), s - s_2) + \\ &+ \chi_0 \delta(y - e(5\pi/4), s - s_2) + \chi_0 \delta(y - e(7\pi/4), s - s_2),\end{aligned}$$

¹ Пакет численного решения задач выпуклой оптимизации
https://web.stanford.edu/~boyd/papers/pdf/cvxpy_ewriting.pdf

Таблица 2. Гарантированные ошибки оценки

Модель	Переменная		
	Γ_{ii}	$\Gamma_{ij}, i \neq j$	ν_{0i}
Порядок величин	$[0,7; 1,3] \cdot 10^{-3}$	$[4; 6] \cdot 10^{-3}$	$[2; 3] \cdot 10^{-7}$
Требуемая точность	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	$5 \cdot 10^{-8}$
Модель-1	$2,01 \cdot 10^{-4}$	$5,67 \cdot 10^{-4}$	$7,01 \cdot 10^{-6}$
Модель-2	$2,23 \cdot 10^{-6}$	$3,45 \cdot 10^{-6}$	$5,16 \cdot 10^{-8}$

где $e(\theta) = (\cos \theta, \sin \theta, 0)^T$ — единичные векторы, соответствующие повороту на угол θ в плоскости $e_1 e_2$, Φ_1, Φ_2 — величины, численно определяемые в решении оптимизационной задачи, а χ_0 — то же значение, что и в решении для оценки диагональных элементов. Аналогично, с точностью до перестановки индексов в соответствующих векторах, определяются оценщики для компонент $\Gamma_{13} = \Gamma_{31}$ и $\Gamma_{23} = \Gamma_{32}$. Таким образом, оптимальный эксперимент для оценки углов перекося между осями чувствительности гироскопов состоит в проведении четырех серий измерений, при каждой из которых вращение происходит вдоль биссектрисы угла между координатными осями с максимальной по величине допустимой угловой скоростью.

Оптимальные оценщики для сдвигов нулей ν_{0i} имеют следующую структуру:

$$\begin{aligned} \Phi_{\nu i}(y, s) &= \phi_1 e_i \delta(y - e_i, s - s_2) + \phi_2 e_i \delta(y + e_i, s - s_2) + \\ &\quad + \phi_3 e_i \delta(y - e_i, s - s_1) + \phi_4 e_i \delta(y + e_i, s - s_1), \\ \chi_{\nu i}(y, s) &= \chi_1 \delta(y - e_i, s - s_2) - \chi_2 \delta(y + e_i, s - s_2) + \chi_3 \delta(y - e_i, s - s_1) - \\ &\quad - \chi_4 \delta(y + e_i, s - s_1), \end{aligned}$$

где значения $\phi_i > 0$, $\chi_i > 0$ численно определяются при решении задачи и близки к величине $1/4$.

Как видим, полученные при помощи гарантирующего оценивания направления вращений и весовые коэффициенты имеют простую геометрическую структуру: оптимальные режимы движения предполагают вращение с максимальной допустимой угловой скоростью, по направлению совпадающей (с точностью до погрешности начальной выставки) либо с осями чувствительности гироскопов, либо с биссектрисами углов между этими осями. Похожий результат был получен при применении метода гарантирующего оценивания в задаче калибровки блока акселерометров [11]. Однако в отличие от указанной задачи в данном случае математическая модель измерений существенно сложнее, зависит от большего количества параметров, поэтому аналитически обосновать оптимальную структуру оценщика сложно и для поиска решений применяются численные методы.

Основной характеристикой качества получаемых решений является гарантированная ошибка оценки, не зависящая от конкретных реализаций погрешностей и измерений. В табл. 2 приведены соответствующие оптимальным оценщикам гарантированные ошибки оценки для компонент Γ, ν_0 .

Таблица 3. Средние погрешности оценки

Модель	Переменная		
	Γ_{ii}	$\Gamma_{ij}, i \neq j$	ν_{0i}
Порядок величин	$[0,7; 1,3] \cdot 10^{-3}$	$[4; 6] \cdot 10^{-3}$	$[2; 3] \cdot 10^{-7}$
Модель-1	$1,25 \cdot 10^{-6}$ (0,13%)	$1,33 \cdot 10^{-6}$ (0,02%)	$4,88 \cdot 10^{-8}$ (22%)
Модель-2	$1,20 \cdot 10^{-6}$ (0,13%)	$1,24 \cdot 10^{-6}$ (0,03%)	$2,89 \cdot 10^{-8}$ (14%)

При заданных в табл. 1 параметрах моделей требуемая точность гарантирующего оценивания достигается для всех компонент матрицы Γ при использовании модели-2, применение модели-1 может приводить к ошибкам порядка 10% от величины оцениваемого параметра, а для компонент ν_{0i} необходимая точность достигается при использовании модели-2.

Приведем результаты оценки компонент Γ , ν_0 , полученные в рамках серии численных экспериментов согласно изложенной выше процедуре при конкретных реализациях систематических погрешностей и шумов в измерениях. “Истинные” значения этих параметров и погрешностей α , β , $\varepsilon(t)$, $\delta\nu(t)$ генерировались с помощью датчика случайных чисел. Для каждого параметра Γ_{ij} , ν_{0i} процедура моделирования и оценки повторялась 20 раз, что позволило оценить типичные отклонения оценки от точного значения. В табл. 3 представим средние абсолютные ($|\Gamma_{ij}^0 - \Gamma_{ij}|$) и относительные ($(|\Gamma_{ij}^0 - \Gamma_{ij}|/|\Gamma_{ij}|$ — указаны в скобках) отклонения оценок, полученных при помощи моделей-1 и -2 от истинных значений. В отличие от гарантированных ошибок эти отклонения не задают верхнюю границу ошибки, но зато характеризуют в совокупности и точность моделей, и точность решения задач оценивания. Это важно, потому что входная информация $\zeta(t)$ строится при помощи нелинейной модели исходного сигнала (3), а оптимальное решение ищется путем применения метода гарантирующего оценивания к линеаризованным моделям 1–2. Тем самым данный эксперимент позволяет охарактеризовать влияние на результат оценивания как погрешностей и шумов в измерениях, так и погрешностей, возникших при переходе к более простым линейным моделям.

Серия численных примеров показывает, что формируемые на основании измерительной информации согласно предложенному алгоритму оценки неизвестных параметров оказываются близки к их “истинным” значениям и отклонения лежат в пределах гарантированных ошибок оценки.

Для иллюстрации того, как точность решения зависит от выбора угловой скорости вращения s_2 , рассмотрим эксперимент, при котором для одних и тех же заранее выбранных значений Γ , ν_0 ищется оценка неизвестных параметров при помощи трех вышеописанных моделей при шестнадцати различных значениях s_2 . Относительная точность получаемых оценок показана на графиках ниже: на рис. 1 показаны ошибки определения Γ_{11} (для Γ_{12} результаты похожие), а на рис. 2 — ошибки определения ν_{01} .

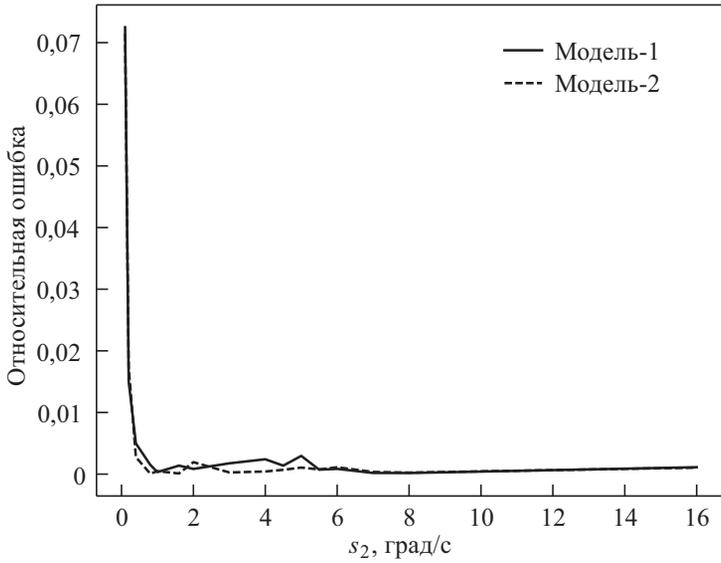


Рис. 1

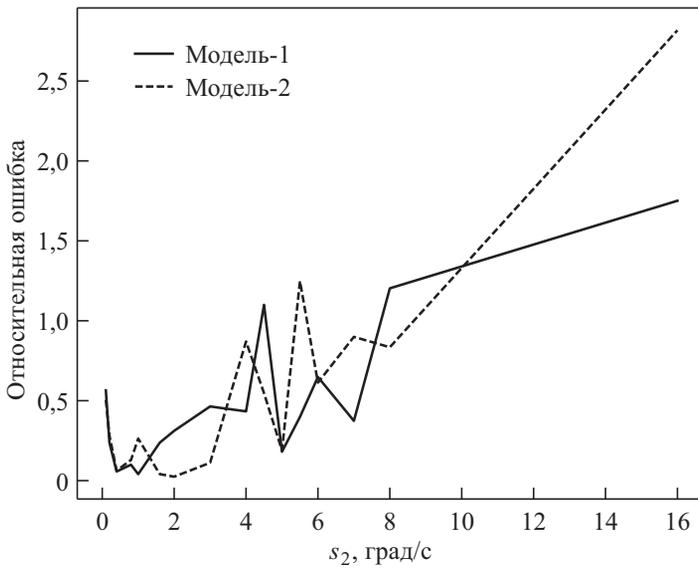


Рис. 2

График показывает, что критическое падение точности наблюдается при минимальных значениях угловой скорости вращения — меньше порога s_{\min} , полученного в разделе 2.2. При оценке матрицы Γ используемые подходы оказываются нечувствительны к увеличению скорости вращений s_2 . Однако при оценивании ν_0 точность ухудшается при росте s_2 , особенно при превышении порога $s_2 > s_{\max} = 3,28 \text{ }^\circ/\text{с} = 0,057 \text{ 1/с}$.

6. Заключение

В статье разработан подход к калибровке блока гироскопов, сочетающий в себе три основные идеи: вывод линейных моделей измерения для осредненных сигналов, применение метода скаляризации, использование гарантирующего подхода оценивания. Проблемы гарантирующего оценивания сведены к дискретным задачам l_1 -аппроксимации, которые решаются с помощью численных алгоритмов. Важным преимуществом предложенной методики является возможность получить оптимальный план эксперимента как результат решения задачи оценивания. Применение гарантирующего оценивания приводит к решениям простой структуры: из большого множества допустимых направлений и величин угловой скорости вращения оптимальным является сочетание двух-четырёх режимов вращения. Представленная в статье методика и соответствующие программные реализации с минимальными изменениями могут быть распространены на более сложные системы, к примеру предполагающие ограниченное количество направлений вращения, учет влияния зависящих от температуры погрешностей датчиков.

Предложенный подход или его модификации могут найти свое применение не только в инерциальной навигации, но и в иных прикладных задачах, требующих определения оптимального набора измерений для оценки неизвестных параметров либо формирования плана эксперимента из множества допустимых сценариев.

ПРИЛОЖЕНИЕ

Доказательство леммы 1.

Так как на исследуемом отрезке времени T вращение происходит вокруг фиксированного направления, можно считать, что при осреднении матрица вращения D_{cir} декомпозируется:

$$(II.1) \quad \bar{D}_{\text{cir}} = \bar{D}_{\text{cir1}} + \bar{D}_{\text{cir2}}, \quad \bar{D}_{\text{cir1}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \bar{D}_{\text{cir2}} = \begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

где c_i — результат осреднения во времени функций $\sin \psi(t)$, $\cos \psi(t)$.

С учетом формул (6), (7), (II.1) вектор u_z при осреднении представляется в виде двух слагаемых, одно из которых пропорционально w , а второе ортогонально $D(0)w$:

$$(II.2) \quad \begin{aligned} \bar{u}_z &= \bar{D}u_x = D' \bar{D}_{\text{cir1}} D_{\text{fix}} u_x + D' \bar{D}_{\text{cir2}} D_{\text{fix}} u_x = \\ &= D' \begin{pmatrix} 0_{1 \times 3} \\ 0_{1 \times 3} \\ w^T \end{pmatrix} u_x + u^\perp = d'_3 w^T u_x + u^\perp = D_{\text{init}} (I_3 + \hat{\beta}) w w^T u_x + u^\perp, \end{aligned}$$

где введено обозначение $u^\perp = D' \bar{D}_{\text{cir2}} D_{\text{fix}} u_x$.

Тот факт, что u^\perp ортогонален направлению $D(0)w = D_{\text{init}}(I_3 + \hat{\beta})w$, можно установить, воспользовавшись формулами (2), (6): $D(0)w = D'D_{\text{cir}}(0)D_{\text{fix}}w = D'(0, 0, 1)^\text{T}$, и в явном виде вычислив скалярное произведение:

$$\begin{aligned} w^\text{T}D(0)^\text{T}u^\perp &= (0, 0, 1)D'^\text{T}D' \begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} D_{\text{fix}}u_x = \\ &= (0, 0, 1)^\text{T} \begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} D_{\text{fix}}u_x = 0. \end{aligned}$$

На примере компоненты c_1 поясним идею, как можно оценить сверху величину, получающуюся при осреднении функции $\cos \psi(t)$. Рассмотрим непрерывный случай осреднения

$$c_1 = \frac{1}{T} \int_0^T \cos \psi(t) dt.$$

Динамика угла ψ описывается дифференциальным уравнением и ограничениями на функции в правой его части:

$$\frac{d\psi(t)}{dt} = s + \varepsilon(t), \quad \psi(0) = \psi_0, \quad |\varepsilon(t)| \leq \varepsilon_{\text{max}}, \quad s + \varepsilon(t) > 0.$$

Произведем замену переменных $t = t(\psi)$, $\varepsilon(\psi) = \varepsilon(t(\psi))$, $|\varepsilon(\psi)| \leq \varepsilon_{\text{max}}$ в интеграле:

$$\int_0^T \cos \psi(t) dt = \int_{\psi_0}^{\psi(T)} \frac{\cos \psi}{s + \varepsilon(\psi)} d\psi.$$

Данный интеграл можно представить в виде суммы интегралов на отрезках знакопостоянства (полупериодах) функции $\cos \psi$ и двух интегралов, соответствующих отрезкам времени в начале и в конце отрезка $[\psi_0, \psi(T)]$. Например, при $\psi_0 < \pi/2$ этот отрезок представляется в следующем виде:

$$[\psi_0, \psi(T)] = [\psi_0, \pi/2] \cup [\pi/2, 3\pi/2] \cup [3\pi/2, 5\pi/2] \cup \dots \cup [\pi/2 + 2\pi n_{\text{cir}}, \psi(T)],$$

где n_{cir} — количество полных оборотов системы вокруг оси вращения, а длина последнего отрезка не превосходит π , т.е. $\pi/2 + 2\pi n_{\text{cir}} \leq \psi(T) \leq 3\pi/2 + 2\pi n_{\text{cir}}$.

Подынтегральная функция на каждом из таких отрезков имеет постоянный знак, а максимальное значение подынтегрального выражения, и, следовательно, интеграла, достигается при $\varepsilon(\psi) = -\text{sign}(\cos \psi) \varepsilon_{\text{max}}$:

$$\int \frac{\cos \psi}{s + \varepsilon(\psi)} d\psi \leq \int \frac{\cos \psi}{\min_{|\varepsilon| \leq \varepsilon_{\text{max}}} (s + \varepsilon)} d\psi = \int \frac{\cos \psi}{s - \text{sign}(\cos \psi) \varepsilon_{\text{max}}} d\psi,$$

что дает возможность оценить каждый интеграл сверху и снизу:

$$\left| \int_{\psi_0}^{\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right| \leq \frac{2}{s - \epsilon_{\max}}, \quad \left| \int_{\pi/2+2\pi n_{\text{cir}}}^{\psi(T)} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right| \leq \frac{2}{s - \epsilon_{\max}},$$

$$\frac{-2}{s - \epsilon_{\max}} \leq \int_{\pi/2}^{3\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \leq \frac{-2}{s + \epsilon_{\max}},$$

$$\frac{2}{s + \epsilon_{\max}} \leq \int_{3\pi/2}^{5\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \leq \frac{2}{s - \epsilon_{\max}}.$$

Следовательно, абсолютное значение интеграла на всем отрезке осреднения оценивается сверху:

$$\left| \int_{\psi_0}^{\psi(T)} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right| \leq \left| \int_{\psi_0}^{\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right| + \left| \int_{\pi/2+2\pi n_{\text{cir}}}^{\psi(T)} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right| +$$

$$+ \left| \sum_{j=1}^{n_{\text{cir}}} \left(\int_{\pi/2}^{3\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi + \int_{3\pi/2}^{5\pi/2} \frac{\cos \psi}{s + \epsilon(\psi)} d\psi \right) \right| \leq$$

$$\leq \frac{4}{s - \epsilon_{\max}} + \left| \sum_{j=1}^{n_{\text{cir}}} \frac{-2}{s + \epsilon_{\max}} + \frac{2}{s - \epsilon_{\max}} \right| \leq \frac{4}{s - \epsilon_{\max}} + \frac{n_{\text{cir}} 4\epsilon_{\max}}{(s + \epsilon_{\max})(s - \epsilon_{\max})}.$$

Угловая скорость вращения и количество полных оборотов системы связаны равенством $sT = 2\pi n_{\text{cir}} + \Delta\psi$ для некоторого $\Delta\psi \leq 2\pi$, что дает оценку величины c_1 сверху:

$$|c_1| = \left| \frac{1}{T} \int_0^T \cos \psi(t) dt \right| \leq \frac{4}{T(s - \epsilon_{\max})} + \frac{n_{\text{cir}} 4\epsilon_{\max}}{T(s^2 - \epsilon_{\max}^2)} =$$

$$= \frac{4}{T(s - \epsilon_{\max})} + \frac{(sT - \Delta\psi) 4\epsilon_{\max}}{2\pi sT s(1 - \epsilon_{\max}^2/s^2)} =$$

$$= \frac{4}{T(s - \epsilon_{\max})} + \frac{2(1 - \Delta\psi/(sT)) \epsilon_{\max}}{\pi(1 - \epsilon_{\max}^2/s^2) s}.$$

Таким образом, получим

$$|c_1| \leq \frac{4}{T(s - \epsilon_{\max})} + C \frac{\epsilon_{\max}}{s},$$

где параметр $C = \frac{2}{\pi(1-\varepsilon_{\max}^2/s^2)}$ является оценкой сверху для дроби $\frac{2(1-\Delta\psi/(sT))}{\pi(1-\varepsilon_{\max}^2/s^2)}$.

Доказательство утверждения 1. Преобразуем подынтегральное выражение функционала в проблеме (18), подставив в него формулы (12),(13) и введя обозначение $v = v(s, y) = D_{\text{init}}(sy + yy^T u_x)$:

$$\begin{aligned} \Phi^T z &= \Phi^T (\Gamma D_{\text{init}}(sy + yy^T u_x) + \nu_0 + r + \delta\nu') = \Phi^T \Gamma v + \Phi^T \nu_0 + \Phi^T \delta\nu' + \\ &+ \Phi^T D_{\text{init}} (-s(\hat{y}\alpha + \hat{y}\beta) + \varepsilon y + y^T u_x (\hat{\alpha} + \hat{\beta})y + yy^T \hat{u}_x \alpha) = (v \otimes \Phi)^T \gamma + \\ &+ \Phi^T \nu_0 + \Phi^T \delta\nu' + \Phi^T D_{\text{init}} (-s(\hat{y}\alpha + \hat{y}\beta) + \varepsilon y - y^T u_x (\hat{y}\alpha + \hat{y}\beta) + yy^T \hat{u}_x \alpha). \end{aligned}$$

Отсюда получим:

$$(П.3) \quad \begin{aligned} \Phi^T z &= (v \otimes \Phi)^T \gamma + \Phi^T \nu_0 + \Phi^T \delta\nu' + \varepsilon \Phi^T D_{\text{init}} y + \\ &+ \Phi^T D_{\text{init}} (-s\hat{y} - y^T u_x \hat{y} + yy^T \hat{u}_x) \alpha + \Phi^T D_{\text{init}} (-s\hat{y} - y^T u_x \hat{y}) \beta. \end{aligned}$$

Здесь воспользовались следующими свойствами матричных операций:

$$\Phi^T \Gamma v = (\Phi^T \otimes v^T) \gamma = (v \otimes \Phi)^T \gamma, \quad \hat{\alpha} y = -\hat{y} \alpha,$$

а также возможностью переноса скалярного произведения $y^T u_x$ в другую часть соответствующей группы множителей: $\hat{\alpha} y y^T u_x = -y^T u_x \hat{y} \alpha$.

Определим матрицы C'_α, C'_β :

$$\begin{aligned} C'_\alpha &= D_{\text{init}} (-s\hat{y} - y^T u_x \hat{y} - y^T u_x \hat{y}), \\ C'_\beta &= D_{\text{init}} (-s\hat{y} - y^T u_x \hat{y}), \end{aligned}$$

тогда правая часть (П.3) представляется в виде функции, линейно зависящей от переменных $q, \alpha, \beta, \varepsilon, \delta\nu'$:

$$(П.4) \quad \begin{aligned} &\Phi^T(y, s)z(y, s) = \\ &= (v \otimes \Phi)^T \gamma + \Phi^T \nu_0 + \Phi^T \delta\nu' + \Phi^T C'_\alpha \alpha + \Phi^T C'_\beta \beta + \varepsilon \Phi^T D_{\text{init}} y. \end{aligned}$$

Подставим формулу (П.4) в функционал исходной проблемы (19):

$$\begin{aligned} I(\Phi) &= \sup_{(q, \alpha, \beta, \varepsilon, \delta\nu') \in \mathcal{B}'} |l(\Phi) - a^T q| = \\ &= \sup_{(q, \alpha, \beta, \varepsilon, \delta\nu') \in \mathcal{B}'} \left| \int \left((v \otimes \Phi)^T \gamma + \Phi^T \nu_0 + \Phi^T \delta\nu' + \Phi^T C'_\alpha \alpha + \right. \right. \\ &\quad \left. \left. + \Phi^T C'_\beta \beta + \varepsilon \Phi^T D_{\text{init}} y \right) dy ds - a^T q \right|. \end{aligned}$$

Так как $q = \text{col}(\gamma, \nu_0)$, то функция $l(\Phi) - a^T q$ линейно зависит от q , причем множитель при q равен

$$\left(\begin{array}{c} \int v \otimes \Phi dyds \\ \int \Phi dyds \end{array} \right) - a.$$

Поэтому если не выполнено условие (22), то имеем $\sup_{q \in \mathbf{R}^{12}} |l(\Phi) - a^T q| = +\infty$ при фиксированном Φ и произвольных допустимых $\alpha, \beta, \varepsilon, \delta\nu'$. Следовательно,

$$\begin{aligned} & \sup_{(q, \alpha, \beta, \varepsilon, \delta\nu') \in \mathcal{B}'} |l(\Phi) - a^T q| = \\ & = \sup_{(q, \alpha, \beta, \varepsilon, \delta\nu') \in \mathcal{B}'} \left| \int (\Phi^T \delta\tilde{\nu} + \Phi^T C'_\alpha \alpha + \Phi^T C'_\beta \beta + \varepsilon \Phi^T D_{\text{init}} y) dyds \right|, \end{aligned}$$

т.е. необходимо найти максимум модуля линейной функции, каждое слагаемое которой зависит только от одной переменной, отсутствующей в прочих слагаемых, поэтому максимум можно искать независимо по каждой из переменных. При фиксированном Φ этот максимум ищется в явном виде:

$$\begin{aligned} \sup_{\alpha: |\alpha_i| \leq \alpha_{\max}} \int \Phi^T C'_\alpha \alpha dyds &= \sup_{\alpha: |\alpha_i| \leq \alpha_{\max}} \int \left(\sum_{i=1}^3 (C'^T_\alpha \Phi)_i \alpha_i \right) dyds = \\ &= \sum_{i=1}^3 \sup_{\alpha_i: |\alpha_i| \leq \alpha_{\max}} \int (C'^T_\alpha \Phi)_i \alpha_i dyds = \\ &= \int \left(\sum_{i=1}^3 \alpha_{\max} \text{sign}((C'^T_\alpha \Phi)_i) (C'^T_\alpha \Phi)_i \right) dyds = \int \|C'_\alpha \Phi\|_1 dyds. \end{aligned}$$

Аналогичная цепочка выкладок справедлива и для остальных слагаемых в функционале (18). Таким образом, явное вычисление точной верхней грани исходного функционала приводит к оптимизационной задаче (21)–(22).

СПИСОК ЛИТЕРАТУРЫ

1. *Лидов М.Л.* К априорным оценкам точности определения параметров по методу наименьших квадратов // Космические исследования. 1964. № 5. С. 713–715.
2. *Красовский Н.Н.* К теории управляемости и наблюдаемости линейных динамических систем // Прикладная математика и механика. 1964. Т. 28. № 1. С. 3–14.
3. *Лидов М.Л.* Минимаксные методы оценивания. М.: Препринт № 71. Ин-т прикл. мат. им. М.В. Келдыша РАН. 2010.
4. *Matasov A.I.* Estimators for Uncertain Dynamic Systems. Dordrecht-Boston-London: Kluwer Academic Publishers, 1999.

5. *Акимов П.А., Деревянкин А.В., Матасов А.И.* Гарантирующий подход и l_1 -аппроксимация в задачах оценивания параметров БИНС при стендовых испытаниях. М.: Изд-во МГУ, 2012.
6. *Козлов А.В., Парусников Н.А., Вавилова Н.Б., Тарыгин И.Е., Голован А.А.* Динамическая стендовая калибровка бескарданных инерциальных навигационных систем в сборе // Известия ЮФУ. Технические науки. 2018. № 1. С. 241–257.
7. *Vavilova N.B., Vasineva I.A., Golovan A.A., Kozlov A.V., Papusha I.A., Parusnikov N.A.* The Calibration Problem in Inertial Navigation // J. Math. Sci. 2021. V. 253. No. 6. P. 818–836.
8. *Веремеенко К.К., Галай И.А.* Разработка алгоритма калибровки инерциальной навигационной системы на двухосном испытательном стенде // Труды МАИ. 2013. № 6. С. 1–14.
9. *Тарыгин И.Е.* Методика калибровки тепловой модели блока чувствительных элементов, состоящего из трех датчиков угловой скорости // Гироскопия и навигация. 2019. № 4. С. 88–102.
10. *Kailath T., Sayed A.H., Hassibi B.* Linear Estimation. New Jersey, Prentice Hall, 2000.
11. *Матасов А.И.* Вариационные задачи для калибровки блока ньютонометров // АиТ. 2019. № 12. С. 59–79.
Matasov A.I. Variational problems for calibrating an accelerometer unit // Autom. Remote Control. 2019. V. 80. No. 12. P. 2135–2151.
12. *Голован А.А., Парусников Н.А.* Математические основы навигационных систем. Ч. I. Математические модели инерциальной навигации. М.: МАКС Пресс, 2011.
13. *Boyd S., Vandenberghe L.* Convex Optimization. Cambridge. Cambridge University Press, 2004.
14. *Boyd S., Parikh N., Chu E., Peleato B., Eckstein J.* Distributed optimization and statistical learning via the alternating direction method of multipliers // Foundations and Trends in Machine learning. 2010. V. 3. No. 1. P. 1–122.

Статья представлена к публикации членом редколлегии О.А. Степановым.

Поступила в редакцию 07.11.2022

После доработки 14.03.2023

Принята к публикации 30.03.2023

Управление в технических системах

© 2023 г. Ю.Г. БУЛЫЧЕВ, д-р техн. наук (profbulychev@yandex.ru)
(АО «Концерн Радиоэлектронные технологии», Москва),
А.А. МОЗОЛЬ, канд. техн. наук (amozol@bk.ru)
(АО «ВНИИ «Градиент», Ростов-на-Дону)

ПЕРИОДО-ВРЕМЕННОЙ МЕТОД ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ДЛЯ РЕШЕНИЯ ЛОКАЦИОННЫХ И НАВИГАЦИОННЫХ ЗАДАЧ

Применительно к локационным и навигационным задачам для однопозиционного пассивного наблюдателя развит беспеленговый метод идентификации параметров полиномиальной модели движения объекта с учетом эволюции невязки между периодическим излученным и принятым квазипериодическим сигналом. Рассматривается прохождение сигнала в произвольной физической среде, при этом не требуются знание периода излученного сигнала и традиционное оценивание текущей частоты Доплера, вызывающей невязку. Метод основан на подсчете числа периодов принимаемого сигнала в заданном интервале наблюдения. Рассмотрены вопросы, связанные с анализом возникающей невязки, наблюдаемостью метода и его точностными характеристиками. Даны полезные практические рекомендации и иллюстративный пример.

Ключевые слова: излучающая цель, периодический сигнал, квазипериодический сигнал, однопозиционный пассивный наблюдатель, беспеленговый метод, временная невязка, периодо-временной метод, полиномиальное движение, параметрическая идентификация, наблюдаемость метода, корреляционная матрица ошибок оценивания, адаптация.

DOI: 10.31857/S0005231023070048, EDN: FCTVPA

1. Введение

Методы пассивной локации и навигации излучающей цели на базе однопозиционного пассивного наблюдателя широко отражены в известной литературе (см., например, [1–20]). Среди них весьма популярны доплеровско-временные беспеленговые методы, оперирующие с периодическими сигналами и ориентированные на возможность измерения непрерывного смещения доплеровской частоты принимаемого сигнала в точке наблюдения, обусловленного движением цели (для задач локации) или движением наблюдателя (для задач навигации); в [6] на стр. 169–173 дан исчерпывающий список литературы по данному вопросу, доступной в открытой печати. При этом измерения могут осуществляться на любой характерной частоте из спектра

излучаемого периодического сигнала (например, на центральной) либо модулирующей функции, а также путем сопоставления моментов прихода фронтов последовательных импульсов с учетом известного периода. Указанные методы основаны на идее «синтеза базы», что приводит в конечном итоге к формированию нескольких точек наблюдения на траектории движения и возможности использования известных методов многопозиционной локации и навигации (например, триангуляционного, разностно-дальномерного, трилатерационного и их комбинаций [21, 22]). При этом, как правило, рассматриваются такие траектории, которые на участке наблюдения либо известны (например, орбитальные с известными параметрами движения), либо достаточной для практики точно аппроксимируются моделью прямолинейного равномерного движения (как с известными, так и неизвестными параметрами движения). При этом принципиальным моментом является учет априорной информации о величине скорости цели или наблюдателя, что для практики зачастую является неприемлемым.

В [20] развивается периодо-временной метод (ПВМ), который снимает ограничение, связанное с получением априорной информации о величине скорости, а также рассматривается вопрос параметрической идентификации применительно к модели криволинейного движения, учитывающего возможный маневр цели или наблюдателя. При этом не требуется предварительная текущая оценка доплеровской частоты, что эквивалентно нахождению производной от временной невязки между периодами излучаемого (периодического) и принимаемого (квазипериодического) сигналов. Однако результаты, полученные в [20], распространяются лишь на радиосигналы (распространяющиеся в виде электромагнитной волны со скоростью света) с известным периодом, а также не исследована зависимость возникающей временной невязки от параметров движения цели. Настоящая статья является дальнейшим развитием известного ПВМ в части устранения указанных недостатков применительно к сигналам, распространяющимся в произвольных физических средах. Не снижая общности рассуждений, а также с целью компактного описания предлагаемого метода ограничимся решением задачи пассивной локации излучающей цели (ИЦ) на базе стационарного однопозиционного пассивного наблюдателя (СОПН). Применение полученных результатов к решению задачи навигации не вызовет принципиальных затруднений.

2. Постановка задачи

Пусть движущаяся ИЦ формирует в текущем времени t периодический сигнал $S_0(t)$ (период сигнала $T_S = \text{const}$ может быть неизвестным), распространяющийся в заданной физической среде в виде волны со скоростью v_S (речь может идти о различных волнах, например электромагнитных или акустических). В точке наблюдения, связанной с СОПН, на интервале наблюдения $[0, T]$ принимается квазидетерминированный сигнал $S(t)$ с переменным периодом.

Согласно ПВМ отрезок наблюдения представляется в виде

$$(2.1) \quad [0, T] = \bigcup_{n=1}^N [t_{n-1}, t_n], \quad t_n > t_{n-1}, \quad t_0 = 0, \quad t_N \leq T,$$

где $t_0 = 0$ — фиксированный момент времени, соответствующий началу принимаемого сигнала (например, поступлению первого импульса), t_n — фиксированный момент времени поступления $M_n = \sum_{p=1}^n \Delta M_p$ периодов принимаемого квазипериодического сигнала (ΔM_p — количество периодов, подсчитываемых на отрезке $[t_{p-1}, t_p]$), при этом в момент времени t_n число M_n периодов целиком укладывается в отрезок $[0, t_n]$.

Теоретические и практические вопросы, связанные с подсчетом указанных периодов, решаются с использованием электронных цифровых частотомеров и подробно изложены в известной технической литературе (см., например, [23, стр. 148–161]).

В точке наблюдения (где расположен СОПН) с учетом движения ИЦ сигнал становится квазипериодическим, поскольку возникает временная невязка $\delta(t)$ между периодами излученного и принятого сигналов

$$(2.2) \quad \delta(t) = v_S^{-1} \Delta R(t) = v_S^{-1} [R(t) - R_0], \quad t \in [0, T],$$

где $R(t)$ — текущая дальность до ИЦ, $R_0 = R(0)$ — начальная дальность.

В прямоугольной декартовой системе координат XYZ (в центре которой находится СОПН) движение ИЦ описывается полиномиальной моделью (для упрощения выкладок и наглядности метода вместо обобщенного конечного полинома с произвольными базисными функциями ограничимся степенным полиномом второй степени с начальным условием $\mathbf{r}_0 = \mathbf{r}(0)$, $\|\mathbf{r}_0\| = R_0$)

$$(2.3) \quad \mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}_0 t + 2^{-1} \mathbf{a}_0 t^2, \quad t \in [0, T],$$

где $\mathbf{r}(t) = \mathbf{r} = [x, y, z]^T$ — вектор положения ($\|\mathbf{r}(t)\| = R(t)$),

$\mathbf{v}_0 = [v_{x0}, v_{y0}, v_{z0}]^T$ — вектор начальной скорости ($v_0 = \|\mathbf{v}_0\|$ — величина скорости),

$\mathbf{a}_0 = [a_{x0}, a_{y0}, a_{z0}]^T$ — вектор ускорения ($a_0 = \|\mathbf{a}_0\|$ — величина ускорения), при этом векторы \mathbf{r}_0 , \mathbf{v}_0 и \mathbf{a}_0 априорно неизвестны.

Если в качестве измеряемого параметра принять величину t_n , то можно воспользоваться следующим векторным уравнением наблюдения:

$$(2.4) \quad \mathbf{h} = \mathbf{t} + \boldsymbol{\xi} = \bar{\mathbf{t}} + \boldsymbol{\delta} + \boldsymbol{\xi},$$

где

$$\mathbf{h} = [h_n, n = \overline{1, N}]^T, \quad \mathbf{t} = [t_n, n = \overline{1, N}]^T, \quad \bar{\mathbf{t}} = [\bar{t}_n, n = \overline{1, N}]^T, \\ \boldsymbol{\delta} = [\delta_n, n = \overline{1, N}]^T, \quad \boldsymbol{\xi} = [\xi_n, n = \overline{1, N}]^T, \quad h_n = h(t_n), \quad \xi_n = \xi(t_n).$$

В (2.4) под $\boldsymbol{\xi} = [\xi_n, n = \overline{1, N}]^T$ понимается гауссовская погрешность измерений с нулевым математическим ожиданием и корреляционной матри-

цей \mathbf{K}_ξ , а измеряемый параметр t_n связан с числом подсчитываемых периодов соотношением

$$(2.5) \quad t_n = M_n T_S + \delta_n = \bar{t}_n + \delta_n = \bar{t}_n + v_S^{-1} [R_n - R_0],$$

где $\delta_n = \delta(t_n)$ — неизвестная временная невязка, $\bar{t}_n = M_n T_S$, $R_n = R(t_n)$, $t_0 = 0$.

Формулу (2.5) можно прокомментировать так (см., например, [6, стр. 154]): за время $\bar{t}_n = M_n T_S$ дальность между ИЦ и СОПН изменится на величину $\Delta R_n = R_n - R_0$, что соответствует временной невязке $\delta_n = v_S^{-1} \Delta R_n$ между периодами излученного и принятого сигналов. Если бы цель была стационарной или двигалась по окружности (в центре которой находится СОПН), то приращение дальности отсутствовало бы и $\delta_n = 0$ для всех n . Именно прохождение волной дополнительного участка пути длиной ΔR_n со скоростью v_S является причиной возникновения невязки δ_n .

Напомним, что при известном периоде T_S в качестве измеряемого параметра можно было принять величину $\delta_n = t_n - M_n T_S$ (именно так формировалось уравнение наблюдения в [6, 20]), при неизвестном периоде T_S доступны для измерения только величины t_n и M_n .

Если расстояние между ИЦ и СОПН уменьшается, то $\delta_n < 0$, в противном случае $\delta_n > 0$. Появление временной невязки $\delta_n = \delta(t_n)$ обусловлено эффектом сжатия или растяжения исходного периодического сигнала в точке наблюдения за счет движения ИЦ.

Требуется с учетом (2.1)–(2.5) разработать метод параметрической идентификации ИЦ с криволинейным (полиномиальным) движением на базе периода-временного СОПН, не требующий знания периода T_S излучаемого сигнала и вычисления текущей частоты Доплера. Метод должен включать в себя решение следующих вопросов:

- получение зависимостей, позволяющих оценить характер эволюции периода принимаемого сигнала (обусловленную движением ИЦ), что является принципиальным для данного метода;

- формирование алгоритма идентификации наклонной дальности и ряда характерных параметров движения ИЦ на точных данных (принимая $\xi_n = 0$, $n = \overline{1, N}$);

- определение условий корректного применения метода на точных данных (т.е. определение условий наблюдаемости метода);

- учет случайных погрешностей измерений;

- решение задачи идентификации на избыточных данных (\mathbf{h}) с учетом шумов измерений (задача сглаживания на основе метода наименьших квадратов (МНК)) и получение соотношения для расчета корреляционной матрицы ошибок идентификации;

- проведение вычислительного эксперимента с целью демонстрации возможностей метода.

3. Исследование эволюции периода сигнала

Набег $\delta(t)$ описывается выражением (для случая прямолинейного равномерного движения)

$$(3.1) \quad \delta(t) = v_S^{-1} \left\{ [R_0^2 + 2tR_0v_0 \cos \gamma_0 + t^2v_0^2]^{1/2} - R_0 \right\}, \quad t \geq 0, \quad \delta(0) = 0,$$

где γ_0 — угол между векторами \mathbf{r}_0 и \mathbf{v}_0 .

При $0 < \gamma_0 \leq \pi/2$ функция $\delta(t)$ является неотрицательной, гладкой и строго выпуклой, $\delta^{(1)}(t) = d\delta(t)/dt = 0$ в точке $t = 0$. При $\pi/2 < \gamma_0 < \pi$ функция $\delta(t)$ является гладкой и строго выпуклой, имеет два корня ($t = 0$ и $t = -2R_0 \cos \gamma_0/v_0$), в точке $t = -R_0 \cos \gamma_0/v_0$ достигает минимального значения ($v_S^{-1}R_0(\sin \gamma_0 - 1)$). При $\gamma_0 = 0$ имеем $\delta(t) = (v_0/v_S)t$, т.е. набег является линейной неотрицательной функцией, не зависящей от R_0 . При $\gamma_0 = \pi$ имеем $\delta(t) = -(v_0/v_S)t$ для $0 \leq t \leq R_0/v_0$, т.е. $\delta(t)$ является линейной функцией и достигает своего минимума ($-R_0/v_S$) в точке $t = R_0/v_0$. Поскольку при $\gamma_0 = 0$ и $\gamma_0 = \pi$ набег $\delta(t)$ не зависит от R_0 , то для этих некорректных случаев, связанных с движением ИЦ вдоль линии визирования, определение дальности с учетом эволюции периода сигнала в точке наблюдения невозможно.

Для более детального исследования $\delta(t)$ найдем несколько первых производных по времени (в точке $t = 0$):

$$(3.2) \quad \begin{cases} \delta_0^{(1)} = v_S^{-1}v_R, \\ \delta_0^{(2)} = (v_S R_0)^{-1}v_\tau^2, \\ \delta_0^{(3)} = -3(v_S R_0^2)^{-1}v_\tau^2v_R, \end{cases}$$

где $v_R = R_0 \cos \gamma_0$ и $v_\tau = v_0 \sin \gamma_0$ — соответственно величины радиальной и тангенциальной скорости.

В итоге можно воспользоваться разложением на основе ряда Тейлора

$$(3.3) \quad \begin{aligned} \delta(t) &= v_S^{-1}t \left(v_R + \frac{v_\tau^2 t}{2R_0} - \frac{v_\tau^2 v_R t^2}{2R_0^2} + \dots \right) = \\ &= v_S^{-1}t \left[v_R + \frac{v_\tau^2 t}{2R_0} \left(1 - \frac{v_R t}{R_0} \right) + \dots \right], \end{aligned}$$

из которого следует, что спектральный состав функции $\delta(t)$ существенно зависит от условий наблюдения, и во многих практически важных случаях не удастся пренебречь производными второго и более высоких порядков, особенно для больших интервалов наблюдения и малых дальностей.

Формулы (3.1)–(3.3) весьма полезны при обосновании возможности практической реализации развиваемого ПВМ в каждом конкретном случае с учетом принятых исходных данных.

4. Построение алгоритма параметрической идентификации на точных данных

С учетом (2.3) можно воспользоваться следующей зависимостью

$$(4.1) \quad R^2(t) - R_0^2 = 2t \langle \mathbf{r}_0, \mathbf{v}_0 \rangle + t^2 (v_0^2 + \langle \mathbf{r}_0, \mathbf{a}_0 \rangle) + t^3 \langle \mathbf{v}_0, \mathbf{a}_0 \rangle + 4^{-1} t^4 a_0^2,$$

где $\langle \cdot, \cdot \rangle$ — символ скалярного произведения двух векторов, $\|\cdot\|$ — символ нормы вектора.

Формула (4.1) представляет собой первое базовое соотношение развиваемого ПВМ.

Второе базовое соотношение следует непосредственно из формулы (2.2):

$$(4.2) \quad R^2(t) - R_0^2 = 2v_S R_0 \delta(t) + v_S^2 \delta^2(t).$$

Приравнявая выражения (4.1) и (4.2), после несложных преобразований получаем уравнение

$$(4.3) \quad -2v_S \delta(t) \chi_1 + 2t \chi_2 + t^2 \chi_3 + t^3 \chi_4 + 4^{-1} t^4 \chi_5 = v_S^2 \delta^2(t),$$

где

$$(4.4) \quad \begin{cases} \chi_1 = R_0, \\ \chi_2 = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle, \\ \chi_3 = (v_0^2 + \langle \mathbf{r}_0, \mathbf{a}_0 \rangle), \\ \chi_4 = \langle \mathbf{v}_0, \mathbf{a}_0 \rangle, \\ \chi_5 = a_0^2 \end{cases}$$

— неизвестные коэффициенты, имеющие понятный физический смысл и подлежащие идентификации.

Поскольку величины δ_n неизвестны, то с учетом (2.5) для дискретного времени запишем уравнение относительно неизвестных величин T_S и χ_i , $i = \overline{1, 5}$:

$$(4.5) \quad \begin{aligned} -2v_S (t_n - M_n T_S) \chi_1 + 2t_n \chi_2 + t_n^2 \chi_3 + t_n^3 \chi_4 + 4^{-1} t_n^4 \chi_5 = \\ = v_S^2 [(t_n - M_n T_S)]^2. \end{aligned}$$

После несложных, но громоздких преобразований формулу (4.5) можно представить в виде нового уравнения (относительно коэффициентов A_i)

$$(4.6) \quad \sum_{i=1}^6 B_{in} A_i = D_n,$$

где

$$(4.7) \quad \begin{cases} A_1 = (v_S \chi_1 - \chi_2) v_S^{-2} T_S^{-1}, & A_2 = -\chi_1 v_S^{-1}, \\ A_3 = (v_S^2 - \chi_3) (2v_S^2 T_S)^{-1}, & A_4 = 2^{-1} T_S, \\ A_5 = -\chi_4 (2T_S v_S^2)^{-1}, & A_6 = \chi_5 (8T_S v_S^2)^{-1}, \\ B_{1n} = t_n, & B_{2n} = M_n, & B_{3n} = t_n^2, \\ B_{4n} = M_n^2, & B_{5n} = t_n^3, & B_{6n} = t_n^4, \\ D_n = M_n t_n. \end{cases}$$

Соотношения (4.6) и (4.7) являются основой для идентификации параметров криволинейного движения ИЦ при неизвестном периоде излучаемого сигнала. В (4.6) неизвестными являются коэффициенты A_i , $i = \overline{1, 6}$, которые однозначно связаны с искомыми параметрами движения ИЦ и периодом излучаемого сигнала. Если уравнение (4.6) записать для всех $n = \overline{1, N}$, где $N \geq 6$, то получим систему линейных алгебраических уравнений (СЛАУ) (с прямоугольной матрицей \mathbf{B})

$$(4.8) \quad \mathbf{B}\mathbf{A} = \mathbf{D},$$

где $\mathbf{B} = [B_{in}, n = \overline{1, N}, i = \overline{1, 6}]$, $\mathbf{A} = [a_i, i = \overline{1, 6}]^T$, $\mathbf{D} = [D_n, n = \overline{1, N}]^T$.

Данная СЛАУ позволяет решать задачу оценивания указанных коэффициентов и параметров, а также периода сигнала при избыточных измерениях. При $N > 6$ речь идет о задаче сглаживания на основе МНК с использованием ортогонально-сингулярного разложения [24].

Рассмотрим частный случай, когда ИЦ движется прямолинейно и равномерно, а период сигнала неизвестен. Теперь вместо (4.3) имеем уравнение

$$(4.9) \quad -2v_S \delta(t) \chi_1 + 2t \chi_2 + t^2 \chi_3 = v_S^2 \delta^2(t),$$

где

$$(4.10) \quad \begin{cases} \chi_1 = R_0, \\ \chi_2 = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle, \\ \chi_3 = v_0^2. \end{cases}$$

При этом вместо (4.6) в этом случае имеем

$$(4.11) \quad t_n A_1 + M_n A_2 + t_n^2 A_3 + M_n^2 A_4 = M_n t_n.$$

Если предположить, что период сигнала известен, т.е. известны величины δ_n , то с учетом (4.9) для нахождения параметров прямолинейного равномерного движения ИЦ достаточно решить СЛАУ (относительно χ_i , $i = \overline{1, 3}$)

$$(4.12) \quad -2v_S \delta_n \chi_1 + 2t_n \chi_2 + t_n^2 \chi_3 = v_S^2 \delta_n^2, \quad n = \overline{1, N}.$$

При этом находим дальность R_0 , величину скорости $v_0 = \|\mathbf{v}_0\|$ и угол γ_0 между векторами \mathbf{r}_0 и \mathbf{v}_0 с учетом очевидных соотношений:

$$(4.13) \quad \begin{cases} R_0 = \chi_1, \\ v_0 = \sqrt{\chi_3}, \\ \gamma_0 = \arccos \left[\chi_2 (R_0 v_0)^{-1} \right]. \end{cases}$$

В случае прямолинейного равноускоренного движения ИЦ (когда векторы \mathbf{v}_0 и \mathbf{a}_0 являются коллинеарными) необходимо решить СЛАУ (относительно χ_i , $i = \overline{1, 5}$)

$$(4.14) \quad -2v_S \delta_n \chi_1 + 2t_n \chi_2 + t_n^2 \chi_3 + t_n^3 \chi_4 + 4^{-1} t_n^4 \chi_5 = v_S^2 \delta_n^2.$$

Теперь имеем:

$$(4.15) \quad \begin{cases} \chi_1 = R_0, \\ \chi_2 = R_0 v_0 \cos \gamma_0, \\ \chi_3 = (v_0^2 + R_0 a_0 \cos \gamma_0), \\ \chi_4 = v_0 a_0, \\ \chi_5 = a_0^2. \end{cases}$$

По найденным значениям χ_1, \dots, χ_5 вычисляем следующие параметры движения ИЦ:

$$(4.16) \quad \begin{cases} R_0 = \chi_1, \\ a_0 = \sqrt{\chi_5}, \\ v_0 = \chi_4 a_0^{-1}, \\ \gamma_0 = \arccos \left[\chi_2 (R_0 v_0)^{-1} \right]. \end{cases}$$

Выражения (4.1)–(4.16) составляют математическую основу развиваемого ПВМ.

В следующем разделе проанализируем условия наблюдаемости развиваемого метода, т.е. выявим ситуации, при которых он становится некорректным с вычислительной точки зрения.

5. Анализ наблюдаемости метода

Развиваемый ПВМ можно реализовать на любом наборе узлов из множества $\{t_1, \dots, t_N\}$, что позволяет не только уменьшить объем вычислений, но и в ряде случаев повысить надежность формируемых оценок (особенно при отсутствии достоверной априорной информации о весовых коэффициентах, необходимых для реализации МНК). Для этого введем векторы временных узлов $\mathbf{t}_{[l]} = [t_{[l]p}, p = \overline{1, P_{[l]}}]^T$, где $l = \overline{1, L}$, $t_{[l]p} \in \{t_1, \dots, t_N\}$, $t_{[l]p+1} > t_{[l]p}$.

Здесь L — число наборов, $P_{[l]}$ — количество узлов в l -м наборе, $t_{[l]p}$ — узел с номером $[l]p$ (это натуральное число, принадлежащее множеству $\{1, \dots, N\}$). На основе (4.12) сформируем следующую СЛАУ:

$$(5.1) \quad \mathbf{C}_{[l]}\mathbf{X}_{[l]} = \mathbf{Y}_{[l]},$$

где $\mathbf{Y}_{[l]} = [\delta_{[l]p}^2, p = \overline{1, P_{[l]}}]^T$, $\mathbf{X}_{[l]} = [\chi_{i[l]}, i = \overline{1, 5}]^T$, а матрица $\mathbf{C}_{[l]}$ (размером $P_{[l]} \times 5$) образована строками $v_S^{-2} \left(-2v_S \delta_{[l]p}, 2t_{[l]p}, t_{[l]p}^2, t_{[l]p}^3, 4^{-1}t_{[l]p}^4 \right)$, $p = \overline{1, P_{[l]}}$.

Введение $\mathbf{t}_{[l]}$ позволяет с учетом геометрии наблюдения, характеристик ИЦ и СОПН находить такие наборы узлов, на которых вопрос идентификации решается наиболее качественно (это относится к известной задаче планирования эксперимента [25]).

Не снижая общности рассуждений, ограничимся плоским случаем (принимая $z = 0$) и сигналом с известным периодом, а также зададимся $P_{[l]} = 5$, что соответствует квадратной матрице $\mathbf{C}_{[l]}$. Очевидно, что для корректного применения развиваемого метода, связанного с решением СЛАУ (5.1), необходимо и достаточно выполнение условия $\det \mathbf{C}_{[l]} \neq 0$, что приводит к искомому результату $\mathbf{X}_{[l]} = \mathbf{C}_{[l]}^{-1}\mathbf{Y}_{[l]}$. Для выявления случаев, при которых это условие нарушается, запишем столбцы матрицы $\mathbf{C}_{[l]}$ в виде векторов:

$$\begin{aligned} \mathbf{C}_{[l]1} &= [-2v_S \delta_{[l]p}, p = \overline{1, 5}]^T, & \mathbf{C}_{[l]2} &= [2t_{[l]p}, p = \overline{1, 5}]^T, \\ \mathbf{C}_{[l]3} &= [t_{[l]p}^2, p = \overline{1, 5}]^T, & \mathbf{C}_{[l]4} &= [t_{[l]p}^3, p = \overline{1, 5}]^T, & \mathbf{C}_{[l]5} &= [4^{-1}t_{[l]p}^4, p = \overline{1, 5}]^T. \end{aligned}$$

Несложно заметить, что столбцы $\mathbf{C}_{[l]2}$, $\mathbf{C}_{[l]3}$ и $\mathbf{C}_{[l]4}$ линейно независимы, следовательно, для проверки условия $\det \mathbf{C}_{[l]} \neq 0$ достаточно показать, что столбец $\mathbf{C}_{[l]1}$ не может быть представлен в виде линейной комбинации этих столбцов.

Так как $R_{[l]p} = [x_{[l]p}^2 + y_{[l]p}^2]^{-2}$ (где $R_{[l]p} = R(t_{[l]p})$, $x_{[l]p}^2 = (x_0 + v_{x0}t_{[l]p} + 2^{-1}a_{x0}t_{[l]p}^2)^2$ и $y_{[l]p}^2 = (y_0 + v_{y0}t_{[l]p} + 2^{-1}a_{y0}t_{[l]p}^2)^2$), то нарушение условия $\det \mathbf{C}_{[l]} \neq 0$ эквивалентно тому, что векторы $\boldsymbol{\mu}_{[l]} = [x_{[l]p}^2, p = \overline{1, 5}]^T$ и $\boldsymbol{\eta}_{[l]} = [y_{[l]p}^2, p = \overline{1, 5}]^T$ не связаны условием коллинеарности: $\boldsymbol{\mu}_{[l]} = k\boldsymbol{\eta}_{[l]}$, где k — коэффициент пропорциональности. В противном случае имеем

$$(5.2) \quad R_{[l]p} = [x_{[l]p}^2 + y_{[l]p}^2]^{-2} = [k^2 y_{[l]p}^2 + y_{[l]p}^2]^{-2} = q |y_{[l]p}|,$$

$$(5.3) \quad -2v_S \delta_{[l]p} = -2 [R_{[l]p} - R_0] = -2 [q |y_{[l]p}| - R_0],$$

где $q = (k^2 + 1)^{-2}$.

Из (5.2) и (5.3) следует, что координаты вектора $\mathbf{C}_{[l]1}$ можно представить линейной комбинацией из координат векторов $\mathbf{C}_{[l]2}$, $\mathbf{C}_{[l]3}$ и $\mathbf{C}_{[l]4}$. Физический

смысл условия $\mu_{[l]} = k\eta_{[l]}$ (условие вычислительной некорректности метода) состоит в том, что ИЦ движется прямолинейно вдоль линии визирования СОПН.

Таким образом, для корректности метода необходимо исключить случаи, когда ИЦ движется вдоль указанной линии или в ее окрестности. Это накладывает определенные ограничения на условия наблюдения ИЦ, что необходимо предусмотреть на практике.

Если ограничиться моделью прямолинейного равномерного движения и сигналом с известным периодом (в этом случае в (5.1) надо положить $p = \overline{1, 3}$ и $\mathbf{t}_{[l]} = [t_{[l]1}, t_{[l]2}, t_{[l]3}]^T$), то решение СЛАУ (5.1) при корректном применении метода позволяет определить искомые параметры движения ИЦ

$$(5.4) \quad \left\{ \begin{array}{l} R_{0[l]} = 2^{-1}v_S \left(\frac{\delta_{[l]1}^2 \Delta_{[l]23}^t - \delta_{[l]2}^2 \Delta_{[l]13}^t + \delta_{[l]3}^2 \Delta_{[l]12}^t}{-\delta_{[l]1} \Delta_{[l]23}^t + \delta_{[l]2} \Delta_{[l]13}^t - \delta_{[l]3} \Delta_{[l]12}^t} \right), \\ \langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]} = 2^{-1}v_S^2 \left(\frac{t_{[l]1}^2 \Delta_{[l]23}^\delta - t_{[l]2}^2 \Delta_{[l]13}^\delta + t_{[l]3}^2 \Delta_{[l]12}^\delta}{-\delta_{[l]1} \Delta_{[l]23}^t + \delta_{[l]2} \Delta_{[l]13}^t - \delta_{[l]3} \Delta_{[l]12}^t} \right), \\ v_{0[l]} = \left[\frac{t_{[l]3} \Delta_{[l]12}^\delta - t_{[l]2} \Delta_{[l]13}^\delta + t_{[l]1} \Delta_{[l]23}^\delta}{\delta_{[l]1} \Delta_{[l]23}^t - \delta_{[l]2} \Delta_{[l]13}^t + \delta_{[l]3} \Delta_{[l]12}^t} \right]^{1/2}, \\ \gamma_{0[l]} = \arccos \left[\frac{\langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]}}{R_{0[l]} v_{0[l]}} \right], \end{array} \right.$$

где $\Delta_{[l]12}^t = t_{[l]1}t_{[l]2} (t_{[l]1} - t_{[l]2})$, $\Delta_{[l]12}^\delta = \delta_{[l]1}\delta_{[l]2} (\delta_{[l]1} - \delta_{[l]2})$ и, если не учитывать ошибки измерений и вычислений, $R_{0[l]} = R_0$, $v_{0[l]} = v_0$, $\langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]} = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle$, $\gamma_{0[l]} = \gamma_0$.

Следовательно, появляется возможность определения параметров движения R_0 , v_0 и γ_0 (где $R_0 = \chi_1$, $v_0 = \sqrt{\chi_3}$, $\gamma_0 = \arccos [\chi_2 (R_0 v_0)^{-1}]$), не прибегая к численному решению СЛАУ, что является несомненным достоинством развиваемого ПВМ.

6. Учет случайных погрешностей измерений

Полагая период сигнала известным, для оценки влияния случайных погрешностей измерений ξ на точностные характеристики метода воспользуемся традиционной процедурой расчета элементов корреляционной матрицы $\mathbf{K}_{\chi[l]}$ ошибок оценивания координат вектора χ в линейном приближении [26]. Для этого с учетом СЛАУ (5.1) (полагая для простоты матрицу $\mathbf{C}_{[l]}$ квадратной размером 5×5) воспользуемся представлением $\chi_{[l]} = \mathbf{C}_{[l]}^{-1} \mathbf{Y}_{[l]} = [\chi_k (\delta_{[l]}), k = \overline{1, 5}]^T$ (где $\delta_{[l]} = [\delta_{[l]p}, p = \overline{1, 5}]^T$) и частными производными следующего вида: $\partial \chi_{k[l]} (\delta_{[l]}) / \partial \delta_{[l]p}$. Корреляционная матри-

ца находится по правилу

$$(6.1) \quad \mathbf{K}_{\chi[l]} = \mathbf{F}_{\chi[l]} \mathbf{K}_{\xi} \mathbf{F}_{\chi[l]}^T,$$

где $\mathbf{F}_{\chi[l]} = [\partial \chi_k[l] (\boldsymbol{\delta}_{[l]}) / \partial \delta_{[l]p}]$, $k = \overline{1, 5}$, $p = \overline{1, 5}$.

Выражение (6.1) позволяет априорно на математических ожиданиях измеряемых параметров оценить потенциальные возможности развиваемого ПВМ и выработать практические рекомендации для его наилучшего использования при конкретных условиях наблюдения ИПЦ, а также обоснованно подходить к выбору основных параметров метода (длины интервала наблюдения (T), количества узлов (N) и временных наборов ($\mathbf{t}_{[l]}$)). Так, номер $l^* \in \{1, \dots, L\}$ оптимального набора $\boldsymbol{\delta}_{[l^*]}$, обеспечивающего минимизацию ошибки оценивания, находится по следующему адаптивному правилу:

$$(6.2) \quad l^* = \arg \min_l \|\mathbf{K}_{\chi[l]}\|,$$

где $\|\mathbf{K}_{\chi[l]}\|$ — любая из норм матрицы $\mathbf{K}_{\chi[l]}$, применяемая в задачах оценивания.

При практической реализации развиваемого ПВМ следует учитывать тот фактор, что при больших значениях v_S (например, когда $v_S = c$, где c — скорость света) решение квадратной СЛАУ (4.8) при наличии случайных погрешностей измерений может приводить к некорректным результатам. Поясним этот факт для случая $N = 6$ на примере вычисления скорости v_0 . Поскольку $v_0 = c\sqrt{1 - 2T_S A_3}$, то ошибка $\Delta_3 = \hat{A}_3 - A_3$ (где \hat{A}_3 — вычисленное значение коэффициента A_3 путем решения СЛАУ (4.8) с учетом ошибок измерений) приводит к следующей оценке скорости: $\hat{v}_0 = \sqrt{v_0^2 + 2c^2 T_S \Delta_3}$. То есть корректная оценка скорости возможна лишь при выполнении условия $\Delta_3 > -v_0^2 (2c^2 T_S)^{-1}$, что накладывает очень жесткое ограничение на величину погрешности Δ_3 . Данный эффект относится также ко всем коэффициентам СЛАУ (4.8), кроме A_2 и A_4 .

Для преодоления указанной некорректности (при больших скоростях v_S) рекомендуется двухэтапный подход к идентификации. На первом этапе решается СЛАУ (4.8), из которой потребуются лишь оценка \hat{A}_4 для A_4 . Это позволяет сформировать искомую оценку $\hat{T}_S = 2\hat{A}_4$ для периода T_S , а на ее основе и оценки для невязок $\hat{\delta}_n = t_n - M_n \hat{T}_S$. Все оценки параметров движения ИПЦ находятся на базе СЛАУ (5.1), в которую вместо δ_n подставляется величина $\hat{\delta}_n$.

7. Учет избыточных измерений

Теперь рассмотрим случай избыточных измерений, когда матрица $\mathbf{C}_{[l]}$ и вектор $\mathbf{Y}_{[l]}$ в (5.1) имеют произвольное число строк $P_{[l]} \leq N$, которое, как правило, значительно превышает количество оцениваемых параметров. Для упрощения расчетов будем рассматривать в СЛАУ (5.1) составляющую

$\mathbf{Y}_{[l]} = \left[v_S^2 \delta_{[l]p}^2, p = \overline{1, P_{[l]}} \right]^T$ в качестве вектора вторичных измеряемых параметров, а первичные измерения $h_{[l]1}, \dots, h_{[l]P_{[l]}}$ полагаем некоррелированными. С учетом этого по аналогии с (5.1) корреляционную матрицу ошибок измерений координат вектора $\mathbf{Y}_{[l]}$ можно представить так

$$(7.1) \quad \mathbf{K}_{\mathbf{Y}_{[l]}} = \mathbf{F}_{\delta[l]} \mathbf{K}_{\xi} \mathbf{F}_{\delta[l]}^T.$$

В предположении, что матрица \mathbf{K}_{ξ} является диагональной, имеем $\mathbf{K}_{\mathbf{Y}_{[l]}} = \text{diag} \left[4\delta_{[l]1}^2, 4\delta_{[l]2}^2, \dots, 4\delta_{[l]P_{[l]}}^2 \right]$. При условии достаточно малых ошибок измерений для построения сглаженной оценки $\mathbf{x}_{[l]}^*$ вектора \mathbf{x} можно воспользоваться методом наименьших квадратов [25]

$$(7.2) \quad \mathbf{x}_{[l]}^* = \left(\mathbf{C}_{[l]}^T \mathbf{K}_{\mathbf{Y}_{[l]}}^{-1} \mathbf{C}_{[l]} \right)^{-1} \mathbf{C}_{[l]}^T \mathbf{K}_{\mathbf{Y}_{[l]}}^{-1} \mathbf{h}_{\mathbf{Y}_{[l]}},$$

где $\mathbf{h}_{\mathbf{Y}_{[l]}} = \left[h_{\mathbf{Y}_{[l]p}}, p = \overline{1, P_{[l]}} \right]^T$ — вектор вторичных измерений.

Корреляционную матрицу ошибок оценивания находим так:

$$(7.3) \quad \mathbf{K}_{\mathbf{x}_{[l]}^*} = \left(\mathbf{C}_{[l]}^T \mathbf{K}_{\mathbf{Y}_{[l]}}^{-1} \mathbf{C}_{[l]} \right)^{-1}.$$

Для выбора оптимального набора с номером $l^* \in \{1, \dots, L\}$ применяем адаптивный алгоритм типа (6.2).

Следует отметить, что подход (7.1)–(7.3) не является строго оптимальным, поскольку элементы матрицы $\mathbf{C}_{[l]}$ зависят от результатов наблюдений. Но при определенных ограничениях на погрешности измерений он дает вполне приемлемый результат.

Для более точного сглаживания можно использовать известные процедуры нелинейного оптимального оценивания, которые приводят на практике к трудоемким рекуррентным вычислительным алгоритмам, предполагающим задание достаточно качественного начального условия.

Другой наиболее простой и достаточно надежный способ построения сглаженной оценки $\mathbf{x}_{[l]}^*$ состоит в предварительном сглаживании первичных измерений $h_{[l]1}, \dots, h_{[l]P_{[l]}}$ соответствующим полиномом $\delta_{[l]}^*(t)$ и применении полученных результатов к решению СЛАУ (5.1). Кроме того, можно найти сглаженную оценку дальности для любого $t \in [0, T]$, а именно,

$$(7.4) \quad R_{[l]}(t) = R_{0[l]}^* + c\delta_{[l]}^*(t).$$

Здесь в качестве оптимального принимаем набор с номером $l = l^* \in \{1, \dots, L\}$.

8. Некоторые обобщения и практические рекомендации

Выше рассматривался случай оценивания начальной дальности $R_0 = R(0)$ для момента времени $t = 0$. Однако если ряд Тейлора, используемый для

описания криволинейного движения ИЦ, записать относительно не начального, а любого произвольного $t = t_* \in [0, T]$, то по аналогии с вышеизложенным можно решить задачу идентификации именно для момента времени t^* , в частности найти дальность $R_* = R(t_*)$.

Развитый метод несложно реализовать в виде следующих алгоритмов: по выборке нарастающего объема, на «скользящей сетке» или в виде фильтра [25]. При этом движение ИЦ на интервале наблюдения можно рассматривать как кусочно-полиномиальное (в [20] оно рассматривалось как кусочно-линейное).

При практической реализации метода возникают вопросы (например, выбор степени полинома, описывающего движение ИЦ или количества подсчитываемых импульсов), связанные с организацией измерительного эксперимента. В [25] даются практические рекомендации для решения этих вопросов в полном объеме. Очевидно, что развиваемый метод наиболее эффективен в случае, когда речь идет о больших пройденных расстояниях (т.е. «синтезируется база» достаточного размера), а это задает определенные ограничения на тип ИЦ (в частности на его скорость, на возможности маневра и т.д.), на адекватность используемого полинома на заданном интервале наблюдения и на технические характеристики СОПН.

Для случаев, связанных с движением ИЦ по линии визирования, можно предложить гибридный вариант использования развитого и известного энергетического метода [27]. Доказано, что данный метод, оперирующий с относительным уровнем принимаемого сигнала, реализует при движении ИЦ по линии визирования свои потенциальные возможности. В некотором смысле развитый и энергетический методы «ортогональны» друг другу в плане точности. Следовательно, комбинируя эти методы, можно выровнять рабочую зону гибридного метода и достичь приемлемых точностных характеристик для различных условий наблюдения ИЦ.

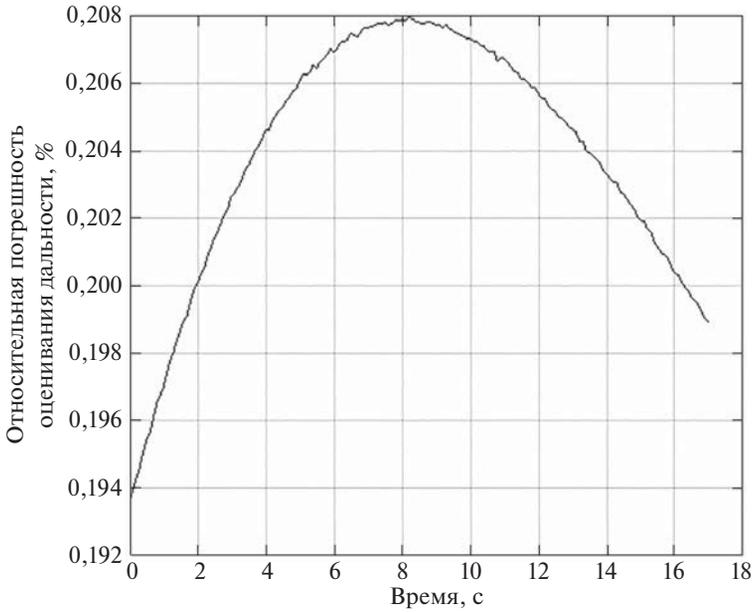
Для более эффективного применения энергетического метода следует использовать процедуры кластеризации и мажоритарной обработки для редукции и отсева ненадежных измерений.

9. Иллюстративный пример

Предположим, что ИЦ осуществляет плоскостное движение: $x(t) = x_0 + v_{x0}t$, $y(t) = y_0 + v_{y0}t$, где $x_0 = y_0 = 11 \cdot 10^3$, $v_{x0} = -5 \cdot 10^2$, $v_{y0} = 6 \cdot 10^2$, $\gamma_0 = 85$. Здесь и далее время и погрешности измерений временных интервалов задаются в секундах (с), координаты и дальность — в метрах (м), скорость — в м/с, ускорение — в м/с², частота — в герцах (Гц), угол — в градусах, относительная погрешность — в процентах.

ИЦ формирует импульсный радиосигнал

$$S_0(t) = \sum_{k=1}^K \text{rect} [(t - kT_S) \tau^{-1}] \cos(2\pi f_0 t),$$



Относительная погрешность оценивания дальности.

где $T_S = 10^{-2}$, $\tau = 10^{-5}$, $f_0 = 10^{10}$. Параметры работы СОПН: $T = 18$, $v_S = c = 3 \cdot 10^8$, $L = 1$ (т.е. использован один единственный набор узлов), $P_{[1]} = 4$ (размер набора), $\Delta M_p = \Delta M = 10$, $\mathbf{K}_\xi = \text{diag} [\sigma^2, \dots, \sigma^2]$, при этом погрешности измерений временного положения фронтов импульсов полагались некоррелированными и задавались по нормальному закону распределения с нулевым математическим ожиданием и значением среднеквадратического отклонения $\sigma = 10^{-9}$.

Реализация метода осуществлялась в два этапа с применением датчика случайных чисел и усреднением по тысяче экспериментов. На первом этапе решалась СЛАУ (4.8) с квадратной матрицей \mathbf{B} размером 4×4 (полагалось $A_5 = A_6 = 0$, так как рассматривается ИЦ с нулевым ускорением), при этом для расчета элементов матрицы \mathbf{B} и столбца \mathbf{D} использован вектор $\mathbf{t}_{[1]} = \bar{\mathbf{t}}_{[1]} + \boldsymbol{\delta}_{[1]} = [t_{[1]p}, p = \overline{1,4}]^T$ с номерами узлов: $[1]1 = 12$, $[1]2 = 65$, $[1]3 = 118$, $[1]4 = 171$, $\bar{\mathbf{t}}_{[1]} = [\bar{t}_{[1]p}, p = \overline{1,4}]^T = [[1]p \cdot 10^{-1}, p = \overline{1,4}]^T$. Из всех четырех оценок неизвестных коэффициентов выбирается только оценка периода сигнала $\hat{T}_{S[1]} = 9,999999731646 \cdot 10^{-3}$ (получена на основе набора $\mathbf{t}_{[1]} = \bar{\mathbf{t}}_{[1]} + \boldsymbol{\delta}_{[1]}$), которой соответствует относительная погрешность $\delta T_{S[1]} = 2,683540941544882 \cdot 10^{-6}$.

На втором этапе с учетом $\hat{\delta}_{[1]p} = t_{[1]p} - M_{[1]p} \hat{T}_{S[1]} = t_{[1]p} - [1]p \Delta M \hat{T}_{S[1]}$ решалась СЛАУ (5.1) с квадратной матрицей $\mathbf{C}_{[1]}$ (размером 3×3), при этом $\chi_4 = \chi_5 = 0$ и использовался набор $\bar{\mathbf{t}}_{[1]} = [\bar{t}_{[1]p}, p = \overline{1,3}]^T = [1, 1; 9, 1; 17, 1]^T$. Сама матрица образована строками $c^{-2} \begin{pmatrix} -2c\hat{\delta}_{[1]p}, 2t_{[1]p}, t_{[1]p}^2 \end{pmatrix}$, $p = \overline{1,3}$. В итоге истинной дальности $R_0 = 1,555634918 \cdot 10^4$ соответствуют оценка $\hat{R}_{0[1]} =$

$= 1,559672203 \cdot 10^4$ и погрешность $\Delta R_{0[1]} = 0,259526489$, истинной скорости $v_0 = 7,810249675 \cdot 10^2$ — оценка $v_{0[1]} = 7,821417156 \cdot 10^2$ и погрешность $\Delta v_{0[1]} = 0,142984942$, истинному углу $\gamma_0 = 84,805571092$ — оценка $\gamma_{0[1]} = 84,761511501$ и погрешность $\Delta \gamma_{0[1]} = 0,051953650$.

На рисунке представлен график зависимости относительной погрешности оценивания дальности, полученный с учетом (7.4).

Для более эффективного использования разработанного в статье метода вопрос выбора величины интервала наблюдения и узлов временной сетки, а также их согласования с динамикой движения ИЦ и величиной погрешностей измерений следует решать в оптимизационной постановке. При решении СЛАУ следует привлекать известные методы регуляризации. Результаты численного эксперимента показывают, что чем больше расстояние между узлами используемой временной сетки, тем меньшее влияние оказывают случайные погрешности измерений на результирующую точность оценивания. Это расстояние должно быть согласовано с динамикой ИЦ, а именно: чем меньше скорость движения ИЦ, тем больше должны быть шаг этой сетки и длительность интервала наблюдения.

10. Заключение

Разработанный ПВМ позволяет идентифицировать модель криволинейного полиномиального движения ИЦ по результатам регистрации временной невязки между периодами излученного сигнала и этими же периодами, подсчитанными в точке наблюдения. Метод не требует знания периода сигнала и предварительной оценки текущей частоты Доплера, а также знания каких-либо априорных данных о параметрах принятой модели движения ИЦ. Исследованы наблюдаемость и основные ограничения метода, условия его наиболее эффективного применения. Получены аналитические соотношения, позволяющие оценить эволюцию временной невязки с учетом характеристик ИЦ и СОПН, а также точностные характеристики метода для различных условий наблюдения.

Метод можно реализовать в различных вариантах: по фиксированной выборке измерений, по выборке измерений нарастающего объема, в виде алгоритма динамической фильтрации (линейной, квазилинейной или нелинейной) и др.

Метод может быть реализован как самостоятельно, так и в составе гибридного метода, объединяющего другие известные подходы пассивной однопозиционной и многопозиционной локации и навигации ИЦ. Поскольку развитый метод позволяет определять дальность, то он может быть использован в дальномерно-дальномерных системах многопозиционной локации при решении известной задачи трилатерации [21, 22].

При наличии в периодо-временных измерениях не только флуктуационных погрешностей, но и сингулярных ошибок целесообразно первоначально подвергнуть эти измерения процедуре обобщенного инвариантно-несмещен-

ного оценивания [28], обеспечивающей компенсацию этих ошибок, достижение эффекта сглаживания и оптимальное оценивание различных числовых характеристик (линейных функционалов, например производных, интегралов, спектральных коэффициентов и т.д.), полезных не только для повышения вычислительной устойчивости метода, но и оценки его эффективности. Для решения СЛАУ с использованием процедуры регуляризации можно применять известный подход [29].

СПИСОК ЛИТЕРАТУРЫ

1. Основы маневрирования кораблей / Под ред. М. Скворцова. М.: Воениздат, 1966.
2. *Шебшаевич В.С.* Введение в теорию космической навигации. М.: Сов. радио, 1971.
3. *Громов Г.Н.* Дифференциально-геометрический метод навигации. М.: Радио и связь, 1986.
4. *Хвоц В.А.* Тактика подводных лодок. М.: Воениздат, 1989.
5. *Соловьев Ю.А.* Спутниковая навигация и ее приложения. М.: ЭкоТрендз, 2003.
6. *Мельников Ю.П., Попов С.В.* Радиотехническая разведка. М.: Радиотехника, 2008.
7. *Ярлыков М.С.* Статистическая теория радионавигации. М.: Радио и связь, 1985.
8. *Сосулин Ю.Г., Костров В.В., Паршин Ю.Н.* Оценочно-корреляционная обработка сигналов и компенсация помех. М.: Радиотехника, 2014.
9. *Бульчев Ю.Г., Манин А.П.* Математические аспекты определения движения летательных аппаратов. М.: Машиностроение, 2000.
10. *Бульчев Ю.Г., Васильев В.В., Джуган Р.В. и др.* Информационно-измерительное обеспечение натурных испытаний сложных технических комплексов. М.: Машиностроение – Полет, 2016.
11. *Гельцер А.А.* Однопозиционный метод определения местоположения источника радиоизлучения с использованием отражений сигналов от множества элементов рельефа и местных предметов // Автореф. дисс. Том. гос. ун-т систем управления и радиоэлектроники. 2012.
12. *Сиренко И.Л., Донец И.В., Рейзенкинд Я.А. и др.* Однопозиционное определение координат и вектора скорости радиоизлучающих объектов // Радиотехника. 2019. № 10 (16). С. 28–32.
13. *Бульчев Ю.Г., Бульчев В.Ю., Ивакина С.С., Насенков И.Г.* Пассивная локация группы движущихся целей одним стационарным пеленгатором с учетом априорной информации // АиТ. 2017. № 1. С. 152–166.
Bulychev Yu.G., Bulychev V.Yu., Ivakina S.S., Nasenkov I.G. Passive Location of a Group of Moving Targets with One Stationary Bearing with Prior Information // Autom. Remote Control. 2017. V. 78. No. 1. P. 125–137.
14. *Бульчев Ю.Г., Бульчев В.Ю., Ивакина С.С., Николас П.И.* Оценка наклонной дальности до цели с полиномиальным законом движения // Вестн. Казан. гос. ун-та. 2013. № 1. С. 67–74.
15. *Бульчев Ю.Г.* Некоторые аспекты идентификации динамических объектов при некорректных условиях наблюдения // АиТ. 2020. № 6. С. 131–152.

- Bulychev Yu.G.* Some Aspects of Identification of Dynamic Objects under Incorrect Observation Conditions // Autom. Remote Control. 2020. V. 81. No. 6. P. 1073–1090.
16. *Дятлов А.П., Дятлов П.А.* Доплеровские обнаружители подвижных объектов с использованием «постороннего» источника излучения // Спец. техника. 2010. № 5. С. 16–22.
 17. *Aidala V.J., Nardone S.C.* Biased Estimation Properties of the Pseudolinear Tracking Filter // IEEE Transact. Aerospas. Electron. Syst. 1982. V. 18. No. 4. P. 432–441.
 18. *Amelin K.S., Miller A.B.* An Algorithm for Refinement of the Position of a Light UAV on the Basis of Kalman Filtering of Bearing Measurements // J. Commun. Techn. Electron. 2014. V. 59. No. 6. P. 622–631.
 19. *Miller A.B.* Development of the Motion Control on the Basis of Kalman Filtering of Bearing-Only Measurements // Autom. Remote Control. 2015. V. 76. No. 6. P. 1018–1035.
 20. *Бульчев Ю.Г., Мозоль А.А.* Однопозиционная пассивная локация источника излучения с криволинейным движением и учетом эволюции периода сигнала в точке приема // РЭ. 2023. Т. 68. № 2. С. 131–137.
 21. *Кондратьев В.С., Котов А.Ф., Марков Л.Н.* Многопозиционные радиотехнические системы. М.: Радио и связь, 1986.
 22. *Черняк В.С.* Многопозиционная радиолокация. М.: Радио и связь, 1993.
 23. *Нефедов В.И., Сигов А.С., Битюков В.К., Самохина Е.В.* Электрорадиоизмерения. М.: Форум: Инфра-М. 2018.
 24. *Лоусон Ч., Хенсон Р.* Численное решение задач метода наименьших квадратов. М.: Наука, 1986.
 25. *Жданюк Б.Ф.* Основы статистической обработки траекторных измерений. М.: Сов. радио, 1978.
 26. *Вентцель Е.С.* Теория вероятностей. М.: Высш. шк., 1999.
 27. *Бульчев Ю.Г., Ивакина С.С., Насенков И.Г.* Метод пассивно-энергетической локации и навигации в стационарной и нестационарной постановках // Радиотехника. 2015. № 6. С. 107–115.
 28. *Бульчев Ю.Г., Елисеев А.В.* Вычислительная схема инвариантно несмещенного оценивания значений линейных операторов заданного класса // ЖВМиМФ. 2008. Т. 48. № 4. С. 580–592.
 29. *Бульчев Ю.Г., Бурлай И.В.* Метод параметрической идентификации систем управления при неточном задании входных данных // АиТ. 1997. № 11. С. 56–65.

Статья представлена к публикации членом редколлегии О.Н. Граничиным.

Поступила в редакцию 16.06.2022

После доработки 04.05.2023

Принята к публикации 21.05.2023

Интеллектуальные системы управления, анализ данных

© 2023 г. А.М. МИХАЙЛОВ, канд. техн. наук (alxmikh@gmail.com),

М.Ф. КАРАВАЙ, д-р. техн. наук (mkaravay@yandex.ru),

В.А. СИВЦОВ (TheDeGe@yandex.ru)

(Институт проблем управления им. В.А. Трапезникова РАН, Москва),

М.А. КУРНИКОВА (mish2109@yandex.ru)

(Национальный медицинский исследовательский центр детской гематологии,
онкологии и иммунологии им. Дмитрия Рогачева, Москва)

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ПО ПОЛНОМУ ПРОФИЛЮ ЭКСПРЕССИИ ГЕНОВ

Рассматривается использование машинного обучения для диагностики заболеваний, основанной на анализе полного профиля экспрессии генов, что отличает данную работу от других подходов, где необходимо проведение предварительного этапа, на котором производится поиск ограниченного числа релевантных генов (десятки и сотни генов). Проведены эксперименты с полными профилями генетической экспрессии (20 531 генов), полученными в результате обработки транскриптомов 801 пациента с известными онкологическими диагнозами (онкология легких, почек, молочной железы, простаты и толстой кишки). Использование индекстрона (индексной системы мгновенного обучения) по новому назначению, т.е. для обработки полных профилей экспрессии, обеспечило точность диагностирования, которая на 99,75 процентов совпала с результатами гистологической верификации.

Ключевые слова: распознавание образов, машинное обучение, обратные образы, профили экспрессии генов, диагностика заболеваний.

DOI: 10.31857/S000523102307005X, **EDN:** FDMPHU

1. Введение

Общим знаменателем существующих подходов к диагностике онкологических заболеваний по профилям экспрессии генов, см., например, методологические и обзорные работы [1–6], является использование ограниченного набора генов. При таком подходе также необходимо проводить предварительный анализ и выделение отдельных генов или комбинаций генов, экспрессионная активность которых наиболее характерна в случае тех или иных заболеваний. При этом достигнутая точность диагностики не превышает 95%, на наборах, включающих до 90 генов [4], что связано, по-видимому, с ограниченным числом используемых генов. Так, в [5] это число снижается даже до 10 генов.

Таким образом, работа с данными проходит в два этапа. Например, в [6] на первом этапе используется анализ принципиальных компонент, в результате которого отбираются 103 гена. Окончательная диагностика производится на втором этапе, когда в результате использования байесовской нейронной сети была обеспечена точность в 93,66%, глубокая нейронная сеть обеспечила 93,41%, а логистическая регрессия — 92,82%. При этом возможно использование третьего этапа, когда отбрасываются сомнительные результаты, определяемые порогом принятия решения, что позволяет почти полностью исключить ошибки диагностирования [6] ценой того, что часть случаев не диагностируется, что считается более приемлемым, чем ошибочный диагноз.

В настоящей работе рассматривается использование машинного обучения для диагностики заболеваний, основанной на анализе полного профиля экспрессии генов, т.е. активности всех 20 531 известных генов, что упрощает работу, так как при этом отпадает необходимость предварительного отбора релевантных генов. В статье рассматриваются подробности работы этой использованной системы машинного обучения.

В биологии уровень экспрессии оценивает транскрипционную активность гена как количество произведенной им матричной РНК (мРНК). Получение профиля экспрессии генов — это малоинвазивная или полностью не инвазивная процедура, как, например, взятие образца слюны, чем определяется комфортность такой процедуры для пациента. При этом качество диагностирования зависит от полноты профиля, т.е. от количества рассмотренных генов. Однако использование полных профилей экспрессии генов — это, скорее, будущее медицинской диагностики, так как в настоящее время дешевое оборудование для массового применения с целью получения таких профилей отсутствует. В настоящее время профили могут быть найдены, в частности, с помощью следующих технологий [7]:

- ПЦР-тесты;
- ДНК-микрочипы (microarrays);
- секвенирование.

Если речь идет об анализе уровня экспрессии относительно небольшого количества генов, может использоваться доступный и сравнительно недорогой метод количественной ПЦР в реальном времени (qPCR). Вторая технология — ДНК-микрочипы, дает более качественную оценку экспрессии генов. Однако когда речь идет о экспрессионном профиле всех генов, то на сегодняшний день для этого необходимо использование секвенирования. Одной из наиболее востребованных платформ для высокопроизводительного секвенирования является оборудование компании Illumina [8]. Система HiSeq 2500 применяет технологию нового поколения SBS, которая поддерживает массово-параллельное секвенирование, используя метод флуоресцентно-меченных нуклеотидов, который позволяет прочтение отдельных оснований по мере их включения в растущие нити ДНК. На рис. 1 представлен вид оборудования, использующего этот метод.



Рис. 1. Система HiSeq 2500.

Illumina HiSeq выполняет две функции, такие как чтение генома и определение уровня экспрессии. Последнее находится путем прочтения не ДНК, а мРНК-кода. При этом по количеству копий мРНК судят об уровне экспрессии. С учетом того, что стоимость РНК-секвенирования падает, в перспективе возможно создание более простого и дешевого оборудования для установки в таких широко распространенных лабораториях, как INVITRO. При массовой доступности такого оборудования речь могла бы идти о перспективной диагностике, поскольку использование полного профиля экспрессии автоматически учитывает все возможные **активные** комбинации генов, которые трудно предугадать и которые влияют на возникновение различных заболеваний. Кроме того, профили экспрессии генов служат важным материалом не только при диагностике, но и в разнообразных научных и клинических исследованиях.

Вторая составляющая рассматриваемой методологии — это машинное обучение, позволяющее автоматически, без анализа генетических комбинаций, обучаться на полных биологических профилях с целью их последующей классификации и диагностирования. Традиционно для машинного обучения используются искусственные нейронные сети [9], являющиеся популярным методом, применяемым для решения множества различных задач предсказания и распознавания образов. Однако обучение таких сетей занимает много времени, требуя больших вычислительных ресурсов, что связано с необходимостью вычисления огромного количества коэффициентов для адаптации многослойной сетевой архитектуры к конкретной задаче. Тем не менее обучение можно намного ускорить и сделать чуть ли не мгновенным, если воспользоваться системой индексации образов [10, 11], напоминающей поисковику типа Google [12], где инкрементное обучение сводится к индексированию новых документов. В 1998 г. был введен термин индексстрон [13] — термин, который используется только как название устройства индексации образов, но не как название метода индексации образов. Отличие индексстрона от поисковиков состоит в том, что вместо инвертирования текстовых документов инвертируются числовые данные. Специфика инвертирования числовых образов рассмотрена в разделе 3.

Обратим внимание, что подход, использованный в [10], имеет общие моменты с предложенным позднее методом TF-IDF [14], разработанным для поиска документов путем вычисления обратных частот слов документов. Метод

TF-IDF используется в поисковых системах, где вычисляются частоты имен документов в полностью инвертированных файлах. Вместе с тем текстовые поисковые системы не будут работать с числовыми данными, поскольку документная частота слова может измениться в зависимости от имеющегося в данный момент набора документов, но и само ключевое слово может полностью исчезнуть даже из-за небольшого шума.

Отметим, что индексный подход к решению задач распознавания образов практически не используется в системах машинного обучения, где большинство методов используют итеративное обучение, градиентный спуск и значительное количество адаптивных коэффициентов, что, собственно, и приводит к медленному обучению. В то же время мозг человека может запомнить новые зрительные образы с одного взгляда. В настоящей статье для обеспечения практически моментального обучения при работе с большими объемами данных используется не метод итеративного обучения, а метод индексации числовых данных. Предыдущие работы по индексации не текстовых, а числовых данных можно найти в [15], где такие быстрые методы использованы для распознавания изображений в кинофильмах. Но в [15] распознавание зашумленных числовых образов, представляющих изображения, сводится к конвертации числовых образов в текстовую форму с дальнейшим использованием уже стандартных методов для поиска текстовой информации. Подход, использованный в данной статье, заключается в применении метода индексного распознавания [10, 11], который делает возможным индексирование зашумленных числовых образов без их промежуточного преобразования в текстовую форму. Отметим также, что целью данной статьи является не разработка нового метода машинного обучения, а использование индексного метода распознавания по новому назначению, а именно для практически мгновенного обучения при работе с большими базами биологических данных.

2. Исходные данные

В эксперименте для сравнения эффективности обучения и классификации данных нейронной сетью и индексным методом был использован набор данных Gene expression cancer RNA-Seq [16]. Этот набор данных содержит 801 строку, каждая из которых содержит 20 531 число с плавающей точкой (см. табл. 1). Каждое число с координатами профиль/ген — это уровень активности соответствующего гена, измеряемой в условных единицах от 0 (нулевая

Таблица 1. Активность генов пациентов

Профиль	Ген 0	Ген 1	Ген 2	...	Ген 20 529	Ген 20 530	Класс (диагноз)
0	0,0	2,017	3,266	...	5,287	0,0	3
1	0,0	0,592	1,588	...	2,094	0,0	1
...
800	0,0	2,325	3,806	...	4,551	0,08	5

активность или отсутствие гена) до максимальной активности 15. При этом числа каждой n -й строчки представляют уровни активности соответствующих генов n -го пациента. Диагнозы всех пациентов в наборе данных [15] известны и были определены другими клиническими методами. В эксперименте использовали 401 нечетную строку для обучения, в процессе которого системе сообщался диагноз, связанный с соответствующей строчкой. В процессе тестирования использовались 400 четных строк, которые классифицировались путем их отнесения к одному из пяти классов, а найденный класс сопоставлялся с известным диагнозом соответствующего пациента.

3. Формальная постановка задачи распознавания и метод ее решения

В настоящей статье индексный метод распознавания, предложенный и рассмотренный в [10] и усовершенствованный в [11] и в других работах, используется по новому назначению, а именно для диагностирования заболеваний путем классификации профилей экспрессии генов. Ниже рассмотрены особенности применения этого метода в данной задаче. Пусть все переменные являются целыми числами и пусть даны N образов, где каждый образ представлен K -мерным вектором признаков

$$(1) \quad \mathbf{x}_n = (x_{n,1}, \dots, x_{n,k}, \dots, x_{n,K}), \quad n = 1, \dots, N.$$

Здесь расстояние Чебышева между любыми двумя образами \mathbf{x}_p и \mathbf{x}_q ($p, q \leq N$) больше некоторого заданного числа R , т.е., $|\mathbf{x}_p - \mathbf{x}_q| > R$, а переменная $x_{n,k}$ ($0 \leq x_{n,k} < X$) является значением k -го признака вектора, представляющего n -й образ. Неравенство $|\mathbf{x}_p - \mathbf{x}_q| > R$ означает, что для векторов $\mathbf{x}_p, \mathbf{x}_q$ существует хотя бы одно измерение k , такое что $|x_{p,k} - x_{q,k}| > R$. В этом случае каждый n -й вектор представляет n -й класс, к которому относятся все векторы \mathbf{x} , такие что $|\mathbf{x} - \mathbf{x}_n| \leq R$. Величина R определяет размер класса и называется радиусом обобщения.

Задача классификации. Для заданного вектора $\mathbf{x} = (x_1, \dots, x_k)$ найти класс n , такой что $|\mathbf{x} - \mathbf{x}_n| \leq R$. Если такой класс не существует, то список (1) дополняется новым вектором $\mathbf{x}_{N+1} = \mathbf{x}$, а число классов увеличивается на 1.

Эту задачу, очевидно, можно решать путем сравнения заданного вектора \mathbf{x} с векторами, представляющими классы, т.е. путем полного перебора классов. Однако решение задачи классификации с помощью метода обратных образов позволяет значительно ускорить поиск, особенно при большом числе классов.

Рассмотрим решение задачи классификации с помощью обратных образов. При этом методе неизвестный образ \mathbf{x} относится к классу m такому, что

$$m : H_R(m|\mathbf{x}) = \max_{n=1}^N H_R(n|\mathbf{x}).$$

Здесь $H_R(m|\mathbf{x})$ — это гистограмма имен классов, содержащихся в обратных образах признаков вектора \mathbf{x} . Таким образом, $H_R(m|\mathbf{x})$ — это условная гистограмма классов, поскольку она зависит от вектора \mathbf{x} .

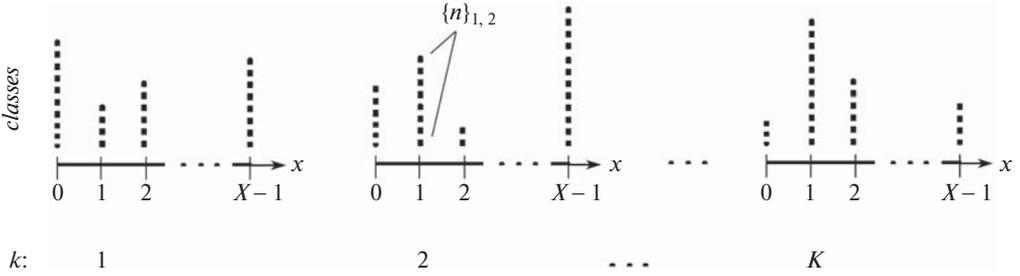


Рис. 2. Обратные образы показаны как K групп колонок.

Заметим, что если радиус обобщения R равен 0, то число классов будет равно числу N векторов списка (1). Если число реально существующих внешне заданных классов меньше N , то используется табличная функция $class(n)$, $n = 1, \dots, N$, устанавливающая соответствие между классами списка (1) и внешними классами, которые при обучении с учителем всегда известны. Обратные образы признаков определяются в [10, 11] как некоторые множества имен классов. Такие множества индексируются с помощью двухмерных индексов. Каждый двухмерный индекс определяется парой чисел (x, k) , где x — значение признака и k — номер измерения. Множество имен $\{n\}$ классов в левой части равенства (2) является обратным образом признака x из измерения k .

$$(2) \quad \{n\}_{x,k} = \{n : x_{n,k} = x\}, \quad k = 1, \dots, K, \quad x = 0, 1, \dots, X - 1.$$

Понятие обратных образов проиллюстрировано графически с помощью рис. 2, где они представлены колонками точек, высоты которых равны мощностям соответствующих обратных образов. Как отмечено выше, элементы колонок — это имена классов. Поскольку обратные множества имеют двухмерные индексы, то все колонки разбиты на K групп, где каждая группа k может содержать максимум X колонок, а каждая колонка — максимум N классов. Здесь X — это диапазон значений признаков. Для решения задачи классификации с помощью обратных образов остается найти гистограмму классов $H_R(n|\mathbf{x})$. Пусть входной образ представлен вектором $\mathbf{x} = (x_1, \dots, x_k)$. Тогда гистограмма классов, содержащихся в обратных образах, может быть найдена с помощью следующего алгоритма:

$$(3) \quad \forall n \in \{n\}_{x(k)+r,k}, \quad H_R(n|\mathbf{x}) = H_R(n|\mathbf{x}) + 1.$$

Здесь $r = -R, \dots, -1, 0, 1, \dots, R$, $k = 1, \dots, K$, $x(k+r), k = x_{k+r,k}$. Таким образом, представленный выше критерий классификации — это гистограмма классов, положение максимума которой требуется найти. Для понимания алгоритма необходимо использование понятия “обратные образы”. Оно определяется выражением (2), снабженным необходимыми комментариями. Обратные образы, содержащие множества классов образов, служат входной информацией цикла (3). Этот цикл гистограммирует классы, выдавая на выходе

Таблица 2. Нормализованные данные

Профиль	Ген 0	Ген 1	Ген 2	...	Ген 20 529	Ген 20 530	Класс (диагноз)
0	0	82	137	...	112	0	3
1	0	24	66	...	36	0	1
...
800	0	95	160	...	95	0	5

искомую гистограмму, положение максимума которой соответствует искомому классу. В терминах программирования — это цикл по макроколонкам r , измерениям k и классам n , находящимся в обратных образах.

При обучении индекстрона если максимум гистограммы при входе \mathbf{x} меньше выбранного некоторого порога, то число классов увеличивается на единицу, $N = N + 1$, и это новое число записывается в соответствующие обратные образы

$$\{n\}_{x^{(k)},k} = \{n\}_{x^{(k)},k} \cup N, \quad k = 1, \dots, K.$$

Начальным условием всегда служит значение $N = 0$. Однако гистограмма классов (3) не вычисляется, если в процессе обучения использовано нулевое значение радиуса обобщения $R = 0$. В то же время при каждой итерации число классов увеличивается на единицу. Графическая иллюстрация данного алгоритма приведена в [17].

При распознавании если максимум гистограммы при входе \mathbf{x} меньше некоторого порога, то входной образ остается нераспознанным.

Заметим, что максимум гистограммы равен размерности K образа \mathbf{x} . Можно показать, что это следует из следующих свойств обратных образов: колонки каждой группы

- содержат строго N разных имен классов: $\sum_{x=0}^{X-1} |\{n\}_{x,k}| = N$,
- не пересекаются: $\{n\}_{x,k} \cap \{n\}_{y,k} = \emptyset$.

При обучении и распознавании образа значение его признаков x_k , $k = 1, \dots, K$, используется в качестве адреса колонки, расположенной в k -м измерении. При этом реальные значения x признаков приводятся к целочисленному диапазону [0–255]:

$$x = 255(x - x_{\min}) / (x_{\max} - x_{\min}).$$

Опыт показывает, что такая дискретизация непрерывных значений, как правило, не снижает точности классификации. В табл. 2 приведен пример нормализации данных табл. 1.

Таким образом, признаки каждого поступившего K -мерного образа выделяют K колонок — по одной колонке в каждой группе. Проблема состоит в том, что, как правило, для числовых признаков пересечение колонок оказывается пустым, так как погрешности измерения и деформации образа искажают

адреса колонок. Поэтому вводятся макроколонки, т.е. наряду с колонкой с адресом x рассматриваются соседние колонки с адресами $x \in [-R, R]$. Возможность использования макроколонок вытекает из приведенных выше свойств обратных образов. Как уже отмечено выше, пересечения колонок находятся путем вычисления гистограммы, определяющей частоты появления классов. При этом входной образ относится к классу, встречающемуся наиболее часто.

4. Программно-аппаратная реализация индекстрона

Распараллеленная версия алгоритма была написана на языке Python и запущена на видеокарте Nvidia GeForce GTX 1660 Super. Эта видеокарта имеет 44 мультипроцессора и 1408 ядер, что позволило распараллелить один процесс на 1408 параллельных подпроцессов. При этом обучение в рассматриваемой задаче может быть распараллелено на 20 531 процессов по числу всех генов профиля, а максимально возможное распараллеливание — это $20\,531 * (2R + 1)$ процессов, так как для каждого гена возможно параллельно проверить все значения в пределах радиуса обобщения.

Распараллеливание было произведено с помощью модуля *cuda*, входящего в библиотеку *numba*. Для обработки потоков (*threads*) мультипроцессорами видеокарты потоки были разделены на блоки по формуле:

$$\begin{aligned} \text{blocks_per_grid} &= \\ &= (\text{number_of_iterations} | + (\text{threads_per_block} - 1)) // \text{threads_per_block}, \end{aligned}$$

где *number_of_iterations* — это количество параллельных итераций.

Перед выполнением программы массивы исходных данных были скопированы из оперативной памяти в память видеокарты с помощью функции *cuda.to_device()*, и по завершении выполнения программы результаты выполнения были скопированы обратно в оперативную память с помощью функции *device_array.copy_to_host()*.

Для распараллеливания функций записи и чтения использовался декоратор. Внутри распараллеливаемых функций вызывалась функция, возвращающая индекс потока, на котором должна осуществляться соответствующая параллельная итерация.

5. Результаты

Точность и количество операций при обучении индекстрона и нейронных сетей в задаче диагностики заболеваний по данным, полученным от 801 пациента, представлены в табл. 3.

Таблица 3. Сравнительные результаты

	Нейронная сеть	Индекстрон
Точность (%)	99,5	99,75
Количество операций	46 млрд операций сложения/умножения	8 млн операций записи в память

Время обучения индекстрона на одноядерном ноутбуке, 1,6 ГГц, составило 0,43 сек. Время обучения четырехслойной нейронной сети на аналогичной аппаратуре увеличивается в соответствии с ростом количества операций в $46 * 109/8 * 106 = 5750$ раз. Заметим, что архитектура индекстрона идеально подходит для распараллеливания, что позволило достичь практически мгновенного обучения за 75 миллисекунд в рассматриваемой задаче. Радиусы обобщения R в режимах обучения и распознавания составляют соответственно 0 и 84, т.е. 33% от диапазона (0–255) изменения признака.

6. Заключение

1. Простота алгоритма индекстрона, в результате которой обучение каждому образу требует только записи в память K целых чисел, а классификация образа имеет порядок сложности $O(hK)$ операций чтения и суммирования, позволяет создавать большие базы экспрессии генов для диагностирования самых разнообразных заболеваний (h — это средняя высота колонок, где $h = N/X$).

2. При широкой доступности оборудования для нахождения профилей генной экспрессии становится возможным создание поисковой диагностической системы типа Google для массового пользования, в которой запросы имеют не текстовую, а численную форму.

3. Проведенные эксперименты подтверждают наличие такой возможности, поскольку быстрое обучение при пополнении базы новыми данными даже на программном уровне занимает всего $0,43/800 = 0,00054$ сек, а точность диагностирования пяти типов онкологических заболеваний составила 99,75%. Однако диагностика многих других заболеваний потребует создания большого числа различных баз данных.

СПИСОК ЛИТЕРАТУРЫ

1. Khan J., Wei J., Ringner M. et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. // Nat Med. (2001). June 7(6): 673–9. <https://doi.org/10.1038/89044>
2. Kumar A., Halder A. Greedy fussy vaguely quantified rough approach for cancer relevant gene selection from gene expression data // Soft Comput. 2022. V. 26. P. 13567–13581. <https://doi.org/10.1007/s00500-022-07312-4>
3. Houssein E., Hassan H., Mustafa al-sayed et. al. Gene Selection for Microarray Cancer Classification based on Manta Rays Foraging Optimization and Support Vector Machines // Arabian Journal for Science and Engineering. 2022. V. 47. P. 2555–2572. <https://doi.org/10/1007/s13369-021-06101-8>
4. Zheng Y., Sun Y., Kuai Y. et al. Gene expression profiling for the diagnosis of multiple primary malignant tumors // Cancer Cell Int. 2021. V. 21, Article no. 47. <https://doi.org/10.1186/s12935-021-01748-8>
5. Ye Q., Wang Q., Qi P. et. al. Development and validation of a 90-gene real-time PCR assay for tumor origin identification // Symposium MXW, 2018.

6. *Joshi P., Dhar R. EpICC: A Bayesian neural network model with uncertainty correction for a more accurate classification of cancer // Sci. Rep 12, (2022). Article no. 14628. <https://doi.org/10.1038/s41598-022-18874-6>*
7. *Steiling K., Christenson S. Tools for genetics and genomics: Gene expression profiling // UpToDate.(2021). Retrieved from <https://www.uptodate.com/contents/tools-for-genetics-and-genomics-gene-expression-profiling>*
8. *СПбГУ Научный парк. Система высокопроизводительного полногеномного секвенирования, 2023. <https://researchpark.spbu.ru/equipment-biobank-rus/equipment-biobank-genom-rus/equipment-biobank-ngsseq-rus/1762-biobank-hiseq-2500-sequencing-system-rus>*
9. *IBM. What are neural networks? // Retrieved from <https://www.ibm.com/cloud/learn/neural-networks>*
10. *Mikhailov A., Pok Y.M. Artificial Neural Cortex // Smart Engineer. Syst. Design. 2001. V. 11. ASME PRESS. N. Y. P. 113–120.*
11. *Mikhailov A., Karavay M. Pattern Inversion as a Pattern Recognition Method for Machine Learning // Cornell University. 2021. Retrieved from <https://arxiv.org/abs/2108.10242>*
12. *Brin S., Page L. The Anatomy of a large-scale hypertextual web search engine // Comput. Networks ISDN Syst. 1998. V. 30. Iss. 1–7. Stanford University, Stanford, CA, 94305, USA. Retrieved from [https://doi.org/10.1016/S069-7552\(98\)00110-X](https://doi.org/10.1016/S069-7552(98)00110-X)*
13. *Mikhailov A. Indextron // Artificial Neural Networks in Engineering Conf. (ANNIE 1998), St. Louis, Missouri, Nov. 4–7, 1998. Proceedings Vol. 8: ANNIE 1998, Publisher: ASME Press, ISBN: 0791800822*
14. *Jones K. A statistical interpretation of term specificity and its application in retrieval // J. Document.: MCB Univer.: MCB Univer. Press, 2004. V. 60. No. 5. P. 493–502. ISSN 0022-0418*
15. *Sivic J., Zisserman A. Efficient visual search of videos cast as text retrieval // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009. V. 31. Issue 4. <https://doi.org/10.1109/TPAMI.2008.111>*
16. *UCI. Machine learning repository // Retrieved from <https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq>*
17. *Mikhailov A., Karavay M. Indextron // Proceedings of the 10th International Conference on Pattern Recognition Application and Methods, 4–6 Feb 2021, Vienna, V.1-978-989-758-486-2. P. 143–149. <https://doi.org/10.5220/0010180301430149>*

Статья представлена к публикации членом редколлегии О.Н. Граничиньым.

Поступила в редакцию 19.07.2022

После доработки 21.03.2023

Принята к публикации 30.03.2023

Интеллектуальные системы управления, анализ данных

© 2023 г. **Р.М. ШАКИРЗЯНОВ**, канд. техн. наук (RMShakirzyanov@kai.ru),
М.П. ШЛЕЙМОВИЧ, канд. техн. наук (MPShleymovich@kai.ru)
(Казанский национальный исследовательский технический
университет им. А.Н. Туполева),
С.В. НОВИКОВА, д-р техн. наук (SVNovikova@kai.ru)
(Казанский национальный исследовательский технический
университет им. А.Н. Туполева,
Мордовский государственный университет им. Н.П. Огарева)

МЕТОД АВТОМАТИЧЕСКОГО ПОЗИЦИОНИРОВАНИЯ БЕСПИЛОТНЫХ АППАРАТОВ НА ОСНОВЕ РАСПОЗНАВАНИЯ СИГНАЛЬНЫХ РАДИАЛЬНО-СИММЕТРИЧНЫХ МАРКЕРОВ ПОДВОДНЫХ ЦЕЛЕЙ

Описывается метод автоматического распознавания целевых точек траекторий беспилотных аппаратов, перемещающихся под водой, таких как автономные подводные аппараты и летающие подводные аппараты самолетоподобных конструкций. В качестве терминальной точки управления рассматривается координата центра объекта, обладающего свойствами радиальной симметрии. Предложен метод построения многомасштабной весовой модели изображения на основе разработанного преобразования быстрой радиальной симметрии и метода Хафа, что обеспечивает устойчивость к шумам и высокую скорость вычисления координат искомой точки. Для случая, когда объект интереса задан контуром определенного цвета, предложена модель на основе хроматической и весовой составляющих. В качестве примера детектирования приведен алгоритм обнаружения базовой подводной станции со световыми маркерами в виде сигнального люминесцирующего кольца.

Ключевые слова: автоматические транспортные системы, беспилотный подводный аппарат, беспилотный летательный подводный аппарат, компьютерное зрение, обработка изображений, обнаружение объектов на изображениях, метод Хафа, быстрое преобразование радиальной симметрии, весовая модель изображения, многомасштабная весовая модель изображения.

DOI: 10.31857/S0005231023070061, **EDN:** FDNZDT

1. Введение

При построении систем автоматического управления беспилотными аппаратами необходимым условием является задание конечного положения аппа-

рата в пространстве. В большинстве работ по автоматическому управлению беспилотными аппаратами конечное положение задается самим исследователем в виде набора конкретных координат [1, 2]. Однако в реальных условиях возникают задачи, когда аппарат должен самостоятельно обнаруживать объект интереса, положение которого будет задавать конечное положение аппарата при управлении. Транспортные системы, где аппарат самостоятельно определяет конечную точку своего движения и рассчитывает траекторию перемещения к ней, принято называть интеллектуальными транспортными системами [3]. Метод определения терминальной точки, как и способ построения оптимальной траектории, при этом не обязательно должен включать в себя элементы искусственного интеллекта (ИИ), такие как модели машинного обучения, нейронные сети и т.п. Однако на практике в случае, если метод расчета не предполагает использования ИИ, систему чаще определяют как «автоматическую» или «автоматизированную».

Наиболее распространенными беспилотными аппаратами в настоящее время являются беспилотные летательные аппараты (БПЛА) [4]. Основным способом позиционирования для БПЛА являются спутниковые системы GPS и ГЛОНАСС [5, 6]. Для беспилотных аппаратов, передвигающихся под водой, так называемых автономных необитаемых подводных аппаратов (АНПА) [7], позиционирование в большинстве случаев осуществляется при помощи ультразвуковых сигналов [8]. Основным недостатком такого подхода является обязательное наличие специальной навигационной аппаратуры на борту аппарата, а также необходимость ответного сигнала от целевой точки позиционирования, что делает аппарат уязвимым для внешнего обнаружения.

Системы позиционирования на основе компьютерного зрения лишены указанных недостатков. Они способны обеспечивать высокую точность распознавания и при этом не выдают своего присутствия внешними сигналами. Кроме того, для таких современных беспилотных систем, как беспилотные летательные подводные аппараты (БПЛПА) [9], визуальное позиционирование при подводном задании цели является единственно возможным. БПЛПА способны преодолевать расстояния сначала по воздуху, а затем погружаться в воду и выполнять маневры под водой. Из-за особенностей распространения и рассеивания сигналов на стыке двух сред воздух-вода позиционирование при помощи ультразвуковых сигналов является труднореализуемым.

Для БПЛА, как и для беспилотных наземных транспортных средств, задача обнаружения объектов на основе компьютерного зрения хорошо изучена [10–12]. Известны работы по визуальному позиционированию точки посадки БПЛА [13], определения положения летательного аппарата на основе предварительно сделанных спутниковых снимков местности и нейросетевых распознавателей [14] и др. Общим недостатком предлагаемых подходов является необходимость дополнительной информации для позиционирования, а также высокие требования к вычислительной мощности оборудования на борту БПЛА.

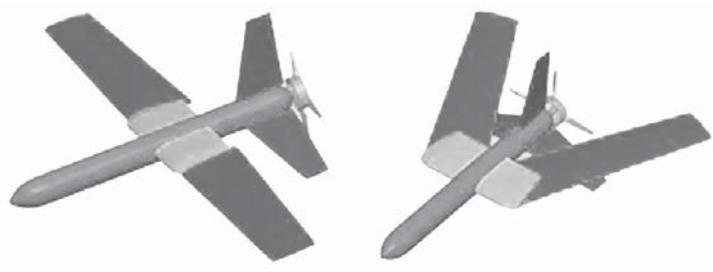


Рис. 1. Конструкция ныряющего самолета со складным крылом (источник — см. [16]).

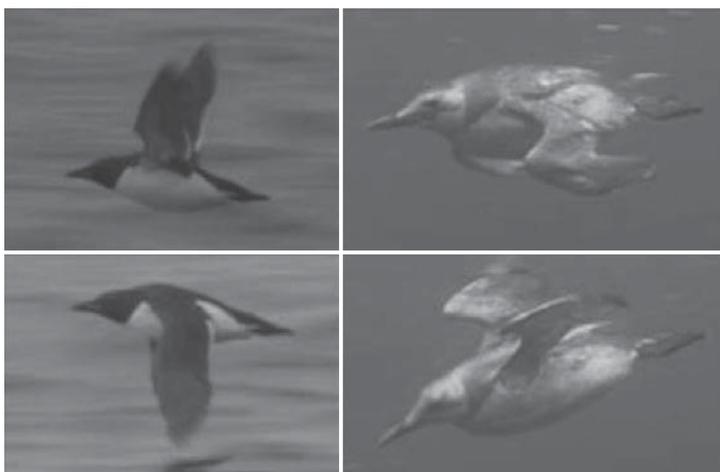


Рис. 2. Механика перемещения кайры в воздухе и под водой, примененная для имитации динамики БПЛА (источник — см. [18]).

Для беспилотных аппаратов, передвигающихся под водой, алгоритмы компьютерного зрения должны удовлетворять дополнительным требованиям из-за особенностей распространения света в воде и дополнительных помех из-за мутности естественных водоемов [15]. В случае БПЛА большое значение также приобретают скорость работы алгоритма из-за относительно высоких скоростей перемещения аппарата и необходимое для детектирования объекта количество вычислений, так как летающие подводные аппараты не способны нести на себе вычислители большой мощности и их энергопотребление высоко относительно АНПА. Поэтому задача разработки алгоритма визуального позиционирования для цели, заданной под водой, с минимальными энерго- и вычислительными затратами особенно актуальна именно для гибридных БПЛА.

Случай, когда конечное целевое положение аппарата находится под водой, более характерен для летательных подводных аппаратов типа «ныряющий самолет» [16], чем, например, для плавающих квадрокоптеров [17]. Самолет-

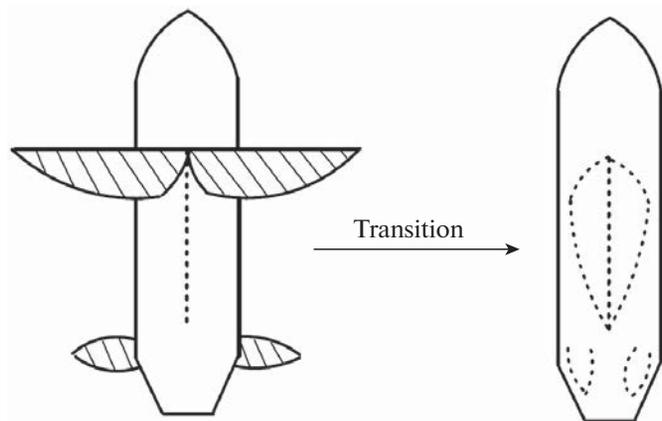


Рис. 3. Конфигурация оперения БПЛА в различных средах (источник — см. [19]).

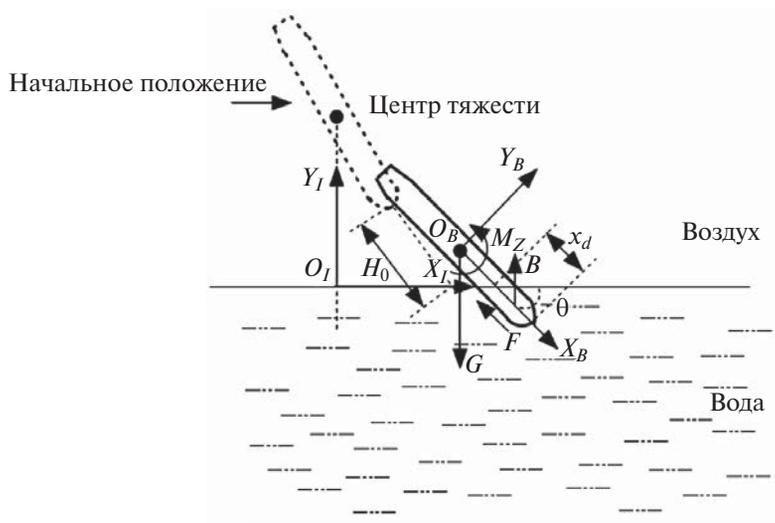


Рис. 4. Силы, влияющие на самолетоподобный БПЛА при нырянии (источник — см. [19]).

топодобные конструкции в основном приспособлены только к прямому переходу «воздух-вода» и испытывают значительные трудности либо вообще не способны осуществлять обратный переход «вода-воздух». В частности, в [16] предложена конструкция ныряющего самолета с отклоняющимися на 65 градусов при нырянии крыльями (см. рис. 1).

Конструкция включает в себя крыло изменяемой стреловидности и двигатель на сжатом углекислом газе. В [18] предложена конструкция ныряющего БПЛА, имитирующего полет и ныряние кайры (рис. 2).

В [19] описана система автоматического управления ныряющего самолета со складывающимся оперением (рис. 3).

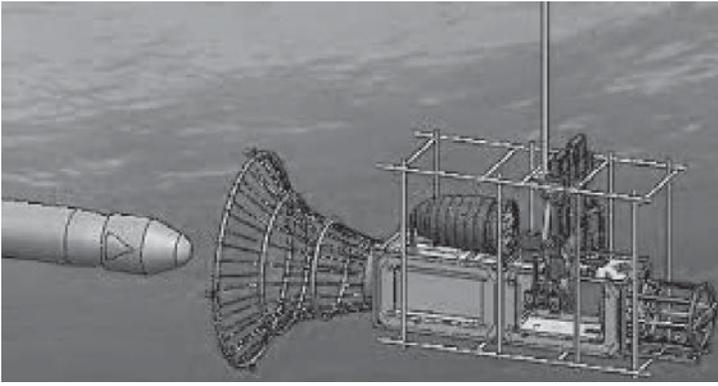


Рис. 5. Пример стыковки автономного необитаемого подводного аппарата с док-станцией (источник — см. [20]).

В [19] строится управление для достижения аппаратом конечной точки под водой (рис. 4).

Критерием оптимальности управления выбрана точность достижения заданной позиции (x_{target}, y_{target}) в конечный момент времени t_f .

$$(1) \quad J = \sqrt{(x_h(t_f) - x_{target})^2 + (y_h(t_f) - y_{target})^2}.$$

Управляющими параметрами выступают начальный угол тангажа θ_0 и начальная скорость вдоль оси OX v_{x_0} , расчет траектории производится без возможности маневрирования. В качестве целевых координат (x_{target}, y_{target}) авторы определяют конкретную точку под водой. Однако в случае адаптивного управления указанные координаты аппарат может вычислять самостоятельно в автоматическом режиме на основе машинного зрения. Таким образом, возникает задача распознавания координат некоего целевого объекта под водой, математический центр тяжести которого будет являться целевой точкой управления (x_{target}, y_{target}) . Данная задача является одним из возможных характерных примеров задачи распознавания в целях решения задачи оптимального управления.

Для подводных беспилотных аппаратов задание объектов интереса под водой является единственно возможным. Таким объектом интереса, в частности, может быть специальная стыковочная док-станция [20], используемая для подзарядки батарей АНПА или/и передачи данных (см. рис. 5).

Особенности позиционирования АНПА и БПЛПА в данном случае будут одни и те же, так как в обоих случаях цель является подводной. То обстоятельство, что цель находится именно под водой, является важным для разработки метода детектирования целевого объекта, так как замутненность воды, наличие взвеси, искажений изображения вследствие преломления света накладывают на метод детектирования жесткие требования по устойчивости к внешним шумам. Оптические методы позиционирования для подводных

целей описаны, например, в [21, 22]. При этом могут использоваться как одна [23], так и две [24] камеры. Основным ограничением предлагаемых методов является их высокое потребление вычислительных и энергетических ресурсов, так как практически все они предназначены для АНПА. Это делает предложенные методы слабоприменимыми для БПЛА. Так, в [20] в качестве математической модели распознавания используется сверточная нейронная сеть CNN. Эффективность подобных сетей высока, если они специальным образом обучаются для довольно узкого заданного класса объектов. Например, показано, что CNN способны распознавать рукописный текст с точностью до 99% [25]. Однако при распознавании произвольных реальных объектов их точность падает до 70% и ниже. Кроме того, CNN требуют мощных вычислительных устройств для обучения, которое может продолжаться от нескольких часов до нескольких суток [26]. Современные модификации могут давать либо более высокую точность, либо более высокую скорость, но пока не способны совмещать в себе оба этих качества. Так, в [27] описаны так называемые двухпроходные алгоритмы типа FASTER RCNN, достигающие точности 73,2% (датасет VOC07¹). Время обучения сети не описано, однако сеть обладает относительно невысокой скоростью распознавания в 17 изображений в секунду. Однопроходные сети типа YOLO достигают скорости распознавания в 155 изображений в секунду, однако их точность при этом всего лишь 52,7% на том же датасете VOC07.

Таким образом, задача разработки алгоритма обнаружения объектов под водой с минимальными вычислительными и временными затратами является актуальной, особенно для БПЛА.

2. Постановка задачи

Рассмотрим задачу распознавания подводного объекта в автоматическом режиме с борта беспилотного аппарата на примере обнаружения базовой док-станции со световыми маркерами в виде сигнального люминесцирующего кольца. Центр кольца задает целевую точку траектории оптимального управления без маневрирования (x_{target}, y_{target}) по примеру задачи (1). В общем случае задача может решаться одновременно для нескольких возможных целей N_{target} .

Дано:

- а) цветное изображение $I = (I_{ij})_{N_x \times N_y}$ в цветовой модели RGB;
- б) множество радиусов окружностей \mathbf{N} , которые требуется найти;
- в) множество допустимых цветов $\overline{\mathbf{C}}$, в оттенках которого могут окрашиваться искомые окружности;
- г) допустимое рассогласование между положениями центров окружностей Δ .

¹ VOC07 — набор данных свободного доступа, содержащий классифицированные изображения (Visual Object Classes-VOC) для 20 классов, редакция 2007 г.

Найти:

д) множество центров окружностей с их радиусами, определяющих терминальную точку оптимального управления $\tilde{A} = \{(x_{target}^k, y_{target}^k), r^k\}$, $k = \overline{1, N_{target}}$.

Задачей настоящего исследования является разработка вычислительного алгоритма, принимающего на вход позиции а)– г) и генерирующего на выходе позицию д). Дополнительным требованием к алгоритму является устойчивость к внешним шумам, а также минимизация вычислительных затрат и времени вычислений.

В случае, когда реальный размер окружности (например, стыковочного кольца) известен, в качестве искомым радиусов изображения следует выбрать ряд уменьшающихся значений от реального до минимально возможного радиуса. Впоследствии по найденному значению радиуса на изображении будет возможно определить расстояние до объекта с применением дополнительных измерительных приборов.

Метод должен обеспечивать распознавание при слабых деформациях окружности в эллипс с соотношением малой и большой полуоси не менее 0,95, что соответствует смещению прямой проекции обзора не более чем на 20° [28].

3. Метод решения задачи распознавания цветных объектов круглой формы

Важной задачей, решаемой средствами компьютерного зрения, является обработка изображений с борта беспилотного аппарата с целью обнаружения и локализации заданных объектов — световых маркеров. Данные объекты характеризуются формой и цветом. Поэтому для их обнаружения и распознавания применяются методы анализа цветовых признаков и признаков формы.

Популярные в последнее время алгоритмы на основе глубоких нейронных сетей для данной постановки задачи будут являться избыточными, так как распознаванию подлежит простая геометрическая фигура, а не сложный многоцветный объект. Кроме того, актуальными проблемами при использовании нейронных сетей остаются сложность в обобщении и длительность этапа обучения. Поэтому классические методы, в которых производится анализ изображения, остаются актуальными.

Многие применяющиеся на практике подходы к анализу формы объектов на изображениях базируются на методах анализа контуров [29]. Световые маркеры для стыковки подводных аппаратов являются цветными объектами, обладающими радиально-симметричной формой. Для их обнаружения и локализации можно применить методы, базирующиеся на анализе свойств радиальной симметрии. В основу предложенного метода легли совместное использование известного метода Хафа [30] и преобразование быстрой радиальной симметрии с использованием построения пирамиды Гаусса.

3.1. Преобразование быстрой радиальной симметрии

В данной работе предлагается эффективный подход к решению рассматриваемой задачи, основанный на применении быстрого преобразования радиальной симметрии FRST (Fast Radial Symmetry Transform). Преобразование FRST позволяет построить весовую модель изображения, с помощью анализа которой можно эффективно локализовать центр радиально-симметричного объекта [31, 32].

Для получения весовой модели изображения необходимо выполнить следующие шаги:

- 1) преобразовать изображение в полутоновый вид;
- 2) вычислить градиенты функции яркости на изображении;
- 3) вычислить значения элементов матриц весов;
- 4) нормировать значения элементов матриц весов;
- 5) вычислить значения элементов матриц обобщенных весов;
- 6) выполнить низкочастотную фильтрацию матриц обобщенных весов;
- 7) вычислить значения элементов матрицы усредненных весов.

На первом шаге осуществляется преобразование изображения в полутоновый вид, в котором каждому пикселю сопоставляется значение яркости. В простейшем случае преобразование заключается в вычислении средних значений интенсивностей красной, зеленой и синей цветовых составляющих. Таким образом, в результате выполнения первого шага изображение представляется в виде функции яркости, аргументами которой являются значения координат пикселей.

На втором шаге выполняется вычисление градиентов функции яркости на изображении с помощью следующих операторов:

$$(2) \quad g_x(\mathbf{p}) := \hat{\mathbf{I}}(i+1, j) - \hat{\mathbf{I}}(i, j);$$

$$(3) \quad g_y(\mathbf{p}) := \hat{\mathbf{I}}(i, j+1) - \hat{\mathbf{I}}(i, j);$$

$$(4) \quad |\mathbf{g}(\mathbf{p})| = \sqrt{g_x^2(\mathbf{p}) + g_y^2(\mathbf{p})},$$

где $\mathbf{g}(\mathbf{p}) = (g_x(\mathbf{p}), g_y(\mathbf{p}))$ — градиент в пикселе \mathbf{p} с координатами (i, j) ; $g_x(\mathbf{p})$, $g_y(\mathbf{p})$ — составляющие градиента для горизонтального и вертикального направлений в пикселе \mathbf{p} соответственно; $\hat{\mathbf{I}}(i, j)$, $\hat{\mathbf{I}}(i+1, j)$, $\hat{\mathbf{I}}(i, j+1)$ — значения яркости в пикселях полутонового изображения с координатами (i, j) , $(i+1, j)$ и $(i, j+1)$ соответственно. В выражениях (2)–(4) знаком $:=$ обозначен оператор присваивания, который будет использоваться и в дальнейшем.

На третьем шаге вычисляются значения элементов матриц весов. Для этого применяется следующая процедура:

- 1) определяется множество целочисленных значений \mathbf{N} , где \mathbf{N} — множество радиусов объектов, которые требуется найти;

2) для каждого значения n из множества \mathbf{N} формируются начальные значения элементов двух матриц весов:

$$(5) \quad \mathbf{M}^n(\mathbf{p}) := 0;$$

$$(6) \quad \mathbf{O}^n(\mathbf{p}) := 0,$$

где p — вектор координат (i, j) .

3) для всех элементов матриц весов, определяемых значениями n из множества \mathbf{N} , применяются следующие операторы:

$$(7) \quad \mathbf{p}_+ := \mathbf{p} + \left\lceil \frac{g_x(\mathbf{p})}{|\mathbf{g}(\mathbf{p})|} n \right\rceil;$$

$$(8) \quad \mathbf{p}_- := \mathbf{p} - \left\lceil \frac{g_x(\mathbf{p})}{|\mathbf{g}(\mathbf{p})|} n \right\rceil;$$

$$(9) \quad \mathbf{M}^n(\mathbf{p}_+) := \mathbf{M}^n(\mathbf{p}_+) + |\mathbf{g}(\mathbf{p})|;$$

$$(10) \quad \mathbf{M}^n(\mathbf{p}_-) := \mathbf{M}^n(\mathbf{p}_-) - |\mathbf{g}(\mathbf{p})|;$$

$$(11) \quad \mathbf{O}^n(\mathbf{p}_+) := \mathbf{O}^n(\mathbf{p}_+) + 1;$$

$$(12) \quad \mathbf{O}^n(\mathbf{p}_-) := \mathbf{O}^n(\mathbf{p}_-) - 1,$$

где $\lceil \cdot \rceil$ — оператор округления вещественного числа до ближайшего не меньшего целого значения.

На четвертом шаге элементы матриц весов нормируются:

$$(13) \quad \mathbf{M}^n(\mathbf{p}) := \frac{|\mathbf{M}^n(\mathbf{p})|}{\max_q \{|\mathbf{M}^n(\mathbf{q})|\}};$$

$$(14) \quad \mathbf{O}^n(\mathbf{p}) := \frac{|\mathbf{O}^n(\mathbf{p})|}{\max_q \{|\mathbf{O}^n(\mathbf{q})|\}}.$$

На пятом шаге вычисляются значения элементов матриц обобщенных весов:

$$(15) \quad \mathbf{F}^n(\mathbf{p}) := (\mathbf{O}^n(\mathbf{p}))^\alpha \mathbf{M}^n(\mathbf{p}),$$

где α — параметр радиальной жесткости.

На шестом шаге выполняется низкочастотная фильтрация матриц обобщенных весов. При этом, как правило, используется гауссов фильтр нижних частот. Фильтрация выражается с помощью оператора свертки

$$(16) \quad \mathbf{S}^n := \mathbf{F}^n * \mathbf{G}^n,$$

где \mathbf{G}^n — матрица коэффициентов гауссова фильтра нижних частот, определенная для значения n из множества \mathbf{N} .

На последнем, седьмом, шаге вычисляются значения элементов матрицы усредненных весов:

$$(17) \quad \mathbf{S}(\mathbf{p}) := \frac{1}{|\mathbf{N}|} \sum_{n \in \mathbf{N}} \mathbf{S}^n(\mathbf{p}).$$

Полученная в результате преобразования FRST матрица усредненных весов представляет собой весовую модель изображения, анализ которой позволяет определить параметры формы объектов, например координаты центра и радиус объектов круглой формы.

3.2. Весовая модель изображения

При построении весовой модели изображения с использованием быстрого преобразования радиальной симметрии необходимо выполнить относительно большое количество вычислений, которое определяется мощностью множества целочисленных значений \mathbf{N} . Элементы указанного множества применяются для анализа существенных изменений яркости на соответствующих расстояниях от текущего пикселя, т.е. имеют смысл радиусов для радиально-симметричных объектов с центрами в пикселях изображения.

Объекты на изображениях реального мира имеют многомасштабную природу, поскольку обладают различными размерами. Для рассматриваемой задачи это означает использование множества \mathbf{N} с большим количеством элементов, которые определяются диапазоном изменения радиусов объектов интереса. Поэтому целесообразно проводить анализ исходного изображения, представленного в многомасштабном виде. Такой многомасштабный анализ позволяет сократить количество вычислений и повысить точность обнаружения объектов на изображении.

Для построения многомасштабной весовой модели изображения необходимо выполнить следующие шаги:

- 1) преобразовать изображение в полутоновый вид;
- 2) построить пирамиду Гаусса;
- 3) вычислить значения элементов матриц усредненных весов для всех уровней пирамиды Гаусса;
- 4) привести матрицы усредненных весов к размерам исходного изображения;
- 5) вычислить значения элементов интегральной матрицы усредненных весов.

На первом шаге осуществляется преобразование изображения в полутоновый вид, в котором каждому пикселю сопоставляется значение яркости.

На втором шаге строится пирамида Гаусса [33]. Она представляет собой множество изображений $\mathbf{P} = \{\mathbf{P}_l \mid l = \overline{0, L-1}\}$, где L — число уровней в пирамиде. В качестве нулевого уровня пирамиды \mathbf{P}_0 рассматривается исходное изображение в полутоновом виде $\hat{\mathbf{I}}$. Остальные уровни пирамиды форми-

руются следующим образом:

$$(18) \quad \mathbf{P}_l := \mathbf{G} * (2 \downarrow [\mathbf{P}_{l-1}]),$$

где \mathbf{G} — матрица коэффициентов гауссова фильтра нижних частот; $2 \downarrow [\cdot]$ — оператор понижения дискретизации изображения, например, посредством удаления каждого второго пикселя в строке и столбце. Элемент пирамиды Гаусса уровня l будет иметь размеры в четыре раза меньше по сравнению с размерами элемента уровня $l - 1$.

На третьем шаге с помощью операторов (5)–(17) формируются весовые модели изображений всех уровней пирамиды Гаусса, т.е. строится множество матриц усредненных весов $\{\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(L-1)}\}$.

На четвертом шаге множество $\{\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(L-1)}\}$ преобразуется в множество $\{\mathbf{R}^{(0)}, \mathbf{R}^{(1)}, \dots, \mathbf{R}^{(L-1)}\}$, каждый элемент которого $\mathbf{R}^{(l)}$ представляет собой результат приведения матрицы $\mathbf{S}^{(l)}$ к размерам исходного изображения, т.е. к размерам $\mathbf{S}^{(0)}$. Указанное преобразование выполняется следующим образом:

$$(19) \quad \mathbf{R}^{(l)} := \mathbf{G} * (2 \uparrow [\mathbf{S}^{(l)}]),$$

где \mathbf{G} — матрица коэффициентов гауссова фильтра нижних частот; $2 \uparrow [\cdot]$ — оператор повышения дискретизации изображения (матрица $\mathbf{S}^{(l)}$ рассматривается как изображение, с каждым пикселем которого ассоциируется весовое значение), например, посредством дублирования каждого пикселя в строке и столбце.

На последнем, пятом, шаге осуществляется сложение полученных матриц $\mathbf{R}^{(0)}, \mathbf{R}^{(1)}, \mathbf{R}^{(L-1)}$ для построения интегральной матрицы усредненных весов:

$$(20) \quad \mathbf{S}(\mathbf{p}) := \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{R}^{(l)}(\mathbf{p}),$$

где \mathbf{p} — пиксель исходного изображения, соответствующий координатам элементов матриц весов.

В результате выполнения рассмотренных шагов формируется многомасштабная весовая модель, анализ которой позволяет определить параметры формы объектов на изображении. Следует отметить, что применение данной модели позволяет повысить точность работы и снизить вычислительную сложность за счет меньшей площади изображения на верхнем уровне пирамиды.

3.3. Хроматическая модель изображения

При анализе изображений световых маркеров с беспилотного подводного аппарата целесообразно использовать модель, содержащую две составляющие — хроматическую и весовую. Первая позволяет учесть цветовые характеристики объектов интереса, а вторая — геометрические особенности их формы.

Тогда модель изображения может быть представлена в виде $\langle \mathbf{C}, \mathbf{S} \rangle$, где \mathbf{C} — хроматическая составляющая, а \mathbf{S} — весовая составляющая. Весовая составляющая модели формируется посредством процедур, описанных в предыдущих разделах. Хроматическая составляющая формируется в результате цветового преобразования и цветовой сегментации.

Для описания хроматической составляющей модели были исследованы преимущества и недостатки нескольких цветовых пространств: RGB, Lab и HSV. Данные пространства сравнивались с точки зрения способов описания для выделения областей изображения, цвет которых лежит в некотором диапазоне [34]. В результате было отмечено, что основным фактором, ограничивающим и усложняющим применение цветовой модели RGB для проведения цветовой сегментации, является единое описание как освещенности, так и цветовой составляющей объекта, что затрудняет четкое задание диапазонов искомых цветов [35]. Неоспоримым преимуществом цветовой модели $L^*a^*b^*$ является отдельный канал L^* , отвечающий за освещенность, и плоскостей (a^* , b^*), отвечающих за цветность. Однако в этой модели диапазоны цветов (хроматические характеристики) описываются в виде геометрической фигуры на плоскости, возможно, имеющей сложное строение. Данный факт усложняет построение модели распределения цвета для проведения сегментации [36]. Цветовая модель HSV в отличие от $L^*a^*b^*$ позволяет задавать диапазон искомых цветов путем указания интервала значений канала H (в большинстве случаев достаточно указать пороговое значение насыщенности, задаваемой в канале S , выше значения, которого чистота цвета становится приемлемой), что значительно упрощает реализацию метода [37].

Таким образом, в качестве цветового пространства для анализа цветных изображений было выбрано цветовое пространство (модель) HSV. Данная модель базируется на особенностях восприятия информации зрительной системой человека и представляется как совокупность цветовых каналов H , S и V , которые определяют тон, насыщенность и значение (яркость) цвета соответственно.

Хроматическая составляющая \mathbf{C} модели изображения представляется следующим образом:

$$(21) \quad \begin{cases} (h(\mathbf{p}), s(\mathbf{p}), v(\mathbf{p})), & \exists k, l, m : h(\mathbf{p}) \in \mathbf{H}^k \wedge s(\mathbf{p}) \in \mathbf{S}^l \wedge v(\mathbf{p}) \in \mathbf{V}^m; \\ (0, 0, 0), & \forall k, l, m : h(\mathbf{p}) \notin \mathbf{H}^k \vee s(\mathbf{p}) \notin \mathbf{S}^l \vee v(\mathbf{p}) \notin \mathbf{V}^m, \end{cases}$$

где \mathbf{p} — пиксель исходного изображения; \mathbf{H}^k , \mathbf{S}^l , \mathbf{V}^m — k -й, l -й и m -й интервалы каналов H , S и V соответственно из множества интервалов цвета заданных объектов.

3.4. Алгоритм обнаружения объектов круглой формы на изображениях световых маркеров

В общем виде процедура обнаружения объектов круглой формы заданного цвета заключается в следующем:

- 1) преобразовать исходное изображение из цветовой модели RGB в цветовую модель HSV;
- 2) построить хроматическую составляющую модели изображения;
- 3) преобразовать исходное изображение в полутоновый вид с учетом хроматической составляющей;
- 4) найти окружности методом Хафа на изображении по хроматической составляющей;
- 5) построить весовую составляющую модели изображения (весовое изображение);
- 6) выполнить пороговую обработку весового изображения;
- 7) найти на весовом изображении контуры;
- 8) вычислить центры масс для каждого контура;
- 9) найти окружности на изображении по весовой составляющей;
- 10) удалить ложно найденные на полутоновом изображении методом Хафа окружности, центры которых не совпадают с допустимой погрешностью с каким-либо центром масс контуров на весовом изображении;
- 11) сформировать множество координат целевых точек управления с указанием радиусов $\tilde{A} = \{(x_{target}^k, y_{target}^k), r^k\}$.

4. Вычислительные эксперименты

Представленный алгоритм был опробован на задаче распознавания круговых светящихся маркеров док-станции подводного базирования под разными углами в условиях замутненной воды. Для вычислительных экспериментов использованы изображения док-станции, полученные в ходе экспериментов Liu S., Ozay M., Okatani T., опубликованных в [20]. Примеры изображений стыковочного блока док-станции в виде подсвеченных кругов, которые распознавались в результате экспериментов, представляет рис. 6.

Изображения получены монокулярной цветной камерой NanoSeaCam разрешением 620 ТВЛ (0,8 Мпикс) с углами обзора $59^\circ \times 44^\circ \times 72^\circ$ (вверх-вниз-в сторону).

4.1. Критерий оценки точности алгоритма

Оценка эффективности разработанного алгоритма проводилась на основе так называемой пиксельно-ориентированной методики, описанной в [38].

Это статистическая мера оценки, основанная на подсчете неправильно классифицированных пикселей. Для ее проведения требуется вычислить вероятности того, что случайно выбранный пиксель на отсегментированном с помощью алгоритма изображении принадлежит либо искомому объекту, либо, соответственно, фону.

В отличие от таких общепринятых метрик, как Accuracy, Precision или Recall с последующим ROC-анализом, пиксельно-ориентированная метрика оценивает не точность классификации объекта (объект обнаружен или не обнаружен), а качество обнаружения. Оценка качества обнаружения включает

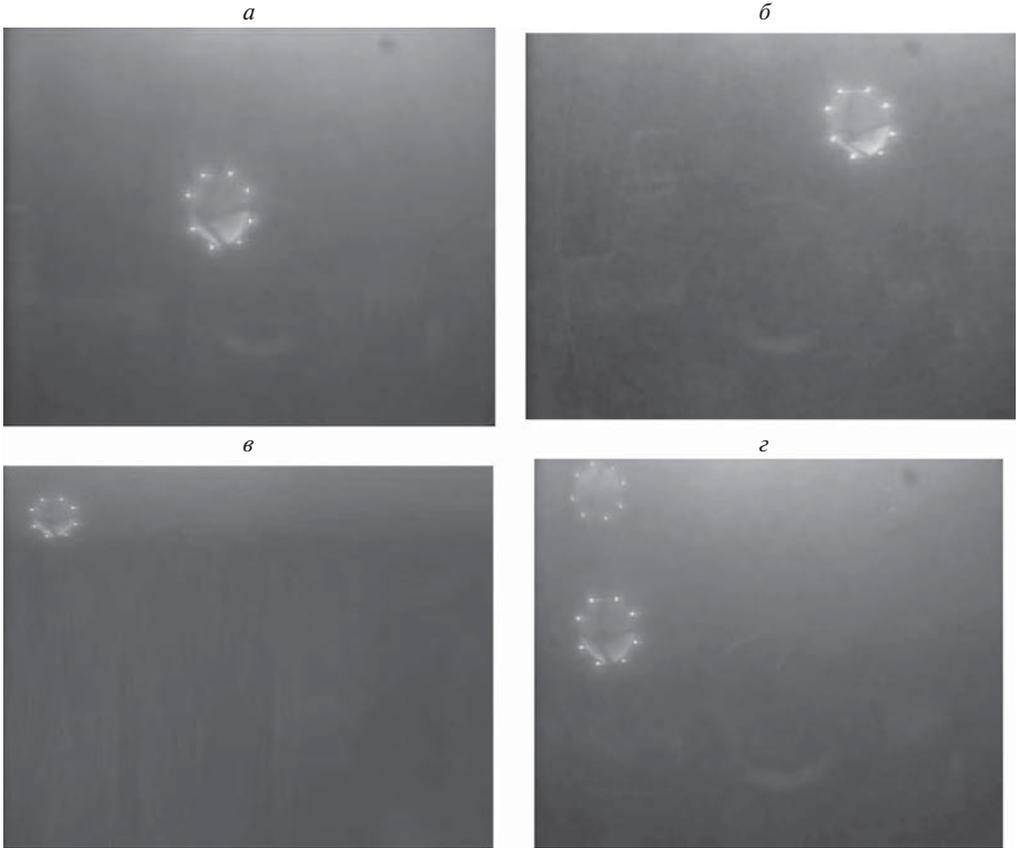


Рис. 6. Варианты изображений окружностей — стыковочного люминесцирующего кольца подводной док-станции: *а)* одиночный объект в центральной части изображения, площадь объекта интереса относительно изображения более 5%; *б)* одиночный объект, смещенный относительно центра изображения, площадь объекта интереса относительно изображения более 5%; *в)* одиночный объект, смещенный относительно центра изображения, площадь объекта интереса относительно изображения менее 5%; *г)* несколько объектов интереса на одном изображении. (источник: датасет в свободном доступе <http://vision.is.tohoku.ac.jp/liushuang/a-vision-based-underwater-docking-system/dataset>).

как оценку точности локализации объекта (определение координат центра), так и точности определения его геометрических характеристик, т.е. радиуса. Поэтому для поставленной задачи пиксельно-ориентированная оценка является предпочтительной [39].

Согласно выбранному методу вероятность ошибки сегментации всего изображения определяется следующим образом:

$$(22) \quad p_{err} = p(b|o)p(o) + p(o|b)p(b).$$

Здесь $p(o)$ — априорная вероятность того, что случайным образом выбранный пиксель исходного изображения принадлежит объекту (отношение площади

Таблица 1. Исходные данные и предварительные результаты экспериментов по распознаванию тестовых изображений

№ эксперимента	Априорные данные				Данные, полученные после применения алгоритма	
	$p(o)$	$p(b)$	Относительная площадь объекта на изображении, %	Оценка относительной площади объекта	Процент ошибочных распознаваний алгоритма по критерию $p(b o)$	Процент ошибочных распознаваний алгоритма по критерию $p(o b)$
1	0,01	0,99	1%	$\leq 5\%$	1	1
2	0,01	0,99	1%	$\leq 5\%$	1	5
3	0,01	0,99	1%	$\leq 5\%$	1	10
4	0,01	0,99	1%	$\leq 5\%$	1	25
5	0,05	0,95	5%	$\leq 5\%$	1	1
6	0,05	0,95	5%	$\leq 5\%$	1	5
7	0,05	0,95	5%	$\leq 5\%$	1	10
8	0,05	0,95	5%	$\leq 5\%$	1	25
9	0,1	0,9	10%	$> 5\%$	1	1
10	0,1	0,9	10%	$> 5\%$	1	5
11	0,1	0,9	10%	$> 5\%$	1	10
12	0,1	0,9	10%	$> 5\%$	1	25
13	0,25	0,75	25%	$> 5\%$	1	1
14	0,25	0,75	25%	$> 5\%$	1	5
15	0,25	0,75	25%	$> 5\%$	1	10
16	0,25	0,75	25%	$> 5\%$	1	25
17	0,5	0,5	50%	$> 5\%$	1	1
18	0,5	0,5	50%	$> 5\%$	1	5
19	0,5	0,5	50%	$> 5\%$	1	10
20	0,5	0,5	50%	$> 5\%$	1	25

объекта к площади всего изображения), $p(b)$ — априорная вероятность того, что случайным образом выбранный пиксель исходного изображения принадлежит фону (отношение площади фона к площади всего изображения), $p(o|b)$ — вероятность того, что пиксель, принадлежащий фону, при сегментации ошибочно отнесен к объекту (отношение части площади фона, ошибочно отнесенной алгоритмом к объекту, ко всей площади фона). В теории статистических гипотез такая ошибка называется ошибкой первого рода, $p(b|o)$ — вероятность того, что пиксель, принадлежащий объекту, будет ошибочно отнесен к фону (отношение части площади объекта, ошибочно отнесенной алгоритмом к фону, к площади объекта) — ошибка второго рода.

Таким образом, $p(o|b)$ и $p(b|o)$ являются совокупными ошибками сегментации.

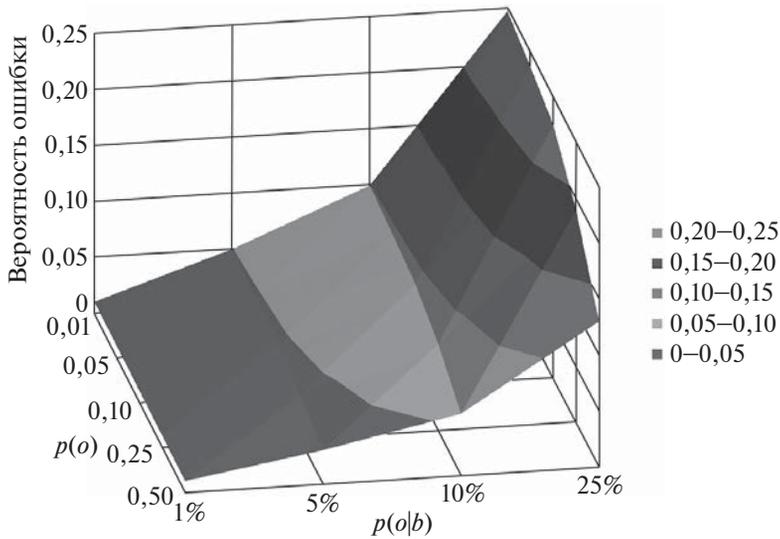


Рис. 7. Вероятность ошибки алгоритма, рассчитанной по способу (22) в зависимости от априорной вероятности $p(o)$ и апостериорной вероятности $p(b|o)$.

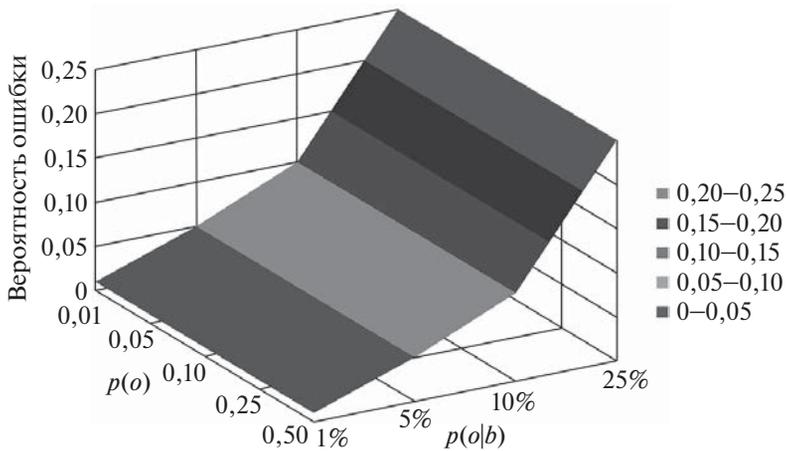


Рис. 8. Вероятность ошибки алгоритма, рассчитанной по способу (23) в зависимости от априорной вероятности $p(o)$ и апостериорной вероятности $p(b|o)$.

Площадь в данном случае понимается как количество пикселей заданной области.

Для случаев, когда искомые объекты малы по размерам относительно всего изображения (занимают менее 5% площади изображения), т.е.

$$\#(o) \ll \#(\mathcal{D}),$$

где $\#(o)$ — мощность множества пикселей, принадлежащих искомому объекту, а $\#(\mathcal{D})$ — мощность множества пикселей всего изображения, вместо

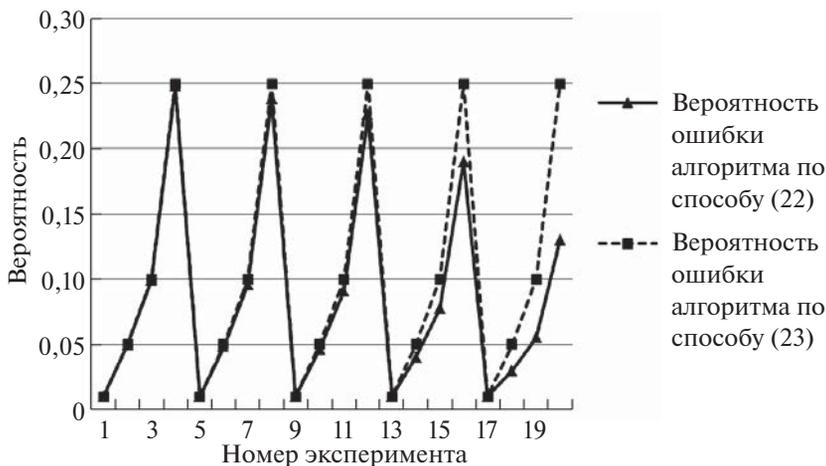


Рис. 9. Сравнительная эффективность способов оценки точности алгоритма.

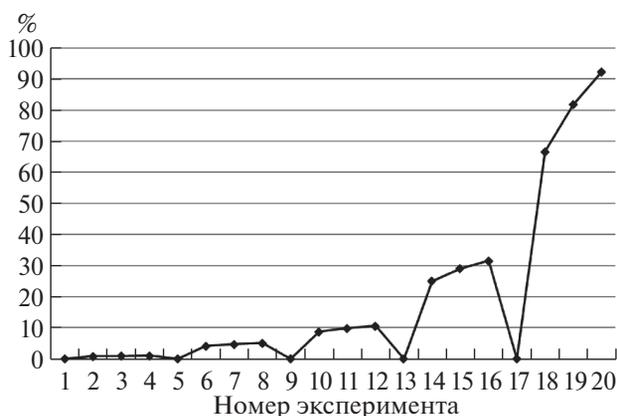


Рис. 10. Разница в точности способа (22) относительно способа (23).

формулы (22) целесообразно применять оценку (23):

$$(23) \quad p_{err} := \frac{\#(o_s) - \#(o)}{\#(\mathcal{D})},$$

где o_s — множество найденных объектов в результате сегментации.

Для определения оптимального способа расчета точности алгоритма было проведено 20 независимых экспериментов по распознаванию кругов на изображениях. Среди них на 8 изображениях объект интереса занимал 5 и менее процентов площади, а на 12 изображениях — более 5%. После обработки изображений разработанным алгоритмом были получены апостериорные значения ошибок распознавания $p(o|b)$ и $p(b|o)$. Априорные и апостериорные данные экспериментов представлены в табл. 1.

Затем на основе полученных данных вычислялась точность согласно способам (22) и (23). Результаты представлены на рис. 7–10.

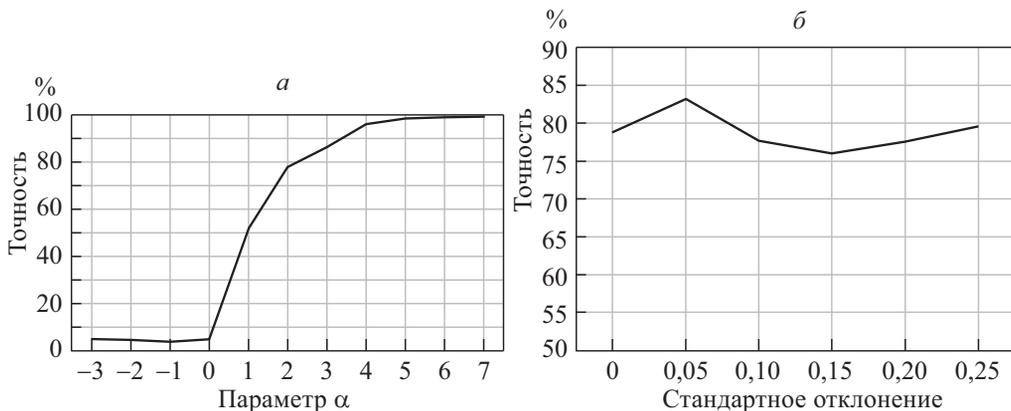


Рис. 11. Влияние параметров алгоритма на точность обнаружения окружностей: а) параметра радиальной жесткости; б) стандартного отклонения.

Результаты показывают, что для объектов с площадью до 5% от площади всего изображения включительно оценки корректности, рассчитанные по обоим способам при зафиксированном уровне ошибки 2-го рода на величине 0,01, в целом имеют небольшие различия. Однако если искомый объект занимает площадь более 5% от площади всего изображения, целесообразно применять оценку по формуле (22). Таким образом, способ (22) является предпочтительным.

Для всех дальнейших исследований точность вычислялась согласно (22).

4.2. Влияние параметров преобразования радиальной симметрии

Исследована зависимость точности обнаружения окружностей от параметра радиальной жесткости α . Было сгенерировано и исследовано 100 изображений окружностей с радиусами от 1 до 100 пикселей на исходном изображении 2400×8400 пикселей. Были проведены эксперименты с изменением параметра жесткости от $\alpha = -3$ до $\alpha = +7$. Результаты представлены на рис. 11,а. Видно, что при увеличении параметра радиальной жесткости α до значения +2 точность обнаружения достигает 75% (0,75). При дальнейшем увеличении α точность 95% достигается при значении +4 и далее асимптотически стремится к 100%. На рис. 11,б показано влияние стандартного отклонения на точность обнаружения окружностей при значении параметра радиальной жесткости α , равном 2. Видно незначительное влияние этого параметра на результат при данных условиях.

4.3. Влияние шумов на точность распознавания

Замутненность воды, наличие в ней мелких органических и неорганических включений, искажения, вызванные преломлением света между водой и объективом камеры, с математической точки зрения описываются как адди-

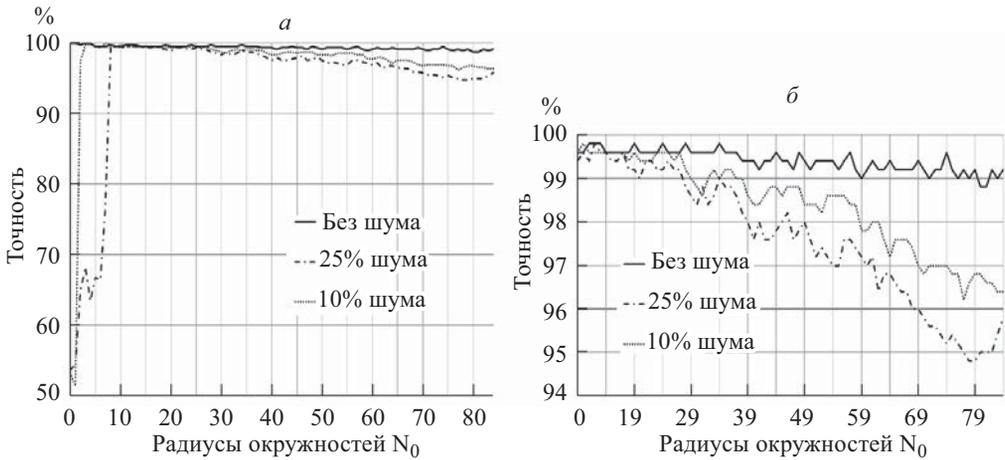


Рис. 12. Влияние импульсного шума на точность распознавания окружностей различного радиуса: а) для всех окружностей; б) для окружностей с радиусом более 10 пикселей.

тивные шумы. В частности, наличие мелкой твердой взвеси в воде моделируется как импульсный шум типа «соль» и «перец». Эффекты искажения и замутненности моделировались как гауссовы шумы и как шумы типа «размытие» с применением масок. Влияние на распознавание каждого типа шумов рассмотрено ниже.

4.3.1. Влияние импульсного шума

Для определения влияния импульсного шума на работу алгоритма было исследовано 84 изображения окружностей с радиусами от 1 до 84 пикселей. Исходное изображение было зашумлено помехами с интенсивностью 25 и 10% площади изображения. Результаты работы алгоритма приведены на рис. 12.

Видно, что импульсный шум значительное влияние оказывает на объекты малых размеров, на результат распознавания более крупных объектов его влияние менее значимо.

4.3.2. Влияние шума Гаусса

Для исследования влияния гауссового шума на результат работы алгоритма исходное изображение было зашумлено с помощью заполнения пикселей изображения нормально распределенными случайными числами с математическим ожиданием $\mu = 0$ и стандартным отклонением $\sigma = 5$. Затем изображение было подвергнуто обработке детектором радиальной симметрии с параметром радиальной жесткости α , равным 2. Результаты представлены на рис. 13.

Гауссов шум оказывает значительное влияние на объекты малых размеров, для крупных объектов влияние такого рода шума несущественно.

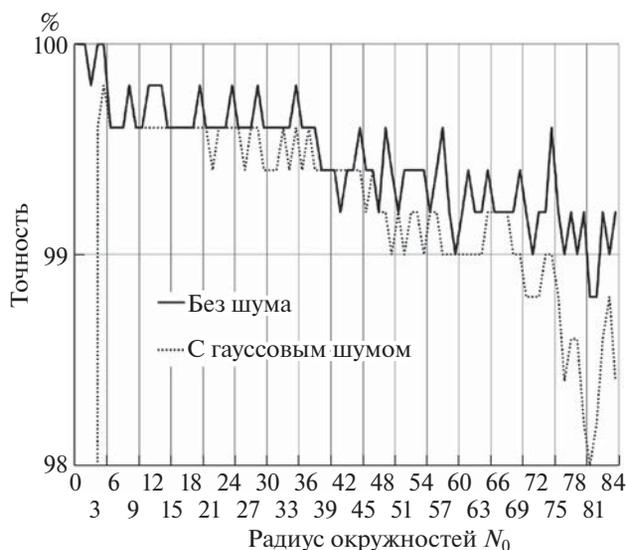


Рис. 13. Влияние гауссового шума на точность распознавания окружностей различного радиуса.

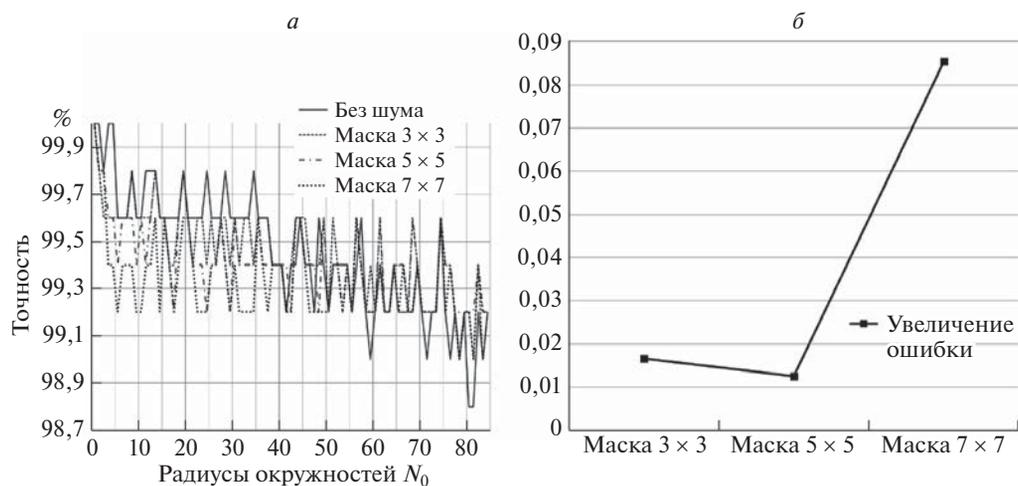


Рис. 14. Влияние размытия изображения масками на точность распознавания кругов: а) сравнительная точность для различных масок размытия; б) относительное среднее увеличение погрешности распознавания для различных масок по сравнению с неискаженным изображением.

4.3.3. Влияние шума типа «размытие»

Для исследования влияния шума типа «размытие» исходное изображение было подвергнуто размытию масками размерности 3×3 , 5×5 и 7×7 . Результаты изменения точности распознавания окружностей представляет рис. 14.

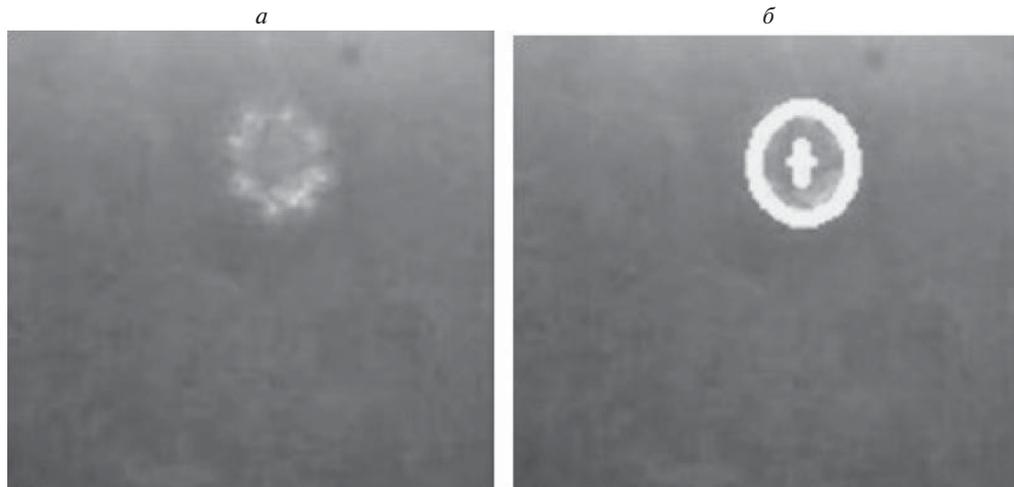


Рис. 15. Пример обнаружения светового маркера: *а)* исходное изображение маркированного кольца док-станции под углом; *б)* распознанная окружность с рассчитанным центром.

Размытие практически не влияет на результат обнаружения объектов, потеря точности составляет не более 0,09%.

Интересным фактом является повышение точности распознавания при наличии шума типа «размытие» по сравнению с незашумленным изображением в нескольких экспериментах. Это связано с проблемой ложного обнаружения, когда алгоритм ошибочно относит к найденным окружностям заданного радиуса также близко расположенные окружности других, близких по значению, радиусов. При наложении шума типа «размытие» вероятность такого ложноположительного обнаружения в некоторых случаях уменьшается.

Полученные результаты исследования влияния шумов различного типа на точность распознавания радиальных изображений показывают, что дополнительно применять алгоритмы повышения резкости не требуется, так как разработанный метод является достаточно устойчивым к шумам в рамках решаемой задачи.

4.4. Детектирование изображения

Изображение светового стыковочного кольца в замутненной воде относится к изображениям с высокой степенью искажения размытием, в меньшей степени — гауссовым и импульсным шумами. Таким образом, для объектов с радиусом более 10 пикселей (крупные объекты) разработанный алгоритм обеспечивает высокую устойчивость ко всем видам шумов, а по основному искажающему критерию — размытию — обеспечивает устойчивость для распознавания подводного стыковочного люминесцирующего кольца всех размеров.

Пример изображения светового маркера в замутненной воде под углом к камере 16° с распознанным центром для стыковки представлен на рис. 15.

В ходе численных экспериментов были распознаны как сами окружности, так и найдены их центры, координаты которых потенциально могут служить координатами конечного оптимального положения беспилотного аппарата при автоматическом управлении. Для изображения, представленного на рис. 15, в частности, имеем: параметры исходного изображения:

- а) цветное изображение размерностью 448×448 пикселей в формате jpeg;
- б) количество искомых окружностей — 1, искомый радиус — 85 пикселей;
- в) множество допустимых цветов — оттенки желтого;
- г) допустимое рассогласование между положениями центров окружностей — 5 пикселей;

найдена окружность с центром в точке:

- $x_{target} = 282$;
- $y_{target} = 298$.

4.5. Сравнение эффективности алгоритмов

В данном разделе проведено сравнение точности и быстродействия разработанного алгоритма с алгоритмами Хафа и быстрой радиальной симметрии, используемыми в отдельности. Такие методы, как, например, нейросетевое распознавание, требуют дополнительных вычислительных затрат и затрат памяти вычислительного устройства на этапе обучения. В первую очередь это связано с требованием наличия релевантного набора данных для обучения, включающих все возможные сочетания входных параметров для получения адекватной модели. Данное обстоятельство делает применение подобных алгоритмов машинного обучения малоприменимыми для низкопроизводительных вычислительных устройств, установленных на борт БПЛА или подобных беспилотных аппаратов.

4.5.1. Сравнение точности обнаружения радиальных объектов

Для иллюстрации эффективности метода обнаружения радиальных объектов на изображениях на основе предложенного алгоритма были проведены вычислительные эксперименты для двух типов изображений окружностей — с равномерным и случайным изменением радиусов.

В табл. 2 и 3 приведены результаты сравнения точности алгоритмов при варьировании минимального расстояния между центрами окружностей. Изображение считается правильно распознанным, если величина отклонения найденных центров от реальных не превосходит 15 пикселей.

Результаты сравнения показывают, что точность разработанного алгоритма в среднем на 35% выше точности классических методов, используемых в отдельности.

Таблица 2. Сравнение точности распознавания для радиальных объектов с равномерным увеличением радиусов

Минимальное расстояние между центрами окружн., пикс.	Реальное кол-во окружностей на изображении	Найдено по методу Хафа	Найдено по разработанному алгоритму
5	7	21	8
10	7	7	7
15	7	7	7
20	7	7	7
25	7	7	7
30	7	7	7
35	7	7	7
40	6	7	6
45	7	8	8
50	6	6	6

Таблица 3. Сравнение точности распознавания для радиальных объектов со случайным изменением радиусов

Минимальное расстояние между центрами окружн., пикс.	Реальное кол-во окружностей на изображении	Найдено по методу Хафа	Найдено по разработанному алгоритму
5	8	152	12
10	8	53	8
15	8	30	8
20	8	22	8
25	8	14	8
30	8	10	8
35	8	10	8
40	8	10	8
45	8	8	8
50	8	6	8

4.5.2. Сравнение скорости обнаружения радиальных объектов

Для относительного сравнения скорости разработанного алгоритма был использован компьютер со следующими характеристиками:

- процессор: Intel Core i5-3230M 2.60 GHz;
- ОЗУ: 8 Гб;
- операционная система: Windows 10, разрядность 64.

Эксперименты проводились для трех тестовых наборов изображений: 2 полных датасета «LISA Traffic Light Dataset» и «Bosch Small Traffic Lights Dataset» набора датасетов открытого типа Kaggle, а также 200 выборочных изображений из различных датасетов набора Kaggle, релевантных проводи-

Таблица 4. Характеристики датасетов изображений для экспериментов по сравнению скорости алгоритмов

Имя датасета	Краткое описание
LISA Traffic Light Dataset	100 изображений из набора размером 1280×960 в формате JPG
Bosch Small Traffic Lights Dataset	200 изображений из набора различных размеров в формате PNG
Выборочный набор из Kaggle	200 изображений из набора различных размеров в формате PNG

Таблица 5. Сравнение скорости распознавания радиальных объектов различных диапазонов окружностей

Диапазон окружностей для распознавания	Время работы алгоритма FRST, мсек	Время работы разработанного алгоритма, мсек	Выигрыш разработанного алгоритма по скорости, раз
1–10	89 988	50 569	1,8
1–20	171 902	79 753	2,1
1–30	253 435	111 020	2,3
1–50	465 673	128 970	3,6

тому исследованию. Описание тестового набора изображений из этих баз приведено в табл. 4.

Были проведены вычислительные эксперименты по распознаванию окружностей в четырех диапазонах: от 1 до 10 пикселей, от 1 до 20 пикселей, от 1 до 30 пикселей и от 1 до 50 пикселей. В табл. 5 приведены результаты оценки быстродействия предложенного алгоритма многомасштабной модели изображения окружностей по сравнению с базовым методом на основе преобразования радиальной симметрии FRST.

Очевидно, что с ростом диапазона искомых окружностей возрастает преимущество в скорости работы предлагаемого метода по сравнению с базовым.

Необходимо заметить, что использованный для проведения экспериментов компьютер обладал избыточной мощностью для практической реализации предложенного алгоритма и был использован исключительно для целей сравнения авторского алгоритма с классическим. В реальных условиях алгоритму достаточно мощности бортовых вычислителей БПЛПА за счет простоты реализуемых вычислений.

5. Заключение

Предложенный подход может быть использован для детектирования и локализации объектов заданных формы и цвета на изображениях световых маркеров для обнаружения объекта интереса беспилотного подводного ли-

бо летательного подводного аппарата. Описанный метод прост в реализации, устойчив к воздействию помех и может быть дополнен различными операциями обработки изображений на всех этапах для повышения эффективности работы, что важно для обнаружения объекта в водной среде с пониженной видимостью.

Кроме того, следует отметить, что процедура формирования многомасштабной весовой модели изображения обладает свойствами естественного параллелизма. Это определяет возможность повышения скорости обработки за счет использования аппаратно-программных средств высокопроизводительных параллельных вычислений, что важно для эффективного управления беспилотным аппаратом в реальном времени и уменьшения его энергопотребления.

Ограничением применимости алгоритма в условиях замутненности воды является размер детектируемого радиального объекта. В частности, объекты радиусом менее 10 пикселей плохо поддаются детектированию разработанным алгоритмом. Более крупные объекты распознаются с высокой точностью свыше 95%. При этом следует задавать параметр радиальной жесткости α равным +4. Для эффективного обнаружения объектов уровень импульсного шума на изображениях не должен превышать 25%, в то же время гауссов шум имеет малое влияние на обнаружение объектов, радиус которых превышает 5 пикселей. Разработанный алгоритм устойчив к шумам типа «размытие» с усредняющими фильтрами. В силу вышесказанного алгоритм не требует дополнительных методов для повышения четкости изображения, что является его неоспоримым преимуществом перед аналогами.

Данное преимущество получено за счет совместного применения нескольких базовых алгоритмов, таких как метод Хафа, преобразование быстрой радиальной симметрии и построение пирамиды Гаусса. Полученная в результате единая модель превосходит каждый из базовых алгоритмов в отдельности, а также при их попарном применении в быстродействии и в требованиях к вычислительным ресурсам.

СПИСОК ЛИТЕРАТУРЫ

1. *Liquo T., Shenmin S., Xiaoyan Y., Jianwen S.* An overview of marine recovery methods of UAV for small ships // J. Harbin Institute Technol. 2019. V. 51. No. 10. P. 1–10.
2. *Chamola V., Kotesh P., Agarwal A., Naren, Gupta N., Guizani M.* A Comprehensive Review of Unmanned Aerial Vehicle Attacks and Neutralization Techniques // Ad Hoc Networks. 2021. V. 111. P. 102324. <https://doi.org/10.1016/j.adhoc.2020.102324>
3. ГОСТ Р 56829-2015 Интеллектуальные транспортные системы. Термины и определения, ГОСТ Р от 10 декабря 2015 года № 56829-2015.
4. *Albeaino G., Gheisari M., Franz B.* A Systematic Review of Unmanned Aerial Vehicle Application Areas and Technologies in the AEC Domain // J. of Inform. Technol. Construct. 2019. V. 24. P. 381–405.

5. *Hajiyev C., Soken H.E., Vural S.V.* Navigation Systems for Unmanned Aerial Vehicles // State Estimation and Control for Low-cost Unmanned Aerial Vehicles. 2015. P. 25–49. https://doi.org/10.1007/978-3-319-16417-5_3
6. *Moiseev V.S., Shafigullin R.R., Gushchina D.S.* Rational Placement and Required Number of Information Unmanned Aerial Systems for On-Line Monitoring of Large Territories // Russian Aeronautics. 2012. V. 55. P. 223–229. <https://doi.org/10.3103/S1068799812030014>
7. *Neira J., Sequeiros C., Huaman R., Machaca E., Fonseca P., Nina W.* Review on Unmanned Underwater Robotics, Structure Designs, Materials, Sensors, Actuators, and Navigation Control // J. Robot. 2021. P. 1–26. <https://doi.org/10.1155/2021/5542920>
8. *Колесников М.П., Мартынова Л.А., Пашкевич И.В., Шелест П.С.* Метод позиционирования автономного необитаемого подводного аппарата в процессе приведения к причальному устройству // Изв. ТулГУ. Технические науки. 2015. № 11-2. С. 38–49.
9. *Qiu S., Cui W.* An Overview on Aquatic Unmanned Aerial Vehicles // Ann. Rev. Res. 2019. V. 5. No. 3. P. 555663. <https://doi.org/10.19080/ARR.2019.05.555663>
10. *Popescu D., Ichim L.* Image Recognition in UAV Application Based on Texture Analysis // Advanced Concepts for Intelligent Vision Systems. ACIVS 2015. Lecture Notes in Computer Science. 2015. V. 9386. https://doi.org/10.1007/978-3-319-25903-1_60
11. *Samadzadegan F., Dadrass Javan F., Ashtari Mahini F., Gholamshahi M.* Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery // Aerospace. 2022. V. 9. No. 1. P. 31. <https://doi.org/10.3390/aerospace9010031>
12. *Fujiyoshi H., Hirakawa T., Yamashita T.* Deep Learning-Based Image Recognition for Autonomous Driving // IATSS Res. 2019. V. 43. No. 1. P. 244–252. <https://doi.org/10.1016/j.iatssr.2019.11.008>
13. *Севостьянов И.Е., Девитт Д.В.* Система визуального позиционирования многороторных беспилотников для совершения высокоточной автономной посадки // Science Time. 2021. № 90. С. 38–42.
14. *Степанов Д.Н.* Методы и алгоритмы определения положения и ориентации беспилотного летательного аппарата с применением бортовых видеокамер // Программные продукты и системы. 2014. № 1. С. 150–157.
15. *Deltheil C., Didier L., Hospital E., Brutzman D.P.* Simulating an Optical Guidance System for the Recovery of an Unmanned Underwater Vehicle // IEEE J. Ocean. Engineer. 2000. V. 25. No. 4. P. 568–574. <https://doi.org/10.1109/48.895364>
16. *Guo D., Bacciaglia A., Simpson M., Bil C., Marzocca P.* Design and Development a Bimodal Unmanned System // AIAA Scitech 2019 Forum. 2019. P. 1–7. <https://doi.org/10.2514/6.2019-2096>
17. *Pinheiro P.M., Neto A.A., Grando R.B., Silva C.B. da, Aoki Vivian M., Cardoso D.S., Horn A.C., Drews P.L.J.* Trajectory Planning for Hybrid Unmanned Aerial Underwater Vehicles with Smooth Media Transition // J. Intelligent Robot. Syst. 2022. V. 104. No. 46. <https://doi.org/10.1007/s10846-021-01567-z>

18. *Lock R.J., Vaidyanathan R., Burgess S.C., Loveless J.* Development of a Biologically Inspired Multi-Modal Wing Model for Aerial-Aquatic Robotic Vehicles through Empirical and Numerical Modelling of the Common Guillemot, *Uria Aalge* // *Bioinspirat. Biomimetics*. 2010. V. 5. No. 4. P. 1–15.
<https://doi.org/10.1088/1748-3182/5/4/046001>
19. *Wu Y., Li L., Su X., Gao B.* Dynamics Modeling and Trajectory Optimization for Unmanned Aerial-Aquatic Vehicle Diving into the Water // *Aerospace Sci. Technol.* 2019. V. 89. P. 220–229. <https://doi.org/10.1016/j.ast.2019.04.004>
20. *Liu S., Ozay M., Okatani T., Xu H., Sun K., Lin Y.* Detection and Pose Estimation for Short-Range Vision-Based Underwater Docking // *IEEE Access*. 2019. V. 30. No. 7. P. 2720–2749. <https://doi.org/10.1109/ACCESS.2018.2885537>
21. *Cowen S., Briest S., Dombrowski J.* Underwater Docking of Autonomous Undersea Vehicles Using Optical Terminal Guidance // *Oceans '97. MTS/IEEE Conference Proceedings, Halifax, NS, Canada*. 1997. V. 2. P. 1143–1147.
<https://doi.org/10.1109/OCEANS.1997.624153>
22. *Negre A., Pradalier C., Dunbabin M.* Robust Vision-Based Underwater Homing Using Self-Similar Landmarks // *J. Field Robot.* 2008. V. 25. No. 6–7. P. 360–377.
<https://doi.org/10.1002/rob.20246>
23. *Ghosh S., Ray R., Vadali S.R.K., Shome S.N., Nandy S.* Reliable Pose Estimation of Underwater Dock Using Single Camera: A Scene Invariant Approach // *Machine Vision Appl.* 2016. V. 27. No. 2. P. 221–236.
<https://doi.org/10.1007/s00138-015-0736-4>
24. *Li Y., Jiang Y., Cao J., Wang B., Li Y.* AUV Docking Experiments Based on Vision Positioning Using Two Cameras // *Ocean Engineer.* 2015. V. 110. P. 163–173.
25. *Чичкарев Е., Сергиенко А., Балалаева Е.* Использование моделей машинного обучения и сетей глубокого обучения для распознавания рукописных чисел и букв русского и латинского алфавитов // *InterConf. Prague, Czech Republic*. 2021. P. 363–380. <https://doi.org/10.51582/interconf.21-22.11.2021.044>
26. *Сирота А.А., Митрофанова Е.Ю., Милованова А.И.* Анализ алгоритмов поиска объектов на изображениях с использованием различных модификаций сверточных нейронных сетей // *Вестник ВГУ. Серия: Системный анализ и информационные технологии*. 2019. № 3. С. 123–137.
<https://doi.org/10.17308/sait.2019.3/1313>
27. *Zou Z., Shi Z., Guo Y., Ye J.* Object Detection in 20 Years: A Survey // *Proceedings of the IEEE*. 2019. No. 111. P. 257–276.
28. *Клетеник Д.В.* Сборник задач по аналитической геометрии. М.: Наука, Физматлит, 1998.
29. *Gonzalez R.C., Woods R.E.* *Digital Image Processing*. Third Edition. London: Pearson, 2007.
30. *Hough P.V.C.* Machine Analysis of Bubble Chamber Pictures // *2nd International Conference on High-Energy Accelerators and Instrumentation, HEACC 1959*. CERN, Geneva, Switzerland. 1959. P. 554–558.
31. *Lyasheva S., Shleyimovich M., Shakirzyanov R.* The Image Analysis Using Fast Radial Symmetry Transform in Control Systems Base on the Computer Vision // *International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon*. 2019. P. 1–6. <https://doi.org/10.1109/FarEastCon.2019.8934298>

32. *Shakirzyanov R.M.* Detection of Traffic Signals Using Color Segmentation and a Radial Symmetry Detector // Bulletin of the Voronezh State Technical University. 2020. V. 16. No. 6. P. 25–33.
33. *Adelson E., Burt P., Anderson C., Ogden J.M., Bergen J.* Pyramid Methods in Image Processing // RCA Engineer. 1984. V. 29. No. 6. P. 33–41.
34. *O'Malley R., Jones E., Glavin M.* Rear-Lamp Vehicle Detection and Tracking in Low-Exposure Color Video for Night Conditions // IEEE Transactions on Intelligent Transportation Systems. 2010. V. 11. No. 2. P. 453–462.
<https://doi.org/10.1109/TITS.2010.2045375>
35. *Друки А.А.* Алгоритмы выделения лиц на статических RGB изображениях и в видеопотоке // Изв. ТПУ. 2012. № 5. С. 65–69.
36. *Darge A., Rajendran R.S., Zerihum D., Chung P.Y.K.* Multi Color Image Segmentation using L*A*B* Color Space // Int. J. Advanced Engineer., Management Sci. 2019. V. 5. P. 346–352. <https://doi.org/10.22161/ijaems.5.5.8>
37. *Forsyth D., Ponce J.* Computer Vision: A Modern Approach. London: Pearson, 2012.
38. *Droogenbroeck V.M., Barnich O.* Design of Statistical Measures for the Assessment of Image Segmentation Schemes // Proceedings of 11th International Conference on Computer Analysis of Images and Patterns (CAIP2005), Lecture Notes in Computer Science. Rocancourt, France, 2005. V. 3691. P. 280–287.
39. *Кольцов П.П., Осипов А.С., Куцаев А.С., Кравченко А.А., Котович Н.В., Захаров А.В.* О количественной оценке эффективности алгоритмов анализа изображений // Компьютерная оптика. 2015. Т. 39, № 4. С. 542–556.

Статья представлена к публикации членом редколлегии О.А. Степановым.

Поступила в редакцию 19.11.2022

После доработки 16.03.2023

Принята к публикации 28.04.2023

Оптимизация, системный анализ и исследование операций

© 2023 г. А.А. ГАЛЯЕВ, д-р техн. наук (galaev@ipu.ru),

П.В. ЛЫСЕНКО, канд. физ.-мат. наук (pavellysen@ipu.ru),

Л.М. БЕРЛИН (berlin.lm@phystech.edu)

(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

СТАТИСТИЧЕСКАЯ СЛОЖНОСТЬ КАК КРИТЕРИЙ ЗАДАЧИ ОБНАРУЖЕНИЯ ПОЛЕЗНОГО СИГНАЛА¹

Рассматриваются три варианта функции статистической сложности, которая используется в качестве критерия в задаче обнаружения полезного сигнала в сигнально-шумовой смеси. Получены вероятностные распределения, максимизирующие рассматриваемые варианты статистической сложности, и сделаны выводы об эффективности использования того или иного варианта для задач обнаружения. На примере синтезированных сигналов продемонстрировано сравнение рассмотренных информационных характеристик и проиллюстрированы аналитические результаты. Предложен способ выбора порога информационного критерия, превышение которого позволяет сделать вывод о появлении полезного сигнала в сигнально-шумовой смеси. Выбор порога априорно зависит от максимальных значений, полученных аналитически. В результате сложность на основе полной вариации меры показала наилучшую способность обнаружения полезного сигнала.

Ключевые слова: статистическая сложность, обнаружение сигнала, информационная дивергенция.

DOI: 10.31857/S0005231023070073, EDN: FDVSTA

1. Введение

Понятие информационной энтропии было впервые введено в статье Клода Шеннона [1] в 1948 г. Эта работа положила начало новой области науки, получившей название теории информации [2]. Развитие теории информации дало возможность аналитического и практического исследования многих прикладных областей науки и техники. Были введены в обиход и получили свою интерпретацию такие термины, как энтропии Гиббса и фон Неймана, расстояние Кульбака–Лейблера, дивергенция Йенсена–Шеннона, информационные дивергенции и некоторые другие, которые стали использоваться в качестве критериев различных оптимизационных задач распознавания [3], классификации [4], фильтрации и отвечать за существо и качество принимаемых на их основе решений.

¹ Работа выполнена при финансовой поддержке Российского научного фонда (грант № 23-19-00134).

К концу прошлого века различные информационные критерии, в основном информационной энтропии Шеннона [1], начали активно применяться в задачах цифровой обработки сигналов, в частности в задаче обнаружения полезного сигнала в шумовой обстановке [5]. Появилось понятие спектральной энтропии [6], связанной с Фурье-спектром рассматриваемого сигнала, что оказалось особенно актуально в анализе акустических сигналов [7]. Кроме того, энтропийный подход успешно применяется в анализе временных рядов в медицинской сфере, например ЭКГ или ЭЭГ [8]. Позднее была предложена функция статистической сложности как развитие понятия энтропии [9–11]. Однако в статьях авторов не приводится аналитического исследования свойств предложенной ими функции, что оказывается особенно важным при решении задачи различения двух гипотез.

Нужно отметить, что существует несколько классических способов решения задачи обнаружения. Первый из них основан на решении задачи оптимальной фильтрации и требует знания свойств сигнала: периодичности, полосы частот и др. [12]. Второй способ опирается на лемму Неймана–Пирсона, решает задачу различения двух гипотез, определяет факт превышения оптимального порога при заданной вероятности ложной тревоги и требует оценки статистических свойств выборочных распределений шума и смеси сигнала и шума [13]. Третий способ эквивалентен решению задачи о разладке при изменении неизвестных статистических характеристик распределений шума и смеси сигнала и шума. Схожую постановку имеет задача поиска аномалий [14]. Все эти способы демонстрируют качественную и надежную работу при превышении сигнала над шумом, но при малых отношениях сигнал/шум или в тяжелой помеховой обстановке зачастую дают неправильный ответ.

Данная статья посвящена проблеме обнаружения полезного сигнала в сигнально-шумовой смеси и сочетает в себе все три ранее перечисленных способа решения задачи обнаружения. Предлагается использовать вариант леммы Неймана–Пирсона для задачи различения двух гипотез [13], который справедлив, в том числе когда вероятность ошибки различения близка к единице и зависит от полной вариации меры двух распределений нулевой и первой гипотез. На основании аналитического выражения функции этой ошибки формализуется критерий задачи обнаружения сигнала как один из вариантов статистической сложности [15], который учитывает детерминированную природу сигнала, примешанного к шуму. Особенностью статистической сложности является то, что она мультипликативна и состоит из двух множителей, один из которых обращается в нуль на детерминированных синусоидальных сигналах одной частоты, а в физике — на объектах заданной структуры, например кристаллах [9], а другой обращается в нуль на равномерных функциях распределений [10], соответствующих, например, белому шуму. Затем введенный критерий сравнивается с уже известными двумя вариантами статистической сложности на основе квадрата евклидова расстояния и дивергенции Йенсена–Шеннона, устанавливаются их свойства и проводится оптимизация как функции многих переменных на множестве дискретных распределе-

ний. В результате удается выделить семейства оптимальных распределений и вычислить максимумы функций статистических сложностей.

Статья имеет следующую структуру. В разделе 1 приведен литературный обзор и освещено актуальное состояние исследований по теме статьи. Раздел 2 посвящен связи рассматриваемых в дальнейшем информационных критериев с классическим критерием задачи обнаружения сигналов. В разделе 3 исследуются свойства трех типов статистической сложности. В разделе 4 аналитические результаты предыдущего раздела подкрепляются численным моделированием на синтезированных сигналах. Раздел 5 объединяет полученные в статье результаты и перечисляет планы на будущее.

2. Лемма Неймана–Пирсона и статистическая сложность

Задача обнаружения сигнала $s(n)$ традиционно сводится к задаче различения двух гипотез

$$\begin{cases} \Gamma_0 : x(n) = w(n), \\ \Gamma_1 : x(n) = s(n) + w(n), \quad n = 1, \dots, N. \end{cases}$$

Гипотеза Γ_0 связана с принятием решения о приеме только шума, а гипотеза Γ_1 — о приеме смеси полезного сигнала и шума, где последовательности $\{x(n)\}$, $n = 1, \dots, N$ — временной ряд из принятых данных, $\{s(n)\}$ — полезный сигнал, $\{w(n)\}$ — аддитивный случайный шум, N — длина временного ряда данных.

Случайные величины временного ряда $(x(1), \dots, x(n), \dots, x(N))$ принимают значения $(x_1, \dots, x_n, \dots, x_N) \in \mathbb{R}^N$.

Для получения аналитического выражения для оценки вероятности ошибки различения гипотез можно применить одну из разновидностей леммы Неймана–Пирсона [13, 16].

Лемма 1 [Неймана–Пирсона]. Пусть имеется произвольная, называемая решающим правилом или тестом, измеримая функция многих переменных $(x_1, \dots, x_N) \in \mathbb{R}^N$, такая что

$$d(x_1, \dots, x_N) = \begin{cases} 1, & \text{верна гипотеза } \Gamma_0, \\ 0, & \text{верна гипотеза } \Gamma_1, \end{cases}$$

по которой находятся

$$\begin{aligned} \alpha(d) &= \text{Вероятность (принять } \Gamma_0 \mid \text{верна } \Gamma_1), \\ \beta(d) &= \text{Вероятность (принять } \Gamma_1 \mid \text{верна } \Gamma_0). \end{aligned}$$

Тогда решающее правило d^ оптимально, если*

$$(1) \quad \alpha(d^*) + \beta(d^*) = \inf_d [\alpha(d) + \beta(d)] = \mathcal{E}r(N; \Gamma_0, \Gamma_1) - \text{функция ошибок,}$$

где инфимум берется по всевозможным тестам.

Здесь $\alpha(\cdot)$ – вероятность ложной тревоги, а $\beta(\cdot)$ – вероятность пропуска полезного сигнала.

Для функции ошибок справедлива точная формула следующего вида:

$$(2) \quad \mathcal{E}r(N; \Gamma_0, \Gamma_1) = 1 - \frac{1}{2} \|P_0^{(N)} - P_1^{(N)}\| = 1 - TV(P_0, P_1),$$

где $P_0^{(N)}$ – многомерная функция распределения статистики наблюдений по гипотезе Γ_0 , $P_1^{(N)}$ – многомерная функция распределения статистики наблюдений по гипотезе Γ_1 , а $TV(P_0, P_1)$ – полная вариация меры со знаком, $\|Q\| = 2 \sup_A |Q(A)|$. Таким образом, если множества-носители мер P_0, P_1 не пересекаются, то возможно безошибочное различение гипотез. Если же меры $P_0^{(N)}$ и $P_1^{(N)}$ близки, то $\|P_0^{(N)} - P_1^{(N)}\| \approx 0$, и тогда $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$.

Для задачи обнаружения детерминированного полезного сигнала, например при малом отношении сигнал/шум, интересен случай $\|P_0^{(N)} - P_1^{(N)}\| = 2TV(P_0, P_1) \approx 0$ и возможность разумной оценки этой величины. Поэтому когда вероятность суммарной ошибки различения двух гипотез близка к единице, появляется возможность использовать аналитическое выражение $TV(P_0, P_1)$ для конструирования критерия в задаче обнаружения полезного сигнала в смеси. Но сначала обратимся к уже известным критериям и установим их свойства.

Чаще всего для удобства математического исследования полезный сигнал и шум моделируются случайными гауссовыми процессами с разными параметрами, в этом случае задача нахождения момента появления сигнала $s(n)$ в принимаемой последовательности отсчетов называется задачей о разладке [13].

Здесь и далее рассматриваются дискретные распределения вероятностей $p = (p_1, \dots, p_i, \dots, p_N)$, которые по определению обладают следующими свойствами:

$$(3) \quad \forall p_i \in [0, 1], \quad \sum_{i=1}^N p_i = 1.$$

Для формализации критериев, которые учитывают не только случайную составляющую сигналов, но и детерминированную, введем понятия функции дисбаланса D и статистической сложности C распределения. Простейшим примером дисбаланса является евклидово расстояние в пространстве дискретных вероятностных распределений [10].

Определение 1. Дисбаланс D_{SQ} имеет смысл дисперсии распределения относительно равномерного, а именно

$$(4) \quad D_{SQ}(p) = \sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 = \sum_{i=1}^N p_i^2 - \frac{1}{N}.$$

Определение 2. Статистическая сложность, определяемая через выражение дисбаланса по определению 1, равна

$$(5) \quad C_{SQ}(p) = H(p) \cdot D_{SQ}(p),$$

где

$$(6) \quad H(p) = \frac{1}{\log N} \left(- \sum_i^N p_i \log p_i \right)$$

– нормированная энтропия Шеннона [1].

При вычислении суммы (6) принимается, что $\frac{0}{\log 0} = 0$ по непрерывности, и это предположение справедливо для всех последующих уравнений.

Из определения 1 следует, что дисбаланс вида (4) и сложность вида (5) удобно применять при оценке и сравнении сигналов, имеющих спектральное распределение, близкое к равномерному. В общем случае вместо равномерного распределения $q_i = 1/N$ при $i = 1, \dots, N$ в формулу (4) может входить произвольное дискретное распределение.

Выше введена формула (4), предложенная в [10] для вычисления дисбаланса относительно равномерного распределения, однако в большинстве работ в качестве такой величины используется дивергенция $JSD(p||q)$ Йенсена–Шеннона [17], так что

Определение 3. Дисбаланс Йенсена–Шеннона равен

$$(7) \quad D_{JSD}(p) = JSD(p||q),$$

где $q = (1/N, \dots, 1/N)$ – равномерное распределение.

Определение 4. Статистическая сложность, определяемая через выражение дисбаланса согласно определению 3, выражается как

$$(8) \quad C_{JSD}(p) = H(p) \cdot D_{JSD}(p).$$

Замечание 1. Выше отмечено, что $\sqrt{D_{SQ}}$ является евклидовой метрикой на пространстве дискретных распределений. В то же время $\sqrt{D_{JSD}}$ также является метрикой и пропорционален метрике Фишера.

Поскольку в лемме 1 Неймана–Пирсона получено, что функция ошибок различения двух гипотез зависит от полной вариации $TV(p, q)$, то введем еще одно понятие дисбаланса.

Определение 5. Дисбаланс по полной вариации меры равен

$$(9) \quad D_{TV}(p) = TV^2(p, q),$$

где $q = (1/N, \dots, 1/N)$.

Определение 6. Статистическая сложность, определяемая через выражение дисбаланса согласно определению 5, равна

$$(10) \quad C_{TV}(p) = H(p) \cdot D_{TV}(p).$$

Представленные выше функции информационных дивергенций, определяющие разные варианты функции дисбаланса, могут быть объединены общим понятием *f-дивергенции* [18]:

$$(11) \quad D_f(p||q) = \sum_{x \in \mathbb{R}^N} q(x) f\left(\frac{p(x)}{q(x)}\right).$$

Выбор функции f порождает целое семейство различных дивергенций:

- Дивергенция Кульбака–Лейблера $D_{KL}(p, q)$ получается из (11) выбором $f(x) = x \log(x)$, $x > 0$.
- Дивергенция Йенсена–Шеннона получается из (11) выбором

$$(12) \quad f(x) = x \log \frac{2x}{x+1} + \log \frac{2}{x+1}, \quad x > 0.$$

- Полная вариация получается при $f(x) = \frac{1}{2}|1 - x|$:

$$(13) \quad TV(p, q) = \frac{1}{2} \sum_{x \in \mathbb{R}^N} |p(x) - q(x)|;$$

TV также является метрикой на пространстве распределений вероятностей. Полная вариация связана с дивергенцией Йенсена–Шеннона следующим соотношением:

$$(14) \quad JSD(p||q) \leq TV(p, q).$$

Из соотношения (14) следует, что дивергенция Йенсена–Шеннона оценивается сверху полной вариацией.

Далее исследуем возможность применения каждого варианта статистической сложности как критерия для индикации появления сигнала, но сначала установим их свойства.

3. Оптимизация статистической сложности

3.1. Оптимизация C_{SQ}

Поставим задачу максимизации функции статистической сложности на множестве дискретных распределений $p = (p_1, \dots, p_N)$

$$(15) \quad C_{SQ}(p) = \frac{1}{\log N} \left(- \sum_{i=1}^N p_i \log p_i \right) \cdot \left(\sum_{i=1}^N \left(p_i - \frac{1}{N} \right)^2 \right) \rightarrow \max_p$$

при условии нормировки

$$(16) \quad \sum_{i=1}^N p_i = 1.$$

Для формулировки леммы о максимальном значении статистической сложности понадобится вспомогательный результат.

Лемма 2. Пусть $0 < x \leq y \leq z \leq 1$, тогда $f(x, y, z) = x^y y^{-x} z^x x^{-z} y^z z^{-y} \geq 1$, причем равенство возможно только при либо $x = y$, либо $y = z$.

Доказательство. Введем новую функцию $g(x, y, z) = \ln f(x, y, z)$,

$$g(x, y, z) = y \ln x - x \ln y + x \ln z - z \ln x + z \ln y - y \ln z.$$

Тогда требуется доказать, что $g(x, y, z) \geq 0$ при $0 < x \leq y \leq z \leq 1$.

По теореме Куна–Таккера решение задачи условной оптимизации функции трех переменных либо находится во внутренней точке ограничения, либо на его границе. Необходимые условия безусловного экстремума функции $g(x, y, z)$ принимают вид

$$(17) \quad \begin{aligned} \frac{\partial g}{\partial x} &= \ln z - \ln y + \frac{y-z}{x} = 0, \\ \frac{\partial g}{\partial y} &= \ln x - \ln z + \frac{z-x}{y} = 0, \\ \frac{\partial g}{\partial z} &= \ln y - \ln x + \frac{x-y}{z} = 0. \end{aligned}$$

Сложим все уравнения последней системы и в результате получим уравнение

$$\frac{y-z}{x} + \frac{z-x}{y} + \frac{x-y}{z} = 0,$$

которое можно переписать в виде

$$\frac{(y-z)(x-y)(z-x)}{xyz} = 0.$$

Это означает, что при выполнении одного из равенств либо $x = y$, либо $y = z$ функция $g(x, y, z)$, возможно, имеет минимум. Пусть $x = y$, тогда третье уравнение из (17) выполняется тождественно, а первое и второе уравнения идентичны и могут быть записаны в виде

$$\ln \eta = \eta - 1, \quad \eta = \frac{z}{y}.$$

Последнее уравнение имеет только один корень $\eta = 1$, т.е. $y = z$.

Вычислим вторые производные и запишем матрицу Гессе:

$$(18) \quad G(x, y, z) = \begin{pmatrix} \frac{z-y}{x^2} & \frac{1}{x} - \frac{1}{y} & \frac{1}{z} - \frac{1}{x} \\ \frac{1}{x} - \frac{1}{y} & \frac{x-z}{y^2} & \frac{1}{y} - \frac{1}{z} \\ \frac{1}{z} - \frac{1}{x} & \frac{1}{y} - \frac{1}{z} & \frac{y-x}{z^2} \end{pmatrix}.$$

Миноры матрицы Гессе равны

$$(19) \quad \begin{aligned} M_1(x, y, z) &= \frac{z - y}{x^2}, \\ M_2(x, y, z) &= -\frac{(x - y)^2 + (z - x)^2 + (y - z)^2}{2x^2y^2}, \\ M_3(x, y, z) &= 0. \end{aligned}$$

Матрица Гессе не является знакоопределенной, более того, ее детерминант равен нулю. Поэтому рассмотрим малую окрестность экстремальной точки.

В малой окрестности $x = y = z$ при условии, что $\delta x \leq \delta y \leq \delta z$, вариация δg функции $g(x, y, z)$ записывается в виде

$$\begin{aligned} \delta g &= (x + \delta y) \ln(x + \delta x) - (x + \delta x) \ln(x + \delta y) + (x + \delta x) \ln(x + \delta z) - \\ &- (x + \delta z) \ln(x + \delta x) + (x + \delta z) \ln(x + \delta y) - (x + \delta y) \ln(x + \delta z) = \\ &= (\delta z - \delta x)(\delta y - \delta x)(\delta z - \delta y) + o(((\delta x)^2 + (\delta y)^2 + (\delta z)^2)^{3/2}) \geq 0, \end{aligned}$$

где ненулевыми оказываются члены при кубах вариаций независимых переменных, а сама вариация функции $g(x, y, z)$ в силу условий леммы оказывается положительной. В случае, когда, например, $\delta y = \delta x$, имеем $g(x, y, z) \equiv 0$, а $f(x, y, z) \equiv 1$. Поэтому экстремум функции является ее нестрогим минимумом. ■

Лемма 3. Максимум статистической сложности (15) достигается на распределении вида

$$(20) \quad \begin{cases} p_i = \frac{1 - p_{\max}}{N - 1}, & i = \overline{1, N} \setminus k, \\ p_k = p_{\max}, \end{cases}$$

где $p_{\max} = \text{const}$, т.е. при появлении отдельной дискреты в произвольном разряде k над равномерным распределением.

Доказательство. Без потери общности положим $k = N$. Из уравнения (16) выразим одну переменную p_N из набора p_i через все остальные:

$$(21) \quad p_N = 1 - \sum_{i=1}^{N-1} p_i.$$

Перепишем уравнение (15) в виде

$$(22) \quad \begin{aligned} C_{SQ} &= -\frac{1}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times \\ &\times \left(\sum_{i=1}^{N-1} \left(p_i - \frac{1}{N} \right)^2 + \left(p_N - \frac{1}{N} \right)^2 \right). \end{aligned}$$

Необходимым условием экстремума функции во внутренней точке области (симплекса 3) является равенство нулю всех частных производных по p_i :

$$(23) \quad \frac{\partial C_{SQ}}{\partial p_i} = 0, \quad i = 1, \dots, N-1.$$

Подстановка функции (22) в (23) дает (с учетом того, что $\frac{\partial p_N}{\partial p_i} = -1$):

$$(24) \quad \begin{aligned} \frac{\partial C_{SQ}}{\partial p_i} = & -\frac{1}{\log N} (\log p_i - \log p_N) \cdot \left(\sum_{i=1}^{N-1} \left(p_i - \frac{1}{N} \right)^2 + \left(p_N - \frac{1}{N} \right)^2 \right) - \\ & - \frac{2}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \cdot (p_i - p_N) = 0, \quad i = 1, \dots, N-1. \end{aligned}$$

В более удобном виде уравнения имеют вид

$$(25) \quad \frac{\partial C_{SQ}}{\partial p_i} = \frac{1}{\log N} (-\log p_i + \log p_N) \cdot D + 2H \cdot (p_i - p_N) = 0, \\ i = 1, \dots, N-1.$$

Запишем разность любых двух уравнений из системы выше для индексов i и j :

$$(26) \quad \frac{\partial C_{SQ}}{\partial p_i} - \frac{\partial C_{SQ}}{\partial p_j} = \frac{1}{\log N} (-\log p_i + \log p_j) \cdot D + 2H \cdot (p_i - p_j) = 0.$$

Учитывая, что величины D и H положительны, из уравнений (25) и (26) можно составить следующие равенства при условии, что рассматриваемые вероятности p_j , $j = 1, \dots, N-1$ не равны p_N :

$$(27) \quad \frac{\log p_i - \log p_j}{\log p_N - \log p_j} - \frac{p_i - p_j}{p_N - p_j} = 0,$$

$$(28) \quad (p_N - p_j) \log p_i + (p_i - p_N) \log p_j + (p_j - p_i) \log p_N = 0,$$

$$(29) \quad p_i^{p_N - p_j} \cdot p_j^{p_i - p_N} \cdot p_N^{p_j - p_i} = 1.$$

Применяя лемму 2, делаем вывод, что последнее уравнение может быть выполнено при равенстве $p_i = p_j$.

Таким образом, получено, что каждая из вероятностей p_i может принимать одно из двух различных значений, которые задают распределение вида

$$(30) \quad \begin{cases} p_i = \frac{1 - p_{\max}}{K}, & \forall i = 1, \dots, K, \\ p_i = p_N = \frac{p_{\max}}{N - K}, & \forall i = K + 1, \dots, N. \end{cases}$$

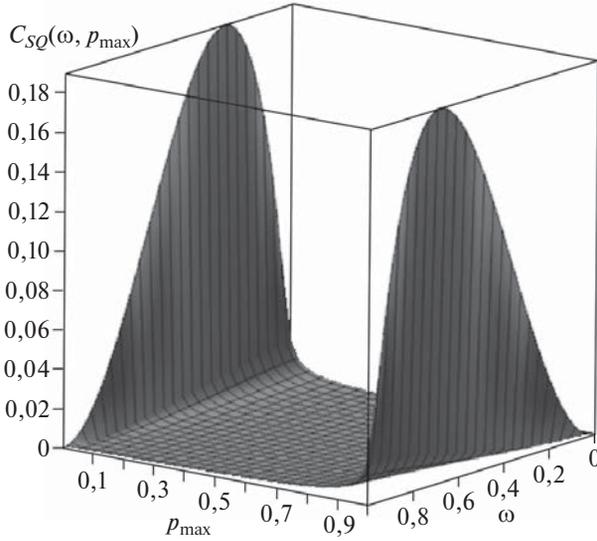


Рис. 1. Поверхности уровня статистической сложности $C_{SQ}(\omega, p_{\max})$.

Теперь нужно показать, что максимуму сложности соответствуют значения $K = 1$ и $K = N - 1$. Для этого вычислим значение дисбаланса (4) на распределении (30), которое обозначим через $D^{(K)}(\omega, p_{\max})$:

$$(31) \quad D^{(K)}(\omega, p_{\max}) = \frac{1}{N} \frac{(p_{\max} + \omega - 1)^2}{\omega(1 - \omega)}, \quad \omega = \frac{K}{N}.$$

В свою очередь, энтропия равна

$$(32) \quad H^{(K)}(\omega, p_{\max}) = 1 - \frac{1}{\log N} \left((1 - p_{\max}) \log \frac{1 - p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1 - \omega} \right).$$

Максимум $C_{SQ}(\omega, p_{\max})$ при $N \leq 100$ был исследован численно, и он достигался при $K = 1$. Из выражения для $D^{(K)}(\omega, p_{\max})$ (31) видно, что при $N \geq 101$ и при изменении с $K = 1$ на $K = 2$ или с $K = N - 1$ на $K = N - 2$ его значение меняется почти в два раза, в то время как энтропия (32) меняется незначительно. Таким образом, распределение вероятностей (30), доставляющее функции сложности экстремальное значение при $K = 1$ или $K = N - 1$, имеет вид (20). ■

Для наглядности на рис. 1 показан график $C_{SQ} = C_{SQ}(\omega, p_{\max})$ при $N = 1024$, на котором ω меняется непрерывно (хотя K меняется дискретно).

Следствие 1. Подставим значения p_i и $p_N = p_{\max}$ из (20) в (22) и рассмотрим сложность C_{SQ} как функцию от p_{\max} . При достаточно больших значениях N она примет вид

$$C_{SQ} \approx (1 - p_{\max}) \cdot p_{\max}^2.$$

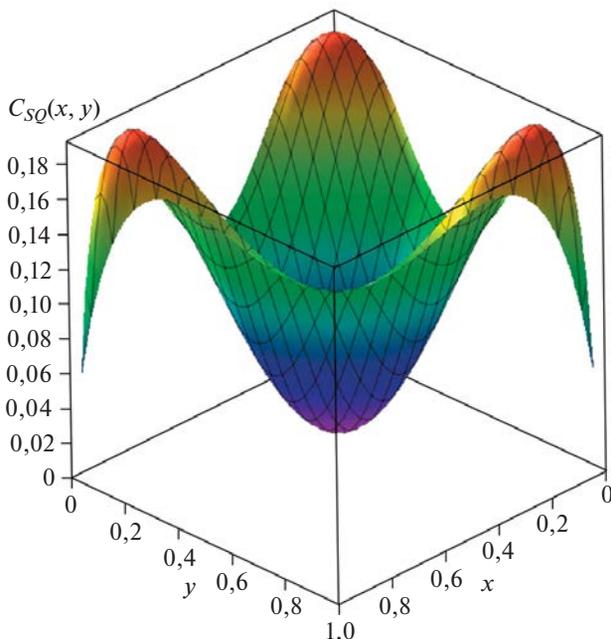


Рис. 2. Поверхности уровня статистической сложности $C_{SQ}(x, y)$ для $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Откуда следует, что эта функция принимает максимальное значение $C_{SQ}^* \approx 4/27$ при $p_{\max} = 2/3$.

Следствие 2. Минимальное значение $C_{SQ} = 0$ достигается на равномерном распределении $p = (1/N, \dots, 1/N)$.

Действие леммы 3 для случая, когда дискретное распределение $p = \{p_1, p_2, p_3\}$ состоит только из трех отсчетов, продемонстрировано на рис. 2. Сложность зависит от двух переменных, поскольку одна из вероятностей может быть выражена через другие. Здесь C_{SQ} имеет три одинаковых ярко выраженных максимума и три одинаковых локальных минимума, относящихся к ситуациям $p_1 = p_2$, $p_2 = p_3$, $p_1 = p_3$ при выполнении необходимых условий экстремума, а также глобальный минимум при $p_1 = p_2 = p_3$.

В табл. 1 показано изменение оптимальных параметров $C_{SQ}(w, p_{\max})$ при росте N .

Таблица 1. Оптимальные параметры $C_{SQ}(w, p_{\max})$ при различных значениях N

N	$C_{SQ}(w^*, p_{\max}^*)$	p_{\max}^*	w^*	$N - K^*$
3	0,1932	0,8315	0,6666	1
256	0,1994	0,7044	0,9960	1
512	0,1942	0,7008	0,9980	1
1024	0,1898	0,6979	0,9990	1
2048	0,1861	0,6955	0,9995	1

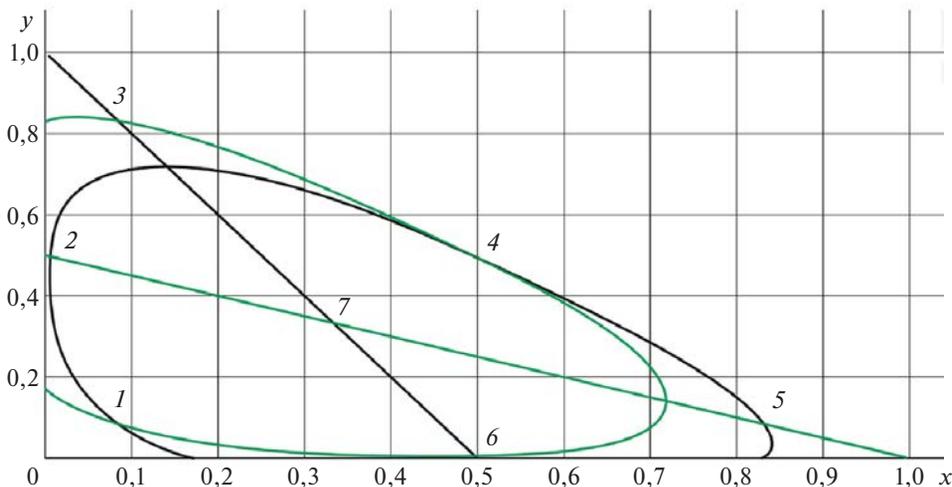


Рис. 3. Кривые, соответствующие необходимым условиям экстремума C_{SQ} для $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Необходимые условия экстремума C_{SQ} для дискретного распределения $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$ выписываются согласно (25) следующим образом:

$$(33) \begin{cases} (-\log x + \log(1 - x - y)) \left(\left(x - \frac{1}{3}\right)^2 + \left(y - \frac{1}{3}\right)^2 + \left(1 - x - y - \frac{1}{3}\right)^2 \right) - \\ - 2(x \log x + y \log y + (1 - x - y) \log(1 - x - y))(-1 + y) = 0, \\ (-\log y + \log(1 - x - y)) \left(\left(x - \frac{1}{3}\right)^2 + \left(y - \frac{1}{3}\right)^2 + \left(1 - x - y - \frac{1}{3}\right)^2 \right) - \\ - 2(x \log x + y \log y + (1 - x - y) \log(1 - x - y))(-1 + x) = 0. \end{cases}$$

Неявные уравнения системы (33) описывают кривые, представленные на рис. 3.

Черная и зеленые кривые отвечают первому и второму неявным уравнениям системы (33) соответственно. На рис. 3 отмечены семь точек экстремума, для которых подсчитано значение статистической сложности. Все полученные данные приведены в табл. 2.

Таблица 2. Экстремальные точки статистической сложности при $N = 3$

p	1	2	3	4	5	6	7
p_1	0,08425	0,006	0,08425	0,497	0,8315	0,497	0, (3)
p_2	0,08425	0,497	0,8315	0,497	0,08425	0,006	0, (3)
p_3	0,8315	0,497	0,08425	0,006	0,08425	0,497	0, (3)
C_{SQ}	0,1932	0,1062	0,1932	0,1062	0,1932	0,1062	0

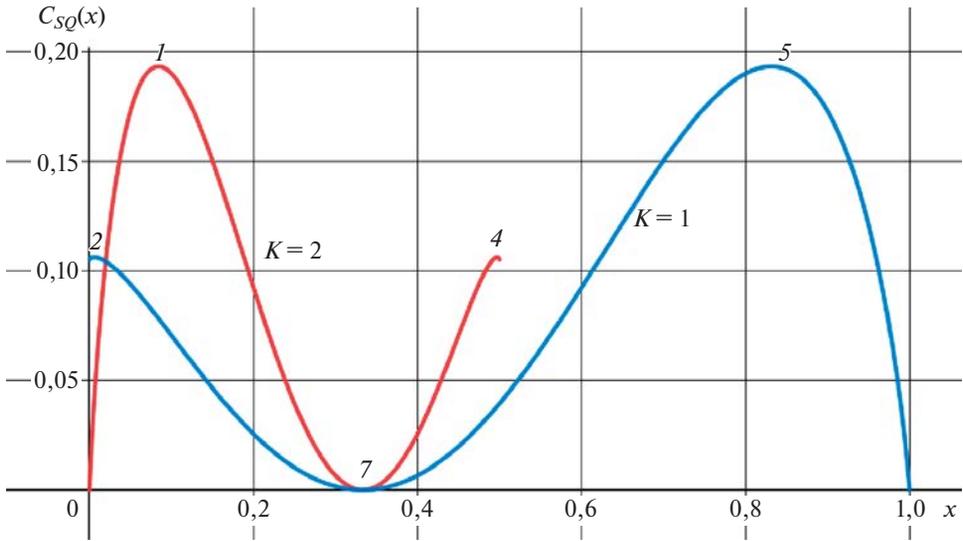


Рис. 4. Статистическая сложность C_{SQ} для $p = \{p_1 = x, p_2 = x, p_3 = 1 - 2x\}$.

Первая, третья и пятая точки максимума соответствуют одному и тому же значению максимума функции. Вторая, четвертая и шестая точки минимума соответствуют одному и тому же значению минимума функции. Стоит отметить, что экстремальные точки отвечают случаю $K = N - 1 = 2$ за исключением глобального минимума, где все вероятности равны между собой, и тем самым описывают три локальных минимума, один глобальный минимум и три равных между собой максимума статистической сложности на рис. 2. Отдельно можно построить статистическую сложность в случае $p = \{p_1 = x, p_2 = x, p_3 = 1 - 2x\}$.

На рис. 4 отмечены точки экстремума в соответствии с табл. 2, которые охватывают все случаи $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

3.2. Оптимизация C_{JSD}

Применим аналогичный подход к статистической сложности, использующей дивергенцию Йенсена–Шеннона в качестве дисбаланса. Сложность в этом случае выражается как

$$(34) \quad C_{JSD}(p) = H(p) \cdot JSD(p||q), \quad q_j = 1/N, \quad j = 1, \dots, N.$$

Раскрыв $JSD(p||q)$, получаем

$$(35) \quad C_{JSD}(p) = H(p) \cdot \left(H(m) - \frac{1}{2}(H(p) + H(q)) \right) \cdot \log N, \quad m = \frac{p+q}{2}.$$

Лемму, подобную лемме 3, доказать не удастся, однако можно выписать необходимые условия экстремума для статистической сложности вида (35).

Следуя той же логике, что и в лемме 3, перепишем последнее уравнение в переменных p_i и p_N :

$$(36) \quad C_{JSD}(p) = - \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times \\ \times \left(H(m) - \frac{1}{2} \left(-\frac{1}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) + 1 \right) \right),$$

где

$$(37) \quad H(m) = -\frac{1}{\log N} \left(\sum_{i=1}^{N-1} \frac{p_i + \frac{1}{N}}{2} \log \frac{p_i + \frac{1}{N}}{2} + \frac{p_N + \frac{1}{N}}{2} \log \frac{p_N + \frac{1}{N}}{2} \right).$$

Тогда с учетом (21)

$$(38) \quad \frac{\partial H(m)}{\partial p_i} = -\frac{1}{\log N} \left(\frac{1}{2} \log \frac{p_i + \frac{1}{N}}{2} - \frac{1}{2} \log \frac{p_N + \frac{1}{N}}{2} \right), \quad i = 1, \dots, N-1.$$

Объединяя все частные производные, получаем необходимые условия экстремума в виде

$$(39) \quad \frac{\partial C_{JSD}(p)}{\partial p_i} = \frac{H(p)}{2} \cdot \left(- \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) + (\log p_i - \log p_N) \right) - \\ - (\log p_i - \log p_N) \cdot JSD(p||q) = 0, \quad i = 1, \dots, N-1.$$

После упрощения остаются уравнения

$$(40) \quad H(p) \cdot \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) + \\ + (\log p_i - \log p_N) \cdot (2JSD(p||q) - H(p)) = 0, \quad i = 1, \dots, N-1.$$

Тогда разность уравнений (40) для индексов i, j будет иметь вид

$$(41) \quad (\log p_i - \log p_j) \cdot (2JSD(p||q) - H(p)) + \\ + H(p) \cdot \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_j + \frac{1}{N}}{2} \right) = 0.$$

Замечание 2. Из вида системы уравнений (41) следует, что система выполняется, если $p_i = p_j$, что является одним из необходимых условий экстремума функции (35). Вследствие нелинейности системы, состоящей из уравнений (41), у нее могут быть и другие корни.

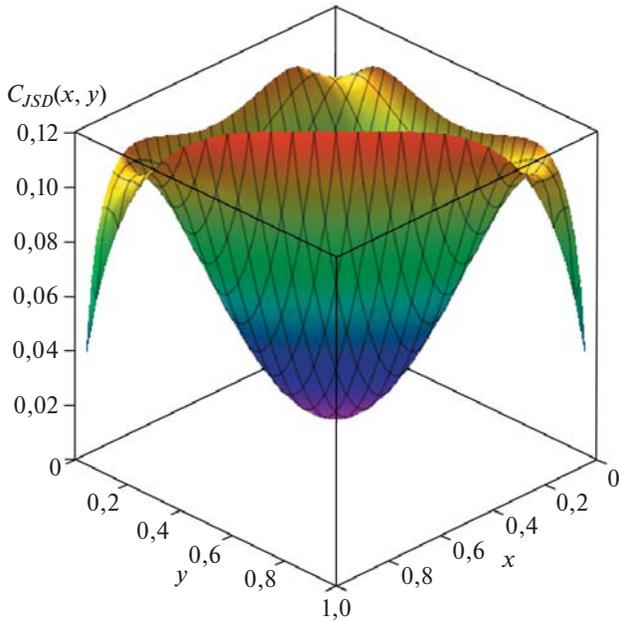


Рис. 5. Поверхности уровня статистической сложности $C_{JS D}$ для $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Для иллюстрации этого факта на рис. 5 приведен график поверхности уровня статистической сложности вида (35), когда дискретное распределение $p = \{p_1, p_2, p_3\}$ состоит только из трех отсчетов.

На рис. 5 видно, что в точках, удовлетворяющих $p_1 = p_2$, $p_2 = p_3$ и $p_1 = p_3$, при выполнении необходимых условий экстремума находятся седловые точки поверхности.

Ранее было установлено, что распределение (30) доставляет экстремум C_{SQ} при $K = N - 1$. Далее будет показано, что оно же доставляет экстремум сложности, основанной на полной вариации меры $TV(p, q)$. Поэтому предлагается найти максимум $C_{JS D}$ на этом распределении и провести сравнение полученных оптимальных параметров распределения при фиксированных N . Выпишем сложность в явном виде и получим

$$(42) \quad C_{JS D}^{(K)} = H^{(K)} \cdot \left(H^{(K)}(m) - \frac{1}{2}(H^{(K)} + 1) \right) \cdot \log N,$$

где $H^{(K)}$ соответствует (32), а $H^{(K)}(m)$ дается следующей формулой:

$$(43) \quad H^{(K)}(m) = 1 - \frac{1}{\log N} \left(\frac{(1 - p_{\max} + \omega)}{2} \log \frac{(1 - p_{\max} + \omega)}{2\omega} + \frac{(1 + p_{\max} - \omega)}{2} \log \frac{(1 + p_{\max} - \omega)}{2 - 2\omega} \right).$$

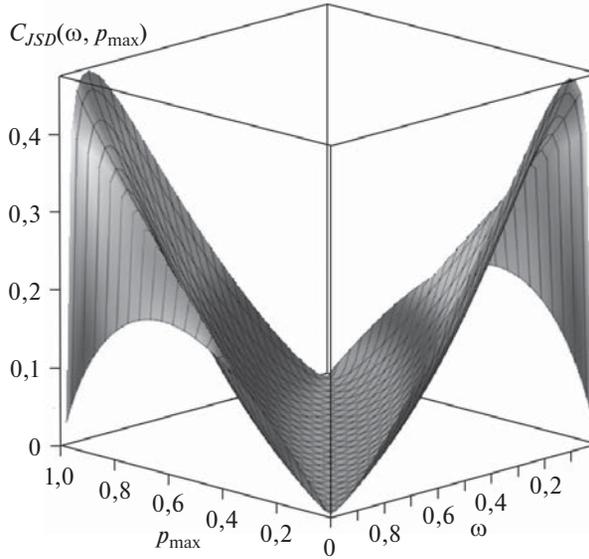


Рис. 6. Поверхности уровня статистической сложности $C_{JSD}(\omega, p_{\max})$.

В табл. 3 показано изменение оптимальных параметров C_{JSD} при росте N .

Для наглядности на рис. 6 показан график $C_{JSD} = C_{JSD}(\omega, p_{\max})$ при $N = 1024$, на котором ω меняется непрерывно (хотя K меняется дискретно).

Результаты, приведенные в табл. 3, демонстрируют, что для выбранного класса распределений (30) оптимальным является набор из N дискрет, где K равны между собой, а остальные нулевые. Стоит обратить внимание, что на полученном распределении C_{JSD} не равна нулю, в том числе благодаря слагаемому $H^{(K)}(m)$, которое отвечает уже “сдвинутому” распределению, состоящему из K элементов, равных $\frac{1}{K} + \frac{1}{N}$, и $N - K$ отсчетов по $\frac{1}{2N}$ каждый.

Таблица 3. Оптимальные параметры $C_{JSD}(\omega, p_{\max})$ при различных значениях N

N	$C_{JSD}(\omega^*, p_{\max}^*)$	p_{\max}^*	ω^*	$N - K^*$
3	0,1266	1	0,4083	1 или 2
256	0,4482	1	0,8703	33
512	0,4790	1	0,8897	56
1024	0,5065	1	0,9051	97
2048	0,5312	1	0,9171	170

3.3. Оптимизация C_{TV}

Приступим к анализу статистической сложности на основе полной вариации

$$(44) \quad C_{TV}(p) = -\frac{1}{4 \log N} \left(\sum_{i=1}^N p_i \log p_i \right) \cdot \left(\sum_{i=1}^N \left| p_i - \frac{1}{N} \right| \right)^2.$$

Предложение 1. Согласно выражению для функции ошибок (2) из леммы 1 Неймана–Пирсона и определению (44) в качестве критерия предлагается использовать C_{TV} для решения задачи различения двух гипотез и индикации появления детерминированной составляющей полезного сигнала в шуме.

Справедлива следующая лемма.

Лемма 4. Максимум статистической сложности (44) достигается на семействе распределений (30).

Доказательство. С учетом симметричности функции (44) и симплекса (3) без ограничения общности найдется целое $K \in \{1, \dots, N - 1\}$, при котором максимум в данной лемме достигается на части симплекса (3), определяемой ограничениями $p_i \leq 1/N$ для $i = 1, \dots, K$ и $p_i \geq 1/N$ для $i = K + 1, \dots, N$. Перепишем уравнение (44) в виде

$$(45) \quad C_{TV} = -\frac{1}{4 \log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \times \\ \times \left(\sum_{i=1}^K \left(-p_i + \frac{1}{N} \right) + \sum_{i=K+1}^N \left(p_i - \frac{1}{N} \right) \right)^2.$$

Тогда для $i = 1, \dots, K$ необходимые условия экстремума принимают вид

$$(46) \quad \frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N} (\log p_i - \log p_N) \cdot D_{TV} - 2H(p) \sqrt{D_{TV}} = 0, \\ i = 1, \dots, K,$$

а для $i = K + 1, \dots, N$ справедливо

$$(47) \quad \frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N} (\log p_i - \log p_N) \cdot D_{TV} = 0, \quad i = K + 1, \dots, N.$$

Составим разность двух уравнений из (46) для индексов i и j . Откуда следует, что если $D_{TV} \neq 0$, то $p_i = p_j$ при $i = 1, \dots, K$. В то время как из (47) следует, что $p_i = p_N$ при $i = K + 1, \dots, N$. Снова получили, что семейство распределений (30) доставляет максимум функции сложности, теперь уже C_{TV} . ■

Далее требуется определить оптимальные значения K и p_{\max} . Для этого вычислим значение дисбаланса $D_{TV}^{(K)}$ на распределении (30):

$$(48) \quad D_{TV}^{(K)} = (p_{\max} + \omega - 1)^2, \quad \omega = \frac{K}{N}.$$

В свою очередь, энтропия равна

$$(49) \quad H^{(K)} = 1 - \frac{1}{\log N} \left((1 - p_{\max}) \log \frac{1 - p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1 - \omega} \right).$$

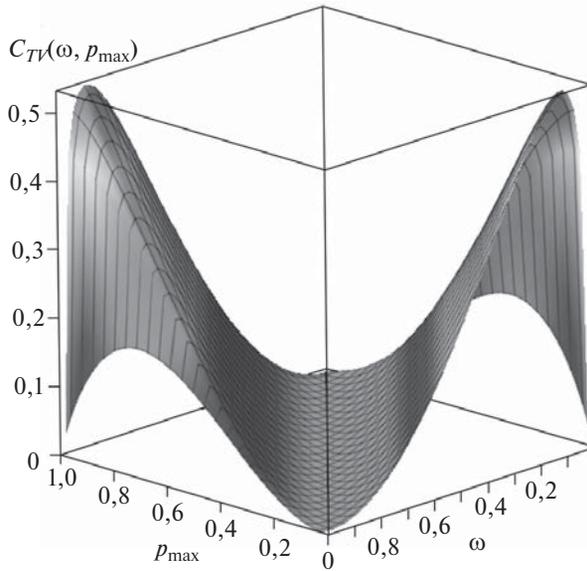


Рис. 7. Поверхности уровня статистической сложности $C_{TV}(\omega, p_{\max})$.

Для наглядности на рис. 7 представлен график $C_{TV} = C_{TV}(\omega, p_{\max})$ при $N = 1024$, на котором ω меняется непрерывно (хотя K меняется дискретно).

Составим необходимые условия (46) и (47) экстремума статистической сложности C_{TV} , выписанной через переменные p_{\max} , ω .

$$(50) \quad \left\{ \begin{array}{l} f_1^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1) \left(\left(1 - \frac{(1-p_{\max}) \log \frac{1-p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1-\omega}}{\log N} \right) - \right. \\ \left. - \frac{(p_{\max} + \omega - 1)}{2 \log N} \left(\log \frac{p_{\max}}{1-\omega} - \log \frac{1-p_{\max}}{\omega} \right) \right) = 0, \\ f_2^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1) \left(\left(1 - \frac{(1-p_{\max}) \log \frac{1-p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1-\omega}}{\log N} \right) - \right. \\ \left. - \frac{(p_{\max} + \omega - 1)}{2 \log N} \left(\frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega} \right) \right) = 0. \end{array} \right.$$

Пересечения кривых, соответствующих неявным уравнениям (50), отвечают точкам экстремума C_{TV} . Составим разность двух необходимых условий экстремума

$$(51) \quad \begin{aligned} f_3^N(p_{\max}, \omega) &:= f_1^N(p_{\max}, \omega) - f_2^N(p_{\max}, \omega) = \\ &= \frac{(p_{\max} + \omega - 1)^2}{\log N} \left(-\log \frac{p_{\max}}{1-\omega} + \log \frac{1-p_{\max}}{\omega} + \frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega} \right) = 0. \end{aligned}$$

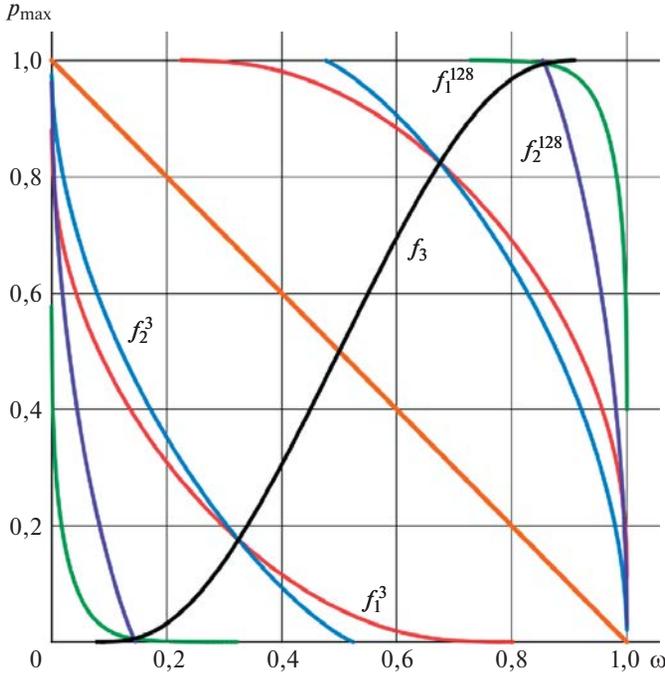


Рис. 8. Кривые необходимых условий экстремума C_{TV} для $N = 3$ и $N = 128$. Уравнения $f_1^N(p_{\max}, \omega)$ и $f_2^N(p_{\max}, \omega)$ имеют общее решение $(p_{\max} + \omega - 1 = 0)$ (оранжевая кривая).

Построим неявные кривые уравнений $f_1^N(p_{\max}, \omega)$, $f_2^N(p_{\max}, \omega)$, $f_3^N(p_{\max}, \omega)$ для некоторых значений N . Для удобства индекс N у $f_3^N(p_{\max}, \omega)$ может быть опущен, так как неявная кривая уравнения (51) не зависит от N .

Согласно рис. 8 статистическая сложность имеет, помимо точек минимума $(p_{\max} + \omega - 1 = 0)$, где $C_{TV} = 0$, еще и две точки максимума для каждого значения N : (p_{\max}^*, ω^*) и $(1 - p_{\max}^*, 1 - \omega^*)$, которые лежат на кривой $f_3(p_{\max}, \omega)$.

В табл. 4 приведены оптимальные значения параметров формул (48) и (49), доставляющие максимум статистической сложности C_{TV} .

Таблица 4. Оптимальные параметры $C_{TV}(\omega, p_{\max})$ при различных значениях N

N	$C_{TV}(\omega^*, p_{\max}^*)$	p_{\max}^*	ω^*	$N - K^*$
3	0,1289	0,8241	0,6751	1 или 2
256	0,4789	0,9976	0,8752	32
512	0,5120	0,9991	0,8901	56
1024	0,5410	0,9997	0,9022	100
2048	0,5667	0,9999	0,9122	180

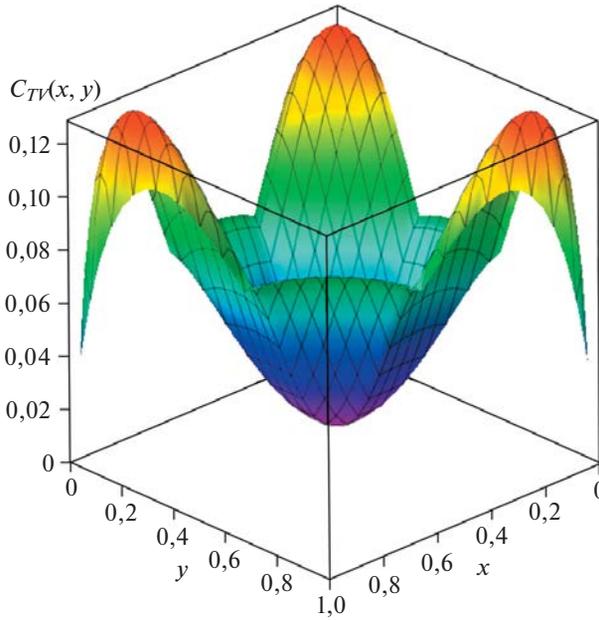


Рис. 9. Поверхности уровня статистической сложности $C_{TV}(x, y)$ для $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

Дополнительно случай $N = 3$ показан на рис. 9, где приведен график поверхности уровня C_{TV} , когда дискретное распределение $p = \{p_1, p_2, p_3\}$ состоит из трех отсчетов.

4. Моделирование и сравнение статистических сложностей

Проведем анализ оптимальных параметров, доставляющих максимум различным типам статистической сложности, и сравним значения в табл. 1, 3 и 4. Основной интерес представляют максимальные значения сложностей и оптимальные значения K . Максимальные значения $C_{TV}(\omega^*, p_{\max}^*) \in [0, 1]$, $C_{JSD}(\omega^*, p_{\max}^*) \in [0, 1]$ близки друг другу и растут с увеличением N . Оптимальные значения K для этих двух видов сложностей также близки.

Для демонстрации аналитических результатов, полученных в предыдущих разделах, показано применение трех вариантов статистической сложности в задаче индикации полезного сигнала в шумовой смеси для синтезированных сигналов. Применен алгоритм из [15], основанный на вычислении дискретных распределений p по спектральному представлению временных рядов.

Синтезированный сигнал длительностью 10 секунд представляет собой сумму конечного числа косинусоидальных колебаний в смеси с белым шумом:

$$(52) \quad x(t) = I(t) \sum_{i=1}^K A_i \cos(2\pi f_i t + \Delta\phi_i) + w(t), \quad t \in [0, 10],$$

где A_i , f_i , $\Delta\phi_i$ – амплитуды, частоты и случайные фазы гармонических колебаний соответственно, $w(t)$ – белый шум, а $I(t)$ – индикаторная функция наличия полезного сигнала в сигнально-шумовой смеси.

$I(t)$ выбрана таким образом, что гармонические сигналы присутствуют в середине итоговой последовательности $x(t)$.

$$(53) \quad I(t) = \begin{cases} 0, & t \in [0, 3), \\ 1, & t \in [3, 7], \\ 0, & t \in (7, 10]. \end{cases}$$

Работа алгоритма осуществляется следующим образом:

1. Синтезированный с частотой дискретизации f_s сигнал делится на короткие окна, содержащие по $N = 2048$ отсчетов.
2. Далее вычисляется спектр для каждого окна с использованием алгоритма БПФ.
3. На основе спектра путем его нормирования вычисляются дискретные плотности p_i , $i = 1, \dots, N$.
4. Для полученного набора p_i вычисляются информационные характеристики $C_{SQ}(p)$, $C_{JSD}(p)$, $C_{TV}(p)$.
5. Полученная последовательность значений информационных характеристик выводится вместе с сигналом по временной оси.

Стоит отметить, что параметры f_s и N подобраны так, чтобы исключить эффект растекания спектра, т.е. получить в спектре четкие дискреты, соответствующие K гармоническим функциям из формулы (52). Отношения амплитуд сигнала и дисперсии шума выбрано близким к единице.

Порог γ для решающего правила предлагается выбрать как 25% от максимального значения критерия для выбранного N из табл. 1, 3, 4:

$$(54) \quad \begin{aligned} \gamma_{CQ} &= 0,25 \cdot 0,1861 = 0,0465; \\ \gamma_{JSD} &= 0,25 \cdot 0,5312 = 0,1328; \\ \gamma_{TV} &= 0,25 \cdot 0,5667 = 0,1417. \end{aligned}$$

Удобство выбора такого порога заключается в том, что он не зависит от конкретной реализации шума и основан на аналитически полученных максимальных значениях функций статистической сложности.

На всех графиках синим цветом обозначена амплитуда исходного сигнала, а красным – статистическая сложность, которая рассчитана по алгоритму, описанному выше. Горизонтальная ось отвечает за время в секундах, а вертикальные – за величину амплитуды сигнала (слева) и критерия (справа). Черной пунктирной линией очерчено значение выбранного порога γ .

В первом эксперименте количество синусоидальных сигналов и соответственно дискрет в спектре равно $K = 3$ при $N = 2048$. На рис. 10 показаны зависимости статистических сложностей от времени для синтезированного сигнала.

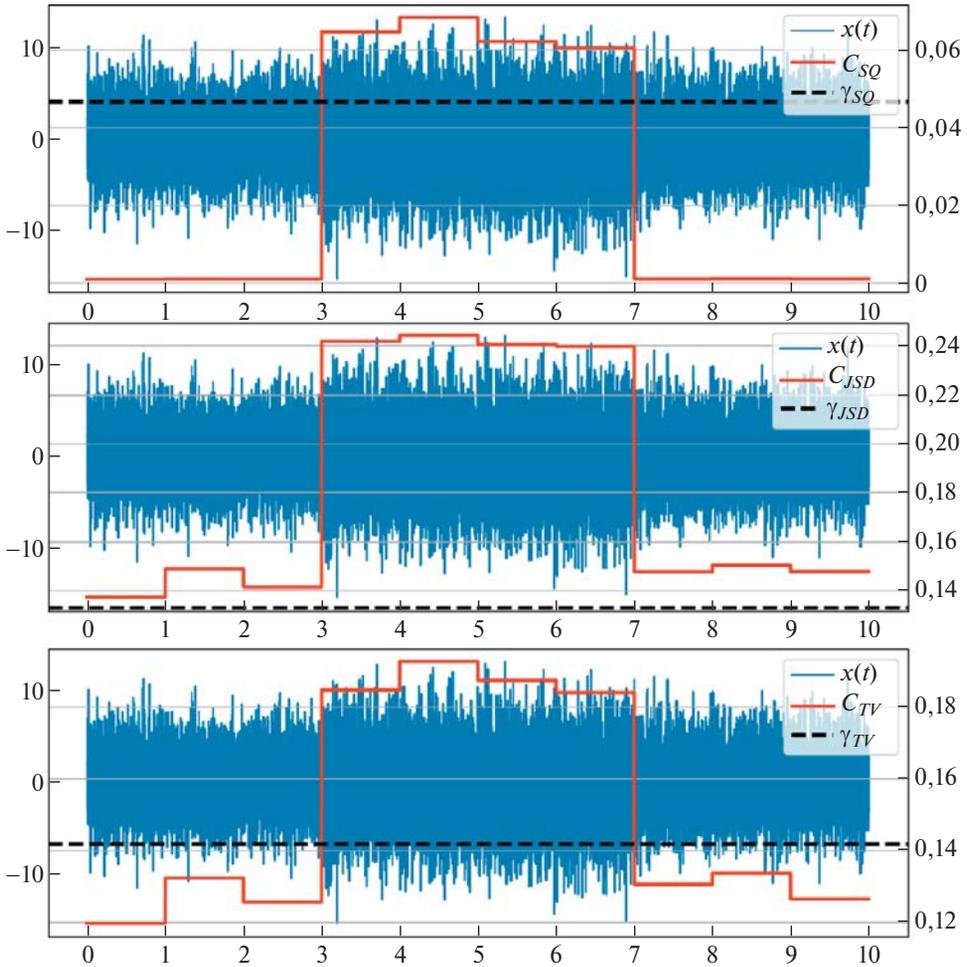


Рис. 10. Три дискреты, $K = 3$.

Как видно, значения C_{SQ} и C_{TV} превышают выбранный порог на интервале присутствия сигнала, что позволяет с уверенностью сделать вывод о его появлении. Что касается C_{JSD} , то априорный выбор порога оказался неудачен в силу того, что истинное значение ее максимума неизвестно, как показано в подразделе 3.2. Если изменить значение порога в большую сторону на 20%, то детектирование на основе C_{JSD} будет таким же успешным, как и на основе C_{TV} .

Во втором эксперименте количество дискрет в спектре $K = 30$. В этом случае C_{SQ} перестает показывать удовлетворительный результат в смысле превышения выбранного порога, поскольку функция C_{SQ} сильно деградирует с ростом K , однако все еще позволяет провести индикацию, как видно на рис. 11. Функция сложности C_{TV} все еще уверенно превышает порог, как и в первом эксперименте, а C_{JSD} превышает порог на всем сигнале, как и в первом эксперименте.

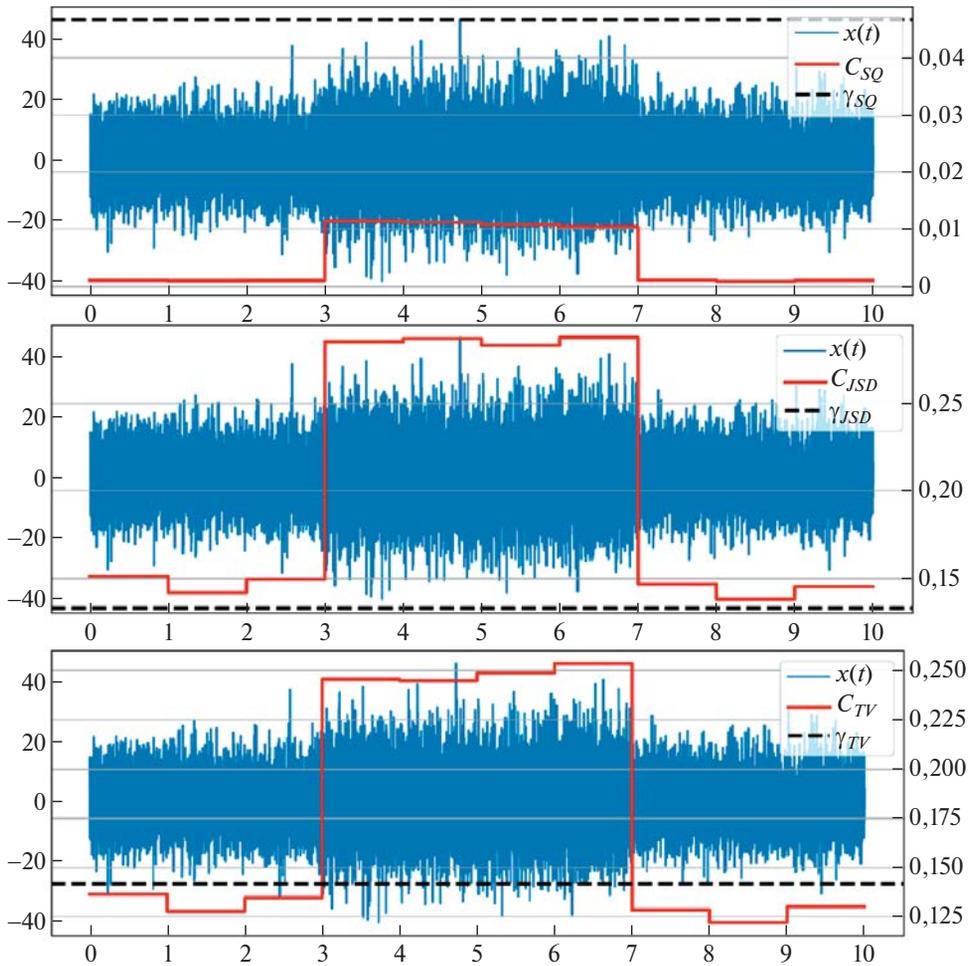


Рис. 11. Тридцать дискрет, $K = 30$.

Таким образом, можно сделать вывод о том, что C_{TV} наиболее удобна в практическом смысле, поскольку хорошо работает на сигналах с большим количеством дискрет в спектре и позволяет принимать решение о появлении полезного сигнала, основываясь на довольно простом правиле, связанном с выбором порога исходя из теоретически максимального значения для статистической сложности.

5. Заключение

В работе дано теоретическое обоснование использования статистической сложности в качестве критерия для решения задачи различения двух гипотез в условиях, когда вероятность ошибки их различения близка к единице. Рассмотрены три варианта вычисления статистической сложности для разных функций дисбаланса. Введены новые понятия дисбаланса и статисти-

стической сложности на основе полной вариации меры двух распределений. Проведено сравнение информационных критериев и найден класс дискретных распределений, на которых достигается максимум различных типов статистической сложности. Найденны значения максимумов при фиксированных значениях количества отсчетов распределения. Показано, что статистическая сложность C_{TV} на основе полной вариации непосредственно связана с задачей различения двух гипотез, а статистическая сложность C_{JSD} на основе энтропии Йенсена–Шеннона дает близкую оценку C_{TV} на выборочных распределениях. В свою очередь C_{SQ} наиболее перспективна для обнаружения отдельной дискретности над равномерным распределением. Предложен способ выбора порога для решающего правила обнаружения полезного сигнала, учитывающий максимальные полученные значения критериев, и показана результативность такого подхода на синтезированных сигналах.

Будущая работа будет посвящена исследованию информационных критериев на основе двух- и многомерных распределений, а также рассмотрению типовых акустических сигналов с реалистичным фоновым шумом.

СПИСОК ЛИТЕРАТУРЫ

1. *Shannon C.E.* A Mathematical Theory of Communication // Bell Syst. Tech. J. 1948. V. 27. P. 379–423.
2. *Gray R.M.* Entropy and Information Theory. New York: Springer, 2011.
<https://doi.org/10.1007/978-1-4419-7970-4>
3. *Holub A., Perona P., Burl M.C.* Entropy-based Active Learning for Object Recognition // Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE. 2008. P. 1–8.
<https://doi.org/10.1109/CVPRW.2008.4563068>
4. *Osisanwo F.Y., Akinsola J.E.T., Awodele O. et al.* Supervised Machine Learning Algorithms: Classification and Comparison // Int. J. Comput. Trends Technol. (IJCTT). 2017. V. 48. No. 3. P. 128–138.
<https://doi.org/10.14445/22312803/IJCTT-V48P126>
5. *Shen J., Hung J., Lee L.* Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments // Proc. 5th International Conference on Spoken Language Processing (ICSLP). 1998. <https://doi.org/10.21437/icslp.1998-527>
6. *Ribeiro M., Henriques T., Castro L., Souto A., Antunes L., Costa-Santos C., Teixeira A.* The Entropy Universe // Entropy. 2021. V. 222. No. 2. art. 222.
<https://doi.org/10.3390/e23020222>
7. *Ramirez J., Segura J.C., Benitez C. et al.* A New Kullback-Leibler VAD for Speech Recognition in Noise // IEEE Signal Proc. Lett. 2004. V. 11. No. 2. P. 266–269.
<https://doi.org/10.1109/LSP.2003.821762>
8. *Horie T., Burioka N., Amisaki T., Shimizu E.* Sample Entropy in Electrocardiogram During Atrial Fibrillation // Yonago Acta Medica. 2018. V. 61. No. 1. P. 49–57.
<https://doi.org/10.33160/yam.2018.03.007>
9. *Lamberti P.W., Martin M.T., Plastino A., Rosso O.A.* Intensive Entropic Non-Triviality Measure // Phys. A: Stat. Mech. Appl. 2004. V. 334. No. 1. P. 119–131.
<https://doi.org/10.1016/j.physa.2003.11.005>

10. *Lopez-Ruiz R.* Shannon Information, LMC Complexity and Renyi Entropies: A Straightforward Approach // *Biophys. Chem.* 2005. V. 115. No. 3. P. 215–218. <https://doi.org/10.1016/j.bpc.2004.12.035>
11. *Zunino L., Soriano M.C., Rosso O.A.* Distinguishing Chaotic and Stochastic Dynamics from Time Series by Using a Multiscale Symbolic Approach // *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.* 2012. V. 86. No. 4. P. 1–5. <https://doi.org/10.1103/PhysRevE.86.046210>
12. *Ronald L.A., Duncan W.M.* Signal Analysis: Time, Frequency, Scale, and Structure. N.J.: IEEE Press, 2004.
13. *Шуряев А.Н.* Вероятностно-статистические методы в теории принятия решений. М.: МЦНМО: НМУ, 2020.
14. *Kishan G.M., Chilukuri K.M., HuaMing Huang.* Anomaly Detection Principles and Algorithms. Cham: Springer. 2017. <https://doi.org/10.1007/978-3-319-67526-8>
15. *Berlin L.M., Galyaev A.A., Lysenko P.V.* Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments // *Sensors.* 2023. V. 23. No. 4. art. 2133. <https://doi.org/10.3390/s23042133>
16. *Johnson P., Moriarty J., Peskir G.* Detecting Changes in Real-Time Data: A User's Guide to Optimal Detection // *Philos. Trans. Royal Soc. A.* 2017. V. 375. P. 16. art. 2100. <https://doi.org/10.1098/rsta.2016.0298>
17. *Li Z., Li Y., Zhang K.A.* Feature Extraction Method of Ship-Radiated Noise Based on Fluctuation-Based Dispersion Entropy and Intrinsic Time-Scale Decomposition // *Entropy.* 2019. V. 21. No. 7. art. 693. <https://doi.org/10.3390/e21070693>
18. *Sason I.* On f-Divergences: Integral Representations, Local Behavior, and Inequalities // *Entropy.* 2018. V. 20. No. 5. art. 383. <https://doi.org/10.3390/e20050383>

Статья представлена к публикации членом редколлегии А.В. Назиньм.

Поступила в редакцию 21.03.2023

После доработки 19.04.2023

Принята к публикации 11.05.2023

Оптимизация, системный анализ и исследование операций

© 2023 г. Э.Х. ГИМАДИ, д-р физ.-мат. наук (gimadi@math.nsc.ru)

(Институт математики им. С.Л. Соболева СО РАН, Новосибирск),

А.А. ШТЕПА (shoomath@gmail.com)

(Новосибирский национальный исследовательский государственный университет)

ОБ АСИМПТОТИЧЕСКОЙ ТОЧНОСТИ ПОИСКА МИНИМУМА СУММЫ ВЕСОВ РАЗНОРЕБЕРНЫХ ОСТОВНЫХ ДЕРЕВЬЕВ ФИКСИРОВАННОГО ДИАМЕТРА¹

Рассматривается труднорешаемая задача поиска нескольких реберно-непересекающихся (разнореберных) остовных деревьев минимального суммарного веса с фиксированным диаметром в полном неориентированном графе со случайными весами ребер из нескольких классов непрерывных распределений: равномерное, смещенное усеченно-экспоненциальное, смещенное усеченно-нормальное. Для решения этой задачи предлагается приближенный алгоритм с трудоемкостью $O(n^2)$, где n — количество вершин в графе. Приводятся условия асимптотической точности для этого алгоритма в случае каждого из рассматриваемых вероятностных распределений.

Ключевые слова: минимальное остовное дерево с ограниченным диаметром, приближенный алгоритм, вероятностный анализ, асимптотическая точность.

DOI: 10.31857/S0005231023070085, EDN: FDWYOE

1. Введение

Задача поиска минимального остовного дерева (Minimum Spanning Tree, MST) — одна из хорошо известных задач дискретной оптимизации. Она состоит в нахождении остовного дерева (связного ациклического подграфа на всех вершинах) минимального веса в данном реберно-взвешенном графе $G = (V, E)$. Полиномиальная разрешимость этой задачи была доказана построением полиномиальных алгоритмов Борувки (1926), Краскала (1956) и Прима (1957). Эти алгоритмы имеют трудоемкости $O(u \log n)$, $O(u \log u)$ и $O(n^2)$ соответственно, где $u = |E|$ и $n = |V|$. Интересно отметить, что математическое ожидание веса MST в графе со случайными весами ребер может быть неожиданно маленьким. Например, для полного графа с весами ребер из класса равномерно распределенных случайных величин на интервале $(0; 1)$

¹ Работа выполнена в рамках государственного задания Института математики им. С.Л. Соболева СО РАН (проект FWNF-2022-0019).

вес MST с высокой вероятностью близок к константе 2,02 [1]. Похожие результаты получены в [2, 3].

Одним из возможных обобщений вышеуказанной задачи является диаметр-ограниченный вариант задачи поиска MST. Диаметр дерева — это количество ребер в самом длинном простом пути в дереве, соединяющем пару вершин. Эта задача состоит в следующем: дан реберно-взвешенный граф и параметр $d = d_n$, необходимо найти в этом графе MST, имеющее диаметр, ограниченный сверху или снизу параметром d . Обе задачи в общей постановке являются *NP*-трудными.

Вариант задачи с ограничением на диаметр сверху является полиномиально разрешимым для значений диаметра два и три, но *NP*-трудным для любого диаметра от 4 до $(n - 1)$, даже для весов ребер, равных 1 или 2 [4, стр. 206]. Задача с ограничением снизу является *NP*-трудной, потому что она содержит в качестве подзадачи для $d = n - 1$ задачу “гамильтонов путь” [4].

Недавно авторы данной статьи начали изучать другую модификацию задачи поиска MST с ограниченным диаметром, когда диаметр этого дерева равен заданному числу. Примечательно, что алгоритм решения такой задачи может быть преобразован в алгоритм для решения задачи с диаметром, ограниченным как сверху, так и снизу. Тем самым область применения такой задачи покрывает область применения задач с ограничением диаметра как сверху, так и снизу.

Для задачи поиска минимального остовного дерева с диаметром, ограниченным сверху, известны приложения в беспроводных сетях [5], в разработке коммуникационных сетей [6], в алгоритмах сжатия информации [7] и распределенного взаимного исключения [8] (подробное описание см., например, в [9]).

Задача отыскания нескольких реберно-непересекающихся остовных деревьев минимального суммарного веса с диаметром, ограниченным снизу, в полном графе может встретиться в проблемах теории надежности коммуникационных сетей, когда необходимо для множества выбранных объектов построить t -связный минимальный по весу граф, исключив при этом конфигурации, для которых после выхода из строя небольшого количества вершин графа резко ухудшается структура сети, т.е. наложить условие на диаметр снизу для каждого из конструируемых остовных деревьев, формирующих t -связный граф. Стоит отметить, что в [10, 11] проведен вероятностный анализ приближенного алгоритма для этой задачи и получены условия его асимптотической точности.

В [12, 13] представлен вероятностный анализ полиномиального алгоритма и предложены условия его асимптотической точности для решения задачи поиска одного и нескольких MST с фиксированным диаметром дерева в случае полного ориентированного графа. К сожалению, анализ этого алгоритма оказывается неприемлемым для случая неориентированного графа, поскольку вероятностные свойства весов ребер, рассматриваемых по ходу работы

алгоритма, нельзя считать независимыми, в отличие от случая ориентированного графа.

В данной статье рассматривается задача нахождения m реберно-непересекающихся (разнореберных) остовных деревьев минимального суммарного веса с фиксированным диаметром $d = d_n$ в полном неориентированном графе (обозначим эту задачу как m - d -UMST). Предлагаются приближенный полиномиальный алгоритм решения этой задачи и достаточные условия его асимптотической точности. Вероятностный анализ алгоритма проведен в случае полного реберно-взвешенного неориентированного графа G без петель в предположении, что веса ребер — положительно значащие независимые случайные величины. Рассматриваются функции распределения (ф.р.) весов ребер графа G трех классов: $\text{UNI}(a_n; b_n)$ — равномерное распределение на конечном отрезке $[a_n; b_n]$, а также усеченно-смещенные распределения: экспоненциальное $\text{EXP}(a_n, \lambda_n)$ и нормальное $\text{NORM}(a_n, \sigma_n)$ на не ограниченном сверху полуинтервале $[a_n; \infty)$. Функции плотности этих распределений имеют вид

$$(1) \quad p(x) = \begin{cases} \frac{1}{b_n - a_n} & \text{при } a_n \leq x \leq b_n \text{ для } \text{UNI}(a_n; b_n); \\ \frac{1}{\lambda_n} \exp\left(-\frac{x - a_n}{\lambda_n}\right) & \text{при } a_n \leq x < \infty \text{ для } \text{EXP}(a_n, \lambda_n); \\ \frac{2}{\sigma_n \sqrt{2\pi}} \exp\left(-\frac{(x - a_n)^2}{2\sigma_n^2}\right) & \text{при } a_n \leq x < \infty \text{ для } \text{NORM}(a_n, \sigma_n); \\ 0 & \text{иначе.} \end{cases}$$

2. Нахождение нескольких реберно-непересекающихся остовных деревьев минимального суммарного веса с фиксированным диаметром в неориентированном графе

В данном разделе сначала формулируется исследуемая задача, а затем предлагается приближенный алгоритм ее решения.

Пусть даны полный n -вершинный реберно-взвешенный неориентированный граф $G = (V, E)$ и положительные целые числа $m \geq 2$, $d \geq 4$ такие, что $m(d + 1) \leq n$. Задача m - d -UMST состоит в отыскании m реберно-непересекающихся остовных деревьев T_1, \dots, T_m таких, что диаметр каждого из них равен $d = d_n$ и их суммарный вес минимален. Для решения этой задачи предлагается следующий детерминированный алгоритм.

Описание алгоритма \mathcal{A}

Предварительный шаг 0. В графе G выбрать произвольное $(n - m(d + 1))$ -вершинное подмножество V' и произвольно разбить оставшиеся $m(d + 1)$ вершин на m подмножеств с $(d + 1)$ вершиной в каждом V_1, V_2, \dots, V_m .

Шаг 1. В каждом подграфе $G(V_s)$, $s = 1, \dots, m$, начиная с произвольной его вершины, построить $(d + 1)$ -вершинный гамильтонов путь P_s , используя жадную эвристику “иди в ближайшую непосещенную вершину”.

Положить $T_s = P_s$, $s = 1, \dots, m$.

Шаг 2. Далее предполагается без потери общности, что параметр d нечетен (см. замечание 1 ниже). Каждую пару путей P_i и P_j , $1 \leq i < j \leq m$, соединить специальным образом так, чтобы построенный подграф состоял из двух $2(d + 1)$ -вершинных реберно-непересекающихся поддеревьев с диаметром, равным d . Каждый путь P_s , $1 \leq s \leq m$, рассматривается в виде двух половин (подпутей) P_s^1 и P_s^2 , каждая из которых содержит одну *концевую* вершину и $\frac{d-1}{2}$ *внутренних* вершин, суммарно $\frac{d+1}{2}$ вершин в каждой половине.

Построение множества ребер, соединяющих пути P_i и P_j , описано в следующих пунктах 2.1–2.6.

2.1. Соединить каждую внутреннюю вершину P_i^1 кратчайшим ребром с внутренней вершиной P_j^1 и добавить это ребро в T_j .

2.2. Соединить каждую внутреннюю вершину P_i^2 кратчайшим ребром с внутренней вершиной P_j^2 и добавить это ребро в T_j .

2.3. Соединить каждую внутреннюю вершину P_j^1 кратчайшим ребром с внутренней вершиной P_i^2 и добавить это ребро в T_i .

2.4. Соединить каждую внутреннюю вершину P_j^2 кратчайшим ребром с внутренней вершиной P_i^1 и добавить это ребро в T_i .

2.5. Соединить каждую концевую вершину пути P_i кратчайшим ребром с внутренней вершиной пути P_j и добавить это ребро в T_j .

2.6. Соединить каждую концевую вершину пути P_j кратчайшим ребром с внутренней вершиной пути P_i и добавить это ребро в T_i .

Шаг 3. В цикле по $s = 1, \dots, m$ каждую вершину подграфа $G(V')$ соединить кратчайшим ребром с внутренней вершиной пути P_s и добавить это ребро в соответствующее дерево T_s .

Построение всех m реберно-непересекающихся остовных деревьев T_1, \dots, T_m завершено (см. пример на рис. 1–3).

Замечание 1. В случае четного параметра d нужно немного модифицировать алгоритм. На шаге 1 для каждого пути для первой выбранной вершины необходимо найти ближайшую вершину v_s , $s = 1, \dots, m$ за d действий, эту вершину запомнить, а первую выбранную вершину для каждого пути “отложить” в сторону. Далее, выполнить все шаги алгоритма для $d' = d - 1$, где первая вершина каждого пути — это найденная вершина v_s , и после выполнения шага 3 соединить “отложенные” вершины с уже найденными ближайшими к ним вершинами v_s , $s = 1, \dots, m$. В итоге создаются остовные деревья, диаметр каждого из которых равен в точности d , при этом оценка для трудоемкости алгоритма останется прежней.

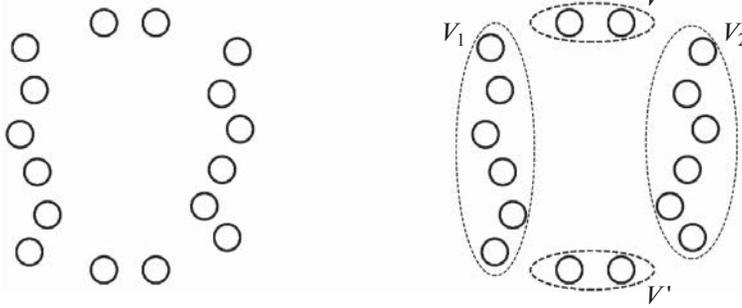


Рис. 1. Изначальные вершины графа и шаг 0 работы алгоритма \mathcal{A} на 16-вершинном полном графе, $m = 2$, $d = 5$.

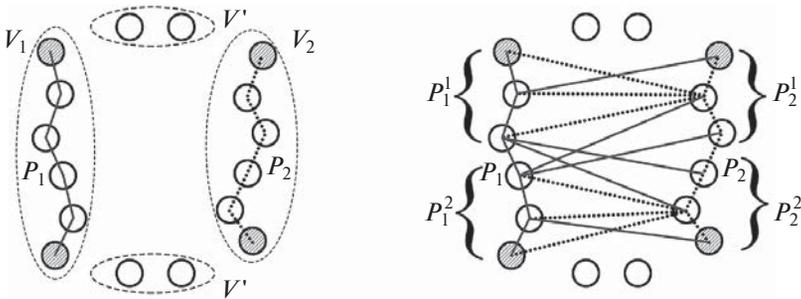


Рис. 2. Шаги 1 и 2 работы алгоритма \mathcal{A} на 16-вершинном полном графе, $m = 2$, $d = 5$. Заштрихованные вершины — это концевые вершины. Сплошные ребра принадлежат дереву T_1 , точечные ребра — дереву T_2 .

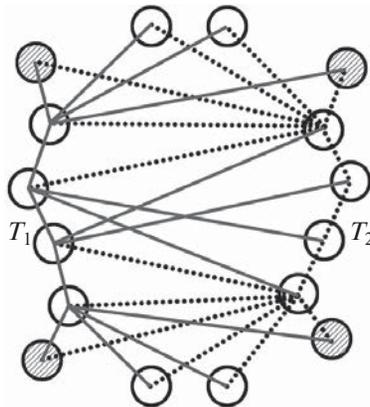


Рис. 3. Шаг 3 работы алгоритма \mathcal{A} на 16-вершинном полном графе, $m = 2$, $d = 5$. Заштрихованные вершины — это концевые вершины. Сплошные ребра принадлежат дереву T_1 , точечные ребра — дереву T_2 .

Введем обозначения: $W_{\mathcal{A}}$ — общий вес всех остовных деревьев T_1, \dots, T_m , построенных алгоритмом \mathcal{A} , а W_1, W_2 и W_3 — это суммарные веса ребер, полученные на шагах 1, 2 и 3 соответственно, тогда $W_{\mathcal{A}} = W_1 + W_2 + W_3$.

Сформулируем два утверждения, касающиеся алгоритма \mathcal{A} .

Утверждение 1. Алгоритм \mathcal{A} строит допустимое решение задачи m - d -UMST.

Утверждение 2. Время работы алгоритма \mathcal{A} оценивается величиной $\mathcal{O}(n^2)$.

3. Вероятностный анализ алгоритма \mathcal{A}

Через $F_{\mathcal{A}}(I)$ и $OPT(I)$ обозначим соответственно приближенное (полученное с помощью некоторого алгоритма \mathcal{A}) и оптимальное значение целевой функции задачи на входе I .

Определение 1. Алгоритм \mathcal{A} имеет гарантированные оценки $(\varepsilon_n, \delta_n)$ качества работы на множестве случайных входов задачи размера n (где n — параметр, характеризующий объем входных данных в описании задачи, см. [4]), если

$$(2) \quad \mathbb{P}\{|F_{\mathcal{A}}(I) - OPT(I)| > \varepsilon_n OPT(I)\} \leq \delta_n,$$

где $\varepsilon_n = \varepsilon_{\mathcal{A}}(n)$ — оценка относительной погрешности решения, полученного с помощью алгоритма \mathcal{A} , $\delta_n = \delta_{\mathcal{A}}(n)$ — оценка вероятности несрабатывания этого алгоритма, которая равна доле случаев, когда алгоритм не гарантирует решения с анонсированной относительной погрешностью ε_n или вовсе не получает какого-либо решения для задачи.

Определение 2 [14]. Приближенный алгоритм \mathcal{A} называется асимптотически точным на классе входных данных задачи, если существуют гарантированные оценки качества работы этого алгоритма такие, что для всех входов I размера n

$$\varepsilon_n \rightarrow 0 \quad \text{и} \quad \delta_n \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty.$$

Далее, случайную величину, равную минимуму среди k случайных независимых одинаково распределенных величин η , обозначим через η_k .

Согласно работе алгоритма для шагов 1–3 справедливы соотношения:

$W_1 = \sum_{s=1}^m \sum_{k=1}^d \eta_k$, так как на шаге 1 строятся m путей P_1, \dots, P_m с d ребрами в каждом.

$W_2 = C_m^2 (4 \frac{d-1}{2} \eta_{(d-1)/2} + 4\eta_{(d-1)})$, так как для каждой из $C_m^2 = \frac{m(m-1)}{2}$ пар путей присоединение новых ребер к искомому множеству остовных деревьев происходит в соответствующих пунктах 2.1–2.6 шага 2 следующим образом:

— во-первых, каждая из $\frac{d-1}{2}$ внутренних вершин одной половины пути соединяется кратчайшим ребром с одной из $\frac{d-1}{2}$ внутренних вершин половины другого пути;

— во-вторых, концевая вершина каждого пути соединяется кратчайшим ребром с одной из $d - 1$ внутренних вершин другого пути.

Множитель 4 возникает, поскольку для каждой пары путей в пунктах 2.1–2.4 внутренние вершины из двух половин одного пути соединяются с внутренними вершинами из двух половин другого пути, а в пунктах 2.5–2.6 соответствующие кратчайшие ребра присоединяются к каждой из 4 концевых вершин рассматриваемой пары путей.

$W_3 = m(n - m(d + 1))\eta_{(d-1)}$, так как на шаге 3 каждая из $n - m(d + 1)$ вершин множества V' соединяется кратчайшими ребрами с внутренними вершинами каждого из m путей P_s , $1 \leq s \leq m$.

Замечание 2. Стоит отметить, что в случае четного параметра d :

$$W_1 = m \sum_{k=1}^{d'} \eta_k, \quad W_2 = C_m^2 \left(4 \frac{d' - 1}{2} \eta_{(d'-1)/2} + 4\eta_{(d'-1)} \right),$$

$$W_3 = m(n - m(d' + 1)) \eta_{(d'-1)},$$

где $d' = d - 1 \geq 3$. Таким образом, заменив везде d на d' , можно провести тот же вероятностный анализ и доказать все предложенные утверждения.

В дальнейшем перейдем от случайных величин η для каждого из рассматриваемых распределений к *нормализованным* случайным величинам, т.е. величинам $\xi = \frac{\eta - a_n}{\beta_n}$, где $\beta_n = \begin{cases} b_n - a_n & \text{на входах UNI}(a_n; b_n); \\ \lambda_n & \text{на входах EXP}(a_n, \lambda_n); \\ \sigma_n & \text{на входах NORM}(a_n, \sigma_n). \end{cases}$

Рассмотрим случайные величины W_1, W_2, W_3 :

$$W_1 = m \sum_{k=1}^d \eta_k = m \sum_{k=1}^d (\beta_n \xi_k + a_n) = mda_n + \beta_n m \sum_{k=1}^d \xi_k = mda_n + \beta_n W'_1,$$

$$W_2 = C_m^2 \left(4 \frac{d-1}{2} \eta_{(d-1)/2} + 4\eta_{(d-1)} \right) =$$

$$= C_m^2 \left(4 \frac{d-1}{2} (\beta_n \xi_{(d-1)/2} + a_n) + 4(\beta_n \xi_{(d-1)} + a_n) \right) =$$

$$= m(m-1)(d+1)a_n + \beta_n m(m-1) \left((d-1)\xi_{(d-1)/2} + 2\xi_{(d-1)} \right) =$$

$$= (m^2(d+1) - md - m) a_n + \beta_n W'_2,$$

$$W_3 = m(n - m(d+1))\eta_{(d-1)} = m(n - m(d+1)) (\beta_n \xi_{(d-1)} + a_n) =$$

$$= m(n - m(d+1))a_n + \beta_n m(n - m(d+1))\xi_{(d-1)} =$$

$$= (mn - m^2(d+1)) a_n + \beta_n W'_3,$$

где W'_1, W'_2, W'_3 — нормализованные случайные величины для W_1, W_2, W_3 соответственно, а β_n — параметр распределения.

Получается следующее соотношение для суммы весов построенных остовных деревьев: $W_{\mathcal{A}} = m(n-1)a_n + \beta_n W'_{\mathcal{A}}$, где $W'_{\mathcal{A}} = W'_1 + W'_2 + W'_3$.

Лемма 1. Алгоритм \mathcal{A} для решения задачи m - d -UMST на n -вершинном полном графе с весами ребер из соответствующего класса распределения ($UNI(a_n; b_n)$, $EXP(a_n, \lambda_n)$ или $NORM(a_n, \sigma_n)$) является алгоритмом со следующими оценками относительной погрешности ε_n и вероятности несрабатывания δ_n :

$$(3) \quad \varepsilon_n = \frac{2\beta_n}{m(n-1)a_n} \widehat{\mathbb{E}W'_A}, \quad \delta_n = \mathbb{P}\{\widetilde{W}'_A > \widehat{\mathbb{E}W'_A}\},$$

где β_n — параметр соответствующего распределения, $\widehat{\mathbb{E}W'_A}$ — это некоторая верхняя оценка для математического ожидания $\mathbb{E}W'_A$, $\widetilde{W}'_A = W'_A - \mathbb{E}W'_A$.

В дальнейшем для вероятностного анализа алгоритма \mathcal{A} оказывается полезным следующее утверждение из теории вероятностей.

Теорема 1 [15]. Рассмотрим независимые случайные величины X_1, \dots, \dots, X_n . Зададим положительные константы T и h_1, \dots, h_n такие, что для всех $k = 1, \dots, n$ и $0 \leq t \leq T$ верно неравенство

$$(4) \quad \mathbb{E}e^{tX_k} \leq e^{\frac{h_k t^2}{2}}.$$

Пусть $S = \sum_{k=1}^n X_k$ и $H = \sum_{k=1}^n h_k$. Тогда

$$\mathbb{P}\{S > x\} \leq \begin{cases} \exp\left\{-\frac{x^2}{2H}\right\}, & \text{если } 0 \leq x \leq HT, \\ \exp\left\{-\frac{Tx}{2}\right\}, & \text{если } x \geq HT. \end{cases}$$

Также для дальнейшего анализа будет полезным следующее

Утверждение 3. Для любого целого $d \geq 3$ верна оценка сверху

$$\sum_{k=1}^d \frac{1}{k} \leq \ln d + \frac{3}{4}.$$

Параметр d предполагается нечетным и определенным на множестве значений двух полуинтервалов: случай 1 ($\ln n \leq d < \frac{n}{\ln n}$) и случай 2 ($\frac{n}{\ln n} \leq d < \frac{n}{m}$).

3.1. Распределение $UNI(a_n; b_n)$

Для распределения $UNI(a_n; b_n)$ перейдем от величин η и η_k к нормализованным случайным величинам $\xi = \frac{\eta - a_n}{b_n - a_n}$ и $\xi_k = \frac{\eta_k - a_n}{b_n - a_n}$.

Лемма 2. Для $\mathbb{E}W'_A$ верно неравенство

$$\mathbb{E}W'_A \leq m \ln d + \frac{2mn}{d}.$$

Лемма 3. Пусть заданы константы $T = 1$ и $h_k = \frac{1}{(k+1)^2}$. Тогда в рамках распределения $\text{UNI}(a_n; b_n)$ для смещенных случайных величин $\tilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$ верны неравенства $\mathbb{E}e^{t\tilde{\xi}_k} \leq e^{\frac{h_k t^2}{2}}$ в теореме Петрова [15, с. 54–55] для каждого $0 \leq t \leq T$ и $1 \leq k \leq d$.

Лемма 4. В случае $\ln n \leq d < \frac{n}{m}$ для суммы констант $h_k = \frac{1}{(k+1)^2}$, которые соответствуют добавленным ребрам в создаваемые деревья, верна верхняя оценка

$$H \leq \frac{mn}{d}.$$

Лемма 5. В случае $\ln n \leq d < \frac{n}{\ln n}$ верна следующая верхняя оценка:

$$\mathbb{E}W'_A \leq \frac{3mn}{d} = \widehat{\mathbb{E}W'_A}.$$

Лемма 6. Для случая $\frac{n}{\ln n} \leq d < \frac{n}{m}$ верно неравенство

$$\mathbb{E}W'_A \leq 3m \ln n = \widehat{\mathbb{E}W'_A}.$$

С помощью предыдущих лемм можно доказать главный результат этого подраздела.

Теорема 2. Пусть параметр $d = d_n$ определен так, что $\ln n \leq d < \frac{n}{m}$. Тогда алгоритм \mathcal{A} для решения задачи m - d -UMST с весами ребер из $\text{UNI}(a_n; b_n)$ асимптотически точен с вероятностью несрабатывания $\delta_n = n^{-m} \rightarrow 0$ при $n \rightarrow \infty$ и следующими условиями на разброс весов ребер графа G :

$$(5) \quad \frac{b_n}{a_n} = \begin{cases} o(d), & \text{если } \ln n \leq d < \frac{n}{\ln n}, \\ o\left(\frac{n}{\ln n}\right), & \text{если } \frac{n}{\ln n} \leq d < \frac{n}{m} \text{ и } m < \ln n. \end{cases}$$

3.2. Распределение $\text{EXP}(a_n, \lambda_n)$

От случайных величин η, η_k перейдем к нормализованным случайным величинам $\xi = \frac{\eta - a_n}{\lambda_n}$ и $\xi_k = \frac{\eta_k - a_n}{\lambda_n}$. В терминах этих переменных запишем ф.р. $\mathfrak{P}_\xi(x) = 1 - e^{-x}$ и плотность распределения согласно (1)

$$p(\xi) = \begin{cases} e^{-\xi}, & \text{если } 0 \leq \xi < \infty, \\ 0 & \text{иначе} \end{cases}$$

случайной величины ξ , а также ф.р. случайной величины ξ_k :

$$(6) \quad \mathfrak{P}_{\xi_k}(x) = 1 - (1 - \mathfrak{P}_\xi(x))^k.$$

Лемма 7. Математическое ожидание случайной величины ξ_k равно $\mathbb{E}\xi_k = 1/k$.

Лемма 8. В случае распределения $\text{EXP}(a_n, \lambda_n)$ верна следующая верхняя оценка для математического ожидания веса решения, полученного алгоритмом \mathcal{A} :

$$\mathbb{E}W'_A \leq m \ln d + \frac{2mn}{d-1} = \widehat{\mathbb{E}W'_A}.$$

Лемма 9. Пусть $T = \frac{1}{2}$, $h_k = \frac{3}{k^2}$. Тогда при любом $1 \leq k \leq d$ и $0 \leq t \leq T$ для смещенных случайных величин $\tilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$ верны условия $\mathbb{E}e^{t\tilde{\xi}_k} \leq e^{\frac{h_k t^2}{2}}$ теоремы Петрова [15, с. 54–55].

Лемма 10. Пусть $\ln n \leq d < \frac{n}{m}$. Тогда при достаточно больших n верна следующая верхняя оценка для суммы констант $h_k = \frac{3}{k^2}$, которые соответствуют добавленным ребрам в создаваемые деревья:

$$H \leq \frac{3mn}{d-1}.$$

Лемма 11. В случае $\ln n \leq d < \frac{n}{\ln n}$ верна верхняя оценка

$$\mathbb{E}W'_A \leq \frac{3mn}{d-1} = \widehat{\mathbb{E}W'_A}.$$

Лемма 12. Для случая $\frac{n}{\ln n} \leq d < \frac{n}{m}$ верно следующее неравенство:

$$\mathbb{E}W'_A \leq 5m \ln n = \widehat{\mathbb{E}W'_A}.$$

Используя введенные выше леммы, можно доказать следующую теорему.

Теорема 3. Пусть параметр $d = d_n$ определен так, что $\ln n \leq d < \frac{n}{m}$. Тогда алгоритм \mathcal{A} для решения задачи m - d -UMST с весами ребер из $\text{EXP}(a_n, \lambda_n)$ асимптотически точен с вероятностью несрабатывания $\delta_n = n^{-m} \rightarrow 0$ при $n \rightarrow \infty$, если имеют место следующие условия на разброс весов ребер графа G :

$$(7) \quad \frac{\lambda_n}{a_n} = \begin{cases} o(d), & \text{если } \ln n \leq d < \frac{n}{\ln n}, \\ o\left(\frac{n}{\ln n}\right), & \text{если } \frac{n}{\ln n} \leq d < \frac{n}{m} \text{ и } m < \ln n. \end{cases}$$

3.3. Распределение $\text{NORM}(a_n, \sigma_n)$

Для распределения $\text{NORM}(a_n, \sigma_n)$ вместо случайных величин η и η_k для соответствующих весов графа введем нормализованные случайные величины $\xi = \frac{\eta - a_n}{\sigma_n}$ и $\xi_k = \frac{\eta_k - a_n}{\sigma_n}$.

Для случайной величины ξ верны следующие плотность распределения согласно (1) и ф.р.:

$$p(\xi) = \begin{cases} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{\xi^2}{2}\right), & \text{если } 0 \leq \xi < \infty, \\ 0 & \text{иначе.} \end{cases} \quad \mathfrak{F}(x) = \sqrt{\frac{2}{\pi}} \int_0^x \exp\left(-\frac{u^2}{2}\right) du.$$

Определение 3. Говорят, что ф.р. $\mathfrak{F}_1(x)$ мажорирует ф.р. $\mathfrak{F}_2(x)$, если $\mathfrak{F}_1(x) \geq \mathfrak{F}_2(x)$ для всех x .

Утверждение 4. Ф.р. $\mathfrak{F}(x)$ нормальной случайной величины с параметром σ_n мажорирует экспоненциальную ф.р. с параметром $\lambda_n = 2\sigma_n$:

$$(8) \quad \mathfrak{F}(x) \geq \mathfrak{P}(x/2) \quad \forall x \geq 0.$$

Лемма 13 [16]. Пусть χ_1, \dots, χ_k — независимые одинаково распределенные случайные величины с ф.р. $F(x)$, а $\widehat{F}(x)$ — ф.р. случайной величины $\chi = \min_{i=1, \dots, k} \chi_i$, ζ_1, \dots, ζ_k — независимые одинаково распределенные случайные величины с ф.р. $G(x)$, аналогично, $\widehat{G}(x)$ — ф.р. случайной величины $\zeta = \min_{i=1, \dots, k} \zeta_i$. Тогда при любом x

$$F(x) \leq G(x) \Rightarrow \widehat{F}(x) \leq \widehat{G}(x).$$

Лемма 14 [16]. Пусть $P_\vartheta, P_\omega, P_\zeta, P_\chi$ — ф.р. случайных величин $\vartheta, \omega, \zeta, \chi$ соответственно, причем ϑ и ζ независимы, а также ω и χ независимы. Тогда

$$(\forall x P_\vartheta(x) \leq P_\omega(x)) \wedge (\forall y P_\zeta(y) \leq P_\chi(y)) \Rightarrow (\forall z P_{\vartheta+\zeta}(z) \leq P_{\omega+\chi}(z)).$$

Лемма 15 [16]. Пусть ф.р. $F(x)$ и $P(x)$ случайных величин таковы, что $F(x) \geq P(x)$. Тогда для алгоритма \mathcal{A} на входах с ф.р. $F(x)$ справедливы те же оценки качества $(\varepsilon_{\mathcal{A}}, \delta_{\mathcal{A}})$, что и в случае входов с ф.р. $P(x)$.

Положим $F(x) = \mathfrak{F}(x)$ и $P(x) = \mathfrak{P}(x/2)$. Из утверждения 4 и лемм 13–15 для смещенного усеченно-нормального распределения следует

Теорема 4. Пусть параметр $d = d_n$ определен так, что $\ln n \leq d < \frac{n}{m}$. Тогда алгоритм \mathcal{A} для решения задачи m - d -UMST на n -вершинном полном неориентированном графе с весами ребер, принимающими значения из неограниченного сверху полуинтервала $[a_n; \infty)$ согласно $\text{NORM}(a_n, \sigma_n)$, асимптотически точен с вероятностью несрабатывания $\delta_n = n^{-m} \rightarrow 0$ при $n \rightarrow \infty$ и следующими условиями на разброс весов ребер графа G :

$$\frac{\sigma_n}{a_n} = \begin{cases} o(d), & \text{если } \ln n \leq d < \frac{n}{\ln n}, \\ o\left(\frac{n}{\ln n}\right), & \text{если } \frac{n}{\ln n} \leq d < \frac{n}{m} \text{ и } m < \ln n. \end{cases}$$

4. Заключение

В этой работе представлен детерминированный приближенный алгоритм, который за время $\mathcal{O}(n^2)$, где n — количество вершин в графе, решает задачу минимизации суммарного веса нескольких реберно-непересекающихся остовных деревьев с фиксированным диаметром в реберно-взвешенном полном неориентированном графе. Проведен анализ этого алгоритма в случае распределения весов ребер графа согласно равномерному $\text{UNI}(a_n; b_n)$ и усеченно-смещенным распределениям вероятности: экспоненциальному $\text{EXP}(a_n, \lambda_n)$ и нормальному $\text{NORM}(a_n, \sigma_n)$. Получены достаточные условия асимптотической точности этого алгоритма для каждого из распределений. Интересным

представляется изучение задачи на входных данных с дискретными распределениями. Также было бы полезно рассмотреть задачу поиска нескольких реберно-непересекающихся остовных деревьев максимального суммарного веса с фиксированным или ограниченным диаметром.

ПРИЛОЖЕНИЕ

Доказательство утверждения 1. Каждая из полученных реберно-непересекающихся конструкций состоит из n вершин и $(n - 1)$ ребер, так как сначала на шаге 1 строится $(d + 1)$ -вершинный путь и затем на шагах 2 и 3 к нему присоединяются все остальные вершины графа, не увеличивая диаметр построенного остовного дерева. В конце получается m таких конструкций — остовных деревьев, представляющих допустимое решение задачи m - d -UMST.

Доказательство утверждения 2. Предварительный шаг 0 требует $\mathcal{O}(n)$ элементарных операций.

На шаге 1 каждый путь строится за время $\mathcal{O}(d^2)$, таким образом, полностью этот шаг выполняется за время $\mathcal{O}(md^2)$ или $\mathcal{O}(nd)$ (так как $m(d+1) \leq n$).

На пп. 2.1–2.4 каждая пара путей (P_i, P_j) , $1 \leq i < j \leq m$ соединяется за время $\mathcal{O}(d^2)$, и для всех $\frac{m(m-1)}{2}$ пар путей требуется $\mathcal{O}(m^2d^2)$ действий или $\mathcal{O}(n^2)$ элементарных операций.

Пп. 2.5–2.6 выполняются с трудоемкостью $\mathcal{O}(md)$.

Шаг 3 требует времени $\mathcal{O}(mdn)$ или $\mathcal{O}(n^2)$ на соединение $|G(V')| < n$ вершин кратчайшим ребром с внутренними вершинами путей P_s в каждом остовном дереве T_s , $1 \leq s \leq m$.

Таким образом, общая трудоемкость алгоритма \mathcal{A} оценивается величиной $\mathcal{O}(n^2)$.

Доказательство леммы 1. Запишем неравенство (2) для гарантированных оценок качества алгоритма применительно к рассматриваемому случаю задачи на минимум.

$$\begin{aligned}
 \mathbb{P}\{W_{\mathcal{A}} - OPT(I) > \varepsilon_n OPT\} &= \mathbb{P}\{W_{\mathcal{A}} > (1 + \varepsilon_n)OPT\} \leq \\
 &\leq \mathbb{P}\{W_{\mathcal{A}} > (1 + \varepsilon_n)m(n - 1)a_n\} = \\
 &= \mathbb{P}\{m(n - 1)a_n + \beta_n W'_{\mathcal{A}} > (1 + \varepsilon_n)m(n - 1)a_n\} = \\
 &= \mathbb{P}\left\{W'_{\mathcal{A}} - \mathbb{E}W'_{\mathcal{A}} > \frac{\varepsilon_n m(n - 1)a_n}{\beta_n} - \mathbb{E}W'_{\mathcal{A}}\right\} = \\
 &= \mathbb{P}\left\{\widetilde{W}'_{\mathcal{A}} > \frac{\varepsilon_n m(n - 1)a_n}{\beta_n} - \mathbb{E}W'_{\mathcal{A}}\right\} \leq \\
 &\leq \mathbb{P}\left\{\widetilde{W}'_{\mathcal{A}} > \frac{\varepsilon_n m(n - 1)a_n}{\beta_n} - \widehat{\mathbb{E}W'_{\mathcal{A}}}\right\} = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \delta_n,
 \end{aligned}$$

предпоследнее равенство верно для $\varepsilon_n = \frac{2\beta_n \widehat{\mathbb{E}W'_{\mathcal{A}}}}{m(n-1)a_n}$.

Доказательство утверждения 3. Легко понять, что

$$\sum_{k=1}^d \frac{1}{k} \leq 1 + \frac{1}{2} + \frac{1}{3} + \int_3^d \frac{dx}{x} = \frac{11}{6} + \ln d - \ln 3 \leq \ln d + \frac{3}{4}.$$

Доказательство леммы 2. Нетрудно установить, что $\mathbb{E}\xi_k = \frac{1}{k+1}$ для входов $\text{UNI}(a_n; b_n)$. Оценим сверху каждое из математических ожиданий случайных величин W'_1, W'_2 и W'_3 .

$$\mathbb{E}W'_1 = \sum_{s=1}^m \sum_{k=1}^d \mathbb{E}\xi_k = m \sum_{k=1}^d \frac{1}{k+1} \leq m \ln d,$$

последнее неравенство верно при $d \geq 3$ в силу утверждения 3, и соотношения

$$\sum_{k=1}^d \frac{1}{k+1} = \sum_{k=1}^d \frac{1}{k} - 1 + \frac{1}{d+1} \leq \sum_{k=1}^d \frac{1}{k} - \frac{3}{4} \leq \ln d.$$

$$\begin{aligned} \mathbb{E}W'_2 &= C_m^2 \left(4 \frac{d-1}{2} \mathbb{E}\xi_{(d-1)/2} + 4 \mathbb{E}\xi_{(d-1)} \right) = \\ &= \frac{m(m-1)}{2} \left(\frac{4(d-1)/2}{(d-1)/2+1} + \frac{4}{d} \right) \leq 2m^2; \end{aligned}$$

$$\mathbb{E}W'_3 = m(n - m(d+1)) \mathbb{E}\xi_{(d-1)} = m \frac{n - m(d+1)}{d} \leq \frac{mn}{d} - m^2.$$

Сложив три неравенства и принимая во внимание, что $m(d+1) \leq n$, получаем

$$\mathbb{E}W'_A = \mathbb{E}(W'_1 + W'_2 + W'_3) \leq m \ln d + 2m^2 + \frac{mn}{d} - m^2 \leq m \ln d + \frac{2mn}{d}.$$

Доказательство леммы 3. Оценим $\mathbb{E}e^{t\xi_k}$ сверху, используя формулу

$$\mathbb{E}e^{t\xi_k} = \sum_{i=0}^{\infty} \frac{t^i}{(k+1) \cdots (k+i)}$$

из монографии [17, с. 129], а также обозначения $\alpha = \frac{t}{k+1}$ и

$$Q_{k,t} = \frac{(k+1)}{(k+2)(1 - \frac{t}{k+3})} \leq Q_{k,T} = \frac{(k+1)(k+3)}{(k+2)^2} < 1$$

для всяких $t \leq T$ и натуральных k :

$$\begin{aligned} \mathbb{E}e^{t\xi_k} &= \sum_{i=0}^{\infty} \frac{t^i}{(k+1) \cdots (k+i)} \leq 1 + \alpha + \alpha^2 Q_{k,t} \leq 1 + \alpha + \alpha^2 \leq e^{\alpha + \frac{\alpha^2}{2}} = \\ &= e^{t\mathbb{E}\xi_k} e^{\frac{h_k t^2}{2}}, \end{aligned}$$

так как $\mathbb{E}\xi_k = \frac{1}{k+1}$ для входов $\text{UNI}(a_n; b_n)$.

Следовательно,

$$\mathbb{E}e^{t(\xi_k - \mathbb{E}\xi_k)} = \mathbb{E}e^{t\tilde{\xi}_k} \leq e^{\frac{h_k t^2}{2}},$$

где $\tilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$.

Доказательство леммы 4. В случае $\ln n \leq d < \frac{n}{m}$ параметр H равен сумме величин H_1 , H_2 и H_3 согласно шагам алгоритма \mathcal{A} . С учетом обозначений и оценок, полученных ранее, приходим к следующему:

$$H_1 = m \sum_{k=1}^d h_k = m \sum_{k=1}^d \frac{1}{(k+1)^2} < \psi m,$$

где $\psi \approx 0,645$. Здесь используется оценка Эйлера для суммы обратных квадратов $1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots = \frac{\pi^2}{6} < 1,645$.

$$\begin{aligned} H_2 &= 4C_m^2 \left(\frac{d-1}{2} h_{(d-1)/2} + h_{(d-1)} \right) \leq 2m^2 \left(\frac{(d-1)/2}{((d-1)/2+1)^2} + \frac{1}{d^2} \right) = \\ &= 2m^2 \left(\frac{2(d-1)}{(d+1)^2} + \frac{1}{d^2} \right) \leq 4m^2 \frac{d}{(d+1)^2}. \end{aligned}$$

Последнее неравенство верно для $d \geq 3$.

$$H_3 = m(n - m(d+1))h_{(d-1)} \leq \frac{mn}{d^2} - m^2 \frac{d}{(d+1)^2}.$$

Поскольку $n \geq m(d+1)$ и $m \geq 2$, получаем

$$\begin{aligned} H &= H_1 + H_2 + H_3 < \psi m + 4m^2 \frac{d}{(d+1)^2} + \left(\frac{mn}{d^2} - m^2 \frac{d}{(d+1)^2} \right) \leq \\ &\leq \frac{mn}{d} \left(\frac{d\psi}{n} + \frac{1}{d} \right) + 3m^2 \frac{d}{(d+1)^2} \leq \left(\frac{\psi d}{2(d+1)} + \frac{1}{d} + \frac{3d^2}{(d+1)^3} \right) \frac{mn}{d}. \end{aligned}$$

Нетрудно убедиться, что выражение в круглых скобках меньше 1 для всякого $d \geq 3$, откуда получаем оценку $H \leq \frac{mn}{d}$.

Доказательство леммы 5. С учетом неравенств $\ln d \leq \ln n$ и $d < \frac{n}{\ln n}$ справедливо:

$$\mathbb{E}W'_A \leq m \ln d + \frac{2mn}{d} \leq m \ln n + \frac{2mn}{d} < m \frac{n}{d} + \frac{2mn}{d} = \frac{3mn}{d} = \widehat{\mathbb{E}W'_A}.$$

Доказательство леммы 6. Вследствие леммы 2 и так как $\ln d \leq \ln n$, а также $\frac{n}{d} \leq \ln n$, получаем

$$\mathbb{E}W'_A \leq m \ln d + \frac{2mn}{d} \leq 3m \ln n = \widehat{\mathbb{E}W'_A}.$$

Доказательство теоремы 2. Вначале отметим, что в ходе работы алгоритма \mathcal{A} используются случайные величины типа ξ_k , $1 \leq k \leq d$. В случае графа с весами ребер из $\text{UNI}(a_n; b_n)$ эти случайные величины со смещением удовлетворяют условиям $\mathbb{E}e^{t\tilde{\xi}_k} \leq e^{\frac{h_k t^2}{2}}$ теоремы Петрова [15, с. 54–55] для констант $T = 1$ и $h_k = \frac{1}{(k+1)^2}$ (см. лемма 3).

Проведем доказательство для двух случаев возможных полуинтервалов значений параметра d .

$$\text{Случай 1: } \ln n \leq d < \frac{n}{\ln n}.$$

Согласно лемме 5 и формуле (3) для относительной погрешности получаем

$$\varepsilon_n = \frac{2(b_n - a_n)}{m(n-1)a_n} \widehat{\mathbb{E}W'_{\mathcal{A}}} = \frac{2(b_n - a_n)}{m(n-1)a_n} \frac{3mn}{d} \leq \frac{6n}{(n-1)} \frac{b_n/a_n}{d}.$$

Видим, что $\varepsilon_n \rightarrow 0$ при $n \rightarrow \infty$, если верны следующие условия на разброс весов ребер G : $\frac{b_n}{a_n} = o(d_n)$.

Используя леммы 1 и 5, оцениваем вероятность несрабатывания:

$$\delta_n = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\left\{\widetilde{W}'_{\mathcal{A}} > \frac{3mn}{d}\right\}.$$

Из леммы 4 и неравенства $d < \frac{n}{\ln n}$ следует $TH \leq \frac{mn}{d} < \frac{3mn}{d} = x$. Согласно теореме Петрова [15, с. 54–55] получаем оценку для вероятности несрабатывания алгоритма \mathcal{A} : $\delta_n = \mathbb{P}\{W'_{\mathcal{A}} > x\} \leq \exp\left\{-\frac{Tx}{2}\right\}$.

Поскольку $\ln n < \frac{n}{d}$ и $\frac{Tx}{2} = \frac{3mn}{2d} > m \ln n$, то

$$\delta_n = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > x\} \leq \exp\left\{-\frac{Tx}{2}\right\} < \exp(-m \ln n) = \frac{1}{n^m} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Таким образом, в случае 1 алгоритм \mathcal{A} дает асимптотически точное решение задачи m - d -UMST на графе с весами ребер из $\text{UNI}(a_n; b_n)$.

$$\text{Случай 2: } \frac{n}{\ln n} \leq d < \frac{n}{m}.$$

Согласно лемме 6 и формуле (3) для относительной погрешности получаем

$$\varepsilon_n = \frac{2(b_n - a_n)}{(n-1)a_n} \widehat{\mathbb{E}W'_{\mathcal{A}}} = \frac{2(b_n - a_n)}{m(n-1)a_n} 3m \ln n \leq \frac{6(b_n/a_n) \ln n}{(n-1)}.$$

Понятно, что $\varepsilon_n \rightarrow 0$ при $n \rightarrow \infty$, если $\frac{b_n}{a_n} = o\left(\frac{n}{\ln n}\right)$.

Теперь, используя лемму 1 и лемму 6, оценим вероятность несрабатывания $\delta_n = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > 3m \ln n\}$. При $T = 1$ с учетом $d \geq \frac{n}{\ln n}$ и леммы 4 верно следующее неравенство: $TH \leq \frac{mn}{d} < 3m \ln n = x$. Поскольку $\frac{Tx}{2} > m \ln n$, то согласно теореме Петрова [15, с. 54–55] приходим к следующей оценке для вероятности несрабатывания алгоритма \mathcal{A} : $\delta_n = \mathbb{P}\{\widetilde{W}'_{\mathcal{A}} > x\} \leq \exp\left\{-\frac{Tx}{2}\right\} \leq \exp(-m \ln n) = \frac{1}{n^m} \rightarrow 0$.

Из этого следует, что в случае 2 алгоритм \mathcal{A} также дает асимптотически точное решение задачи m - d -UMST на n -вершинном полном неориентированном графе с весами ребер из $\text{UNI}(a_n; b_n)$.

Таким образом, можно сделать вывод, что для значений параметра d в обоих случаях при условиях (5) имеем оценки: относительной погрешности $\varepsilon_n \rightarrow 0$ и вероятности несрабатывания $\delta_n \rightarrow 0$ при $n \rightarrow \infty$.

Доказательство леммы 7. С учетом (6) получаем

$$\begin{aligned} \mathbb{E}\xi_k &= \int_0^{\infty} x d\mathfrak{P}_{\xi_k}(x) = \int_0^{\infty} xk(1 - \mathfrak{P}_{\xi}(x))^{k-1} d\mathfrak{P}_{\xi}(x) = \int_0^{\infty} xke^{-kx} dx = \\ &= -xe^{-kx} \Big|_0^{\infty} + \int_0^{\infty} e^{-kx} dx = -\frac{1}{k} e^{-kx} \Big|_0^{\infty} = \frac{1}{k}. \end{aligned}$$

Доказательство леммы 8. Оценим сверху каждое из математических ожиданий случайных величин W'_1 , W'_2 и W'_3 :

$$\mathbb{E}W'_1 = \sum_{s=1}^m \sum_{k=1}^d \mathbb{E}\xi_k = m \sum_{k=1}^d \frac{1}{k} \leq m \left(\ln d + \frac{3}{4} \right)$$

с учетом утверждения 3 и леммы 7;

$$\begin{aligned} \mathbb{E}W'_2 &= C_m^2 \left(4 \frac{d-1}{2} \mathbb{E}\xi_{(d-1)/2} + 4\mathbb{E}\xi_{d-1} \right) = \\ &= 2m(m-1) \left(1 + \frac{1}{d-1} \right) \leq \frac{2d}{d-1} m^2 - 2m; \end{aligned}$$

$$\mathbb{E}W'_3 = m(n - m(d+1))\mathbb{E}\xi_{(d-1)} = m \frac{n - m(d+1)}{d-1} = \frac{mn}{d-1} - \frac{d+1}{d-1} m^2.$$

Сложив левые и правые части трех соотношений для $\mathbb{E}W'_1$, $\mathbb{E}W'_2$, $\mathbb{E}W'_3$, и принимая во внимание, что $m \leq \frac{n}{d+1}$, получим:

$$\begin{aligned} \mathbb{E}W'_A &\leq m \ln d - \frac{5}{4}m + \frac{mn}{d-1} + m^2 \leq \\ &\leq m \ln d + \frac{mn}{d-1} + \frac{mn}{d+1} \leq m \ln d + \frac{2mn}{d-1}. \end{aligned}$$

Доказательство леммы 9. Согласно формуле (6) для величины $\mathbb{E}e^{t\xi_k}$ верно

$$\begin{aligned} \mathbb{E}e^{t\xi_k} &= \int_0^{\infty} e^{tx} d\mathfrak{P}_{\xi_k}(x) = \int_0^{\infty} e^{tx} ke^{-kx} dx = \int_0^{\infty} ke^{-(k-t)x} dx = \\ &= -\frac{k}{k-t} e^{-(k-t)x} \Big|_0^{\infty} = \frac{1}{1-t/k} = \sum_{s=0}^{\infty} \left(\frac{t}{k} \right)^s \leq 1 + \frac{t}{k} + \left(\frac{t}{k} \right)^2 \frac{1}{1-t/k}. \end{aligned}$$

С учетом верного в условиях леммы неравенства $\frac{t}{k} \leq \frac{1}{2}$ оценим величину $\mathbb{E}e^{t\xi_k}$ сверху:

$$\begin{aligned} \mathbb{E}e^{t\xi_k} &\leq 1 + \frac{t}{k} + 2\left(\frac{t}{k}\right)^2 = 1 + \frac{t}{k} + \frac{1}{2}\left(\frac{t}{k}\right)^2 + \frac{3}{2}\left(\frac{t}{k}\right)^2 \leq \\ &\leq \left(1 + \frac{t}{k} + \frac{1}{2}\left(\frac{t}{k}\right)^2\right) \left(1 + \frac{3}{2}\left(\frac{t}{k}\right)^2\right) \leq e^{t/k} \exp\left(\frac{3}{2}\left(\frac{t}{k}\right)^2\right) = e^{t\mathbb{E}\xi_k} \exp\left(\frac{h_k t^2}{2}\right), \end{aligned}$$

так как согласно лемме 7 $\mathbb{E}\xi_k = \frac{1}{k}$ для $\text{EXP}(a_n, \lambda_n)$. Отсюда непосредственно следует выполнение условий $\mathbb{E}e^{\tilde{t}\xi_k} \leq e^{\frac{h_k t^2}{2}}$ теоремы Петрова с константами $T = 1/2$, $h_k = 3/k^2$.

Доказательство леммы 10. В случае $\ln n \leq d < \frac{n}{m}$ параметр H равен сумме величин H_1 , H_2 и H_3 согласно шагам алгоритма \mathcal{A} . С учетом ранее введенных обозначений и оценок получаем

$$H_1 = m \sum_{k=1}^d h_k = m \sum_{k=1}^d \frac{3}{k^2} < 3(1 + \psi)m < 5m,$$

где ψ равно оценке Эйлера для суммы обратных квадратов за вычетом 1 ($\psi \approx 0,645$).

$$\begin{aligned} H_2 &= 4C_m^2 \left(\frac{d-1}{2} h_{(d-1)/2} + h_{(d-1)} \right) = \\ &= 6m(m-1) \left(\frac{2}{d-1} + \frac{1}{(d-1)^2} \right) \leq 6m^2 \frac{2d-1}{(d-1)^2}. \end{aligned}$$

$$\begin{aligned} H_3 &= m(n - m(d+1))h_{(d-1)} = \\ &= \frac{3}{(d-1)^2} (mn - m^2(d+1)) \leq \frac{3mn}{(d-1)^2} - 3m^2 \frac{d+1}{(d-1)^2}. \end{aligned}$$

С учетом $n \geq m(d+1)$ и $m \geq 2$ получаем

$$\begin{aligned} H &= H_1 + H_2 + H_3 < 5m + \frac{3m^2}{(d-1)^2} \left((4d-2) - (d+1) \right) + \frac{3mn}{(d-1)^2} = \\ &= 3m \left(\frac{5}{3} + \frac{3m}{d-1} \right) + \frac{3mn}{(d-1)^2} \leq \\ &\leq \frac{3n}{d+1} \left(\frac{5}{3} + \frac{3m}{d-1} \right) + \frac{3mn}{(d-1)^2} \leq \frac{3n}{d+1} \left(\frac{5m}{6} + \frac{3m}{d-1} \right) + \frac{3mn}{(d-1)^2} = \\ &= \frac{3mn}{d-1} \left(\frac{5(d-1)}{6(d+1)} + \frac{3}{d+1} + \frac{1}{d-1} \right) \leq \frac{3mn}{d-1}. \end{aligned}$$

Последний знак неравенства обусловлен тем, что при достаточно больших n выражение в круглых скобках меньше 1, поскольку $d \geq \ln n$.

Доказательство леммы 11. С учетом $\ln d \leq \ln n$, $d < \frac{n}{\ln n}$ и леммы 8 приходим к

$$\mathbb{E}W'_{\mathcal{A}} \leq m \ln d + \frac{2mn}{d-1} \leq m \frac{n}{d} + \frac{2mn}{d-1} \leq \frac{3mn}{d-1} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Доказательство леммы 12. Согласно лемме 8, а также неравенствам $\ln d \leq \ln n$ и $n \leq d \ln n$ получаем

$$\mathbb{E}W'_{\mathcal{A}} \leq m \ln d + \frac{2mn}{d-1} \leq m \ln n + \frac{2md}{d-1} \ln n \leq 5m \ln n = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Доказательство теоремы 3. Вначале отметим, что случайные величины $\tilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$ удовлетворяют условиям $\mathbb{E}e^{t\tilde{\xi}_k} \leq e^{\frac{h_k t^2}{2}}$ теоремы Петрова для констант $T = 1/2$ и $h_k = \frac{3}{k^2}$ (см. лемму 9).

Проведем доказательство для двух случаев возможных полуинтервалов значений параметра d .

$$\text{Случай 1: } \ln n \leq d < \frac{n}{\ln n}.$$

Согласно лемме 11 и формуле (3) для относительной погрешности получаем

$$\varepsilon_n = \frac{2\lambda_n}{m(n-1)a_n} \frac{3mn}{(d-1)} \leq \frac{6n}{(n-1)} \frac{\lambda_n/a_n}{(d-1)}.$$

Видим, что при $n \rightarrow \infty$ $\varepsilon_n \rightarrow 0$, если $\frac{\lambda_n}{a_n} = o(d_n)$.

Теперь, используя леммы 1 и 11, оценим вероятность несрабатывания:

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > \frac{3mn}{d-1}\right\} = \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > \frac{3mn}{d-1}\right\}.$$

Для каждого ребра с весом, соответствующим случайной величине ξ_k , определим константы $T = 1/2$ и $h_k = \frac{3}{k^2}$.

Из леммы 10 следует $TH \leq \frac{3mn}{2(d-1)} < \frac{3mn}{d-1} = x$.

Согласно теореме Петрова получаем следующую оценку для вероятности несрабатывания алгоритма \mathcal{A} :

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leq \exp\left\{-\frac{Tx}{2}\right\}.$$

Поскольку $\frac{n}{d} > \ln n$, то $\frac{Tx}{2} = \frac{3mn}{2(d-1)} > m \ln n$. Из этого следует, что

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leq \exp\left\{-\frac{Tx}{2}\right\} < \exp(-m \ln n) = \frac{1}{n^m} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Таким образом, в случае 1 алгоритм \mathcal{A} дает асимптотически точное решение задачи m - d -UMST на n -вершинном полном неориентированном графе с весами ребер из $\text{EXP}(a_n, \lambda_n)$.

$$\text{Случай 2: } \frac{n}{\ln n} \leq d < \frac{n}{m}.$$

Согласно лемме 12 и формуле (3) для относительной погрешности ε_n получаем

$$\varepsilon_n = \frac{2\lambda_n}{(n-1)a_n} \widehat{\mathbb{E}W'_A} = \frac{2\lambda_n}{m(n-1)a_n} 5m \ln n \leq \frac{10(\lambda_n/a_n) \ln n}{n-1}.$$

Видим, что $\varepsilon_n \rightarrow 0$ при $n \rightarrow \infty$, если выполнены условия $\frac{\lambda_n}{a_n} = o\left(\frac{n}{\ln n}\right)$.

Теперь, используя леммы 1 и 12, можем оценить вероятность несрабатывания

$$\delta_n = \mathbb{P}\{\widetilde{W}'_A > \widehat{\mathbb{E}W'_A}\} = \mathbb{P}\{\widetilde{W}'_A > 5m \ln n\}.$$

Полагая константы h_k , как и в случае 1, зададим $T = 1/2$ и $x = 5m \ln n$.

Принимая во внимание лемму 10 и величины x, T, H , а также, что $d \geq \frac{n}{\ln n}$, приходим к следующему неравенству: $TH \leq \frac{3mn}{2d} < 5m \ln n = x$.

Поскольку $\frac{Tx}{2} > m \ln n$, согласно теореме Петрова получаем следующую оценку для вероятности несрабатывания алгоритма \mathcal{A} :

$$\delta_n = \mathbb{P}\{\widetilde{W}'_A > x\} \leq \exp\left\{-\frac{Tx}{2}\right\} \leq \exp(-m \ln n) = \frac{1}{n^m} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Из этого следует, что в случае 2 алгоритм \mathcal{A} также дает асимптотически точное решение задачи m - d -UMST на n -вершинном полном неориентированном графе с весами ребер из $\text{EXP}(a_n, \lambda_n)$.

Таким образом, можно сделать вывод, что для значений параметра d в обоих случаях при условиях (7) имеем оценки: относительной погрешности $\varepsilon_n \rightarrow 0$ и вероятности несрабатывания $\delta_n \rightarrow 0$ при $n \rightarrow \infty$.

Доказательство утверждения 4. Приведем доказательство этого утверждения, как в [18]. Разность между левой и правой частями неравенства (8) обозначим как

$$h(x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{u^2}{2}} du - (1 - e^{-\frac{x}{2}}).$$

Легко проверить, что для функции $h(x)$ и ее производной

$$h'(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} - \frac{1}{2} e^{-\frac{x}{2}}$$

справедливо: $h(0) = 0$, $\lim_{x \rightarrow \infty} h(x) = 0$, $h'(x) > 0$.

Поскольку на положительной полуоси равенство $h'(x) = 0$ верно только в единственной точке $x_0 = \frac{1}{2}(1 + \sqrt{1 + 12 \ln(2) - 4 \ln(\pi)}) \geq 0$, можно сделать вывод, что $h(x) \geq 0$ при $x \geq 0$, откуда следует справедливость утверждения.

СПИСОК ЛИТЕРАТУРЫ

1. *Frieze A.* On the Value of a Random MST Problem // Discrete Applied Mathematics. 1985. V. 10. P. 47–56. [https://doi.org/10.1016/0166-218X\(85\)90058-7](https://doi.org/10.1016/0166-218X(85)90058-7)
2. *Angel O., Flaxman A.D., Wilson D.B.* A Sharp Threshold for Minimum Bounded-Depth and Bounded-Diameter Spanning Trees and Steiner Trees in Random Networks // Combinatorica. 2012. V. 32. P. 1–33. <https://doi.org/10.1007/s00493-012-2552-z>
3. *Cooper C., Frieze A., Ince N., Janson S., Spencer J.* On the Length of a Random Minimum Spanning Tree // Combinatorics, Probability and Computing. 2016. V. 25. No. 1. P. 89–107. <https://doi.org/10.1017/S0963548315000024>
4. *Garey M.R., Johnson D.S.* Computers and Intractability. 1979. Freeman, San Francisco. 340 p.
5. *Clementi A.E.F., Ianni M.D., Monti A., Rossi, G., Silvestri R.* Experimental Analysis of Practically Efficient Algorithms for Bounded-Hop Accumulation in Ad-Hoc Wireless Networks // Proc. 19th IEEE Int. Parallel Distributed Processing Symposium (IPDPS'05). 2005. P. 8–16. <https://doi.org/10.1109/IPDPS.2005.210>
6. *Bala K., Petropoulos K., Stern T.E.* Multicasting in a Linear Lightwave Network // Proc. of IEEE INFOCOM'93. 1993. P. 1350–1358. <https://doi.org/10.1109/INFCOM.1993.253399>
7. *Bookstein A., Klein S.T.* Compression of Correlated Bit-Vectors // Inform. Syst. 1996. V. 16. No. 4. P. 110–118.
8. *Raymond K.* A Tree-Based Algorithm for Distributed Mutual Exclusion // ACM Trans. on Comput. Syst. 1989. V. 7. No. 1. P. 61–77. <https://doi.org/10.1145/58564.59295>
9. *Gruber M.* Exact and Heuristic Approaches for Solving the Bounded Diameter Minimum Spanning Tree Problem. Vienna University of Technology. PhD Thesis. 2009.
10. *Gimadi E.Kh., Serdyukov A.I.* A Probabilistic Analysis of Approximation Algorithm for the Minimum Weight Spanning Tree Problem with a Bounded Below Diameter / Oper. Res. Proceed. V. 99. In: K. Inderfurth et. al (eds.). Springer, Berlin. 2000. P. 63–68. https://doi.org/10.1007/978-3-642-58300-1_12
11. *Gimadi E.Kh., Istomin A.M., Shin E.Yu.* On Algorithm for the Minimum Spanning Tree Problem Bounded Below // Proc. conference DOOR 2016. Vladivostok. Russia. CEUR-WS. V. 1623. 2016. P. 11–17.
12. *Gimadi E.Kh., Shin E.Yu.* On Given Diameter MST Problem on Random Input Data / In: I. Bykadorov, V. Strusevich, T. Tchemisova (eds.). MOTOR 2019. Communications in Computer and Information Science. V. 1090. Springer, Cham. 2019. P. 30–38. https://doi.org/10.1007/978-3-030-33394-2_3
13. *Gimadi E.Kh., Shevyakov A.S., Shtepa A.A.* A Given Diameter MST on a Random Graph / In: N. Olenov, Y. Evtushenko, M. Khachay, V. Malkova (eds.). Optimization and Applications – 11th International Conference OPTIMA. 2020. LNCS. V. 12422. P. 110–121. https://doi.org/10.1007/978-3-030-62867-3_9
14. *Гимади Э.Х., Глебов Н.И., Перепелица В.А.* Алгоритмы с оценками для задач дискретной оптимизации // Проблемы кибернетики. 1975. Вып. 31. С. 35–42.
15. *Petrov V.V.* Limit Theorems of Probability Theory. Sequences of Independent Random Variables. 1995. Clarendon Press. Oxford. 304 p.

16. Гимади Э.Х., Глазков Ю.В. Об асимптотически точном алгоритме решения одной модификации трехиндексной планарной задачи о назначениях // Дискретный анализ и исследование операций. 2006. Сер. 2. Т. 13. № 1. С. 10–26.
17. Гимади Э.Х., Хачай М.Ю. Экстремальные задачи на множествах перестановок. Екатеринбург: Изд-во УМЦ УПИ, 2016. 219 с.
18. Гимади Э.Х., Шин Е.Ю. Вероятностный анализ алгоритма нахождения в графе минимального остовного дерева с ограниченным снизу диаметром // Дискретный анализ и исследование операций. 2015. Т. 22. № 4. С. 5–20.
<https://doi.org/10.17377/daio.2015.22.474>

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

Поступила в редакцию 23.01.2023

После доработки 27.03.2023

Принята к публикации 28.04.2023



31.08.1931 — 25.04.2023

DOI: 10.31857/S0005231023070097, **EDN:** FEWNYL

В конце апреля нынешнего года в предместье г. Бостона (США) ушел из жизни выдающийся советский и российский ученый Лев Ильич Розоноэр.

Лев Ильич пришел в Институт автоматике и телемеханики (ИАТ, ныне ИПУ РАН) в 1955 г., сразу по окончании МЭИ, и проработал в Институте более 40 лет — до отъезда по семейным обстоятельствам в США в 1996 г.

Л.И. Розоноэр принадлежит к числу тех ученых, кто своими работами закладывал основы теории управления. Мировую известность ему принесли работы по теории оптимального управления и, в первую очередь, формулировка и доказательство корректности принципиально новой идеи, которая впоследствии была развита математиками из Института им. В.А. Стеклова и получила название принципа максимума Понтрягина. Эта идея была представлена Л.И. Розоноэром в его кандидатской диссертации в 1960 г., за которую ему предлагали сразу присудить докторскую степень (докторскую диссертацию Л.И. Розоноэр защитил через 4 года). Основные результаты своей кандидатской диссертации Лев Ильич опубликовал в трех номерах журнала «Автоматика и телемеханика». Другие важные результаты Л.И. Розоноэра в период его работы в ИПУ были связаны с термодинамикой (оптимальное управление термодинамическими процессами) и теорией систем (агрегирова-

ние информации в больших системах, модели биологической эволюции). Эти результаты публиковались не только в России, но и в ведущих зарубежных журналах.

Долгие годы Л.И. Розоноэр был членом редколлегии журнала «Автоматика и телемеханика». Десятки его фундаментальных работ были впервые опубликованы на страницах нашего журнала.

Лев Ильич был замечательным учителем. Стоило ему углубиться в изучение какой-то новой науки, как он немедленно начинал рассказывать студентам МФТИ или сотрудникам ИПУ РАН о том, что узнавал сам. Эти лекции привлекали множество слушателей.

Уникальность Л.И. Розоноэра как ученого заключается не только в отдельных его результатах (какими бы значимыми они ни были), сколь в беспрецедентной широте научных интересов. Уже в ИПУ он активно участвовал в работах по распознаванию образов, математической логике, теории алгоритмов и конечных автоматов. Но по-настоящему эта широта мысли проявилась в период жизни Л.И. Розоноэра в США. В 2018 г. в российском издательстве «Физматлит» вышла его книга «Последние тексты. Теория систем; физика; человек, наука, социум». Лев Ильич включил в нее только те результаты, которые были получены им после 2000 г. И эти результаты оказались неожиданными даже для его коллег в ИПУ. Чего стоит, например, его работа, озаглавленная в духе Канта «Как возможна наука о духе». В ней он попытался заложить основы научного подхода к изучению человеческого духа — таинственного феномена, которым до сих пор интересовались только философия и религия.

Однако своим главным результатом Лев Ильич считал гипотезу о случайном ходе времени на микроуровне как причине декогеренции в квантовой механике. Декогеренция — это спонтанно возникающий в квантовомеханических системах фазовый сдвиг, из-за которого законы микромира переходят в законы макромира. Идея о том, что время на микроуровне представляет собой случайный процесс, настолько необычна, что она вызвала замешательство у физиков, которым еще только предстоит с ней разобраться. И это будет непросто: ведь специалисты по квантовой механике далеко не всегда также глубоко разбираются в теории случайных процессов.

Институт проблем управления может гордиться тем, что в нем вырос и долгие годы работал ученый такого масштаба!

Коллектив ИПУ им. В.А. Трапезникова РАН

СОДЕРЖАНИЕ

Стохастические системы

- Горцев А.М., Нежелская Л.А.** Аналитическое исследование однолинейной системы массового обслуживания с входящим МАР-потокм событий 3
- Хацкевич В.Л.** О непрерывных случайных процессах с нечеткими состояниями 23

Управление в технических системах

- Акимов П.А., Матасов А.И.** Метод гарантирующего оценивания при калибровке блока гироскопов 41
- Бульчев Ю.Г., Мозоль А.А.** Периодно-временной метод параметрической идентификации для решения локационных и навигационных задач 66

Интеллектуальные системы управления, анализ данных

- Михайлов А.М., Каравай М.Ф., Сивцов В.А., Курникова М.А.** Машинное обучение для диагностики заболеваний по полному профилю экспрессии генов 83
- Шакирзянов Р.М., Шлеймович М.П., Новикова С.В.** Метод автоматического позиционирования беспилотных аппаратов на основе распознавания сигнальных радиально-симметричных маркеров подводных целей 93

Оптимизация, системный анализ и исследование операций

- Галяев А.А., Лысенко П.В., Берлин Л.М.** Статистическая сложность как критерий задачи обнаружения полезного сигнала 121
- Гимади Э.Х., Штепа А.А.** Об асимптотической точности поиска минимума суммы весов разноразмерных остовных деревьев фиксированного диаметра . . 146
- Некролог 167

C O N T E N T S

Stochastic Systems

- Gortsev A.M., Nezhelskaya L.A.** Analytical Research of a Single-Line Queuing System with an Incoming MAP-Flow of Events 3
- Khatskevich V.L.** On Continuous Random Processes with Fuzzy States 23

Control in Technical Systems

- Akimov P.A., Matasov A.I.** Guaranteed Estimation Method in the Gyro Unit Calibration Problem 41
- Bulychev Yu.G., Mozol' A.A.** Period-Time Method of Parametric Identification for Solution of Location and Navigation Tasks 66

Intellectual Control Systems, Data Analysis

- Mikhailov A.M., Karavay M.F., Sivtsov V.A., Kurnikova M.A.** Machine Learning for Medical Diagnostics using all Genes Expression Profiles 83
- Shakirzyanov R.M., Shleymovich M.P., Novikova S.V.** Method for Unmanned Vehicles Automatic Positioning based on Signal Radially Symmetric Markers Recognition of Underwater Targets 93

Optimization, System Analysis, and Operations Research

- Galyaev A.A., Lysenko P.V., Berlin L.M.** Statistical Complexity as a Criterion for the Task of Detecting a Useful Signal 121
- Gimadi E.Kh., Shtepa A.A.** On Asymptotically Optimal Approach for Finding of the Minimum Total Weight of Edge-Disjoint Spanning Trees with a Given Diameter 146
- Obituary 167