РОССИЙСКАЯ АКАДЕМИЯ НАУК

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Журнал основан в январе 1965 г. **ISSN: 0555-2923** **Выходит 4 раза в год**

Том 56, 2020 **Вып. 3**

Июль–Август–Сентябрь

М о с к в а

С О Д Е Р Ж А Н И Е

CONTENTS

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.72

© 2020 г. **Б. Накибоглу**

ГРАНИЦА СФЕРИЧЕСКОЙ УПАКОВКИ ДЛЯ КАНАЛОВ БЕЗ ПАМЯТИ[1]

Границы сферической упаковки с полиномиальными по длине блока коэффициентами при экспоненте выводятся с помощью метода Августина для кодов в двух семействах каналов без памяти: (возможно, нестационарных) каналов без памяти с (возможно, многими) аддитивными ограничениями по стоимости и стационарных каналов без памяти с выпуклыми ограничениями на композицию (т.е. эмпирическое распределение, тип) входных кодовых слов. Получен также вариант границы Галлагера, показывающий, что эти границы сферической упаковки точны в смысле экспоненциальной по длине блока скорости убывания вероятности ошибки при довольно широких предположениях.

Ключевые слова: метод Августина, экспонента сферической упаковки, экспонента ошибки, функция надежности, каналы без памяти, гауссовские каналы, пуассоновские каналы.

DOI: 10.31857/S0555292320030018

§ 1. Введение

Большинство известных доказательств границы сферической упаковки (ГСУ) относятся либо к стационарным каналам с конечными множествами входов [2–14], либо к стационарным каналам со специальной структурой шума, например, с пуассоновским или гауссовским шумом [15–22]. Доказательства ГСУ, основанные на методе Августина, являются исключениями: в работах [23–25] не предполагается ни конечность множества входов, ни специальная структура шума, ни стационарность канала. Однако в [23; 24, § 31; 25] установлена ГСУ для каналов-произведений, а не для каналов без памяти, поэтому доказательства ГСУ для кодов с ограничениями по композиции[2] в стационарных каналах [9–20], в том числе для важного специального случая кодов с ограничениями по стоимости [15–20], не охвачены работами [23; 24, § 31; 25]. В [24, § 36] Августин доказал ГСУ для (возможно, нестационарных) каналов без памяти с ограничениями по стоимости в предположении ограниченной функции стоимости. Условия в [24, теорема 36.6] позволили охватить все ранее рассмотренные модели [2–22], кроме гауссовских [17–20].

Теорема 2, представленная в § 3, устанавливает ГСУ при условиях, полностью охватывающих все модели, изучавшиеся в [2–25], используя результаты работы [26], в которой были исследованы августиновские информационные величины. Мы применяем работу [26] и августиновские информационные величины аналогично тому, как в [25] применялась работа [27] и информационные величины Реньи. Для каналов-

[1] Работа была частично представлена на Международном симпозиуме IEEE по теории информации 2017 г. [1].

[2] Как указано в [12, с. 183], ГСУ для кодов с постоянной композицией содержалась в [9] с неполным доказательством. Первое полное доказательство ГСУ для кодов с постоянной композицией было дано в [10].

произведений в [25, теорема 2] были улучшены предыдущие результаты Августина из [23; 24, § 31], а именно установлена ГСУ с коэффициентом при экспоненте, полиномиальным по длине блока n, в предположении, что пропускные способности Реньи порядка $1/2$ каналов-компонент имеют порядок $O(\ln n)$. Для каналов без памяти с ограничениями по стоимости теорема 2 улучшает коэффициент, установленный в [24, теорема 36.6], аналогичным образом: с $e^{-O(\sqrt{n})}$ до $e^{-O(\ln n)}$. Коэффициент в теореме 2, однако, уступает коэффициентам, полученным в [3–6] для различных симметричных каналов, в [13] для кодов с постоянной композицией в дискретных стационарных каналах-произведениях, в [17] для стационарных гауссовских каналов, и в [21, 22] для некоторых некогерентных каналов с замиранием. Определение оптимального коэффициента в духе [3–5] остается открытой задачей в общем случае[3]. Как и [24, теорема 36.6], теорема 2 справедлива и для нестационарных каналов, но в отличие от [24, теорема 36.6] в теореме 2 функции стоимости не предполагаются ограниченными.

В большинстве ранее полученных ГСУ [2–22] предполагалась стационарность канала. Имея стационарный канал-произведение, можно построить стационарный канал без памяти, налагая ограничения на композицию (т.е. тип, эмпирическое распределение) входных кодовых слов. Ограничения по стоимости можно рассматривать как выпуклый частный случай этих более общих ограничений по композиции. Такая интерпретация, рассматриваемая вместе с техникой выбрасывания с учетом композиции, является одним из основных мотивов к изучению кодов с постоянной композицией. Однако техника выбрасывания с учетом композиции полезна только тогда, когда множество входов канала конечно. Тем не менее, если множество ограничений на композицию кодовых слов выпукло, можно получить ГСУ с полиномиальным коэффициентом, используя августиновские информационные величины (см. теорему 1 в § 3). Доказательство теоремы 1 использует августиновский центр множества ограничений, а не августиновское среднее наиболее частой композиции кода. Отметим, что наиболее частая композиция кода может даже состоять из не более одного кодового слова, когда множество входов бесконечно. Условия теоремы 1 достаточно общие, чтобы охватить условия всех известных нам предыдущих доказательств ГСУ для каналов без памяти, кроме тех, что основаны на методе Августина [23–25]. Теоремы 1 и 2 дают асимптотические ГСУ, но в их доказательстве используются неасимптотические ГСУ, представленные в леммах 9 и 10.

Из ГСУ следует, что экспоненциальная скорость убывания оптимальной вероятности ошибки с ростом длины блока, т.е. функция надежности, или экспонента ошибки, ограничена сверху экспонентой сферической упаковки (ЭСУ). Для рассматриваемых каналов без памяти августиновский вариант границы Галлагера показывает, что ЭСУ также ограничивает функцию надежности снизу при условии, что разрешено списочное декодирование. Августиновский вариант границы Галлагера представлен в п. 2.4. Но чтобы показать, что ЭСУ является нижней границей функции надежности при списочном декодировании, можно также использовать стандартные результаты, такие как [29, 30], с небольшими изменениями. Таким образом, августиновский вариант интересен нам не тем, что из него вытекают результаты о функции надежности, а тем, каким образом они из него вытекают. Отличительной особенностью августиновского варианта является то, что он устанавливает результаты о достижимости через информацию Августина, а не информацию Реньи, используемой в стандартной форме границы Галлагера [29]. Для этого августиновский вариант опирается на свойство августиновского среднего как неподвижной точки (формула (7)). Стоит отметить, что в [30] неявно использовалось это утверждение о неподвижной точке, хотя и иным способом.

[3] В работе [28] выведена улучшенная ГСУ (оптимальная в смысле коэффициента в неособых случаях) для всех случаев, рассмотренных в [3–6, 13, 17, 21, 22], с использованием августиновских информационных величин и результатов работы [26].

Прежде чем приступить к детальному обсуждению вопроса, укажем на одну тонкость при выводе ГСУ, на которую обычно не обращают внимания. В [31] утверждалось, что там доказана ГСУ для произвольных стационарных каналов-произведений без использования каких-либо рассуждений с постоянной композицией[4]. Однако установленная в [31, теорема 19] верхняя граница на функцию надежности строго больше ЭСУ во многих каналах. Это было показано численно в [13, с. 1594 и Приложение А]. Аналитическое подтверждение этого наблюдения приведено в Приложении А настоящей статьи. Проблематичным шагом в [31] было применение техники множителей Лагранжа (см. [13, сноска 8]). Доказательство теоремы 19 в [31] опирается на теорему 16 из [31], которая верна лишь для множителя Лагранжа s, соответствующего распределению на входе p, такому что $E_{\rm sp}(R,W) = E_{\rm sp}(R,W,p)$. Однако для произвольного распределения на входе p соответствующий множитель Лагранжа может быть, а может и не быть равным множителю для оптимального распределения на входе p. Именно по этой причине верхняя граница на функцию надежности, установленная в [31, теорема 19], не равна ЭСУ в общем случае вопреки тому, что затем утверждалось в [32, лемма 1] и [33, теорема 10.1.4]. Коротко говоря, доказательство теоремы 19 в [31] по умолчанию подразумевает минимаксное равенство, неверное в общем случае. Для стационарных каналов без памяти с конечными алфавитами на входе эту трудность можно преодолеть, используя рассуждения с постоянной композицией. Однако в этом случае доказательство, приведенное в [31], становится просто повторением доказательства из [10]. Недавно в [34] был предложен вывод ГСУ для стационарных каналов с одним ограничением по стоимости, использующий подход из [31]. Однако, как и в [31], доказательство в [34] подразумевает минимаксное равенство, неверное в общем случае. В частности, в [34, формула (26)] утверждается, что Q^n не зависит от $\boldsymbol{x}_m$. Чтобы утверждать это, следует добавить в [34, формулы (25) и (26)] дополнительный супремум по $\boldsymbol{x}_m$ в качестве самой внутренней оптимизации. Но с этим дополнительным супремумом станет неверным объяснение, приведенное в [34, с. 931]. С учетом сказанного в Приложении А не верится, что доказательство из [34] можно спасти без привлечения существенных новых идей, таких как техника выбрасывания с учетом композиции аналогично [10] или выбрасывания с учетом стоимости кодовых слов аналогично [18]. Коротко говоря, ни в [31], ни в [34] нет успешного доказательства ГСУ для стационарных каналов без памяти, даже для случая конечного множества входов.

В оставшейся части этого параграфа мы введем обозначения, опишем модель канала и определим коды со списочным декодированием. В § 2 вначале приведем краткие сведения о дивергенции Реньи, августиновских информационных величинах и ЭСУ, а затем выведем августиновский вариант границы Галлагера. В § 3 вначале сформулируем наши главные асимптотические результаты, т.е. ГСУ в теоремах 1 и 2, а затем выведем неасимптотические ГСУ, из которых они следуют. В § 4 мы выведем ЭСУ для отдельных гауссовских и пуассоновских каналов и подтвердим эквивалентность определения из § 2 и ранее использованных для этих каналов. В § 5 мы вкратце обсудим, почему метод Августина работает, и сравним наши результаты с результатами Августина из [24], а также обсудим применения метода Августина и августиновских информационных величин в родственных задачах.

1.1. Обозначения и соглашения. Скалярное произведение двух векторов μ и q в $\mathbb{R}^\ell$, т.е. $\sum\limits_{i=1}^{\ell} \mu^i q^i$, обозначается через $\mu \cdot q$. Вектор из всех единиц размерности ℓ будем обозначать через $\mathbb{1}$ для любого $\ell \in \mathbb{Z}_+$, размерность ℓ будет ясна из контекста. Замыкание, внутреннюю область и выпуклую оболочку множества $\mathcal{S}$ будем

[4] В [31, с. 413] сказано: “Важной особенностью нижней границы, которую мы выведем, является то, что не делается никаких предположений о постоянной композиции кодовых слов, даже в качестве промежуточного шага”.

обозначать через $\operatorname{cl}\mathcal{S}$, $\operatorname{int}\mathcal{S}$ и $\operatorname{ch}\mathcal{S}$ соответственно; соответствующая топология или структура векторного пространства будет ясна из контекста.

Для произвольного множества $\mathcal{Y}$ обозначим через $\mathcal{P}(\mathcal{Y})$ множество всех вероятностных мер, не равных нулю лишь на конечном числе элементов из $\mathcal{Y}$. Для любой $p \in \mathcal{P}(\mathcal{Y})$ назовем множество всех $y \in \mathcal{Y}$, для которых $p(y) > 0$, носителем p и будем обозначать его через $\operatorname{supp} p$. Для любого измеримого пространства $(\mathcal{Y}, \mathcal{Y})$ обозначим множество всех вероятностных мер на нем через $\mathcal{P}(\mathcal{Y})$, а множество всех конечных мер – через $\mathcal{M}^+(\mathcal{Y})$. Интеграл от измеримой функции f по мере μ будем обозначать через $\int f\,\mu(dy)$ или $\int f(y)\,\mu(dy)$. Интегралы по мере Лебега на вещественной прямой обозначаем через $\int f\,dy$ или $\int f(y)\,dy$. Если μ – вероятностная мера, то интеграл от f по мере μ будем называть математическим ожиданием f и обозначать через $\mathbf{E}_\mu[f]$ или $\mathbf{E}_\mu[f(y)]$.

Некоторые символы будут использоваться в различных смыслах, однако конкретный смысл будет всегда ясен из контекста. Декартово произведение множеств [35, с. 38] будем обозначать знаком $\times$. Через $|\cdot|$ обозначаем абсолютную величину вещественных чисел и мощность множеств. Знаком $\leqslant$ обозначается как обычное отношение для вещественных чисел, так и соответствующее поточечное неравенство для функций и векторов. Для двух мер μ и q на измеримом пространстве $(\mathcal{Y}, \mathcal{Y})$ пишем $\mu \leqslant q$, если $\mu(\mathcal{E}) \leqslant q(\mathcal{E})$ для всех $\mathcal{E} \in \mathcal{Y}$. Произведение топологий [35, с. 38], σ-алгебр [35, с. 118] и мер [35, теорема 4.4.4] будем обозначать знаком $\otimes$. Используем краткие обозначения $\mathcal{X}_1^n$ для декартова произведения множеств $\mathcal{X}_1, \ldots, \mathcal{X}_n$ и $\mathcal{Y}_1^n$ для произведения σ-алгебр $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$.

1.2. Модель канала. *Канал* W – это функция из *множества входов* $\mathcal{X}$ в множество всех вероятностных мер на *пространстве выходов* $(\mathcal{Y}, \mathcal{Y})$:

$$W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y}). \tag{1}$$

Множество $\mathcal{Y}$ называется *множеством выходов*, а $\mathcal{Y}$ – *σ-алгеброй выходных событий*. Множество всех каналов из множества входов $\mathcal{X}$ в пространство выходов $(\mathcal{Y}, \mathcal{Y})$ обозначим через $\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$. Для любых $p \in \mathcal{P}(\mathcal{X})$ и $W \in \mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$ вероятностная мера с маргинальным распределением на $\mathcal{X}$, равным p, и условным распределением при условии x, равным $W(x)$, обозначается через $p \circledast W$. Структуры, описанной в (1), самой по себе в общем случае недостаточно для существования и единственности меры $p \circledast W$ с требуемыми свойствами для всех $p \in \mathcal{P}(\mathcal{X})$. Существование и единственность такой меры $p \circledast W$ гарантировано для всех $p \in \mathcal{P}(\mathcal{X})$, если W – функция вероятностей переходов из $(\mathcal{X}, \mathcal{X})$ в $(\mathcal{Y}, \mathcal{Y})$, т.е. элемент $\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$, а не $\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$.

Канал W называется *дискретным каналом*, если как $\mathcal{X}$, так и $\mathcal{Y}$ – конечные множества. Для любого $n \in \mathbb{Z}_+$ и любых каналов $W_t\colon \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$, $t \in \{1, \ldots, n\}$, *канал-произведение* $W_{[1,n]}\colon \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ *длины* n определяется следующим образом:

$$W_{[1,n]}(x_1^n) = \bigotimes_{t=1}^{n} W_t(x_t), \quad \forall x_1^n \in \mathcal{X}_1^n.$$

Канал $U\colon \mathcal{Z} \to \mathcal{P}(\mathcal{Y}_1^n)$ называется *каналом длины* n *без памяти*, если существует канал-произведение $W_{[1,n]}$, для которого $U(z) = W_{[1,n]}(z)$ для всех $z \in \mathcal{Z}$ и при этом $\mathcal{Z} \subset \mathcal{X}_1^n$. Канал-произведение является *стационарным*, если $W_t = W$ при всех $t \in \{1, \ldots, n\}$ для некоторого $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. Для такого канала обозначим через $\Upsilon(x)$, где $\Upsilon(x) \in \mathcal{P}(\mathcal{X})$, композицию (т.е. эмпирическое распределение, тип) каждого из $x_1^n \in \mathcal{X}_1^n$.

Для любого $\ell \in \mathbb{Z}_+$ ℓ-мерная *функция стоимости* ρ – это функция из множества входов в $\mathbb{R}^\ell$, ограниченная снизу, т.е. функция вида $\rho\colon \mathcal{X} \to \mathbb{R}^\ell_{\geqslant z}$ для некоторого

$z \in \mathbb{R}$. Без ограничения общности будем предполагать[5], что

$$\inf_{x\in\mathcal{X}} \rho^i(x) \geqslant 0, \quad \forall i \in \{1,\dots,\ell\}.$$

Множество всех ограничений по стоимости, которые выполняются хотя бы для одного элемента из $\mathcal{X}$, обозначим через $\Gamma_\rho^{\mathrm{ex}}$, а множество всех ограничений по стоимости, выполненных хотя бы для одного элемента из $\mathcal{P}(\mathcal{X})$, – через Γ_ρ:

$$\Gamma_\rho^{\mathrm{ex}} \triangleq \bigl\{\varrho \in \mathbb{R}^\ell_{\geqslant 0} : \exists x \in \mathcal{X}, \text{ такой что } \rho(x) \leqslant \varrho\bigr\},$$
$$\Gamma_\rho \triangleq \Bigl\{\varrho \in \mathbb{R}^\ell_{\geqslant 0} : \exists p \in \mathcal{P}(\mathcal{X}), \text{ такой что } \sum_x p(x)\rho(x) \leqslant \varrho\Bigr\}.$$

Тогда оба множества $\Gamma_\rho^{\mathrm{ex}}$ и Γ_ρ имеют непустые внутренние области, причем Γ_ρ является выпуклой оболочкой $\Gamma_\rho^{\mathrm{ex}}$, т.е. $\Gamma_\rho = \operatorname{ch}\Gamma_\rho^{\mathrm{ex}}$.

Функция стоимости на канале-произведении называется аддитивной, если ее можно представить в виде суммы функций стоимости на каналах-компонентах. Для заданных $W_t\colon \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ и $\rho_t\colon \mathcal{X}_t \to \mathbb{R}^\ell_{\geqslant 0}$, $t \in \{1,\dots,n\}$, обозначим результирующую аддитивную функцию стоимости на $\mathcal{X}_1^n$ для канала $W_{[1,n]}$ через $\rho_{[1,n]}$, т.е.

$$\rho_{[1,n]}(x_1^n) = \sum_{t=1}^n \rho_t(x_t), \quad \forall x_1^n \in \mathcal{X}_1^n.$$

1.3. Коды со списочным декодированием. Пара (Ψ, Θ) называется (M, L)-*кодом* в канале $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, если

- *функция кодирования* Ψ – функция из множества сообщений $\mathcal{M} \triangleq \{1, 2, \dots, M\}$ в множество входов $\mathcal{X}$;
- *функция декодирования* Θ – измеримая функция из пространства выходов $(\mathcal{Y}, \mathcal{Y})$ в множество $\widehat{\mathcal{M}} \triangleq \{\mathcal{L} : \mathcal{L} \subset \mathcal{M} \text{ и } |\mathcal{L}| = L\}$.

Для заданного (M, L)-кода (Ψ, Θ) в канале $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *условная вероятность ошибки* P_{e}^m для $m \in \mathcal{M}$ и *средняя вероятность ошибки* P_{e} определяются как

$$P_{\mathrm{e}}^m \triangleq \mathbf{E}_{W(\Psi(m))}[\mathbb{1}_{\{m \notin \Theta(y)\}}],$$
$$P_{\mathrm{e}} \triangleq \frac{1}{M}\sum_{m\in\mathcal{M}} P_{\mathrm{e}}^m.$$

Будем говорить, что функция кодирования Ψ, и следовательно, соответствующий код, удовлетворяют ограничению по стоимости ϱ, если $\bigvee\limits_{m\in\mathcal{M}} \rho(\Psi(m)) \leqslant \varrho$. Будем говорить, что функция кодирования Ψ, и следовательно, соответствующий код, на стационарном канале-произведении удовлетворяют ограничению на эмпирическое распределение $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, если композиции всех кодовых слов принадлежат $\mathcal{A}$, т.е. если $\Upsilon(\Psi(m)) \in \mathcal{A}$ для всех $m \in \mathcal{M}$.

§ 2. Предварительные сведения

Дивергенция Реньи, скошенные вероятностные меры и каналы, а также августиновские информационные величины являются ключевыми понятиями в нашем последующем анализе. Эти понятия вводятся в пп. 2.1 и 2.2, а более подробное обсуждение можно найти в [26,36]. В п. 2.3 мы определяем ЭСУ и выводим ее хорошо

[5] Августин в [24, § 33] накладывал дополнительное условие $\bigvee\limits_{x\in\mathcal{X}} \rho(x) \leqslant \mathbb{1}$, однако это условие исключает из рассмотрения некоторые важные случаи, такие как гауссовские каналы.

известные свойства для нашей общей модели канала, а в п. 2.4 выводится августиновский вариант границы Галлагера.

2.1. Дивергенции Реньи и скошенные вероятностные меры и каналы.

Определение 1. Для любых $\alpha \in \mathbb{R}_+$ и $w, q \in \mathcal{M}^+(\mathcal{Y})$ *дивергенцией Реньи порядка* α *между* w *и* q называется величина

$$D_\alpha(w \,\|\, q) \triangleq \begin{cases} \dfrac{1}{\alpha-1} \ln \displaystyle\int \left(\frac{dw}{d\nu}\right)^{\alpha} \left(\frac{dq}{d\nu}\right)^{1-\alpha} \nu(dy), & \alpha \neq 1, \\ \displaystyle\int \frac{dw}{d\nu} \left[\ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu}\right] \nu(dy), & \alpha = 1, \end{cases}$$

где ν – любая мера, такая что $w \prec \nu$ и $q \prec \nu$.

За исчерпывающим обзором сведений о дивергенции Реньи, используемых на протяжении всей статьи, отсылаем читателя к [36]. Заметим, что дивергенция Реньи порядка 1 – это дивергенция Кульбака – Лейблера. Дивергенцию Реньи для других порядков также можно охарактеризовать через дивергенцию Кульбака – Лейблера (см. [36, теорема 30]). Эта характеризация связана с еще одним ключевым для нас понятием, а именно понятием скошенной вероятностной меры.

Определение 2. Для любого $\alpha \in \mathbb{R}_+$ и любых вероятностных мер $w, q \in \mathcal{P}(\mathcal{Y})$, таких что $D_\alpha(w \,\|\, q) < \infty$, *скошенной вероятностной мерой* w^q_α *порядка* α называется

$$\frac{dw^q_\alpha}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(w\|q)} \left(\frac{dw}{d\nu}\right)^{\alpha} \left(\frac{dq}{d\nu}\right)^{1-\alpha}. \tag{2}$$

Условная дивергенция Реньи и скошенный канал – прямые обобщения понятий дивергенции Реньи и скошенной вероятностной меры, которые позволят нам кратко записывать некоторые соотношения на протяжении дальнейшего анализа.

Определение 3. Для любых $\alpha \in \mathbb{R}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $Q\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $p \in \mathcal{P}(\mathcal{X})$ *условной дивергенцией Реньи порядка* α *для распределения на входе* p называется

$$D_\alpha(W \,\|\, Q \,|\, p) \triangleq \sum_{x \in \mathcal{X}} p(x) D_\alpha(W(x) \,\|\, Q(x)).$$

Если существует $q \in \mathcal{P}(\mathcal{Y})$, такая что $Q(x) = q$ для всех $x \in \mathcal{X}$, то будем обозначать $D_\alpha(W \,\|\, Q \,|\, p)$ через $D_\alpha(W \,\|\, q \,|\, p)$.

Определение 4. Для любых $\alpha \in \mathbb{R}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $Q\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *скошенным каналом* W^Q_α *порядка* α называется функция из $\{x : \ D_\alpha(W(x) \,\|\, Q(x)) < \infty\}$ в $\mathcal{P}(\mathcal{Y})$, такая что

$$\frac{dW^Q_\alpha(x)}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|Q(x))} \left(\frac{dW(x)}{d\nu}\right)^{\alpha} \left(\frac{dQ(x)}{d\nu}\right)^{1-\alpha}. \tag{3}$$

Если существует $q \in \mathcal{P}(\mathcal{Y})$, такая что $Q(x) = q$ для всех $x \in \mathcal{X}$, будем обозначать W^Q_α через W^q_α.

Следующий оператор $\mathrm{T}_{\alpha,p}(\cdot)$ неявно рассматривался Фано в [9, гл. 9], Арутюняном в [10] и Полтыревым в [30], а в явном виде Августином в [24, § 34], но во всех четырех источниках лишь для порядков, меньших единицы.

Определение 5. Для любых $\alpha \in \mathbb{R}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $p \in \mathcal{P}(\mathcal{X})$ оператор Августина $\mathrm{T}_{\alpha,p}(\cdot)\colon \mathcal{Q}_{\alpha,p} \to \mathcal{P}(\mathcal{Y})$ порядка α для распределения на входе p определяется

как

$$\mathrm{T}_{\alpha,p}(q) \triangleq \sum_x p(x) W_\alpha^q(x), \quad \forall q \in \mathcal{Q}_{\alpha,p}, \tag{4}$$

где $\mathcal{Q}_{\alpha,p} \triangleq \{q \in \mathcal{P}(\mathcal{Y}) : \ D_\alpha(W \,\|\, q \,|\, p) < \infty\}$, а скошенный канал W_α^q определен в (3).

2.2. Августиновские информационные величины.

Определение 6. Для любых $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $p \in \mathcal{P}(\mathcal{X})$ *информацией Августина порядка α для распределения на входе p* называется величина

$$I_\alpha(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p). \tag{5}$$

Инфимум в (5) достигается на единственной вероятностной мере, которая обозначается через $q_{\alpha,p}$ и называется *августиновским средним порядка α для распределения на входе p*. При этом для августиновского среднего порядка α справедливы следующие тождества:

$$\begin{aligned} &D_{1\vee\alpha}(q_{\alpha,p} \,\|\, q) \geqslant D_\alpha(W \,\|\, q \,|\, p) - I_\alpha(p; W) \geqslant D_{1\wedge\alpha}(q_{\alpha,p} \,\|\, q), \\ &\forall q \in \mathcal{P}(\mathcal{Y}),\ \alpha \in \mathbb{R}_+, \end{aligned} \tag{6}$$

$$\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \quad \forall \alpha \in \mathbb{R}_+. \tag{7}$$

Эти свойства установлены в [26, лемма 13(b)–(d)]; ранее они были получены Августином [24, лемма 34.2] для порядков, меньших единицы. На протяжении всей статьи мы будем ссылаться на утверждения об августиновских информационных величинах из [26]. Более подробный обзор предшествующих результатов об августиновских информационных величинах можно также найти в [26].

Определение 7. Для любых $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ *августиновской пропускной способностью порядка α канала W для множества ограничений $\mathcal{A}$* называется величина

$$C_{\alpha,W,\mathcal{A}} \triangleq \sup_{p \in \mathcal{A}} I_\alpha(p; W).$$

Когда множество ограничений $\mathcal{A}$ совпадает со всем $\mathcal{P}(\mathcal{X})$, августиновская пропускная способность порядка α обозначается через $C_{\alpha,W}$, т.е. $C_{\alpha,W} \triangleq C_{\alpha,W,\mathcal{P}(\mathcal{X})}$.

Используя определения информации Августина и августиновской пропускной способности, получаем следующее выражение для $C_{\alpha,W,\mathcal{A}}$:

$$C_{\alpha,W,\mathcal{A}} = \sup_{p \in \mathcal{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p).$$

Если множество $\mathcal{A}$ выпукло, то согласно [26, теорема 1] порядок супремума и инфимума можно поменять:

$$\sup_{p \in \mathcal{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{A}} D_\alpha(W \,\|\, q \,|\, p). \tag{8}$$

Если к тому же $C_{\alpha,W,\mathcal{A}}$ конечна, то из [26, теорема 1] следует, что существует единственная вероятностная мер $q_{\alpha,W,\mathcal{A}}$, называемая *августиновским центром порядка α канала W для множества ограничений $\mathcal{A}$*, такая что

$$C_{\alpha,W,\mathcal{A}} = \sup_{p \in \mathcal{A}} D_\alpha(W \,\|\, q_{\alpha,W,\mathcal{A}} \,|\, p).$$

Множество всех вероятностных мер, удовлетворяющих ограничению по стоимости ϱ, будем обозначать через $\mathcal{A}(\varrho)$, т.е.

$$\mathcal{A}(\varrho) \triangleq \{p \in \mathcal{P}(\mathcal{X}) : \mathbf{E}_p[\rho] \leqslant \varrho\}.$$

Допуская некоторую вольность в обозначениях, множества ограничений, определяемые через ограничения по стоимости, будем обозначать через $C_{\alpha,W,\varrho}$ вместо $C_{\alpha,W,\mathcal{A}(\varrho)}$. Чтобы иметь возможность применять технику выпуклого сопряжения без существенных видоизменений, продолжим определение августиновской пропускной способности на недопустимые значения ограничений по стоимости, т.е. на ϱ, не лежащие в Γ_ρ:

$$C_{\alpha,W,\varrho} \triangleq \begin{cases} \sup\limits_{p\in\mathcal{A}(\varrho)} I_\alpha(p;W), & \text{если } \varrho \in \Gamma_\rho, \\ -\infty, & \text{если } \varrho \in \mathbb{R}^\ell_{\geqslant 0} \setminus \Gamma_\rho, \end{cases} \qquad \forall \alpha \in \mathbb{R}_+.$$

Чтобы получить характеризацию $C_{\alpha,W,\varrho}$ с помощью техники выпуклого сопряжения, вначале определим понятия информации и пропускной способности Августина–Лежандра. Эти понятия были впервые введены в [1, раздел III-A] и [26, п. 5.2] как обобщения аналогичных понятий из [12, гл. 8].

Определение 8. Для любых $\alpha \in \mathbb{R}_+$, канала W вида $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon \mathcal{X} \to \mathbb{R}^\ell_{\geqslant 0}$, $p \in \mathcal{P}(\mathcal{X})$ и $\lambda \in \mathbb{R}^\ell_{\geqslant 0}$ *информацией Августина–Лежандра порядка* α *для распределения на входе* p *и множителя Лагранжа* λ называется величина

$$I^\lambda_\alpha(p;W) \triangleq I_\alpha(p;W) - \lambda \cdot \mathbf{E}_p[\rho].$$

Определение 9. Для любых $\alpha \in \mathbb{R}_+$, канала W вида $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon \mathcal{X} \to \mathbb{R}^\ell_{\geqslant 0}$ и $\lambda \in \mathbb{R}^\ell_{\geqslant 0}$ *пропускной способностью Августина–Лежандра порядка* α *для множителя Лагранжа* λ называется величина

$$C^\lambda_{\alpha,W} \triangleq \sup_{p\in\mathcal{P}(\mathcal{X})} I^\lambda_\alpha(p;W).$$

С точностью до некоторых перемен знака функция $C^\lambda_{\alpha,W}$ является выпукло сопряженной к $C_{\alpha,W,\varrho}$ в силу аналогичного соотношения между $I^\lambda_\alpha(p;W)$ и $I_\alpha(p;W)$ (см. [26, формулы (72)–(74) и (76)]):

$$C^\lambda_{\alpha,W} = \sup_{\varrho\geqslant 0} C_{\alpha,W,\varrho} - \lambda \cdot \varrho, \quad \forall \lambda \in \mathbb{R}^\ell_{\geqslant 0}.$$

Тогда $C_{\alpha,W,\varrho}$ можно выразить через $C^\lambda_{\alpha,W}$, по крайней мере, для внутренних точек Γ_ρ:

$$C_{\alpha,W,\varrho} = \inf_{\lambda\geqslant 0} C^\lambda_{\alpha,W} + \lambda \cdot \varrho.$$

Кроме того, согласно [26, лемма 29] существует непустое выпуклое компактное множество величин $\lambda_{\alpha,W,\varrho}$, таких что $C_{\alpha,W,\varrho} = C^{\lambda_{\alpha,W,\varrho}}_{\alpha,W} + \lambda_{\alpha,W,\varrho} \cdot \varrho$, при условии, что $C_{\alpha,W,\varrho}$ конечна.

С другой стороны, используя определения величин $I_\alpha(p;W)$, $I^\lambda_\alpha(p;W)$ и $C^\lambda_{\alpha,W}$, получаем следующее выражение для $C^\lambda_{\alpha,W}$:

$$C^\lambda_{\alpha,W} = \sup_{p\in\mathcal{P}(\mathcal{X})} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho].$$

Для функции $C^{\lambda}_{\alpha,W}$ выполняется минимаксное соотношение, аналогичное (8) (см. [26, теорема 2]). Впрочем, это соотношение лучше объяснить с использованием понятия радиуса Августина – Лежандра, определяемого следующим образом.

О п р е д е л е н и е 10. Для любых $\alpha \in \mathbb{R}_+$, канала $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$ и $\lambda \in \mathbb{R}^{\ell}_{\geqslant 0}$ *радиусом Августина – Лежандра порядка* α *канала* W *для множителя Лагранжа* λ называется величина

$$S^{\lambda}_{\alpha,W} \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathcal{X}} D_{\alpha}(W(x)\,\|\,q) - \lambda\cdot\rho(x).$$

Тогда согласно [26, теорема 2] для любых $\alpha \in \mathbb{R}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ с $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$ и $\lambda \in \mathbb{R}^{\ell}_{\geqslant 0}$ справедливо равенство

$$C^{\lambda}_{\alpha,W} = S^{\lambda}_{\alpha,W}. \tag{9}$$

Если к тому же $C^{\lambda}_{\alpha,W}$ конечна, то существует единственная вероятностная мера $q^{\lambda}_{\alpha,W} \in \mathcal{P}(\mathcal{Y})$, называемая *центром Августина – Лежандра порядка* α *канала* W *для множителя Лагранжа* λ, такая что

$$C^{\lambda}_{\alpha,W} = \sup_{x\in\mathcal{X}} D_{\alpha}(W(x)\,\|\,q^{\lambda}_{\alpha,W}) - \lambda\cdot\rho(x).$$

Информационные величины Августина – Лежандра определяются стандартным применением техники выпуклого сопряжения. Однако начиная с [29, теоремы 8 и 10], т.е. с вариантов границы Галлагера с ограничениями по стоимости, обычным инструментом применения техники выпуклого сопряжения для вычисления экспоненты ошибки вместо информационных величин Августина – Лежандра становятся информационные величины Реньи – Галлагера (см., например, [18–20]). Краткое обсуждение информационных величин Реньи – Галлагера содержится в Приложении B, более подробно см. в [26].

2.3. Экспонента сферической упаковки.

О п р е д е л е н и е 11. Для любых $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ и $R \in \mathbb{R}_{\geqslant 0}$ экспонентой сферической упаковки называется величина

$$E_{\mathrm{sp}}(R, W, \mathcal{A}) \triangleq \sup_{\alpha\in(0,1)} \frac{1-\alpha}{\alpha}\left(C_{\alpha,W,\mathcal{A}} - R\right). \tag{10}$$

ЭСУ для случая $\mathcal{A} = \mathcal{P}(\mathcal{X})$ будем обозначать через $E_{\mathrm{sp}}(R, W)$. Кроме того, допуская некоторую вольность в обозначениях, ЭСУ в случае $\mathcal{A} = \{p\}$ будем обозначать через $E_{\mathrm{sp}}(R, W, p)$, а в случае $\mathcal{A} = \{p:\ \mathbf{E}_p[\rho] \leqslant \varrho\}$ – через $E_{\mathrm{sp}}(R, W, \varrho)$.

Л е м м а 1. *Для любых* $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *и* $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ *функция* $E_{\mathrm{sp}}(R, W, \mathcal{A})$ *не возрастает и выпукла по* R *на* $\mathbb{R}_{\geqslant 0}$, *конечна на* $(C_{0^+,W,\mathcal{A}}, \infty)$ *и непрерывна на* $[C_{0^+,W,\mathcal{A}}, \infty)$, *где* $C_{0^+,W,\mathcal{A}} = \lim\limits_{\alpha\downarrow 0} C_{\alpha,W,\mathcal{A}}$. *В частности,*

$$E_{\mathrm{sp}}(R, W, \mathcal{A}) = \begin{cases} \infty, & R < C_{0^+,W,\mathcal{A}}, \\ \sup\limits_{\alpha\in(0,1)} \dfrac{1-\alpha}{\alpha}\left(C_{\alpha,W,\mathcal{A}} - R\right), & R = C_{0^+,W,\mathcal{A}}, \\ \sup\limits_{\alpha\in[\phi,1)} \dfrac{1-\alpha}{\alpha}\left(C_{\alpha,W,\mathcal{A}} - R\right), & R = C_{\phi,W,\mathcal{A}} \text{ для некоторого } \phi \in (0,1), \\ 0, & R \geqslant C_{1,W,\mathcal{A}}. \end{cases} \tag{11}$$

Лемма 1 вытекает из свойств непрерывности и монотонности функции $C_{\alpha,W,\mathcal{A}}$, установленных в [26, лемма 23]; ее доказательство приведено в Приложении C. Доказательство леммы 1 аналогично доказательству утверждения [25, лемма 13], которое опирается на [25, лемма 8], а не на [26, лемма 23].

Можно выразить $E_{\rm sp}(R,W,\mathcal{A})$ через $E_{\rm sp}(R,W,p)$, используя определения величин $C_{\alpha,W,\mathcal{A}}$, $E_{\rm sp}(R,W,\mathcal{A})$ и $E_{\rm sp}(R,W,p)$:

$$E_{\rm sp}(R,W,\mathcal{A}) = \sup_{\alpha\in(0,1)}\sup_{p\in\mathcal{A}}\frac{1-\alpha}{\alpha}\left(I_\alpha(p;W)-R\right) =$$

$$= \sup_{p\in\mathcal{A}}\sup_{\alpha\in(0,1)}\frac{1-\alpha}{\alpha}\left(I_\alpha(p;W)-R\right) = \sup_{p\in\mathcal{A}}E_{\rm sp}(R,W,p). \tag{12}$$

Утверждение леммы 1 для $E_{\rm sp}(R,W,p)$ выполнено по определению, но его можно значительно усилить для значений R в интервале $(\lim\limits_{\alpha\downarrow0} I_\alpha(p;W), I_1(p;W)]$, используя простейшие свойства информации Августина.

Лемма 2. *Пусть $W\colon\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ и $p\in\mathcal{P}(\mathcal{X})$ таковы, что $I_{0+}(p;W)\neq I_1(p;W)$, где $I_{0+}(p;W)\triangleq\lim\limits_{\alpha\downarrow0}I_\alpha(p;W)$. Тогда для любой скорости $R\in(I_{0+}(p;W),I_1(p;W)]$ существует единственный порядок $\alpha^*\in(0,1]$, такой что*

$$R = I_1(p;W_{\alpha^*}^{q_{\alpha^*,p}}). \tag{13}$$

Порядки α^*, определенные в (13), образуют возрастающую непрерывную биективную функцию скорости R из $(I_{0+}(p;W),I_1(p;W)]$ в $(0,1]$, такую что

$$E_{\rm sp}(R,W,p) = D_1(W_{\alpha^*}^{q_{\alpha^*,p}}\,\|\,W\,|\,p), \tag{14}$$

$$\frac{\partial}{\partial R}E_{\rm sp}(R,W,p) = \frac{\alpha^*-1}{\alpha^*}. \tag{15}$$

Таким образом, $E_{\rm sp}(R,W,p)$ конечна, выпукла, непрерывно дифференцируема, возрастает по R на $(I_{0+}(p;W),I_1(p;W))$ и удовлетворяет соотношению

$$E_{\rm sp}(I_{0+}(p;W),W,p) = \lim_{\alpha\downarrow0}D_1(W_\alpha^{q_{\alpha,p}}\,\|\,W\,|\,p). \tag{16}$$

Кроме того, если $E_{\rm sp}(I_{0+}(p;W),W,p)$ конечна, то существует канал $V\colon\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, для которого $I_1(p;V)=I_{0+}(p;W)$ и $D_1(V\,\|\,W\,|\,p)=E_{\rm sp}(I_{0+}(p;W),W,p)$.

Доказательство. Заметим, что $I_1(p;W_\alpha^{q_{\alpha,p}})$ – возрастающая и непрерывная функция порядка α согласно [26, лемма 17(a),(f)], поскольку $I_{0+}(p;W)\neq I_1(p;W)$ по условию. Кроме того, $\lim\limits_{\alpha\downarrow0}I_1(p;W_\alpha^{q_{\alpha,p}}) = I_{0+}(p;W)$ в силу [26, лемма 17(g)]. Тогда по теореме о промежуточном значении [37, теорема 4.23] существует единственный порядок α^*, удовлетворяющий условию (13). Функция, определенная в (13), – возрастающая непрерывная биективная функция скорости R из $(I_{0+}(p;W),I_1(p;W)]$ в $(0,1]$ как обратная к возрастающей непрерывной биективной функции $\alpha\rightsquigarrow I_1(p;W_\alpha^{q_{\alpha,p}})$ из $(0,1]$ в $(I_{0+}(p;W),I_1(p;W)]$.

С другой стороны, $I_\alpha(p;W)$ непрерывно дифференцируема по α в силу [26, лемма 17(e)]; тогда из [26, формулы (35) и (46)] получаем

$$\frac{\partial}{\partial\alpha}\frac{1-\alpha}{\alpha}(I_\alpha(p;W)-R) = \frac{1}{\alpha^2}\big(R-I_1(p;W_\alpha^{q_{\alpha,p}})\big). \tag{17}$$

Следовательно, для любой $R\in(I_{0+}(p;W),I_1(p;W)]$ супремум в определении величины $E_{\rm sp}(R,W,p)$ достигается на порядке α^*, удовлетворяющем (13). Тогда (14) вытекает из [26, формула (35)]. Кроме того, для любых $R\in(I_{0+}(p;W),I_1(p;W))$

и $\overline{R} \geqslant 0$ имеем

$$E_{\mathrm{sp}}(\overline{R}, W, p) \geqslant \frac{1-\alpha^*(R)}{\alpha^*(R)}(I_{\alpha^*}(p;W) - \overline{R}) =$$
$$= E_{\mathrm{sp}}(R, W, p) + \frac{1-\alpha^*(R)}{\alpha^*(R)}(R - \overline{R}). \tag{18}$$

Рассуждая аналогично и меняя ролями R и $\overline{R}$, получаем, что для любых $\overline{R} \in$ $\in (I_{0^+}(p;W), I_1(p;W))$ и $R \geqslant 0$

$$E_{\mathrm{sp}}(R, W, p) \geqslant E_{\mathrm{sp}}(\overline{R}, W, p) + \frac{1-\alpha^*(\overline{R})}{\alpha^*(\overline{R})}(\overline{R} - R). \tag{19}$$

Так как α^* возрастает и непрерывна по скорости, из (18) и (19) получаем (15) для всех $R \in (I_{0^+}(p;W), I_1(p;W)]$.

В случае $R = I_{0^+}(p;W)$ заметим, что $\dfrac{1-\alpha}{\alpha}(I_\alpha(p;W) - R)$ убывает по α на $(0,1)$ в силу (17) и [26, лемма 17(f),(g)]. Таким образом,

$$E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) = \lim_{\alpha \downarrow 0} \frac{1-\alpha}{\alpha}(I_\alpha(p;W) - I_{0^+}(p;W)).$$

Тогда (16) вытекает из теоремы о среднем [37, теорема 5.10] и [26, формула (46)]. Кроме того, если $E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) = \gamma$ для $\gamma \in \mathbb{R}_+$, то $D_1(W_\alpha^{q_\alpha,p}(x) \,\|\, W(x)) \leqslant \dfrac{\gamma}{p(x)}$ в силу неотрицательности дивергенции Реньи. Отсюда

$$\int G\left(\frac{dW_\alpha^{q_\alpha,p}(x)}{dW(x)}\right) W(dy\,|\,x) \leqslant \frac{\gamma}{p(x)} + \frac{1}{e} + 1$$

для $G(\tau) = \tau \mathbb{1}_{\{0 \leqslant \tau < e\}} + \tau \ln \tau \mathbb{1}_{\{\tau \geqslant e\}}$, поскольку $\tau \ln \tau \geqslant -1/e$. Тогда элементы семейства $\left\{\dfrac{dW_\alpha^{q_\alpha,p}(x)}{dW(x)}\right\}_{\alpha \in (0,1)}$ равномерно $W(x)$-интегрируемы в силу необходимого и достаточного условия равномерной интегрируемости Валле–Пуссена [38, теорема 4.5.9]. Таким образом, любая последовательность элементов $\left\{W_\alpha^{q_\alpha,p}(x)\right\}_{\alpha \in (0,1)}$ имеет сходящуюся подпоследовательность в топологии сходимости на множествах [38, теорема 4.7.25]. Для каждого $x \in \operatorname{supp} p$ пусть $V(x)$ – предельная точка вышеуказанной подпоследовательности последовательности $\left\{W_{1/\kappa}^{q_{1/\kappa},p}(x)\right\}_{\kappa \in \mathbb{Z}_+}$. Тогда из соотношений (13), (14) и полунепрерывности снизу дивергенции Реньи по ее аргументам в топологии сходимости на множествах [36, теорема 15] следует, что $I_1(p;V) \leqslant I_{0^+}(p;W)$ и $D_1(V \,\|\, W \,|\, p) \leqslant \gamma$. С другой стороны в силу определения $E_{\mathrm{sp}}(R,W,p)$ и [26, лемма 13(e)] имеем

$$E_{\mathrm{sp}}(R,W,p) = \sup_{\alpha \in (0,1)} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) \leqslant$$
$$\leqslant \sup_{\alpha \in (0,1)} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) = \begin{cases} D_1(V \,\|\, W \,|\, p), & R \geqslant I_1(p;V), \\ \infty, & R < I_1(p;V). \end{cases}$$

Таким образом, $I_1(p;V)$ не может быть меньше, чем $I_{0^+}(p;W)$, так как $E_{\mathrm{sp}}(R,W,p)$ равна бесконечности при всех $R < I_{0^+}(p;W)$. Значит, $I_1(p;V) = I_{0^+}(p;W)$. Следовательно, $D_1(V \,\|\, W \,|\, p)$ не может быть меньше, чем γ, поскольку $E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) =$ $= \gamma$. Отсюда $D_1(V \,\|\, W \,|\, p) = \gamma$. ▲

Лемма 2 дает простое объяснение альтернативному выражению для $E_{\mathrm{sp}}(R,W,p)$, известному как ЭСУ в форме Арутюняна [10].

Лемма 3. *Для любых* $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$ *и* $R \in \mathbb{R}_+$

$$E_{\mathrm{sp}}(R, W, p) = \inf_{V\colon I_1(p;V) \leqslant R} D_1(V \,\|\, W \,|\, p). \tag{20}$$

Доказательство. Если $R \in [I_1(p;W), \infty)$, то равенство (20) выполнено, так как выражение в правой части (20) равно нулю, что получается подстановкой $V = W$ с учетом неотрицательности дивергенции Реньи.

С другой стороны, в силу определения $E_{\mathrm{sp}}(R, W, p)$ (см. [26, лемма 13(e)]) и неравенства максимина-минимакса имеем

$$\begin{aligned} E_{\mathrm{sp}}(R, W, p) &= \sup_{\alpha\in(0,1)} \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) \leqslant \\ &\leqslant \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} \sup_{\alpha\in(0,1)} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) = \\ &= \inf_{V\colon I_1(p;V)\leqslant R} D_1(V \,\|\, W \,|\, p). \end{aligned}$$

Таким образом, (20) имеет место, когда $E_{\mathrm{sp}}(R, W, p)$ равна бесконечности, т.е. для всех $R \in [0, I_{0+}(p;W))$ и, возможно, для $R = I_{0+}(p;W)$, а также когда $E_{\mathrm{sp}}(R, W, p)$ конечна в силу леммы 2. ▲

Из выражения в форме Арутюняна вытекает следующее достаточное условие оптимальности порядка α в определении ЭСУ, данном в (10).

Лемма 4. *Для любых* $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ *и* $R \in (C_{0+,W,\mathcal{A}}, C_{1,W,\mathcal{A}})$, *если существуют* $\alpha^* \in (0,1)$ *и функция* V_p *меры* p *из* $\mathcal{A}$ *в* $\mathcal{P}(\mathcal{Y}\,|\,\mathcal{X})$, *удовлетворяющие двум неравенствам*

$$D_1(V_p \,\|\, q_{\alpha^*,W,\mathcal{A}} \,|\, p) \leqslant R, \quad \forall p \in \mathcal{A}, \tag{21}$$

$$D_1(V_p \,\|\, W \,|\, p) \leqslant \frac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R), \quad \forall p \in \mathcal{A}, \tag{22}$$

то $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \dfrac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R)$.

Для некоторых каналов $V_p = W_{\alpha^*}^{q_{\alpha^*},W,\mathcal{A}}$ удовлетворяет как (21), так и (22); для таких каналов значение ЭСУ можно определить с помощью леммы 4. Однако для произвольных канала, скорости и соответствующего оптимального порядка α^* в (10) функция V_p, удовлетворяющая как (21), так и (22), может и не существовать, как, например, для Z-канала, обсуждаемого в Приложении А. Если бы достаточное условие оптимальности порядка α^*, указанное в (21) и (22), было также и необходимым, то доказательство Блейхута в [31] было бы верным; это, однако, не так в общем случае, как показано в Приложении А. Следует отметить, что для каналов, удовлетворяющих необходимому условию, указанному в (21) и (22), можно выводить ГСУ с помощью подхода из [31].

Доказательство леммы 4. Заметим, что в силу (6)

$$D_1(V_p \,\|\, q_{\alpha^*,W,\mathcal{A}} \,|\, p) = I_1(p;V_p) + D_1\Bigg(\sum_x p(x)V_p(x) \,\|\, q_{\alpha^*,W,\mathcal{A}}\Bigg).$$

Таким образом, $I_1(p;V_p) \leqslant R$ для всех $p \in \mathcal{A}$, поскольку дивергенция Реньи неотрицательна. Тогда $E_{\mathrm{sp}}(R, W, p) \leqslant \dfrac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R)$ для всех $p \in \mathcal{A}$ по лемме 3. Тогда $E_{\mathrm{sp}}(R, W, \mathcal{A}) \leqslant \dfrac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R)$ в силу (12). С другой стороны, $E_{\mathrm{sp}}(R, W, \mathcal{A}) \geqslant$ $\geqslant \dfrac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R)$ по определению. Итак, $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \dfrac{1-\alpha^*}{\alpha^*}(C_{\alpha^*,W,\mathcal{A}} - R)$. ▲

2.4. Августиновский вариант границы Галлагера. Экспонента сферической упаковки является верхней границей экспоненциальной скорости убывания оптимальной вероятности ошибки с ростом длины блока, т.е. функции надежности, для каналов без памяти, удовлетворяющих довольно широким предположениям, как при списочном декодировании, так и без него, в силу ГСУ, приведенных в теоремах 1 и 2 в § 3. При списочном декодировании ЭСУ является также и нижней границей экспоненциальной скорости убывания оптимальной вероятности ошибки по длине блока [8, упражнение 5.20; 12, упражнение 10.28; 39]. Это последнее наблюдение можно проверить с помощью стандартных результатов, таких как [29, 30], хотя и с небольшими изменениями. Мы же проверим его, используя вариант границы Галлагера в терминах информации Августина. Напомним, что обычно граница Галлагера выводится через информацию Реньи, а не информацию Августина. Ключевую роль в нашем доказательстве играет свойство неподвижной точки (7). Это вариант границы Галлагера мы называем августиновским, так как, насколько нам известно, именно Августин впервые получил результат о достижимости [24, лемма 36.1] с помощью свойства неподвижной точки, описанного в (7).

Лемма 5. *Для любых $M, L \in \mathbb{Z}_+$, таких что $L < M$, любых $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$ и $\alpha \in \left[\frac{1}{1+L}, 1\right)$ существует (M, L)-код с функцией кодирования вида $\Psi\colon \mathcal{M} \to \mathcal{B}$, для которого*

$$\ln P_{\mathrm{e}} \leqslant \frac{\alpha - 1}{\alpha}\left[\tau - \ln\frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha},$$

где $\tau = \inf\limits_{x \in \mathcal{B}} D_\alpha(W(x) \,\|\, q_{\alpha,p})$.

Мы не даем отдельного доказательства леммы 5, поскольку она является частным случаем приведенной ниже леммы 6 с множителем Лагранжа, равным нулю. Прежде чем сформулировать лемму 6, укажем на немедленное следствие леммы 5 для кодов с постоянной композицией в стационарных каналах без памяти. Напомним, что для любой фиксированной композиции p при длине блока n вероятность множества всех последовательностей с композицией p (множества $\mathcal{T}_{p,n}$) для независимой и одинаково распределенной выборки (т.е. $p_t = p$ для всех t) согласно [12, с. 26] удовлетворяет следующему равенству для некоторого $\xi \in [0, 1]$:

$$\left(\bigotimes_{t=1}^{n} p_t\right)(\mathcal{T}_{p,n}) = e^{-\xi\frac{|\operatorname{supp} p|}{12\ln 2}}(2\pi n)^{-\frac{|\operatorname{supp} p|-1}{2}}\sqrt{\prod_{x\colon p(x)>0}\frac{1}{p(x)}}.$$

Тогда, учитывая равенство $q_{\alpha, \bigotimes\limits_{t=1}^{n} p_t} = \bigotimes\limits_{t=1}^{n} q_{\alpha,p_t}$, установленное в [26, лемма 14], и полагая $\mathcal{B}$ равным $\mathcal{T}_{p,n}$, получаем

Следствие 1. *Для любых $n \in \mathbb{Z}_+$, $\phi \in (0, 1)$, семейства $\{W_t\}_{t\in\mathbb{Z}_+}$, такого что $W_t = W$ для некоторого $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, вероятностной меры $p \in \mathcal{P}(\mathcal{X})$, такой что $np(x) \in \mathbb{Z}_{\geqslant 0}$ для всех $x \in \mathcal{X}$, целых чисел M, L, таких что $\frac{1}{L+1} < \eta$ и $\frac{1}{n}\ln\frac{M}{L} = I_\eta(p; W)$ для некоторого $\eta \in [\phi, 1)$, существует (M, L)-код в канале $\bigotimes\limits_{t=1}^{n} W_t$ с функцией кодирования вида $\Psi\colon \mathcal{M} \to \mathcal{T}_{p,n}$, для которого*

$$\ln P_{\mathrm{e}} \leqslant -nE_{\mathrm{sp}}\Big(\frac{1}{n}\ln\frac{M}{L}, W, p\Big) + \\ + \frac{1}{\eta}\left[1 - \eta + \frac{|\operatorname{supp} p|}{12\ln 2} + \frac{|\operatorname{supp} p| - 1}{2}\ln(2\pi n) + \frac{1}{2}\sum_{x\colon p(x)>0}\ln p(x)\right].$$

Лемма 6. *Для любых* $\ell, M, L \in \mathbb{Z}_+$, *таких что* $L < M$, *любых* $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$ *и* $\alpha \in \left[\frac{1}{1+L}, 1\right)$ *существует* (M, L)*-код с функцией кодирования вида* $\Psi\colon \mathcal{M} \to \mathcal{B}$, *для которого*

$$\ln P_{\mathrm{e}} \leqslant \frac{\alpha - 1}{\alpha}\left[\tau + \left(\inf_{x\in\mathcal{B}} \lambda \cdot \rho(x)\right) - \ln\frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha}, \tag{23}$$

где $\tau = \inf\limits_{x\in\mathcal{B}} D_\alpha(W(x) \,\|\, q_{\alpha,p}) - \lambda \cdot \rho(x)$.

Следует отметить, что можно получить аналогичную границу через информацию Реньи – Галлагера, см. лемму 12 в Приложении B. Прежде чем перейти к доказательству леммы 6, вкратце обсудим ее следствия. Для каналов без памяти с ограничениями по стоимости с аддитивными функциями стоимости можно применить лемму 6 для $\lambda = 0$ и распределений-произведений на входе, удовлетворяющих ограничению по стоимости $\varrho - \epsilon$. Тогда, применяя слабый закон больших чисел с учетом [26, леммы 14, 27 и 28], можно сделать вывод, что функция надежности ограничена снизу экспонентой сферической упаковки при довольно широких предположениях. В случае $\ell = 1$ этот результат можно усилить, применяя теорему Берри – Эссеена и [25, лемма 19] при соответствующих предположениях. Если получить результат, аналогичный [25, лемма 19], для сумм случайных векторов, то следующее утверждение можно будет немедленно обобщить на случай $\ell > 1$.

Следствие 2. *Пусть* $n \in \mathbb{Z}_+$, $n \geqslant 21$, $\phi \in (0,1)$, $W_{[1,n]}\colon \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ – *канал-произведение длины* n *с аддитивной функцией стоимости* $\rho_{[1,n]}\colon \mathcal{X}_1^n \to \mathbb{R}_{\geqslant 0}$, *такой что* $\inf\limits_{x_t\in\mathcal{X}_t} \rho_t(x_t) = 0$ *для всех* $t \in [1,n]$ *и*

$$C_{\alpha,\widetilde{W}_{[1,n]},n\varrho} \geqslant C_{\alpha,W_{[1,n]},n\varrho} - \epsilon, \quad \forall \alpha \in [\phi, 1),\ \varrho \in \left[\widetilde{\varrho}, \widetilde{\varrho} + \frac{9\varsigma}{n}\right], \tag{24}$$

для некоторых $\widetilde{\varrho} \in \mathbb{R}_+$, $\varsigma \in \mathbb{R}_{\geqslant 0}$, $\epsilon \in \mathbb{R}_{\geqslant 0}$ *и канала-произведения* $\widetilde{W}_{[1,n]}\colon \widetilde{\mathcal{X}}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ *с множествами входов* $\widetilde{\mathcal{X}}_t = \{x_t \in \mathcal{X}_t : \rho_t(x_t) \leqslant \varsigma\}$, *такого что* $\widetilde{W}_{[1,n]}(x_1^n) = W_{[1,n]}(x_1^n)$ *для всех* $x_1^n \in \widetilde{\mathcal{X}}_1^n$. *Тогда для любых* $\delta \in (0, \widetilde{\varrho})$, $\varrho \in \left(\widetilde{\varrho} + \frac{3e\varsigma}{n}, \widetilde{\varrho} + \frac{9\varsigma}{n}\right)$ *и целых чисел* M, L, *таких что* $\frac{1}{L+1} < \eta$ *и* $\ln\frac{M}{L} = C_{\eta,W_{[1,n]},n\varrho}$ *для некоторого* $\eta \in [\phi, 1)$, *существует* (M, L)*-код с функцией кодирования вида* $\Psi\colon \mathcal{M} \to \left\{x_1^n \in \widetilde{\mathcal{X}}_1^n : \rho_{[1,n]}(x_1^n) \leqslant n\varrho\right\}$, *для которого*

$$\begin{aligned}\ln P_{\mathrm{e}} &\leqslant -E_{\mathrm{sp}}\left(\ln\frac{M}{L}, W_{[1,n]}, n\varrho\right) + \\ &+ \frac{1-\eta}{\eta}\left(\frac{C_{\eta,W_{[1,n]},n\varrho}}{n}\frac{6\varsigma e}{\delta} + 2\epsilon + 1\right) + \frac{\ln 4n}{2\eta} + \frac{1}{n}.\end{aligned} \tag{25}$$

Заметим, что если бы область значений ϱ не зависела от ς в (24), а интервал значений α был бы компактным подмножеством в $(0,1)$, то можно было бы найти ς для любого положительного ϵ с помощью теоремы Дини [35, теорема 2.4.10], используя конструкцию, аналогичную той, что была использована в доказательстве [27, лемма 15(g)], поскольку $C_{\alpha,W,\varrho}$ непрерывна по (α,ϱ) на $(0,1)\times\mathbb{R}_+$ согласно [26, лемма 27(c)]. Это, однако, не так, и могут существовать каналы-произведения, для которых неравенство (24) не выполнено ни для какой тройки $(\epsilon, \varsigma, \widetilde{\varrho})$. Кроме того, даже когда условия следствия 2 выполнены, точность границы (25) зависит от значений констант, удовлетворяющих этим условиям. Тем не менее, легко видеть, что если либо все функции стоимости каналов-компонент ограничены, либо $C_{\phi,W_{[1,n]},n\varrho}$ имеет порядок $\Theta(n)$ (т.е. растет линейно по n), а условие (24) выполнено для ϵ и ς, имеющих порядок $O(\ln n)$ (т.е. растущих не более чем линейно по $\ln n$), то коэффициент

в (25) равен $e^{-O(\ln n)}$, т.е. убывает полиномиально по n. Доказательство следствия 2 дано в Приложении С.

Доказательство леммы 6. Установим существование кода с требуемыми свойствами методом случайного кодирования. Для любых $p \in \mathcal{P}(\mathcal{X})$ и $\mathcal{B} \subset \mathcal{X}$, таких что $p(\mathcal{B}) > 0$, определим $p_{\mathcal{B}} \in \mathcal{P}(\mathcal{X})$ как

$$p_{\mathcal{B}}(x) \triangleq \frac{\mathbb{1}_{\{x \in \mathcal{B}\}} p(x)}{p(\mathcal{B})}, \quad \forall x \in \mathcal{X}. \tag{26}$$

Рассмотрим ансамбль кодов, в котором распределения сообщений по элементам $\mathcal{B}$ совместно независимы и $\Psi(m) = x$ с вероятностью $p_{\mathcal{B}}(x)$ для всех m из множества сообщений. Декодер выбирает L сообщений с наибольшим отношением $\frac{f_{\Psi(m)}}{h_{\Psi(m)}}$ для функций f_x и h_x, определяемых следующим образом:

$$f_x(y) \triangleq \frac{dW(x)}{dq}(y), \quad \forall x \in \mathcal{X},\ y \in \mathcal{Y}, \tag{27}$$

$$h_x \triangleq e^{\frac{\alpha-1}{\alpha}\lambda \cdot \rho(x)}, \quad \forall x \in \mathcal{X}, \tag{28}$$

где мера $q \in \mathcal{P}(\mathcal{Y})$, которую мы выберем позже, такова, что $q_{\alpha,p} \prec q$. В случае равенства значений декодер выбирает сообщение (или сообщения) с наименьшим индексом (наименьшими индексами). Чтобы оценить математическое ожидание вероятности ошибки по ансамблю, рассмотрим математическое ожидание условной вероятности ошибки сообщения с наибольшим индексом. Ошибка возникает только тогда, когда отношения $\frac{f}{h}$, соответствующие L или более сообщениям, по крайней мере столь же велики, как $\frac{f}{h}$, соответствующее переданному сообщению. Вероятность этого можно оценить, используя зависящее от y вспомогательное пороговое значение $\gamma(y)$, и получить следующую оценку математического ожидания P_{e} по ансамблю[6]:

$$\begin{aligned}
&\mathbf{E}[P_{\mathrm{e}}] \leqslant \sum_x p_{\mathcal{B}}(x)\, \mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} \leqslant \gamma\}} f_x\right] + \\
&+ \binom{M-1}{L} \sum_x p_{\mathcal{B}}(x)\, \mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} > \gamma\}} \left[\sum_z p_{\mathcal{B}}(z) \mathbb{1}_{\{\frac{f_z}{h_z} \geqslant \frac{f_x}{h_x}\}}\right]^L f_x\right].
\end{aligned} \tag{29}$$

Наложим ограничения $\alpha \leqslant 1$, чтобы оценить первое слагаемое в этой сумме, а также $\alpha \geqslant \frac{1}{1+L}$, чтобы оценить второе:

$$\begin{aligned}
&\mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} \leqslant \gamma\}} f_x\right] \leqslant h_x\, \mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} \leqslant \gamma\}} \left(\frac{f_x}{h_x}\right)^{\alpha} \gamma^{1-\alpha}\right], \\
&\mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} > \gamma\}} \left[\sum_z p_{\mathcal{B}}(z) \mathbb{1}_{\{\frac{f_z}{h_z} \geqslant \frac{f_x}{h_x}\}}\right]^L f_x\right] \leqslant \\
&\leqslant h_x\, \mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} > \gamma\}} \left[\sum_z p_{\mathcal{B}}(z) \left(\frac{f_z}{h_z}\right)^{\alpha}\right]^L \left(\frac{f_x}{h_x}\right)^{1-L\alpha}\right] \leqslant \\
&\leqslant h_x\, \mathbf{E}_q\left[\mathbb{1}_{\{\frac{f_x}{h_x} > \gamma\}} \left[\sum_z p_{\mathcal{B}}(z) \left(\frac{f_z}{h_z}\right)^{\alpha}\right]^L \left(\frac{f_x}{h_x}\right)^{\alpha} \gamma^{1-\alpha-L\alpha}\right].
\end{aligned}$$

[6] Заметим, что $\sum\limits_{t=L}^{M} \binom{M}{t} s^t (1-s)^{M-t} = \binom{M}{L} s^L \sum\limits_{t=0}^{M-L} \frac{L!\,(M-L)!}{(L+t)!\,(M-L-t)} s^t (1-s)^{M-L-t} \leqslant \binom{M}{L} s^L$ для всех $s \in [0,1]$.

Полагая $\gamma = \left[\sum\limits_x p_{\mathcal{B}}(x)\Big(\frac{f_x}{h_x}\Big)^{\alpha}\right]^{\frac{1}{\alpha}}\Big[\binom{M-1}{L}\Big]^{\frac{1}{L\alpha}}$, получаем

$$\mathbf{E}[P_{\mathrm{e}}] \leqslant \sum_x p_{\mathcal{B}}(x)h_x\,\mathbf{E}_q\left[\left(\frac{f_x}{h_x}\right)^{\alpha}\gamma^{1-\alpha}\right] =$$

$$= \mathbf{E}_q\left[\left[\sum_x p_{\mathcal{B}}(x)\left(\frac{f_x}{h_x}\right)^{\alpha}\right]^{\frac{1}{\alpha}}\right]\Big(\sup_{x\in\mathcal{B}} h_x\Big)\left[\binom{M-1}{L}\right]^{\frac{1-\alpha}{L\alpha}}.$$

С другой стороны, используя формулу Стирлинга $\sqrt{2\pi n}(n/e)^n \leqslant n! \leqslant e\sqrt{n}(n/e)^n$ и неравенство $\ln z \leqslant z-1$, получаем

$$\frac{1}{L}\ln\binom{M-1}{L} \leqslant \frac{1}{L}\ln\frac{e\sqrt{M-1}}{2\pi\sqrt{L(M-1-L)}} + \ln\frac{M-1}{L} +$$
$$+\frac{M-1-L}{L}\ln\left(1+\frac{L}{M-1-L}\right) \leqslant \ln\frac{M-1}{L}+1.$$

Отсюда вытекает следующая оценка математического ожидания P_{e} по ансамблю:

$$\ln\mathbf{E}[P_{\mathrm{e}}] \leqslant \ln\mathbf{E}_q\left[\left(\sum_x p_{\mathcal{B}}(x)e^{(1-\alpha)\lambda\cdot\rho(x)}\left(\frac{dW(x)}{dq}\right)^{\alpha}\right)^{\frac{1}{\alpha}}\right] +$$
$$+\frac{\alpha-1}{\alpha}\left[\Big(\inf_{x\in\mathcal{B}}\lambda\cdot\rho(x)\Big) - \ln\frac{(M-1)e}{L}\right]. \tag{30}$$

Полагая $q = q_{\alpha,p}$ и используя вначале определение скошенного канала, данное в (3), а затем определение τ, получаем

$$\ln\mathbf{E}[P_{\mathrm{e}}] \leqslant \ln\mathbf{E}_{q_{\alpha,p}}\left[\left(\sum_x p_{\mathcal{B}}(x)e^{(\alpha-1)[D_\alpha(W(x)\|\,q_{\alpha,p})-\lambda\cdot\rho(x)]}\frac{dW_\alpha^{q_{\alpha,p}}(x)}{dq_{\alpha,p}}\right)^{\frac{1}{\alpha}}\right] +$$
$$+\frac{\alpha-1}{\alpha}\left[\Big(\inf_{x\in\mathcal{B}}\lambda\cdot\rho(x)\Big) - \ln\frac{(M-1)e}{L}\right] \leqslant \ln\mathbf{E}_{q_{\alpha,p}}\left[\left(\sum_x p_{\mathcal{B}}(x)\frac{dW_\alpha^{q_{\alpha,p}}(x)}{dq_{\alpha,p}}\right)^{\frac{1}{\alpha}}\right] +$$
$$+\frac{\alpha-1}{\alpha}\left[\tau + \Big(\inf_{x\in\mathcal{B}}\lambda\cdot\rho(x)\Big) - \ln\frac{(M-1)e}{L}\right].$$

Так как существует код с вероятностью ошибки P_{e}, меньшей или равной $\mathbf{E}[P_{\mathrm{e}}]$, то существование кода с функцией кодирования вида $\Psi\colon \mathcal{M}\to\mathcal{B}$, удовлетворяющего условию (23), следует из (4), (7) и (26). ▲

§ 3. Границы сферической упаковки для каналов без памяти

Предположение 1. *Максимальное значение $C_{1/2,U_{\mathcal{A}}^{(t)}}$ для целых t, меньших или равных n, имеет порядок $O(\ln n)$, т.е.*

$$\exists n_0\in\mathbb{Z}_+,\ K\in\mathbb{R}_+,\quad \textit{такие что}\quad \max_{t:\,t\leqslant n} C_{1/2,U_{\mathcal{A}}^{(t)}} \leqslant K\ln(n),\quad \forall n\geqslant n_0,$$

где канал $U_{\mathcal{A}}^{(t)}\colon \mathcal{X}_{\mathcal{A}}^{(t)}\to\mathcal{P}(\mathcal{Y})$ таков, что $U_{\mathcal{A}}^{(t)}(x) = W(x)$ для всех $x\in\mathcal{X}_{\mathcal{A}}^{(t)}$, а $\mathcal{X}_{\mathcal{A}}^{(t)}$ определяется для любых $t\in\mathbb{Z}_+$ и $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ как

$$\mathcal{X}_{\mathcal{A}}^{(t)} \triangleq \{x\in\mathcal{X}:\ \exists p\in\mathcal{A},\ \textit{такая что } p(z)t\in\mathbb{Z},\ \forall z\in\mathcal{X},\ \textit{таких что } p(x)>0\}.$$

Теорема 1. *Пусть* $\{W_t\}_{t\in\mathbb{Z}_+}$ – *стационарная последовательность каналов, такая что* $W_t = W$ *для всех* $t \in \mathbb{Z}_+$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ – *выпуклое множество ограничений, удовлетворяющее предположению* 1, *а* ε, α_0, α_1 – *положительные параметры, такие что* $0 < \alpha_0 < \alpha_1 < 1$. *Тогда для любой последовательности кодов* $\left\{(\Psi^{(n)}, \Theta^{(n)})\right\}_{n\in\mathbb{Z}_+}$ *на каналах-произведениях* $\left\{W_{[1,n]}\right\}_{n\in\mathbb{Z}_+}$, *такой что*

$$C_{\alpha_1,W,\mathcal{A}} \geqslant \frac{1}{n}\ln\frac{M_n}{L_n} \geqslant C_{\alpha_0,W,\mathcal{A}} + \frac{\ln n}{n}[\ln(\ln n)]^{\varepsilon}, \quad \forall n \geqslant n_0, \tag{31}$$

и $\Upsilon(\Psi^{(n)}(m)) \in \mathcal{A}$ *для всех* $m \in \mathcal{M}^{(n)}$, *существуют* $\tau \in \mathbb{R}_+$ *и* $n_1 \geqslant n_0$, *такие что*

$$P_{\mathrm{e}}^{(n)} \geqslant n^{-\tau} e^{-nE_{\mathrm{sp}}\left(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\right)}, \quad \forall n \geqslant n_1. \tag{32}$$

Следует отметить, что условия теоремы 1 выполнены для гауссовских и пуассоновских моделей, которые изучались в [15–22]; таким образом, из теоремы 1 следуют асимптотические ГСУ для таких каналов. Неасимптотический аналог этого утверждения представлен в п. 3.3 между леммой 9 и следствием 3.

Для кодов в каналах без памяти с ограничениями по стоимости можно отказаться от условия стационарности, используя технику выпуклого сопряжения (см. [24, теорема 36.6]). Теорема 2 усиливает [24, теорема 36.6], избавляясь от условия ограниченности функции стоимости и устанавливая ГСУ с коэффициентом, полиномиальным по длине блока n, вместо коэффициента вида $e^{-O(\sqrt{n})}$.

Предположение 2. *Максимальное значение* $C_{1/2,W_t,n\varrho}$ *для целых* t, *меньших или равных* n, *имеет порядок* $O(\ln n)$:

$$\exists n_0 \in \mathbb{Z}_+,\ K \in \mathbb{R}_+, \quad \textit{такие что} \quad \max_{t:\, t\leqslant n} C_{1/2,W_t,n\varrho} \leqslant K\ln(n), \quad \forall n \geqslant n_0.$$

Теорема 2. *Пусть* $\{(W_t, \rho_t)\}_{t\in\mathbb{Z}_+}$ – *последовательность каналов с соответствующими функциями стоимости, удовлетворяющая предположению* 2, *а* ε, α_0, α_1 – *положительные параметры, такие что* $0 < \alpha_0 < \alpha_1 < 1$. *Тогда в каналах-произведениях* $\left\{W_{[1,n]}\right\}_{n\in\mathbb{Z}_+}$ *с аддитивными функциями стоимости* $\left\{\rho_{[1,n]}\right\}_{n\in\mathbb{Z}_+}$ *вида* $\rho_{[1,n]}(x_1^n) = \sum\limits_{t=1}^{n}\rho_t(x_t)$, *для любой последовательности кодов* $\left\{(\Psi^{(n)}, \Theta^{(n)})\right\}_{n\in\mathbb{Z}_+}$, *такой что*

$$C_{\alpha_1,W_{[1,n]},n\varrho} \geqslant \ln\frac{M_n}{L_n} \geqslant C_{\alpha_0,W_{[1,n]},n\varrho} + \varepsilon(\ln n)^2, \quad \forall n \geqslant n_0, \tag{33}$$

и $\rho_{[1,n]}(\Psi^{(n)}(m)) \leqslant n\varrho$ *для всех* $m \in \mathcal{M}^{(n)}$ *при ограничении по стоимости* ϱ *на одно обращение к каналу, таком что* $n\varrho \in \operatorname{int}\Gamma_{\rho_{[1,n]}}$, *существуют* $\tau \in \mathbb{R}_+$ *и* $n_1 \geqslant n_0$, *для которых*

$$P_{\mathrm{e}}^{(n)} \geqslant n^{-\tau} e^{-E_{\mathrm{sp}}\left(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\right)}, \quad \forall n \geqslant n_1. \tag{34}$$

Чтобы установить асимптотическую ГСУ с полиномиальным коэффициентом в теореме 2, мы предполагали, что августиновская пропускная способность порядка $1/2$ с ограничениями по стоимости растет не быстрее чем логарифмически по стоимости. Хотя это условие и выполнено для большинства содержательных случаев, включая различные гауссовские и пуассоновские каналы, изучавшиеся в [15–22], а также многие их нестационарные варианты, существуют и каналы, для которых это условие не выполнено, см. [26, пример 1]. Таким образом, может иметь смысл избавиться от условия логарифмического роста по ограничению по стоимости. Этого можно добиться, если в доказательстве теоремы 2 использовать лемму 11 вместо леммы 10.

Наша основная цель в этом параграфе – доказать две асимптотические ГСУ, сформулированные в теоремах 1 и 2, которые и представляют собой основной результат настоящей статьи. Для этого вначале в п. 3.1 мы изложим результат о невозможности для задачи проверки гипотез с независимыми выборками с помощью теоремы Берри – Эссеена. Затем в п. 3.2 введем понятия усредненной августиновской пропускной способности и усредненной ЭСУ. Мы выведем неасимптотические ГСУ (хотя и в параметрическом виде) через эти усредненные величины для кодов с ограничениями по композиции в стационарных каналах без памяти в п. 3.3 и для кодов с ограничениями по стоимости в (возможно, нестационарных) каналах без памяти в п. 3.4. Достаточно очевидный вывод этих асимптотических ГСУ из неасимптотических с помощью леммы 8 из п. 3.2 представлен в Приложении C.

3.1. Результат о невозможности для задачи проверки гипотез.

Лемма 7. *Для любых* $\kappa \geqslant 3$, $\alpha \in (0,1)$, $n \in \mathbb{Z}_+$ *и произведений мер* $w, q \in \mathcal{P}(\mathcal{Y}_1^n)$ *вида* $w = \bigotimes_{t=1}^{n} w_t$ *и* $q = \bigotimes_{t=1}^{n} q_t$ *положим* $\xi_t \triangleq \ln \frac{dw_{t,\mathrm{ac}}}{dq_t} - \mathbf{E}_{w_\alpha^q}\left[\ln \frac{dw_{t,\mathrm{ac}}}{dq_t}\right]$, *где* $w_{t,\mathrm{ac}}$ – *компонента* w_t, *абсолютно непрерывная по* q_t *и такая, что*

$$g_\kappa \triangleq \left(\sum_{t=1}^{n} \mathbf{E}_{w_\alpha^q}\left[|\xi_t|^\kappa\right]\right)^{1/\kappa}.$$

Тогда любое $\mathcal{E} \in \mathcal{Y}$, *такое что* $q(\mathcal{E}) \leqslant \frac{1}{4\sqrt{n}} e^{-D_1(w_\alpha^q \| q) - \alpha 3 g_\kappa}$, *также удовлетворяет условию*

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geqslant \frac{1}{4\sqrt{n}} e^{-D_1(w_\alpha^q \| w) - (1-\alpha) 3 g_\kappa}. \tag{35}$$

Лемма 7 дает результат о невозможности для задачи проверки гипотез с независимыми выборками в духе [7, теорема 5]. Однако вместо неравенства Чебышева в ней применяется теорема Берри – Эссеена, использованная в [25, лемма 19].

Доказательство леммы 7. Положим $\mathcal{E}_0 = \left\{y : \left|\sum_{t=1}^{n} \xi_t\right| \leqslant 3g_\kappa\right\}$; тогда

$$\left|\ln \frac{dw_\alpha^q}{dq}(y_1^n) - D_1(w_\alpha^q \,\|\, q)\right| \leqslant \alpha 3 g_\kappa, \quad \forall y_1^n \in \mathcal{E}_0,$$
$$\left|\ln \frac{dw_\alpha^q}{dw}(y_1^n) - D_1(w_\alpha^q \,\|\, w)\right| \leqslant (1-\alpha) 3 g_\kappa, \quad \forall y_1^n \in \mathcal{E}_0.$$

Отсюда

$$w_\alpha^q(\mathcal{E} \cap \mathcal{E}_0) \leqslant q(\mathcal{E} \cap \mathcal{E}_0) e^{D_1(w_\alpha^q \| q) + \alpha 3 g_\kappa}, \tag{36}$$
$$w_\alpha^q(\mathcal{E} \cap \mathcal{E}_0) \leqslant w(\mathcal{E} \cap \mathcal{E}_0) e^{D_1(w_\alpha^q \| w) + (1-\alpha) 3 g_\kappa}. \tag{37}$$

С другой стороны, в силу определения $\mathcal{E}_0$ и [25, лемма 19] имеем

$$\mathbf{P}_{w_\alpha^q}[\mathcal{E}_0] \geqslant \frac{1}{2\sqrt{n}}. \tag{38}$$

Неравенство (35) следует из (36)–(38). ▲

3.2. Усреднение по Августину. Наши неасимптотические ГСУ выражаются через усредненную августиновскую пропускную способность и усредненную ЭСУ, которые

определяются для всех $\epsilon \in (0,1)$ и $R \in \mathbb{R}_{\geqslant 0}$ следующим образом:

$$\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} \triangleq \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[1 \vee \left(\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\right] C_{\eta,W,\mathcal{A}}\, d\eta, \tag{39}$$

$$\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}) \triangleq \sup_{\alpha\in(0,1)} \frac{1-\alpha}{\alpha}\big(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R\big). \tag{40}$$

Заметим, что $\lim\limits_{\epsilon\downarrow 0} \widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} = C_{\alpha,W,\mathcal{A}}$ для всех $\alpha \in (0,1)$, поскольку $C_{\alpha,W,\mathcal{A}}$ непрерывна по α на $(0,1)$ согласно [26, лемма 23(d)]. Кроме того, используя монотонность функций $C_{\alpha,W,\mathcal{A}}$ и $\dfrac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ по α (см. [26, лемма 23(a),(b)]), можно показать, что сходимость на компактных подмножествах $(0,1)$ равномерна. Тем не менее, $C_{\alpha,W,\mathcal{A}}$ нельзя аппроксимировать величиной $\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$ равномерно на всем интервале $(0,1)$, поскольку $\lim\limits_{\alpha\uparrow 1} \widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} = \infty$ для всех положительных ϵ, когда $C_{1/2,W,\mathcal{A}}$ положительна, даже если $C_{1,W,\mathcal{A}}$ конечна. Этот факт следует из монотонности функций $C_{\alpha,W,\mathcal{A}}$ и $\dfrac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ по α:

$$\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} \geqslant \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha-\frac{\epsilon}{2}\alpha} \frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta} C_{\eta,W,\mathcal{A}}\, d\eta \geqslant$$

$$\geqslant \frac{\alpha}{2-\epsilon}\left(1 + \frac{\alpha\epsilon}{2(1-\alpha)}\right) C_{\alpha-\epsilon\alpha,W,\mathcal{A}}.$$

Как функция скорости, усредненная ЭСУ равномерно сходится к ЭСУ на любом компактном множестве скоростей, меньших августиновской пропускной способности порядка 1, как показывает следующая

Лемма 8. *Для любых* $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *и* $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, *таких что* $C_{1/2,W,\mathcal{A}} \in \mathbb{R}_+$, *и любых* $\phi \in (0,1)$, $R \in [C_{\phi,W,\mathcal{A}}, \infty)$ *и* $\epsilon \in (0,\phi)$

$$0 \leqslant \widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}) - E_{\mathrm{sp}}(R,W,\mathcal{A}) \leqslant \frac{\epsilon}{1-\epsilon}\frac{R \vee E_{\mathrm{sp}}(R,W,\mathcal{A})}{\phi} \leqslant \tag{41}$$

$$\leqslant \frac{\epsilon}{1-\epsilon}\frac{R}{\phi^2}. \tag{42}$$

Лемма 8 вытекает из монотонности функций $C_{\alpha,W,\mathcal{A}}$ и $\dfrac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ по α, установленной в [26, лемма 23(a),(b)]; доказательство приведено в Приложении C. Доказательство леммы 8 практически совпадает с доказательством леммы 15 из [25], которая устанавливает тот же самый результат для случая $\mathcal{A} = \mathcal{P}(\mathcal{X})$.

3.3. Неасимптотические границы сферической упаковки для кодов с ограничениями по композиции. Коды с ограничениями по композиции в стационарных каналах без памяти активно изучались, но лишь для случая конечного множества входов. Чаще всего предполагается, что все кодовые слова имеют в точности одинаковую композицию. Если это не так, обычно применяется метод выбрасывания с учетом композиций кодовых слов, при котором оставляются слова с наиболее частой композицией. Однако такое выбрасывание приводит к нетривиальному результату только тогда, когда множество входов конечно. Границы, получаемые при анализе наиболее частой композиции в таких случаях, можно выводить через информацию Августина и августиновское среднее. Напротив того, в следующей лемме используются понятия

августиновских пропускной способности и центра, а также усреднения по Августину, чтобы избежать предположений о конечности множества входов или о композиции, имеющей не менее чем полиномиальную по длине блока долю кодовых слов.

Лемма 9. *Пусть* $n \in \mathbb{Z}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A}$ – *выпуклое подмножество* $\mathcal{P}(\mathcal{X})$, κ, ϕ, ϵ – *положительные параметры, такие что* $\kappa \geqslant 3$, $\phi < 1$, $\epsilon < 1$, *и пусть*

$$\gamma \triangleq 3\sqrt[\kappa]{3n}\left(\left[C_{1/2,U_{\mathcal{A}}^{(n)}} + (1-\phi)\ln(1+n)\right] \vee \kappa\right). \tag{43}$$

Если $W_t = W$ *для всех* $t \leqslant n$, *а* $M, L \in \mathbb{Z}_+$ *таковы, что* $\ln\dfrac{M}{L} > n\widetilde{C}^{\epsilon}_{\phi,W,\mathcal{A}} + \dfrac{\gamma}{1-\phi} + \ln\dfrac{8e^3n^{1,5}}{\epsilon}$, *то для любого* (M,L)*-кода в канале* $W_{[1,n]}$, *такого что* $\Upsilon(\Psi(m)) \in \mathcal{A}$ *для всех* $m \in \mathcal{M}$, *справедливо*

$$P_{\mathrm{e}} \geqslant \left(\frac{\epsilon e^{-2\gamma}}{8e^3n^{1,5}}\right)^{1/\phi} e^{-n\widetilde{E}^{\epsilon}_{\mathrm{sp}}\left(\frac{1}{n}\ln\frac{M}{L},W,\mathcal{A}\right)}. \tag{44}$$

Отсутствие в (44) члена, учитывающего скорость кода, можно считать достоинством данной формулы, но это возможно только за счет наличия коэффициента, при котором граница становится тривиальной, т.е. равной нулю, когда ϕ стремится к нулю. Слегка изменив рассуждения в доказательстве, можно получить следующую альтернативную границу:

$$P_{\mathrm{e}} \geqslant \frac{\epsilon e^{-2\gamma}}{8n^{1,5}} e^{-n\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A})}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{1}{n}\ln\frac{8e^3n^{1,5}}{\epsilon e^{-2\gamma}}. \tag{45}$$

С учетом соотношения $C_{\alpha,U_{\mathcal{A}(\varrho)}^{(n)}} \leqslant C_{\alpha,W,n\varrho}$ из леммы 9 и формулы (45) вытекают приведенные ниже следствие 3 и формула (46).

Следствие 3. *Пусть* $n, \ell \in \mathbb{Z}_+$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$, $\varrho \in \Gamma_\rho$, κ, ϕ, ϵ – *положительные параметры, такие что* $\kappa \geqslant 3$, $\phi < 1$, $\epsilon < 1$, *и пусть*

$$\gamma \triangleq 3\sqrt[\kappa]{3n}\left(\left[C_{1/2,W,n\varrho} + (1-\phi)\ln(1+n)\right] \vee \kappa\right).$$

Если $W_t = W$ *для всех* $t \leqslant n$, *а* $M, L \in \mathbb{Z}_+$ *таковы, что* $\ln\dfrac{M}{L} > n\widetilde{C}^{\epsilon}_{\phi,W,\varrho} + \dfrac{\gamma}{1-\phi} + \ln\dfrac{8e^3n^{1,5}}{\epsilon}$, *то для любого* (M,L)*-кода в канале* $W_{[1,n]}$, *такого что* $\sum\limits_{t=1}^{n}\rho(\Psi_t(m)) \leqslant n\varrho$ *для всех* $m \in \mathcal{M}$, *справедливо*

$$P_{\mathrm{e}} \geqslant \left(\frac{\epsilon e^{-2\gamma}}{8e^3n^{1,5}}\right)^{1/\phi} e^{-n\widetilde{E}^{\epsilon}_{\mathrm{sp}}\left(\frac{1}{n}\ln\frac{M}{L},W,\varrho\right)}.$$

Слегка изменив рассуждения, можно получить следующую альтернативную границу:

$$P_{\mathrm{e}} \geqslant \frac{\epsilon e^{-2\gamma}}{8n^{1,5}} e^{-n\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\varrho)}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{1}{n}\ln\frac{8e^3n^{1,5}}{\epsilon e^{-2\gamma}}. \tag{46}$$

Доказательство леммы 9 и формулы (45). Августиновский центр $q_{\alpha,W,\mathcal{A}}$ непрерывен по α на $(0,1)$ в топологии полной вариации на $\mathcal{P}(\mathcal{Y})$ согласно [26, леммы 23(d) и 24]. Таким образом, $q_{\cdot,W,\mathcal{A}}$ – это вероятность перехода из $((0,1),\mathcal{B}((0,1)))$ в $(\mathcal{Y},\mathcal{Y})$. Для каждого $t \leqslant n$ определим усредненный центр $q^{\epsilon}_{\alpha,t}$ как маргинальное распределение на $\mathcal{Y}$ для вероятностной меры $u_{\alpha,\epsilon} \circledast q_{\cdot,W,\mathcal{A}}$, где $u_{\alpha,\epsilon}$ –

равномерное распределение вероятностей на $(\alpha - \epsilon\alpha, \alpha + \epsilon(1-\alpha))$:

$$q^{\epsilon}_{\alpha,t} \triangleq \frac{1}{\epsilon} \int\limits_{\alpha-\alpha\epsilon}^{\alpha+(1-\alpha)\epsilon} q_{\eta,W,\mathcal{A}}\, d\eta, \quad \forall t \in \{1,\dots,n\}. \tag{47}$$

Определим $q_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ и $q_\alpha \in \mathcal{P}(\mathcal{Y}_1^n)$ как

$$q_{\alpha,t} \triangleq \frac{n}{n+1} q^{\epsilon}_{\alpha,t} + \frac{1}{n+1} q_{1/2,U^{(n)}_{\mathcal{A}}}, \quad \forall t \in \{1,\dots,n\},$$
$$q_\alpha \triangleq \bigotimes_{t=1}^{n} q_{\alpha,t}.$$

Для краткости обозначим вероятностную меру $\bigotimes\limits_{t=1}^{n} W(\Psi_t(m))$, порожденную $\Psi(m)$, через w^m. Имеем

$$D_\alpha(w^m \,\|\, q_\alpha) \overset{(a)}{=} \sum_{t=1}^{n} D_\alpha\big(W(\Psi_t(m)) \,\|\, q_{\alpha,t}\big) \overset{(b)}{\leqslant}$$
$$\overset{(b)}{\leqslant} \sum_{t=1}^{n} \left(\ln \frac{n+1}{n} + D_\alpha(W(\Psi_t(m)) \,\|\, q^{\epsilon}_{\alpha,t}) \right) \overset{(c)}{\leqslant} 1 + \sum_{t=1}^{n} D_\alpha\big(W(\Psi_t(m)) \,\|\, q^{\epsilon}_{\alpha,t}\big) \overset{(d)}{\leqslant}$$
$$\overset{(d)}{\leqslant} 1 + \sum_{t=1}^{n} \frac{1}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(W(\Psi_t(m)) \,\|\, q_{\eta,W,\mathcal{A}}\big)\, d\eta \overset{(e)}{=}$$
$$\overset{(e)}{=} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(W \,\|\, q_{\eta,W,\mathcal{A}} \,|\, \Upsilon(\Psi(m))\big)\, d\eta \overset{(f)}{\leqslant}$$
$$\overset{(f)}{\leqslant} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right) D_\eta\big(W \,\|\, q_{\eta,W,\mathcal{A}} \,|\, \Upsilon(\Psi(m))\big) \overset{(g)}{\leqslant}$$
$$\overset{(g)}{\leqslant} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right) C_{\eta,W,\mathcal{A}}\, d\eta \overset{(h)}{=} 1 + n\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}, \tag{48}$$

где (a) справедливо в силу [36, теорема 28], (b) – в силу [26, лемма 1], так как $\frac{n}{n+1} q^{\epsilon}_{\alpha,t} \leqslant q_{\alpha,t}$, (c) – в силу $\ln\tau \leqslant \tau - 1$, (d) – в силу неравенства Йенсена и [36, теорема 12], (e) – согласно определению 3 и определению композиции Υ, (f) – в силу [36, теорема 3 и предложение 2], (g) – в силу [26, теорема 1], так как $\Upsilon(\Psi(m)) \in \mathcal{A}$, и (h) – по определению величины $\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$. Пусть v^m_α – скошенная вероятностная мера порядка α между w^m и q_α, определенная в (2). Тогда v^m_α имеет вид $v^m_\alpha = \bigotimes\limits_{t=1}^{n} W^{q_{\alpha,t}}_\alpha(\Psi_t(m))$ в силу структуры w^m и q_α как произведений мер. Введем случайные величины

$$\xi^m_{\alpha,t} \triangleq \ln \frac{d[W(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}} - \mathbf{E}_{v^m_\alpha}\left[\ln \frac{d[W(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}}\right],$$
$$\xi^m_\alpha \triangleq \sum_{t=1}^{n} \xi^m_{\alpha,t},$$

где $[W(\Psi_t(m))]_{\text{ac}}$ – компонента $W(\Psi_t(m))$, абсолютно непрерывная по $q_{\alpha,t}$. Тогда для всех $\kappa \in \mathbb{R}_+$ и $\alpha \in (0,1)$

$$\mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]^{1/\kappa} \overset{(a)}{\leqslant} 3^{1/\kappa}\frac{\left[(1-\alpha)D_\alpha(W(\Psi_t(m))\,\|\,q_{\alpha,t})\right]\vee\kappa}{\alpha(1-\alpha)} \overset{(b)}{\leqslant}$$
$$\overset{(b)}{\leqslant} 3^{1/\kappa}\frac{\left[(1-\alpha)D_\alpha\big(W(\Psi_t(m))\,\|\,q_{1/2,U_{\mathcal{A}}^{(n)}}\big)+(1-\alpha)\ln(1+n)\right]\vee\kappa}{\alpha(1-\alpha)} \overset{(c)}{\leqslant}$$
$$\overset{(c)}{\leqslant} 3^{1/\kappa}\frac{\left[D_{1/2}\big(W(\Psi_t(m))\,\|\,q_{1/2,U_{\mathcal{A}}^{(n)}}\big)+(1-\alpha)\ln(1+n)\right]\vee\kappa}{\alpha(1-\alpha)} \overset{(d)}{\leqslant}$$
$$\overset{(d)}{\leqslant} 3^{1/\kappa}\frac{\left[C_{1/2,U_{\mathcal{A}}^{(n)}}+(1-\alpha)\ln(1+n)\right]\vee\kappa}{\alpha(1-\alpha)},$$

где (a) справедливо в силу [25, лемма 17], (b) – в силу [26, лемма 1], так как $\frac{q_{1/2,U_{\mathcal{A}}^{(n)}}}{n+1} \leqslant$ $\leqslant q_{\alpha,t}$, (c) – в силу [36, теорема 3 и предложение 2], а (d) – в силу [26, теорема 1], так как $\Psi_t(m) \in \mathcal{X}_{\mathcal{A}}^{(n)}$. Тогда, используя определение γ, данное в (43), получаем

$$\left[\sum_{t=1}^n \mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]\right]^{1/\kappa} \leqslant \frac{\gamma}{3\alpha(1-\alpha)}, \quad \forall\alpha\in[\phi,1). \tag{49}$$

С другой стороны, из [36, теорема 30] следует

$$D_1(v_\alpha^m\,\|\,q_\alpha) = D_\alpha(w^m\,\|\,q_\alpha) - \frac{\alpha}{1-\alpha}D_1(v_\alpha^m\,\|\,w^m), \quad \forall m\in\mathcal{M},\ \alpha\in(0,1). \tag{50}$$

Таким образом, можно оценить $D_1(v_\alpha^m\,\|\,q_\alpha)$, пользуясь неотрицательностью дивергенции Реньи (см. [36, теорема 8]) и соотношением (48):

$$0 \leqslant D_1(v_\alpha^m\,\|\,q_\alpha) \leqslant 1 + n\widetilde{C}^\epsilon_{\alpha,W,\mathcal{A}}.$$

Отсюда

$$\lim_{\alpha\downarrow\phi} D_1(v_\alpha^m\,\|\,q_\alpha) + \frac{\gamma}{1-\alpha} < \ln\frac{M}{L}\frac{\epsilon}{8e^2n^{1,5}},$$
$$\lim_{\alpha\uparrow 1} D_1(v_\alpha^m\,\|\,q_\alpha) + \frac{\gamma}{1-\alpha} = \infty.$$

Функция $D_1(v_\alpha^m\,\|\,q_\alpha)$ непрерывна по α в силу [25, лемма 16], поскольку q_α непрерывна по α в топологии полной вариации[7] на $\mathcal{P}(\mathcal{Y})$. Тогда по теореме о среднем [37, теорема 4.23] для каждого $m\in\mathcal{M}$ существует $\alpha_m\in(\phi,1)$, такое что

$$\left(D_1(v_\alpha^m\,\|\,q_\alpha)+\frac{\gamma}{1-\alpha}\right)\Bigg|_{\alpha=\alpha_m} = \ln\frac{M}{L}\frac{\epsilon}{8e^2n^{1,5}}. \tag{51}$$

Для любого $K\in\mathbb{Z}_+$ в $(0,1)$ найдется замкнутый подынтервал длины $1/K$, в который попали $\left\lceil\frac{M}{K}\right\rceil$ или более таких α_m. Пусть этот интервал имеет вид $[\eta,\eta+1/K]$,

[7] В частности, $\|q_\alpha - q_\eta\| \leqslant \sqrt{8\ln\frac{\epsilon}{\epsilon-(1-\epsilon)|\alpha-\eta|}}$, так как $\|q_\alpha-q_\eta\| \leqslant \sqrt{4D_{1/2}(q_\alpha\,\|\,q_\eta)}$ согласно [36, теорема 31], $D_{1/2}(q_\alpha\,\|\,q_\eta) = nD_{1/2}(q_{\alpha,t}\,\|\,q_{\eta,t})$ в силу [36, теорема 28] и определения q_α, $D_{1/2}(q_{\alpha,t}\,\|\,q_{\eta,t}) \leqslant 2\ln\frac{2}{2-\|q_{\alpha,t}-q_{\eta,t}\|}$ согласно [26, формула (9)], и $\|q_{\alpha,t}-q_{\eta,t}\| \leqslant 2\frac{1-\epsilon}{\epsilon}|\eta-\alpha|$ по определению q_α.

и положим

$$\widetilde{q} \triangleq \bigotimes_{t=1}^{n} \widetilde{q}_t, \quad \widetilde{q}_t = \frac{n}{n+1} q_{\widetilde{\alpha},t}^{\widetilde{\epsilon}} + \frac{1}{n+1} q_{1/2,U_{\mathcal{A}}^{(n)}},$$

где

$$\widetilde{\epsilon} = \frac{1}{K} + \epsilon\Big(1 - \frac{1}{K}\Big), \quad \widetilde{\alpha} = \frac{1-\epsilon}{1-\widetilde{\epsilon}}\eta.$$

Тогда для всех $\alpha \in \left[\eta, \eta + \frac{1}{K}\right]$ согласно определению усредненного центра $q_{\alpha,t}^{\epsilon}$, данному в (47), получаем

$$q_{\alpha,t} \leqslant \frac{\widetilde{\epsilon}}{\epsilon}\widetilde{q}_t, \quad q_\alpha \leqslant \Big(\frac{\widetilde{\epsilon}}{\epsilon}\Big)^n \widetilde{q}.$$

По крайней мере половина сообщений, для которых $\alpha_m \in [\eta, \eta+1/K]$, т.е. по крайней мере $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K} \right\rceil \right\rceil$ сообщений, удовлетворяют условию $\widetilde{q}(\mathcal{E}_m) \leqslant 2\frac{L}{\lceil M/K \rceil}$ в силу неравенства Маркова, так как $\sum\limits_{m\in\widetilde{\mathcal{M}}} \widetilde{q}(\mathcal{E}_m) \leqslant L$ по определению списочного декодирования, где $\widetilde{\mathcal{M}} = \{m :\ \alpha_m \in [\eta, \eta + 1/K]\}$. Тогда по крайней мере $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K} \right\rceil \right\rceil$ сообщений, для которых $\alpha_m \in [\eta, \eta + 1/K]$, удовлетворяют условию

$$q_{\alpha_m}(\mathcal{E}_m) \leqslant \frac{2LK}{M}\left(1 + \frac{1}{K}\frac{1-\epsilon}{\epsilon}\right)^n \leqslant \frac{2LK}{M}\left(1 + \frac{1}{K\epsilon}\right)^n.$$

Заметим, что $\frac{n}{\epsilon} > 1$, так как мы предположили, что $\epsilon < 1$ и $n \geqslant 1$. Полагая $K = \left\lfloor \frac{n}{\epsilon} \right\rfloor$ и применяя соотношение $(1+\tau)^{1/\tau} < e$ и (51), получаем

$$q_{\alpha_m}(\mathcal{E}_m) \leqslant \frac{2LK}{M}\left(1 + \frac{1}{K\epsilon}\right)^{Ke\frac{n}{Ke}} \leqslant \frac{2L}{M}\frac{n}{\epsilon}e^2 \leqslant \frac{1}{4\sqrt{n}} e^{-D_1\left(v_{\alpha_m}^m \| q_{\alpha_m}\right) - \frac{\gamma}{1-\alpha_m}}.$$

Тогда из неравенства (49) и леммы 7 следует, что

$$P_{\mathrm{e}}^m \geqslant \frac{1}{4\sqrt{n}} e^{-D_1\left(v_{\alpha_m}^m \| w^m\right) - \frac{\gamma}{\alpha_m}}. \tag{52}$$

Из (48) и (50)–(52) получаем

$$P_{\mathrm{e}}^m \geqslant \frac{1}{4\sqrt{n}}\left(\frac{\epsilon}{8e^2 n^{1,5}}\frac{M}{L}\right)^{\frac{1-\alpha_m}{\alpha_m}} e^{-\frac{1-\alpha_m}{\alpha_m}\left(1 + n\widetilde{C}_{\alpha_m,W}^{\epsilon}\right) - \frac{2\gamma}{\alpha_m}}. \tag{53}$$

Следовательно, для всех m, удовлетворяющих (51), в силу определения $\widetilde{E}_{\mathrm{sp}}^{\epsilon}(R, W, \mathcal{A})$, данного в (40), справедливо

$$P_{\mathrm{e}}^m \geqslant \frac{e^{-2\gamma/\phi}}{4\sqrt{n}}\left(\frac{\epsilon}{8e^3 n^{1,5}}\right)^{\frac{1-\phi}{\phi}} e^{-n\widetilde{E}_{\mathrm{sp}}^{\epsilon}(R,W,\mathcal{A})}, \quad R = \frac{1}{n}\ln\frac{M}{L}.$$

Поскольку таких сообщений не менее $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K} \right\rceil \right\rceil$, а по построению $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K} \right\rceil \right\rceil \geqslant \frac{M\epsilon}{2n}$, получаем нижнюю границу (44).

Заметим, что $\frac{\epsilon e^{-2\gamma}}{8e^3 n^{1,5}} < 1$, так как $\epsilon \in (0,1)$, $n \geqslant 1$ и $\gamma \geqslant 0$. Таким образом, коэффициент в (44) стремится к нулю при ϕ, стремящемся к нулю. Чтобы избежать этого, можно изменить рассуждения после неравенства (53), добавляя ошибку аппрокси-

мации к скорости в члене с усредненной ЭСУ: Для всех m, удовлетворяющих (51), в силу определения $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\varrho)$, данного в (40), и неравенства (53) справедливо

$$P_{\mathrm{e}}^{m} \geqslant \frac{e^{-2\gamma}}{4\sqrt{n}} e^{-n\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W)}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{2\gamma}{n} - \frac{1}{n}\ln\frac{8e^3 n^{1,5}}{\epsilon}.$$

Поскольку таких сообщений не менее $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil$, а по построению $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil \geqslant \frac{M\epsilon}{2n}$, получаем нижнюю границу (45). ▲

3.4. Неасимптотические границы сферической упаковки для кодов с ограничениями по стоимости. В п. 3.3 нас в основном интересовали стационарные каналы без памяти с выпуклыми ограничениями по композиции, а каналы без памяти с ограничениями по стоимости, рассмотренные в следствии 3, возникли просто как побочный результат. Однако, как будет показано далее, для каналов без памяти с ограничениями по стоимости можно установить ГСУ даже для нестационарных каналов, используя технику выпуклого сопряжения. Единственным недостатком использования техники выпуклого сопряжения будет то, что мы сможем установить ГСУ для ограничений по стоимости, лежащих в $\operatorname{int}\Gamma_\rho$, а не в Γ_ρ.

Лемма 10. *Пусть $n \in \mathbb{Z}_+$, $W_{[1,n]}\colon \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ – канал-произведение длины n с аддитивной функцией стоимости $\rho_{[1,n]}\colon \mathcal{X}_1^n \to \mathbb{R}^{\ell}_{\geqslant 0}$, такой что $\rho_{[1,n]}(x_1^n) = \sum\limits_{t=1}^{n}\rho_t(x_t)$ для некоторого $\rho_t\colon \mathcal{X}_t \to \mathbb{R}^{\ell}_{\geqslant 0}$, ϱ – ограничение по стоимости на одно обращение к каналу, такое что $n\varrho \in \operatorname{int}\Gamma_{\rho_{[1,n]}}$, а κ, ϕ, ϵ – произвольные положительные параметры, такие что $\kappa \geqslant 3$, $\phi < 1$, $\epsilon < 1$, и пусть*

$$\gamma \triangleq 3\left[3\sum_{t=1}^{n}\left(\left[C_{1/2,W_t,n\varrho} + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^{\kappa}\right]^{1/\kappa}. \tag{54}$$

Если M и L – целые числа, такие что $\ln\frac{M}{L} > \widetilde{C}^{\epsilon}_{\phi,W_{[1,n]},n\varrho} + \frac{\gamma}{1-\phi} + \ln\frac{8e^3n^{1,5}}{\epsilon}$, то для любого (M,L)-кода в канале $W_{[1,n]}$, такого что $\bigvee\limits_{m=1}^{M}\rho_{[1,n]}(\Psi(m)) \leqslant n\varrho$, справедливо

$$P_{\mathrm{e}} \geqslant \left(\frac{\epsilon e^{-2\gamma}}{8e^3n^{1,5}}\right)^{1/\phi} e^{-\widetilde{E}^{\epsilon}_{\mathrm{sp}}\left(\ln\frac{M}{L},W_{[1,n]},n\varrho\right)}. \tag{55}$$

Граница (55) не содержит члена, учитывающего скорость в усредненной ЭСУ, но она становится тривиальной, т.е. равной нулю, когда ϕ стремится к нулю. Слегка изменяя рассуждения в доказательстве, можно получить следующую альтернативную границу:

$$P_{\mathrm{e}} \geqslant \frac{\epsilon e^{-2\gamma}}{8n^{1,5}} e^{-\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W_{[1,n]},n\varrho)}, \quad R = \ln\frac{M}{L} - \ln\frac{8e^3n^{1,5}}{\epsilon e^{-2\gamma}}. \tag{56}$$

Границы в лемме 10 и (56) достаточно хороши для доказательства асимптотической ГСУ в теореме 2. Однако для более точного подсчета величину γ можно улучшить. Получающееся выражение, однако, включает в себя дополнительную оптимизацию.

Лемма 11. *Лемма 10 и граница (56) остаются справедливыми, если определение γ, данное в (54), заменить на следующее*:

$$\gamma \triangleq \max_{\substack{\{\varrho_t\}\colon \varrho_t \geqslant 0\ \forall t,\\ \sum\limits_{t=1}^{n}\varrho_t \leqslant n\varrho}} 3\left[3\sum_{t=1}^{n}\left(\left[C^{\lambda}_{1/2,W_t} + \lambda\cdot\varrho_t + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^{\kappa}\right]^{1/\kappa}, \tag{57}$$

где $\lambda = \lambda_{1/2,W_{[1,n]},n\varrho}$. *Если* $W_t = W$ *и* $\rho_t = \rho$ *для всех* $t \in [1,n]$, *то для* γ, *определенного в* (57), *справедливо*

$$\gamma = \max_{\substack{\{\varrho_t\}\colon \varrho_t \geqslant 0\ \forall t,\\ \sum\limits_{t=1}^{n} \varrho_t \leqslant n\varrho}} 3\left[3\sum_{t=1}^{n}\left(\left[C_{1/2,W,\varrho} + \lambda\cdot(\varrho_t - \varrho) + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^{\kappa}\right]^{1/\kappa}. \quad (58)$$

Доказательство леммы 10 и формулы (56). Для каждого $\alpha \in (0,1)$ в силу [26, лемма 29(c)] существует $\lambda_\alpha \in \mathbb{R}^{\ell}_{\geqslant 0}$, такое что

$$C_{\alpha,W_{[1,n]},n\varrho} = C^{\lambda_\alpha}_{\alpha,W_{[1,n]}} + \lambda_\alpha \cdot \varrho n, \quad (59)$$

поскольку $n\varrho \in \operatorname{int}\Gamma_{\rho_{[1,n]}}$. Следовательно, $q_{\alpha,W_{[1,n]},n\varrho} = q^{\lambda_\alpha}_{\alpha,W_{[1,n]}}$ согласно [26, лемма 31]. При этом $q^{\lambda_\alpha}_{\alpha,W_{[1,n]}}$ имеет вид $q^{\lambda_\alpha}_{\alpha,W_{[1,n]}} = \bigotimes\limits_{t=1}^{n} q^{\lambda_\alpha}_{\alpha,W_t}$ согласно [26, лемма 32]. Тогда

$$\|q^{\lambda_\alpha}_{\alpha,W_t} - q^{\lambda_\phi}_{\phi,W_t}\|^2 \overset{(a)}{\leqslant} 4D_{1/2}\big(q^{\lambda_\alpha}_{\alpha,W_t}\,\|\,q^{\lambda_\phi}_{\phi,W_t}\big) \overset{(b)}{\leqslant} 4D_{1/2}\big(q^{\lambda_\alpha}_{\alpha,W_{[1,n]}}\,\|\,q^{\lambda_\phi}_{\phi,W_{[1,n]}}\big) \overset{(c)}{=}$$
$$\overset{(c)}{=} 4D_{1/2}\big(q_{\alpha,W_{[1,n]},n\varrho}\,\|\,q_{\phi,W_{[1,n]},n\varrho}\big) \overset{(d)}{\leqslant} 8\ln\frac{2}{2 - \|q_{\alpha,W_{[1,n]},\varrho} - q_{\phi,W_{[1,n]},\varrho}\|},$$

где (a) справедливо в силу [36, теорема 31], (b) – в силу [36, теоремы 8 и 28], (c) – поскольку $q_{\alpha,W_{[1,n]},n\varrho} = q^{\lambda_\alpha}_{\alpha,W_{[1,n]}}$ $\forall\alpha \in (0,1)$, и (d) – в силу [26, формула (9)]. С другой стороны, августиновский центр $q_{\alpha,W_{[1,n]},n\varrho}$ непрерывен по α на $(0,1)$ в топологии полной вариации на $\mathcal{P}(\mathcal{Y}_1^n)$ согласно [26, леммы 23(d) и 24]. Таким образом, $q^{\lambda_\alpha}_{\alpha,W_t}$ также непрерывна по α в топологии полной вариации на $\mathcal{P}(\mathcal{Y}_t)$. Тогда $q^{\lambda_\cdot}_{\cdot,W_t}$ – вероятность перехода из $((0,1),\mathcal{B}((0,1)))$ в $(\mathcal{Y}_t,\mathcal{Y}_t)$. Определим $q^{\epsilon}_{\alpha,t}$ как маргинальное распределение на $\mathcal{Y}_t$ для вероятностной меры $u_{\alpha,\epsilon} \circledast q^{\lambda_\cdot,W,\varrho}_{\cdot,W_t}$, где $u_{\alpha,\epsilon}$ – равномерное распределение вероятностей на $(\alpha - \epsilon\alpha, \alpha + \epsilon(1-\alpha))$:

$$q^{\epsilon}_{\alpha,t} = \frac{1}{\epsilon}\int\limits_{\alpha-\alpha\epsilon}^{\alpha+(1-\alpha)\epsilon} q^{\lambda_\eta}_{\eta,W_t}\,d\eta, \quad \forall t \in \{1,\dots,n\}.$$

Положим

$$q_{\alpha,t} \triangleq \frac{n}{n+1}q^{\epsilon}_{\alpha,t} + \frac{1}{n+1}q_{1/2,W_t,n\varrho}, \quad \forall t \in \{1,\dots,n\},$$
$$q_\alpha \triangleq \bigotimes_{t=1}^{n} q_{\alpha,t}.$$

Для краткости обозначим вероятностную меру $\bigotimes\limits_{t=1}^{n} W_t(\Psi_t(m))$, порождаемую $\Psi(m)$, через w^m. Имеем

$$D_\alpha(w^m\,\|\,q_\alpha) \overset{(a)}{=} \sum_{t=1}^{n} D_\alpha(W_t(\Psi_t(m))\,\|\,q_{\alpha,t}) \overset{(b)}{\leqslant}$$
$$\overset{(b)}{\leqslant} \sum_{t=1}^{n}\left(\ln\frac{n+1}{n} + D_\alpha\big(W_t(\Psi_t(m))\,\|\,q^{\epsilon}_{\alpha,t}\big)\right) \overset{(c)}{\leqslant} 1 + \sum_{t=1}^{n} D_\alpha\big(W_t(\Psi_t(m))\,\|\,q^{\epsilon}_{\alpha,t}\big) \overset{(d)}{\leqslant}$$

$$\overset{(d)}{\leqslant} 1+\sum_{t=1}^{n}\frac{1}{\epsilon}\int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(W_t(\Psi_t(m))\,\|\,q_{\eta,W_t}^{\lambda_\eta}\big)\,d\eta \overset{(e)}{=}$$

$$\overset{(e)}{=} 1+\frac{1}{\epsilon}\int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(w^m\,\|\,q_{\eta,W_{[1,n]}}^{\lambda_\eta}\big)\,d\eta \overset{(f)}{\leqslant}$$

$$\overset{(f)}{\leqslant} 1+\frac{1}{\epsilon}\int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon}\left(1\vee\frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right) D_\eta\big(w^m\,\|\,q_{\eta,W_{[1,n]}}^{\lambda_\eta}\big)\,d\eta \overset{(g)}{\leqslant}$$

$$\overset{(g)}{\leqslant} 1+\frac{1}{\epsilon}\int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon}\left(1\vee\frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right)\big(C_{\eta,W}^{\lambda_\eta}+\lambda_\eta\cdot\varrho n\big)\,d\eta \overset{(h)}{=} 1+\widetilde{C}_{\alpha,W,\varrho}^{\epsilon}, \tag{60}$$

где (a) выполнено в силу [36, теорема 28], (b) – в силу [26, лемма 1], так как $\frac{n}{n+1}q_{\alpha,t}^{\epsilon}\leqslant q_{\alpha,t}$, (c) – в силу оценки $\ln\tau\leqslant\tau-1$, (d) – в силу неравенства Йенсена и [36, теорема 12], (e) – в силу [36, теорема 28] и [26, лемма 32], (f) – в силу [36, теорема 3 и предложение 2], (g) – в силу [26, теорема 2], так как $\rho_{[1,n]}(\Psi(m))\leqslant n\varrho$, и (h) – в силу (39) и (59). Пусть v_α^m – скошенная вероятностная мера порядка α между w^m и q_α, определенная в (2). Тогда v_α^m – вероятностная мера вида $v_\alpha^m=\bigotimes_{t=1}^{n}[W_t]_\alpha^{q_{\alpha,t}}(\Psi_t(m))$ в силу структуры w^m и q_α как произведений мер. Введем случайные величины

$$\xi_{\alpha,t}^m\triangleq\ln\frac{d[W_t(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}}-\mathbf{E}_{v_\alpha^m}\left[\ln\frac{d[W_t(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}}\right],$$

$$\xi_\alpha^m\triangleq\sum_{t=1}^{n}\xi_{\alpha,t}^m,$$

где $[W_t(\Psi_t(m))]_{\mathrm{ac}}$ – компонента $W_t(\Psi_t(m))$, абсолютно непрерывная по $q_{\alpha,t}$. Тогда для всех $\kappa\in\mathbb{R}_+$ и $\alpha\in(0,1)$

$$\mathbf{E}_{v_\alpha^m}[|\xi_{\alpha,t}^m|^\kappa]^{1/\kappa}\overset{(a)}{\leqslant}3^{1/\kappa}\frac{[(1-\alpha)D_\alpha(W_t(\Psi_t(m))\,\|\,q_{\alpha,t})]\vee\kappa}{\alpha(1-\alpha)}\overset{(b)}{\leqslant}$$

$$\overset{(b)}{\leqslant}3^{1/\kappa}\frac{\big[(1-\alpha)D_\alpha\big(W_t(\Psi_t(m))\,\|\,q_{1/2,W_t,n\varrho}\big)+(1-\alpha)\ln(1+n)\big]\vee\kappa}{\alpha(1-\alpha)}\overset{(c)}{\leqslant}$$

$$\overset{(c)}{\leqslant}3^{1/\kappa}\frac{\big[D_{1/2}\big(W_t(\Psi_t(m))\,\|\,q_{1/2,W_t,n\varrho}\big)+(1-\alpha)\ln(1+n)\big]\vee\kappa}{\alpha(1-\alpha)}\overset{(d)}{\leqslant}$$

$$\overset{(d)}{\leqslant}3^{1/\kappa}\frac{\big[C_{1/2,W_t,n\varrho}+(1-\alpha)\ln(1+n)\big]\vee\kappa}{\alpha(1-\alpha)},$$

где (a) справедливо в силу [25, лемма 17], (b) – в силу [26, лемма 1], так как $\frac{q_{1/2,W_t,n\varrho}}{n+1}\leqslant q_{\alpha,t}$, (c) – в силу [36, теорема 3 и предложение 2], и (d) – в силу [26, теорема 1], так как $\rho_t(\Psi_t(m))\leqslant n\varrho$. Тогда, используя определение γ, данное в (54), получаем

$$\left[\sum_{t=1}^{n}\mathbf{E}_{v_\alpha^m}\big[|\xi_{\alpha,t}^m|^\kappa\big]\right]^{1/\kappa}\leqslant\frac{\gamma}{3\alpha(1-\alpha)},\quad\forall\alpha\in[\phi,1). \tag{61}$$

Оставшаяся часть доказательства аналогична доказательству леммы 9 и границы (45) после формулы (49). Единственные отличия, заслуживающие упоминания, состоят в том, что вместо неравенств (48) и (49) используются (60) и (61), а вместо $\widetilde{q}_t = \dfrac{n}{n+1} q^{\widetilde{\epsilon}}_{\widetilde{\alpha},t} + \dfrac{1}{n+1} q_{1/2,U_{\mathcal{A}}^{(n)}}$ используется $\widetilde{q}_t = \dfrac{n}{n+1} q^{\widetilde{\epsilon}}_{\widetilde{\alpha},t} + \dfrac{1}{n+1} q_{1/2,W_t,\varrho}$. ▲

Доказательство леммы 11 совпадает с доказательством леммы 10 и границы (56) за исключением определения $q_{\alpha,t}$ и оценки для $\mathbf{E}_{v_\alpha^m}\left[|\xi^m_{\alpha,t}|^\kappa\right]^{1/\kappa}$. В частности, мы полагаем $q_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ равной $\dfrac{n}{n+1} q^{\epsilon}_{\alpha,t} + \dfrac{1}{n+1} q^{\lambda_{1/2}}_{1/2,W_t}$ для всех $t \leqslant n$ и оцениваем $\mathbf{E}_{v_\alpha^m}\left[|\xi^m_{\alpha,t}|^\kappa\right]^{1/\kappa}$ следующим образом:

$$
\begin{aligned}
&\mathbf{E}_{v_\alpha^m}\left[|\xi^m_{\alpha,t}|^\kappa\right]^{1/\kappa} \overset{(a)}{\leqslant} 3^{1/\kappa} \frac{\left[(1-\alpha) D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t}\big)\right] \vee \kappa}{\alpha(1-\alpha)} \overset{(b)}{\leqslant} \\
&\overset{(b)}{\leqslant} 3^{1/\kappa} \frac{\left[(1-\alpha) D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q^{\lambda_{1/2}}_{1/2,W_t}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)} \overset{(c)}{\leqslant} \\
&\overset{(c)}{\leqslant} 3^{1/\kappa} \frac{\left[D_{1/2}\big(W_t(\Psi_t(m)) \,\|\, q^{\lambda_{1/2}}_{1/2,W_t}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)} \overset{(d)}{\leqslant} \\
&\overset{(d)}{\leqslant} 3^{1/\kappa} \frac{\left[C^{\lambda_{1/2}}_{1/2,W_t} + \lambda_{1/2} \cdot \rho_t(\Psi_t(m)) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)},
\end{aligned}
$$

где (a) справедливо в силу [25, лемма 17], (b) – в силу [26, лемма 1], так как $\dfrac{q^{\lambda_{1/2}}_{1/2,W_t}}{n+1} \leqslant$ $\leqslant q_{\alpha,t}$, (c) – в силу [36, теорема 3 и предложение 2], а (d) – в силу [26, теорема 2]. Тогда, используя определение γ, данное в (57), получаем

$$
\left[\sum_{t=1}^{n} \mathbf{E}_{v_\alpha^m}\left[|\xi^m_{\alpha,t}|^\kappa\right]\right]^{1/\kappa} \leqslant \frac{\gamma}{3\alpha(1-\alpha)}, \quad \forall \alpha \in [\phi, 1).
$$

Оставшаяся часть доказательства совпадает с доказательством леммы 10 и границы (56).

Если $W_t = W$ и $\rho_t = \rho$ для всех $t \in [1,n]$, то $C_{\alpha,W_{[1,n]},n\varrho} = nC_{\alpha,W,\varrho}$ для всех $\varrho \in \Gamma_\rho$ и $C^{\lambda}_{\alpha,W_{[1,n]}} = nC^{\lambda}_{\alpha,W}$ для всех $\lambda \in \mathbb{R}^\ell_{\geqslant 0}$ согласно [26, леммы 28 и 32]. Тогда (58) следует из равенства $C_{\alpha,W,\varrho} = C^{\lambda_{\alpha,W,\varrho}}_{\alpha,W} + \lambda_{\alpha,W,\varrho} \cdot \varrho$, установленного в [26, лемма 29(c)]. ▲

§ 4. Примеры

Из результатов п. 2.4 и § 3 можно сделать вывод, что ЭСУ управляет экспоненциальной скоростью убывания вероятности ошибки для кодов со списочным декодированием в каналах без памяти при довольно широких предположениях. Однако вычисление самой ЭСУ – это отдельный вопрос, весьма существенный с практической точки зрения. В этом параграфе мы выведем ЭСУ для различных гауссовских и пуассоновских каналов и покажем, что для этих каналов можно получать ее в параметрическом виде, аналогично тому, как сделано в лемме 2 для $E_{\mathrm{sp}}(R,W,p)$. Мы считаем, что эти параметрические выражения проще и интуитивно понятнее, чем обычно используемые равносильные выражения, полученные ранее.

4.1. Гауссовские каналы. Обозначим плотность распределения гауссовской случайной величины с нулевым средним и дисперсией σ^2 через φ_{σ^2}, т.е.

$$
\varphi_{\sigma^2}(z) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}}, \quad \forall z \in \mathbb{R}.
$$

Допуская некоторую вольность в обозначениях, соответствующую вероятностную меру на $\mathcal{B}(\mathbb{R})$ будем также обозначать через φ_{σ^2}.

Пример 1 (скалярный гауссовский канал). Пусть W – скалярный гауссовский канал с дисперсией шума σ^2, и пусть соответствующая функция стоимости ρ квадратична:

$$W(\mathcal{E}\,|\,x) = \int\limits_{\mathcal{E}} \varphi_{\sigma^2}(y-x)\,dy, \quad \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}),$$
$$\rho(x) = x^2, \quad \forall x \in \mathbb{R}.$$

Августиновские пропускная способность и центр с ограничениями по стоимости для этого канала вычислены в [26, пример 4]:

$$C_{\alpha,W,\varrho} = \\ = \begin{cases} \dfrac{\alpha\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \dfrac{1}{\alpha-1}\ln\dfrac{(\theta_{\alpha,\sigma,\varrho})^{\alpha/2}\sigma^{(1-\alpha)}}{\sqrt{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}}, & \alpha \in \mathbb{R}_+ \setminus \{1\}, \\ \dfrac{1}{2}\ln\left(1 + \dfrac{\varrho}{\sigma^2}\right), & \alpha = 1, \end{cases} \tag{62}$$

$$q_{\alpha,W,\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}, \tag{63}$$

$$\theta_{\alpha,\sigma,\varrho} \triangleq \sigma^2 + \frac{\varrho}{2} - \frac{\sigma^2}{2\alpha} + \sqrt{\left(\frac{\varrho}{2} - \frac{\sigma^2}{2\alpha}\right)^2 + \varrho\sigma^2}. \tag{64}$$

Следует отметить, что $C_{\alpha,W,\varrho} = I_\alpha(\varphi_\varrho; W)$ и $q_{\alpha,W,\varrho} = q_{\alpha,\varphi_\varrho}$ для всех положительных порядков α, т.е. гауссовское распределение с нулевым средним и дисперсией ϱ является оптимальным распределением на входе для всех порядков. Таким образом, $E_{\mathrm{sp}}(R, W, \varrho) = E_{\mathrm{sp}}(R, W, \varphi_\varrho)$.

ЭСУ скалярного гауссовского канала можно охарактеризовать с помощью леммы 4. Для этого вначале заметим, что для любого $\theta > 0$ и соответствующей гауссовской вероятностной меры φ_θ скошенный канал $W_\alpha^{\varphi_\theta}$ порядка α, определенный в (3), имеет вид

$$W_\alpha^{\varphi_\theta}(\mathcal{E}\,|\,x) = \int\limits_{\mathcal{E}} \varphi_{\frac{\sigma^2\theta}{\alpha\theta+(1-\alpha)\sigma^2}}\left(y - \frac{\alpha\theta}{\alpha\theta + (1-\alpha)\sigma^2}x\right)dy, \quad \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}). \tag{65}$$

Так как $\theta_{\alpha,\sigma,\varrho}$ является корнем уравнения $\theta^2 - \theta\Big[\varrho + \Big(2 - \frac{1}{\alpha}\Big)\sigma^2\Big] + \Big(1 - \frac{1}{\alpha}\Big)\sigma^4 = 0$ относительно неизвестной θ согласно [26, формулы (132) и (133)], то используя [26, формула (131)], можно непосредственной подстановкой убедиться, что

$$D_1\big(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}},\sigma,\varrho}\,\|\,\varphi_{\theta_{\alpha,\sigma,\varrho}}\,|\,p\big) = \\ = \frac{\alpha^2\theta_{\alpha,\sigma,\varrho}}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}(\mathbf{E}_p[\rho] - \varrho) + \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}, \tag{66}$$

$$D_1\big(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}},\sigma,\varrho}\,\|\,W\,|\,p\big) = \frac{(1-\alpha)^2\sigma^2}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}(\mathbf{E}_p[\rho] - \varrho) + \\ + \frac{(1-\alpha)\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}. \tag{67}$$

Таким образом, для любого $\alpha^* \in (0,1)$ канал $V_p = W_{\alpha^*}^{\varphi_{\theta_{\alpha^*,\sigma,\varrho}},\sigma,\varrho}$ удовлетворяет условиям (21) и (22) леммы 4 для $R = \frac{1}{2}\ln\frac{\alpha^*\theta_{\alpha,\sigma,\varrho} + (1-\alpha^*)\sigma^2}{\sigma^2}$ в силу (62) и ограничения

$\mathbf{E}_p[\rho] \leqslant \varrho$. Кроме того, функция $f(\alpha) \triangleq \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2}{\sigma^2}$ непрерывна и возрастает по α, причем $\lim\limits_{\alpha\downarrow 0} f(\alpha) = 0$ и $f(1) = C_{1,W,\varrho}$. Таким образом, ЭСУ можно записать в следующем параметрическом виде через $\alpha \in [0,1]$ для всех скоростей из $[0, C_{1,W,\varrho}]$:

$$R = \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2}{\sigma^2}, \tag{68}$$

$$E_{\mathrm{sp}}(R,W,\varrho) = \frac{(1-\alpha)\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2)} + \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}. \tag{69}$$

С помощью (66) и (67) можно выразить как скорость, так и ЭСУ через скошенный канал $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$. С другой стороны, $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ является распределением на выходе для распределения на входе φ_ϱ в канале $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$, поскольку $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ является также августиновским средним $q_{\alpha,\varphi_\varrho}$ для распределения на входе φ_ϱ, удовлетворяющего свойству неподвижной точки $\mathrm{T}_{\alpha,\varphi_\varrho}(q_{\alpha,\varphi_\varrho}) = q_{\alpha,\varphi_\varrho}$. Таким образом, используя (66) и (67), можно переписать (68) и (69) следующим образом:

$$R = I_1\big(\varphi_\varrho; W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}\big), \tag{70}$$

$$E_{\mathrm{sp}}(R,W,\varrho) = D_1\big(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}} \,\|\, W \,|\, \varphi_\varrho\big). \tag{71}$$

Чтобы получить выражение для ЭСУ, не зависящее от $\theta_{\alpha,\sigma,\varrho}$ явно, заметим вначале, что из (64) и (68) следует

$$\alpha = \frac{e^{2R}-1}{2}\left(\sqrt{1+\frac{4\sigma^2}{\varrho}\frac{e^{2R}}{e^{2R}-1}}-1\right). \tag{72}$$

С другой стороны, $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ является распределением на выходе в канале $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$ при распределении на входе φ_ϱ. Таким образом, из (65) получаем

$$\frac{\sigma^2\theta_{\alpha,\sigma,\varrho}}{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2} + \left(\frac{\alpha\theta_{\alpha,\sigma,\varrho}}{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2}\right)^2\varrho = \theta_{\alpha,\sigma,\varrho}.$$

Значит,

$$\frac{\sigma^2}{\theta_{\alpha,\sigma,\varrho}} = 1 - \frac{\alpha\varrho}{\alpha\theta_{\alpha,\sigma,\varrho}+(1-\alpha)\sigma^2} = 1 - \frac{\varrho\alpha}{\sigma^2 e^{2R}}, \tag{73}$$

где последнее равенство следует из (68).

Используя в (69) вначале (68) и (73), а затем (72), получаем следующее выражение для ЭСУ:

$$\begin{aligned} E_{\mathrm{sp}}(R,W,\varrho) &= \frac{1}{2}\frac{(1-\alpha)\varrho}{\sigma^2 e^{2R}} + R + \frac{1}{2}\ln\Big(1-\frac{\alpha\varrho}{\sigma^2 e^{2R}}\Big) = \\ &= \frac{\varrho}{4\sigma^2}\left[1+\frac{1}{e^{2R}} - \Big(1-\frac{1}{e^{2R}}\Big)\sqrt{1+\frac{4\sigma^2}{\varrho}\frac{e^{2R}}{e^{2R}-1}}\right] + \\ &+ \frac{1}{2}\ln\left[e^{2R} - \frac{\varrho}{\sigma^2}\left(\frac{e^{2R}-1}{2}\right)\left(\sqrt{1+\frac{4\sigma^2}{\varrho}\frac{e^{2R}}{e^{2R}-1}}-1\right)\right]. \end{aligned} \tag{74}$$

Выражение для ЭСУ, полученное в (74), равносильно формуле (7.4.33) из [8].

Параметрическую характеризацию (68), (69) можно также получить с помощью более прямого подхода, используя дифференцируемость $\theta_{\alpha,\sigma,\varrho}$ и $C_{\alpha,W,\varrho}$ по α. В частности, поскольку $\theta_{\alpha,\sigma,\varrho}$ – корень уравнения

$$\theta^2 - \theta\Big[\varrho + \Big(2 - \frac{1}{\alpha}\Big)\sigma^2\Big] + \Big(1 - \frac{1}{\alpha}\Big)\sigma^4 = 0$$

относительно неизвестной θ, получаем следующее выражение в замкнутом виде для производной августиновской пропускной способности по порядку:

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}C_{\alpha,W,\varrho} = \frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right] + \\
&+ \frac{\theta^2_{\alpha,\sigma,\varrho} - \theta_{\alpha,\sigma,\varrho}\Big[\varrho + \Big(2 - \frac{1}{\alpha}\Big)\sigma^2\Big] + \Big(1 - \frac{1}{\alpha}\Big)\sigma^4}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}\left[\frac{\alpha}{1-\alpha} + \frac{\alpha^2}{\theta_{\alpha,\sigma,\varrho}}\frac{\partial}{\partial\alpha}\theta_{\alpha,\sigma,\varrho}\right] = \\
&= \frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right]. \qquad (75)
\end{aligned}$$

Используя вначале (75), а затем (62), получаем

$$\begin{aligned}
&\frac{d}{d\alpha}\frac{1-\alpha}{\alpha}(C_{\alpha,W,\varrho} - R) = \\
&= \frac{1-\alpha}{\alpha}\left(\frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right]\right) - \\
&- \frac{1}{\alpha^2}(C_{\alpha,W,\varrho} - R) = \frac{1}{\alpha^2}\left(R - \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}\right).
\end{aligned}$$

Теперь проверка знака производной приводит к параметрическому выражению, приведенному в (68) и (69), в силу (62).

Пример 2 (параллельные гауссовские каналы). Пусть $W_{[1,n]}$ – произведение скалярных гауссовских каналов с дисперсией шума σ_i^2 для $i \in \{1,\dots,n\}$, и пусть функция стоимости $\rho_{[1,n]}$ аддитивна и квадратична, т.е.

$$W_{[1,n]}(\mathcal{E} \,|\, x_1^n) = \int\limits_{\mathcal{E}}\left[\prod_{i=1}^n \varphi_{\sigma_i^2}(y_i - x_i)\right] dy_1^n, \quad \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}^n),$$

$$\rho_{[1,n]}(x_1^n) = \sum_{i=1}^n x_i^2, \quad \forall x_1^n \in \mathbb{R}^n.$$

Августиновские пропускная способность и центр канала $W_{[1,n]}$ с ограничениями по стоимости были вычислены в [26, пример 5]:

$$C_{\alpha,W_{[1,n]},\varrho} = \sum_{i=1}^n C_{\alpha,W_i,\varrho_{\alpha,i}}, \qquad (76)$$

$$q_{\alpha,W_{[1,n]},\varrho} = \bigotimes_{i=1}^n \varphi_{\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}}, \qquad (77)$$

$$\varrho_{\alpha,i} = \frac{\big|\alpha - 2\sigma_i^2\lambda_\alpha\big|^+}{2\lambda_\alpha\big(\alpha + 2(\alpha-1)\sigma_i^2\lambda_\alpha\big)}, \qquad (78)$$

где $\theta_{\alpha,\sigma,\varrho}$ определено в (64), а λ_α однозначно определяется[8] условием $\sum\limits_i \varrho_{\alpha,i} = \varrho$. Кроме того, $\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}$ можно выразить через σ_i и λ_α, не используя $\varrho_{\alpha,i}$ явно, следующим образом:

$$\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} = \sigma_i^2 + \left|\frac{1}{2\lambda_\alpha} - \frac{\sigma_i^2}{\alpha}\right|^+. \tag{79}$$

С другой стороны, $\frac{d}{d\varrho_i} C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}} = \lambda_\alpha$ для всех i с положительными $\varrho_{\alpha,i}$, и $\frac{d}{d\varrho_i} C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}} \leqslant \lambda_\alpha$ для всех i. Вычисляя производную сложной функции с учетом (75) и (62), получаем

$$\begin{aligned}
&\frac{d}{d\alpha}\frac{1-\alpha}{\alpha}\big(C_{\alpha,W_{[1,n]},\varrho} - R\big) = \\
&= \frac{1-\alpha}{\alpha}\sum_{i=1}^n \left[\frac{\partial}{\partial\phi} C_{\phi,W_i,\varrho_{\alpha,i}}\Big|_{\phi=\alpha} + \frac{\partial}{\partial\varrho} C_{\alpha,W_i,\varrho}\Big|_{\varrho=\varrho_{\alpha,i}} \frac{d}{d\alpha}\varrho_{\alpha,i}\right] - \frac{1}{\alpha^2}\big(C_{\alpha,W_{[1,n]},\varrho} - R\big) = \\
&= \frac{1}{\alpha^2}\left[R - \frac{1}{2}\sum_{i=1}^n \ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}\right] + \frac{1-\alpha}{\alpha}\lambda_\alpha \sum_{i=1}^n \frac{d}{d\alpha}\varrho_{\alpha,i} = \\
&= \frac{1}{\alpha^2}\left[R - \frac{1}{2}\sum_{i=1}^n \ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}\right].
\end{aligned}$$

Итак, получаем следующее параметрическое выражение для $E_{\mathrm{sp}}(R, W_{[1,n]}, \varrho)$ через α для всех $R \in [0, C_{\alpha,W_{[1,n]},\varrho}]$:

$$R = \frac{1}{2}\sum_{i=1}^n \ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}, \tag{80}$$

$$E_{\mathrm{sp}}(R, W_{[1,n]}, \varrho) = \frac{1}{2}\sum_{i=1}^n \left[\frac{(1-\alpha)\varrho_{\alpha,i}}{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2} + \ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}}\right]. \tag{81}$$

Таким образом, $E_{\mathrm{sp}}(R, W_{[1,n]}, \varrho) = D_1(V_\alpha \,\|\, W_{[1,n]} \,|\, \Phi_\alpha)$ при $R = I_1(\Phi_\alpha; V_\alpha)$, где распределение на входе Φ_α – гауссовское распределение с нулевым средним с диагональной ковариационной матрицей с собственными значениями $\varrho_{\alpha,1}, \ldots, \varrho_{\alpha,n}$, а V_α – скошенный канал порядка α между $W_{[1,n]}$ и q_{α,Φ_α}.

Если $\sigma_i^2 \geqslant \frac{\alpha}{2\lambda_\alpha}$ для некоторого i, то $\varrho_{\alpha,i} = 0$, и соответствующие члены в суммах (80) и (81) нулевые. В [19] с помощью этого наблюдения получено альтернативное параметрическое выражение для ЭСУ. Чтобы вывести его, отметим вначале, что

$$\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2 = \frac{\alpha}{2\lambda_\alpha} \vee \sigma_i^2, \quad \forall i \in \{1, \ldots, n\}$$

согласно (79). Поэтому из (79)–(81) и ограничения $\sum\limits_i \varrho_{\alpha,i} = \varrho$ вытекает следующее параметрическое выражение через $N = \frac{\alpha}{2\lambda_\alpha}$, равносильное выражению из [18] (см.

[8] Ограничение $\sum\limits_i \varrho_{\alpha,i} = \varrho$ задает λ_α однозначно, так как выражение в правой части (78) – невозрастающая функция λ_α.

[8, формулы (7.5.28), (7.5.32), и (7.5.34)], [18, с. 294], [19, формула (20)]):

$$R = \frac{1}{2} \sum_{i:\, \sigma_i^2 \leqslant N} \ln \frac{N}{\sigma_i^2}, \tag{82}$$

$$\varrho = \frac{1}{\alpha} \sum_{i=1}^{n} \frac{\left|N - \sigma_i^2\right|^+}{1 + (\alpha - 1)\dfrac{\sigma_i^2}{N}}, \tag{83}$$

$$E_{\mathrm{sp}}(R, W_{[1,n]}, \varrho) = \frac{(1-\alpha)\varrho}{2N} + \frac{1}{2} \sum_{i:\, \sigma_i^2 \leqslant N} \ln \frac{\alpha N}{N - (1-\alpha)\sigma_i^2}. \tag{84}$$

Заметим, что в этих выражениях не требуется знать значение λ_α или использовать равенство $N = \dfrac{\alpha}{2\lambda_\alpha}$. Как отмечено в [18,19], чтобы вычислить $E_{\mathrm{sp}}(R, W_{[1,n]}, \varrho)$, можно сперва определить N с помощью (82), а затем α с помощью (83).

4.2. Пуассоновские каналы. Пусть $T \in \mathbb{R}_+$ и $a, b \in \mathbb{R}_{\geqslant 0}$, причем $a \leqslant b$. Тогда для пуассоновского канала $\Lambda\colon \mathcal{F} \to \mathcal{P}(\mathcal{Y})$ множество входов $\mathcal{F}$ – это множество всех измеримых функций вида $f\colon (0,T] \to [a,b]$, множество выходов $\mathcal{Y}$ – множество всех неубывающих непрерывных справа функций на $(0,T]$ с целыми значениями, σ-алгебра выходных событий $\mathcal{Y}$ – борелевская σ-алгебра для топологии, порожденной метрикой Скорохода на $\mathcal{Y}$, а $\Lambda(f)$ – пуассоновский точечный процесс с детерминированной функцией интенсивности f для всех $f \in \mathcal{F}$. Допуская некоторую вольность в обозначениях, пуассоновский процесс с постоянной интенсивностью γ будем обозначать через $\Lambda(\gamma)$. Функция стоимости $\rho\colon \mathcal{F} \to \mathbb{R}_{\geqslant 0}$ имеет вид

$$\rho(f) \triangleq \frac{1}{T} \int_0^T f(t)\, dt.$$

В [27, раздел V-C] вычислены пропускные способности и центры Реньи (при отсутствии ограничений) для различных пуассоновских каналов. Эти выражения равны соответствующим августиновским пропускным способностям и центрам, поскольку $C^0_{\alpha,W} = C^{\mathrm{g}0}_{\alpha,W}$ для любого W и $q^0_{\alpha,W} = q^{\mathrm{g}0}_{\alpha,W}$ для любого W с конечной $C^0_{\alpha,W}$ согласно [26, теоремы 2 и 3].

Пример 3 (пуассоновские каналы с заданной средней интенсивностью). В [27, пример 9] рассмотрены каналы $\Lambda^\varrho\colon \mathcal{F}^\varrho \to \mathcal{P}(\mathcal{Y})$, где $\mathcal{F}^\varrho = \{f \in \mathcal{F} : \rho(f) = \varrho\}$ и $\Lambda^\varrho(f) = \Lambda(f)$ для всех $f \in \mathcal{F}^\varrho$. Пропускная способность и центр Реньи канала Λ^ϱ – а следовательно, его августиновские пропускная способность и центр – приведены в [27, формулы (73) и (74)]:

$$C_{\alpha,\Lambda^\varrho} = \begin{cases} \dfrac{\alpha}{\alpha - 1}(\zeta_{\alpha,\varrho} - \varrho)T, & \alpha \neq 1, \\ \left(\dfrac{\varrho - a}{b - a} b \ln \dfrac{b}{\varrho} + \dfrac{b - \varrho}{b - a} a \ln \dfrac{a}{\varrho}\right) T, & \alpha = 1, \end{cases} \tag{85}$$

$$q_{\alpha,\Lambda^\varrho} = \Lambda(\zeta_{\alpha,\varrho}), \tag{86}$$

$$\zeta_{\alpha,\varrho} \triangleq \left(\frac{\varrho - a}{b - a} b^\alpha + \frac{b - \varrho}{b - a} a^\alpha\right)^{1/\alpha}. \tag{87}$$

Так как выражение для $C_{\alpha,\Lambda^\varrho}$ дифференцируемо по α, то вычисляя знак производной, получаем следующее параметрическое выражение для ЭСУ:

$$R = \left(\frac{\varrho - a}{b-a} b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{\zeta_{\alpha,\varrho}} + \frac{b-\varrho}{b-a} a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{\zeta_{\alpha,\varrho}} \right) T, \tag{88}$$

$$E_{\mathrm{sp}}(R, \Lambda^\varrho) =$$

$$= \left(\varrho - \zeta_{\alpha,\varrho} + \frac{\varrho - a}{b-a} b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{b} + \frac{b-\varrho}{b-a} a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{a} \right) T. \tag{89}$$

Имеется альтернативное параметрическое выражение, которое гораздо легче представить, используя скошенные каналы. Чтобы его вывести, заметим вначале, что из [27, формулы (66) и (68)] и определения скошенного канала, данного в (3), следует, что

$$\Lambda_\alpha^{\Lambda(g)}(f) = \Lambda\big(f^\alpha g^{(1-\alpha)}\big). \tag{90}$$

Тогда, используя [27, формула (68)] и (86)–(89), получаем

$$R = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda^\varrho}}(f_{\mathrm{opt}}) \,\|\, q_{\alpha,\Lambda^\varrho}\big), \tag{91}$$

$$E_{\mathrm{sp}}(R, \Lambda^\varrho) = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda^\varrho}}(f_{\mathrm{opt}}) \,\|\, \Lambda(f_{\mathrm{opt}})\big), \tag{92}$$

где f_{opt} – любая функция из $\mathcal{F}^\varrho$ со значениями $\{a, b\}$, т.е. любая функция вида $f_{\mathrm{opt}}\colon (0, T] \to \{a, b\}$, такая что $\rho(f_{\mathrm{opt}}) = \varrho$.

Пример 4 (пуассоновские каналы с ограничениями на среднюю интенсивность). Убедимся вначале, что августиновские пропускная способность $C_{\alpha,\Lambda,\varrho}$ и центр $q_{\alpha,\Lambda,\varrho}$ при наличии ограничений имеют вид[9]

$$C_{\alpha,\Lambda,\varrho} = C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}, \tag{93}$$

$$q_{\alpha,\Lambda,\varrho} = q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}, \tag{94}$$

где $C_{\alpha,\Lambda^\varrho}$ и $q_{\alpha,\Lambda^\varrho}$ определены в (85)–(87), а ϱ_α – убывающая функция порядка α, имеющая вид

$$\varrho_\alpha \triangleq \begin{cases} \alpha^{\frac{\alpha}{1-\alpha}} \Big(\dfrac{b-a}{b^\alpha - a^\alpha} \Big)^{\frac{1}{1-\alpha}} + \dfrac{ab^\alpha - ba^\alpha}{b^\alpha - a^\alpha}, & \alpha \neq 1, \\ e^{-1} b^{\frac{b}{b-a}} a^{-\frac{a}{b-a}}, & \alpha = 1. \end{cases} \tag{95}$$

Чтобы доказать равенства (93) и (94), заметим вначале, что $C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} \leqslant C_{\alpha,\Lambda,\varrho}$, так как $C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} = C_{\alpha,\Lambda,\mathcal{P}(\mathcal{F}^{\varrho\wedge\varrho_\alpha})}$ и $\mathcal{P}(\mathcal{F}^{\varrho\wedge\varrho_\alpha}) \subset \mathcal{A}(\varrho)$, где $\mathcal{A}(\varrho) = \{p \in \mathcal{P}(\mathcal{F}) : \mathbf{E}_p[\rho(f)] \leqslant \varrho\}$. С другой стороны, применяя вначале [27, формула (82)], а затем [27, формула (76)], получаем

$$D_\alpha\big(\Lambda(f) \,\|\, \Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) \leqslant \frac{b-\rho(f)}{b-a} D_\alpha\big(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) +$$

$$+ \frac{\rho(f)-a}{b-a} D_\alpha\big(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) = C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} +$$

$$+ \frac{\varrho\wedge\varrho_\alpha - \rho(f)}{b-a} \left[D_\alpha\big(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) - D_\alpha\big(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) \right], \quad \forall f \in \mathcal{F}.$$

[9] Заметим, что $C_{\alpha,\Lambda,\varrho} = C_{\alpha,\Lambda^{\leqslant\varrho}}$ и $q_{\alpha,\Lambda,\varrho} = q_{\alpha,\Lambda^{\leqslant\varrho}}$ для пуассоновского канала $\Lambda^{\leqslant\varrho}\colon \{f \in \mathcal{F} : \rho(f) \leqslant \varrho\} \to \mathcal{P}(\mathcal{Y})$, рассмотренного в [27, пример 10].

Поскольку $D_\alpha(\Lambda(a)\,\|\,\Lambda(\zeta_{\alpha,\varrho_\alpha})) = D_\alpha(\Lambda(b)\,\|\,\Lambda(\zeta_{\alpha,\varrho_\alpha}))$ в силу [27, формулы (83) и (84)], получаем

$$D_\alpha\big(\Lambda(f)\,\|\,\Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\big) \leqslant C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}} + \\ + \mathbb{1}_{\{\varrho<\varrho_\alpha\}}\frac{\varrho-\rho(f)}{b-a}\left[D_\alpha\big(\Lambda(a)\,\|\,\Lambda(\zeta_{\alpha,\varrho})\big) - D_\alpha\big(\Lambda(b)\,\|\,\Lambda(\zeta_{\alpha,\varrho})\big)\right], \quad \forall f\in\mathcal{F}.$$

С другой стороны, $D_\alpha(\Lambda(a)\,\|\,\Lambda(\zeta_{\alpha,\varrho})) \leqslant D_\alpha(\Lambda(b)\,\|\,\Lambda(\zeta_{\alpha,\varrho}))$ для всех $\varrho\leqslant\varrho_\alpha$ согласно [27, формула (83)], и следовательно,

$$D_\alpha\big(\Lambda(f)\,\|\,\Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha})\,|\,p\big) \leqslant C_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}, \quad \forall p:\ \mathbf{E}_p[\rho]\leqslant\varrho.$$

Таким образом, соотношения (93) и (94) вытекают из [26, лемма 25], поскольку $\Lambda(\zeta_{\alpha,\varrho\wedge\varrho_\alpha}) = q_{\alpha,\Lambda^{\varrho\wedge\varrho_\alpha}}$ согласно (86).

Так как выражение для $C_{\alpha,\Lambda,\varrho}$, данное в (93), дифференцируемо по α, то вычисляя знак производной, можно найти оптимальный порядок для ЭСУ, заданной в (10). В результате получаем следующее параметрическое выражение:

$$R = \left(\frac{\varrho\wedge\varrho_\alpha-a}{b-a}b^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}\ln\frac{b^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}}{\zeta_{\alpha,\varrho\wedge\varrho_\alpha}} + \right. \\ \left. + \frac{b-\varrho\wedge\varrho_\alpha}{b-a}a^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}\ln\frac{a^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}}{\zeta_{\alpha,\varrho\wedge\varrho_\alpha}}\right)T, \tag{96}$$

$$E_{\mathrm{sp}}(R,\Lambda,\varrho) = \left(\varrho\wedge\varrho_\alpha-\zeta_{\alpha,\varrho\wedge\varrho_\alpha} + \frac{\varrho\wedge\varrho_\alpha-a}{b-a}b^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}\ln\frac{b^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}}{b} + \right. \\ \left. + \frac{b-\varrho\wedge\varrho_\alpha}{b-a}a^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}\ln\frac{a^\alpha\zeta_{\alpha,\varrho\wedge\varrho_\alpha}^{1-\alpha}}{a}\right)T. \tag{97}$$

С учетом [27, формула (68)], (86), (90), (94), (96) и (97) получаем следующую параметрическую характеризацию:

$$R = D_1\big(\Lambda_\alpha^{q_\alpha,\Lambda,\varrho}(f_\alpha)\,\|\,q_{\alpha,\Lambda,\varrho}\big), \tag{98}$$

$$E_{\mathrm{sp}}(R,\Lambda,\varrho) = D_1\big(\Lambda_\alpha^{q_\alpha,\Lambda,\varrho}(f_\alpha)\,\|\,\Lambda(f_\alpha)\big), \tag{99}$$

где f_α – любая функция из $\mathcal{F}^{\varrho\wedge\varrho_\alpha}$ со значениями $\{a,b\}$, т.е. любая функция вида $f_\alpha\colon(0,T]\to\{a,b\}$, такая что $\rho(f_\alpha)=\varrho\wedge\varrho_\alpha$.

§ 5. Обсуждение

Мы применили метод Августина для вывода ГСУ для двух семейств каналов без памяти. Для стационарных каналов без памяти с выпуклыми ограничениями по композиции новым в методе Августина является следующее наблюдение:

$$\lim_{\phi\to\alpha}\ \sup_{p\in\mathcal{A}} D_\alpha(W\,\|\,q_{\phi,W,\mathcal{A}}\,|\,p) = C_{\alpha,W,\mathcal{A}}, \quad \forall\alpha\in(0,1). \tag{100}$$

Заметим, что результаты, полученные для выпуклых стационарных каналов без памяти с ограничениями по композиции справедливы также и для стационарных каналов без памяти с ограничениями по стоимости, так как любое ограничение по стоимости для стационарного канала без памяти можно также представить в виде выпуклого ограничения по композиции. Для нестационарных каналов без памяти с ограничениями по стоимости мы применяли соотношение (100) в сочетании с тех-

никой выпуклого сопряжения. Теорема 2 улучшает аналогичный результат Августина, т.е. [24, теорема 36.6], благодаря членам, учитывающим ошибку аппроксимации. Коэффициент при экспоненте в теореме 2 имеет вид $e^{-O(\ln n)}$, а не $e^{-O(\sqrt{n})}$, как в [24, теорема 36.6]. Кроме того, в отличие от [24, теорема 36.6], в теореме 2 функции стоимости не предполагаются ограниченными, благодаря чему эта теорема справедлива и для гауссовских моделей, рассматриваемых в [17–20].

ГСУ для классически-квантовых каналов была установлена в [40]. После этого прорыва интерес к ГСУ для классически-квантовых каналов возобновился [41–44]. Однако метод Августина не применялся ни к каким квантовым теоретико-информационным моделям. Успешное использование метода Августина позволило бы, по меньшей мере, избавиться от предположений стационарности и конечности множества входов.

Августиновский вариант границы Галлагера, обсуждаемый в п. 2.4, известен мало. Нашей основной целью в п. 2.4 было изложить этот подход в простейшем виде. Так, при оценке $\mathbf{E}[P_{\mathrm{e}}]$ мы довольствовались переходом от (29) к (30). При более тщательном анализе, оценивая уклонение случайной величины, равной информации Августина – Лежандра порядка α, т.е. $\ln \dfrac{f_x}{h_x}$ для $q = q_{\alpha,p}$, можно получить более точные оценки, аналогичные оценкам из [45–47]. Граница случайного кодирования для классически-квантовых каналов была получена в [48]. В [42, с. 5606] было высказано предположение, что тот факт, что рассуждения в [48] опираются на коды с независимыми одинаково распределенными символами, может препятствовать модификации доказательства на коды с постоянной композицией, т.е. с ограничениями по композиции. Мы считаем, что августиновский вариант границы Галлагера может оказаться полезным для преодоления этой трудности.

В качестве стороннего замечания укажем, что из результатов п. 2.4 и § 3 следует, что ЭСУ является функцией надежности для определенных каналов с замиранием, т.е. каналов с состояниями, при условии, что разрешено списочное декодирование, даже в нестационарном случае. В частности, как каналы с быстрым замиранием и отсутствием информации о состоянии (т.е. со статистической информацией о состоянии), так и каналы с быстрым замиранием и информацией о состоянии, известной только на приемном конце, являются каналами без памяти с ограничениями по стоимости. Таким образом, для каналов, рассмотренных в [20], а также для каналов, рассмотренных в [49, § 4], функция надежности при списочном декодировании равна ЭСУ. Это также справедливо и для моделей с ограничением по мощности на каждую антенну, рассматриваемых в [50–52], поскольку такие каналы являются каналами без памяти с ограничениями по стоимости, хотя и со многими ограничениями. Однако, как отмечено в § 4, определение ЭСУ для таких каналов является отдельным вопросом.

Оптимальный коэффициент при экспоненте в ГСУ был известен для отдельных каналов с самых истоков теории информации (см., например, [3, 4, 17]). В последние годы интерес к установлению таких точных ГСУ при различных предположениях симметрии возобновился [5, 6, 13, 21, 22]. Параметрические выражения для ЭСУ, данные в (13) и (14), использовались в [14] для вывода подобных границ для кодов с постоянной композицией. Как показано в [28], используя аналогичные параметрические характеризации, улучшенную ГСУ можно получить в каждом случае из рассмотренных в [3–6, 13, 17, 21, 22]. Это является одной из причин, по которой мы представили параметрические выражения, данные в (70), (71), (91), (92), (98) и (99). Оказывается (см. [53]), что с помощью аналогичных параметрических характеризаций при соответствующих предположениях симметрии можно также улучшать сильные обращения теоремы кодирования в смысле коэффициентов при экспоненте.

В работе Блейхута [31] выводилась нижняя граница вероятности ошибки кодов без использования рассуждений с постоянной композицией. Блейхут утверждал, что экспоненциальная скорость убывания этой границы равна ЭСУ – см. [31, теорема 19]. В других публикациях Блейхута также утверждалось равенство вышеуказанных экспонент и ЭСУ, см. [32, лемма 1] и [33, теорема 10.1.4]. Мы покажем, что для для Z-канала экспонента в границе Блейхута равна бесконечности для любых скоростей, меньших пропускной способности канала. Следовательно, и [31, теорема 19], и [32, лемма 1], и [33, теорема 10.1.4] неверны. Что более важно, мы покажем, что даже наилучшая граница, которую можно получить методом Блейхута, строго хуже, чем ГСУ, в смысле экспоненциальной скорости убывания с ростом длины блока.

Для любых $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ и $R \in [0, C_{1,W}]$ положим

$$G(R, W, p, q) \triangleq \inf_{V\colon D_1(V\|q|p) \leqslant R} D_1(V \,\|\, W \,|\, p), \quad \forall p \in \mathcal{P}(\mathcal{X}),\ q \in \mathcal{P}(\mathcal{Y}).$$

Поскольку $D_1(V \,\|\, (pV) \,|\, p) = I_1(p; V)$, где $(pV) = \sum\limits_x p(x)V(x)$, из альтернативного выражения для $E_{\mathrm{sp}}(R, W)$, данного в (20), следует, что

$$E_{\mathrm{sp}}(R, W) = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} G(R, W, p, q), \quad \forall R \in [0, C_{1,W}]. \tag{A.1}$$

Таким образом, из неравенства максимина-минимакса вытекает

$$E_{\mathrm{sp}}(R, W) \leqslant \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q), \quad \forall R \in [0, C_{1,W}].$$

Начальная часть доказательства теоремы 19 в [31] устанавливает следующую границу на вероятность ошибки для кодов в стационарном канале-произведении с каналами-компонентами W, у которых множество входов $\mathcal{X}$ и множество выходов $\mathcal{Y}$ конечны:

$$P_{\mathrm{e}}^{\max, n} \geqslant O(1) e^{-o(n) - n \sup\limits_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q)}, \quad \forall q \in \mathcal{P}(\mathcal{Y}). \tag{A.2}$$

Во второй части доказательства [31, теорема 19] утверждается, что величина $\sup\limits_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R})$ равна $E_{\mathrm{sp}}(R, W)$ для некоторой пары (α_R, p_R), удовлетворяющей следующим равенствам[10]:

$$E_{\mathrm{sp}}(R, W) = E_{\mathrm{sp}}(R, W, p_R) = \frac{1 - \alpha_R}{\alpha_R}(I_{\alpha_R}(p_R; W) - R).$$

С учетом равенства (A.1) во второй части доказательства [31, теорема 19] утверждается, что

$$\sup_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R}) \stackrel{?}{=} \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} G(R, W, p, q). \tag{A.3}$$

Заметим, что с учетом (A.1) из [32, лемма 1] и [33, теорема 10.1.4] также вытекает это равенство. Чтобы опровергнуть равенство (A.3), рассмотрим Z-канал, т.е. дискретный канал с множеством входов $\mathcal{X} = \{1, 2\}$ и множеством выходов $\mathcal{Y} = \{a, b\}$, такой что

$$W = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 - \varepsilon \end{bmatrix}.$$

[10] В явном виде Блейхут упоминает только первое равенство.

Вычислим августиновские пропускную способность и центр порядка α с помощью тождеств $C_{\alpha,W} = D_\alpha(W(1)\,\|\,q_{\alpha,W})$ и $C_{\alpha,W} = D_\alpha(W(2)\,\|\,q_{\alpha,W})$:

$$C_{\alpha,W} = \ln\left(1 + \left(\frac{1-\varepsilon^\alpha}{(1-\varepsilon)^\alpha}\right)^{\frac{1}{1-\alpha}}\right), \tag{A.4}$$

$$q_{\alpha,W}(a) = \frac{(1-\varepsilon)^{\frac{\alpha}{1-\alpha}}}{(1-\varepsilon)^{\frac{\alpha}{1-\alpha}} + (1-\varepsilon^\alpha)^{\frac{1}{1-\alpha}}}. \tag{A.5}$$

Тогда для $q = q_{\alpha,W}$ скошенный канал W_α^q, определенный в (3), имеет вид

$$W_\alpha^{q_{\alpha,W}} = \begin{bmatrix} 1 & 0 \\ \varepsilon^\alpha & 1-\varepsilon^\alpha \end{bmatrix}.$$

Кроме того, непосредственной подстановкой можно убедиться, что августиновский центр $q_{\alpha,W}$ порядка α является неподвижной точкой оператора Августина порядка α, определенного в (4), для априорного распределения $p_{\alpha,W}$, такого что

$$p_{\alpha,W}(1) = \frac{(1-\varepsilon)^{\frac{\alpha}{1-\alpha}} - \varepsilon^\alpha(1-\varepsilon^\alpha)^{\frac{\alpha}{1-\alpha}}}{(1-\varepsilon)^{\frac{\alpha}{1-\alpha}} + (1-\varepsilon^\alpha)^{\frac{1}{1-\alpha}}}.$$

Таким образом, августиновский центр $q_{\alpha,W}$ порядка α равен августиновскому среднему порядка α с априорным распределением $p_{\alpha,W}$, т.е. $q_{\alpha,W} = q_{\alpha,p_{\alpha,W}}$, в силу леммы [26, лемма 13(c),(d)] и равенства $I_\alpha(p_{\alpha,W}; W) = D_\alpha(W\,\|\,q_{\alpha,W}\,|\,p_{\alpha,W})$. Следовательно, также $I_\alpha(p_{\alpha,W}; W) = C_{\alpha,W}$.

Заметим, что функция $C_{\alpha,W}$, данная в (A.4), дифференцируема по α, причем

$$\frac{\partial}{\partial\alpha}C_{\alpha,W} = \frac{q_\alpha(b)}{(1-\alpha)^2}\left(\ln\frac{1-\varepsilon^\alpha}{1-\varepsilon} + \frac{\varepsilon^\alpha}{1-\varepsilon^\alpha}\ln\frac{\varepsilon^\alpha}{\varepsilon}\right) =$$
$$= \frac{1}{(1-\alpha)^2}D_1\big(W_\alpha^{q_{\alpha,W}}\,\|\,W\,|\,p_{\alpha,W}\big).$$

Тогда, применяя вначале тождество $I_\alpha(p_{\alpha,W}; W) = C_{\alpha,W}$, а затем [26, формула (35)], получаем

$$\frac{\partial}{\partial\alpha}\frac{1-\alpha}{\alpha}(C_{\alpha,W} - R) = \frac{1}{\alpha^2}\left(R - C_{\alpha,W} + (1-\alpha)\alpha\frac{\partial}{\partial\alpha}C_{\alpha,W}\right) =$$
$$= \frac{1}{\alpha^2}\left(R - I_\alpha(p_{\alpha,W}; W) + \frac{\alpha}{1-\alpha}D_1\big(W_\alpha^{q_{\alpha,W}}\,\|\,W\,|\,p_{\alpha,W}\big)\right) =$$
$$= \frac{1}{\alpha^2}\left(R - I_1\big(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}}\big)\right).$$

Можно проверить численно, что $I_1(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}})$ – возрастающая по α функция для любого $\varepsilon \in (0,1)$. Таким образом, можно выразить скорость и соответствующую ЭСУ в следующем параметрическом виде:

$$R(\alpha) = I_1\big(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}}\big), \quad E_{\mathrm{sp}}(R(\alpha), W) = D_1\big(W_\alpha^{q_{\alpha,W}}\,\|\,W\,|\,p_{\alpha,W}\big).$$

Таким образом, для $R = R(\phi)$ получаем $(\alpha_R, p_R) = (\phi, p_{\phi,W})$ и $q_{\alpha_R,p_R} = q_{\phi,W}$. Тогда

$$\sup_{p\in\mathcal{P}(\mathcal{X})} G(R,W,p,q_{\alpha_R,p_R}) \geqslant G(R,W,p,q_{\alpha_R,W})\big|_{p:\,p(1)=1} =$$
$$= \inf_{v:\,D_1(v\|q_{\alpha_R,W})\leqslant R} D_1(v\,\|\,W(1)).$$

С другой стороны, $v(b) > 0$ для всех v, таких что $D_1(v \,\|\, q_{\alpha_R,W}) \leqslant R$, поскольку если $v(b) = 0$ и $C_{\alpha_R,W} > R$, то $D_1(v \,\|\, q_{\alpha_R,W}) = \ln \dfrac{1}{q_{\alpha_R,W}(a)} = C_{\alpha_R,W}$. Кроме того, $D_1(v \,\|\, W(1)) = \infty$, если $v(b) > 0$. Итак,

$$\inf_{v\colon D_1(v\|q_{\alpha_R,W})\leqslant R} D_1(v \,\|\, W(1)) = \infty.$$

Таким образом, граница, установленная в первой части доказательства [31, теорема 19], т.е. (A.2), тривиальна для $q = q_{\alpha_R,p_R}$. Кроме того, и [31, теорема 19], и [32, лемма 1], и [33, теорема 10.1.4] неверны, поскольку $E_{\mathrm{sp}}(R,W) < \infty$ для всех $R \in [0, C_{1,W}]$ в случае Z-каналов[11].

Выбирая другие q, из (A.2) можно получать нетривиальные границы, однако из них также нельзя вывести ГСУ. Если $q(a) < e^{-R}$, то $\sup\limits_{p\in\mathcal{P}(\mathcal{X})} G(R,W,p,q_{\alpha_R,p_R}) = \infty$ в силу приведенных выше аргументов. В случае $q(a) \geqslant e^{-R}$ получаем следующее:

$$\begin{aligned}
&\sup_{p\in\mathcal{P}(\mathcal{X})} G(R,W,p,q) \geqslant \inf_{V\colon D_1(V\|q|p_R)\leqslant R} D_1(V \,\|\, W \,|\, p_R) \geqslant \\
&\geqslant \inf_V D_1(V \,\|\, W \,|\, p_R) + \frac{1-\alpha_R}{\alpha_R}(D_1(V \,\|\, q \,|\, p_R) - R) \overset{(a)}{=} \\
&\overset{(a)}{=} \frac{1-\alpha_R}{\alpha_R}(D_{\alpha_R}(W \,\|\, q \,|\, p_R) - R) \overset{(b)}{\geqslant} \frac{1-\alpha_R}{\alpha_R}(I_{\alpha_R}(p_R;W) + D_{\alpha_R}(q_{\alpha_R,p_R} \,\|\, q) - R) = \\
&= E_{\mathrm{sp}}(R,W) + \frac{1-\alpha_R}{\alpha_R} D_{\alpha_R}(q_{\alpha_R,p_R} \,\|\, q) \overset{(c)}{\geqslant} E_{\mathrm{sp}}(R,W) + \frac{1-\alpha_R}{2}\|q - q_{\alpha_R,p_R}\|^2,
\end{aligned}$$

где (a) справедливо в силу [36, теорема 30], (b) – в силу [26, лемма 13(c)], а (c) – в силу [36, теорема 31]. С другой стороны, $q_{\alpha_R,p_R}(a) < e^{-R}$, поскольку $q_{\alpha_R,p_R} = q_{\alpha_R,W}$ и $C_{\alpha_R,W} > R$. Итак,

$$\begin{aligned}
&\inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{P}(\mathcal{X})} G(R,W,p,q) \geqslant E_{\mathrm{sp}}(R,W) + 2(1-\alpha_R)(e^{-R} - q_{\alpha_R,W}(a))^2 = \\
&= \sup_{p\in\mathcal{P}(\mathcal{X})} \inf_{q\in\mathcal{P}(\mathcal{Y})} G(R,W,p,q) + 2(1-\alpha_R)(e^{-R} - q_{\alpha_R,W}(a))^2.
\end{aligned}$$

Следовательно, методом Блейхута в том виде, в каком как он приведен в [31], вывести ГСУ невозможно. Когда множество входов конечно, эту трудность можно преодолеть, используя технику выбрасывания с учетом композиции. Но такой подход применялся Арутюняном в [10] еще до работы [31].

ПРИЛОЖЕНИЕ B: ИНФОРМАЦИОННЫЕ ВЕЛИЧИНЫ РЕНЬИ – ГАЛЛАГЕРА

Информационные величины Реньи – Галлагера порядка 1 равны соответствующим информационным величинам Августина – Лежандра порядка 1 по определению. Поэтому в дальнейшем обсуждении мы ограничиваемся порядками, отличными от единицы.

Определение 12. Для любых $\alpha \in \mathbb{R}_+ \setminus \{1\}$, канала $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$, $p \in \mathcal{P}(\mathcal{X})$ и $\lambda \in \mathbb{R}^{\ell}_{\geqslant 0}$ *информацией Реньи – Галлагера поряд-*

[11] В недавней работе [54] отмечалось, что метод Блейхута можно использовать для вывода ГСУ, если выполнено минимаксное равенство [54, формула (3.63)]. Таким образом, из нашего анализа вытекает, что равенство [54, формула (3.63)] в общем случае выполнено не всегда. Этот факт можно также вывести из отсутствия минимаксного равенства для $G(R,W,p,q)$, не опираясь на результаты работы [54].

ка α для распределения на входе p и множителя Лагранжа λ называется величина

$$I_\alpha^{\mathrm{g}\lambda}(p;W) \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha\bigl(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q\bigr).$$

Если λ – нулевой вектор, то информация Реньи – Галлагера равна информации Реньи. Как и для информации Реньи, для информации Реньи – Галлагера имеется выражение в замкнутом виде через среднюю меру, достигающую инфимума в этом определении.

О п р е д е л е н и е 13. Для любых $\alpha\in\mathbb{R}_+$, канала $W\colon\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon\mathcal{X}\to\mathbb{R}^\ell_{\geqslant 0}$, $p\in\mathcal{P}(\mathcal{X})$ и $\lambda\in\mathbb{R}^\ell_{\geqslant 0}$ *средней мерой порядка α для распределения на входе p и множителя Лагранжа λ* называется

$$\frac{d\mu_{\alpha,p}^\lambda}{d\nu} \triangleq \left[\sum_x p(x)e^{(1-\alpha)\lambda\cdot\rho(x)}\left(\frac{dW(x)}{d\nu}\right)^\alpha\right]^{\frac{1}{\alpha}}. \tag{B.1}$$

Средним Реньи – Галлагера порядка α для распределения на входе p и множителя Лагранжа λ называется

$$q_{\alpha,p}^{\mathrm{g}\lambda} \triangleq \frac{\mu_{\alpha,p}^\lambda}{\|\mu_{\alpha,p}^\lambda\|}.$$

Как $\mu_{\alpha,p}^\lambda$, так и $q_{\alpha,p}^{\mathrm{g}\lambda}$ зависят от множителя Лагранжа λ при $\alpha\in\mathbb{R}_+\setminus\{1\}$. Кроме того, непосредственной подстановкой можно убедиться, что

$$D_\alpha\bigl(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q\bigr) = I_\alpha^{\mathrm{g}\lambda}(p;W) + D_\alpha\bigl(q_{\alpha,p}^{\mathrm{g}\lambda}\,\|\,q\bigr),\quad \alpha\in\mathbb{R}_+\setminus\{1\}.$$

Тогда в силу [26, лемма 2] имеем

$$I_\alpha^{\mathrm{g}\lambda}(p;W) = D_\alpha\bigl(p\circledast We^{\frac{1-\alpha}{\alpha}\lambda\cdot\rho}\,\|\,p\otimes q_{\alpha,W}^{\mathrm{g}\lambda}\bigr) = \frac{\alpha}{\alpha-1}\ln\|\mu_{\alpha,p}^\lambda\|,\quad \alpha\in\mathbb{R}_+\setminus\{1\}. \tag{B.2}$$

Используя определения информации Августина – Лежандра и информации Реньи – Галлагера, с учетом неравенства Йенсена и вогнутости функции натурального логарифма получаем

$$I_\alpha^\lambda(p;W)\geqslant I_\alpha^{\mathrm{g}\lambda}(p;W),\quad \alpha\in(0,1], \tag{B.3}$$

$$I_\alpha^\lambda(p;W)\leqslant I_\alpha^{\mathrm{g}\lambda}(p;W),\quad \alpha\in[1,\infty). \tag{B.4}$$

О п р е д е л е н и е 14. Для любых $\alpha\in\mathbb{R}_+$, канала $W\colon\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ с функцией стоимости $\rho\colon\mathcal{X}\to\mathbb{R}^\ell_{\geqslant 0}$ и $\lambda\in\mathbb{R}^\ell_{\geqslant 0}$ *пропускной способностью Реньи – Галлагера порядка α для множителя Лагранжа λ* называется величина

$$C_{\alpha,W}^{\mathrm{g}\lambda} \triangleq \sup_{p\in\mathcal{P}(\mathcal{X})} I_\alpha^{\mathrm{g}\lambda}(p;W). \tag{B.5}$$

Несмотря на то, что неравенства (B.3) и (B.4) являются строгими для большинства распределений, согласно [26, теорема 3] справедливо равенство

$$C_{\alpha,W}^{\mathrm{g}\lambda} = S_{\alpha,W}^\lambda. \tag{B.6}$$

Таким образом, $C_{\alpha,W}^{\mathrm{g}\lambda} = C_{\alpha,W}^\lambda$ в силу (9). По этой причине при определении оптимальной производительности можно использовать любое из этих семейств. Следующая лемма по существу является переформулировкой теоремы 8 из [29], благодаря

которой стало популярно использование информационных величин Реньи – Галлагера в задачах с ограничениями по стоимости (см. [18–20]).

Лемма 12. *Для любых* $\ell, M, L \in \mathbb{Z}_+$, *таких что* $L < M$, $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\rho\colon \mathcal{X} \to \mathbb{R}^{\ell}_{\geqslant 0}$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$ *и* $\alpha \in \left[\frac{1}{1+L}, 1\right)$ *существует* (M, L)*-код с функцией кодирования вида* $\Psi\colon \mathcal{M} \to \mathcal{B}$, *для которого*

$$\ln P_{\mathrm{e}} \leqslant \frac{\alpha-1}{\alpha}\left[I^{\mathrm{g}\lambda}_{\alpha}(p;W) + \left(\inf_{x\in\mathcal{B}} \lambda\cdot\rho(x)\right) - \ln\frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha}. \tag{B.7}$$

Доказательство леммы 12. Следуем доказательству леммы 6 вплоть до соотношения (30). В силу (26) и (30) имеем

$$\begin{aligned}
\ln \mathbf{E}[P_{\mathrm{e}}] &\leqslant \ln \mathbf{E}_q\left[\left(\sum_{x\in\mathcal{B}} \frac{p(x)}{p(\mathcal{B})} e^{(1-\alpha)\lambda\cdot\rho(x)}\left(\frac{dW(x)}{dq}\right)^{\alpha}\right)^{\frac{1}{\alpha}}\right] + \\
&+ \frac{\alpha-1}{\alpha}\left[\left(\inf_{x\in\mathcal{B}} \lambda\cdot\rho(x)\right) - \ln\frac{(M-1)e}{L}\right] \leqslant \\
&\leqslant \ln \mathbf{E}_q\left[\left(\sum_{x\in\mathcal{X}} p(x) e^{(1-\alpha)\lambda\cdot\rho(x)}\left(\frac{dW(x)}{dq}\right)^{\alpha}\right)^{\frac{1}{\alpha}}\right] + \\
&+ \frac{\alpha-1}{\alpha}\left[\left(\inf_{x\in\mathcal{B}} \lambda\cdot\rho(x)\right) - \ln\frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha}.
\end{aligned}$$

Поскольку существует код с P_{e}, меньшей или равной $\mathbf{E}[P_{\mathrm{e}}]$, существование кода с функцией кодирования вида $\Psi\colon \mathcal{M} \to \mathcal{B}$, удовлетворяющего условию (B.7), вытекает из (B.1) и (B.2). ▲

ПРИЛОЖЕНИЕ C: ОПУЩЕННЫЕ ДОКАЗАТЕЛЬСТВА

Доказательство леммы 1. Функция $E_{\mathrm{sp}}(R, W, \mathcal{A})$ выпукла по R, поскольку поточечный супремум семейства выпуклых функций является выпуклым, а функция $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ выпукла по R для любого $\alpha \in (0,1)$. Аналогично доказывается, что $E_{\mathrm{sp}}(R, W, \mathcal{A})$ не возрастает по R. Непрерывность и конечность устанавливаются далее при доказательстве соотношения (11).

Напомним, что $C_{\alpha,W,\mathcal{A}}$ – неубывающая функция порядка α согласно [26, лемма 23(a)].

- Если $C_{0^+,W,\mathcal{A}} = \infty$, то $C_{1/2,W} = \infty$, и $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \infty$ для всех $R \in \mathbb{R}_{\geqslant 0}$. С другой стороны, $R < C_{0^+,W,\mathcal{A}}$ для всех $R \in \mathbb{R}_{\geqslant 0}$. Следовательно, (11) выполнено, а утверждения о непрерывности и конечности $E_{\mathrm{sp}}(R, W, \mathcal{A})$ бессодержательны.
- Если $C_{0^+,W,\mathcal{A}} < \infty$ и $C_{0^+,W,\mathcal{A}} = C_{1,W,\mathcal{A}}$, то $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \infty$ для всех $R \in \in [0, C_{1,W,\mathcal{A}})$, и $E_{\mathrm{sp}}(R, W, \mathcal{A}) = 0$ для всех $R \geqslant C_{1,W,\mathcal{A}}$. Следовательно, (11) выполнено, и утверждения о непрерывности и конечности $E_{\mathrm{sp}}(R, W, \mathcal{A})$ справедливы.
- Если $C_{0^+,W,\mathcal{A}} < \infty$ и $C_{0^+,W,\mathcal{A}} \neq C_{1,W,\mathcal{A}}$, то $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \infty$ для всех $R \in \in [0, C_{0^+,W,\mathcal{A}})$. Для $R \geqslant C_{0^+,W,\mathcal{A}}$ из неотрицательности $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ следуют ограничения (11).

  В силу (11) величина $E_{\mathrm{sp}}(R, W, \mathcal{A})$ конечна для всех $R > \lim\limits_{\phi\downarrow 0} C_{\phi,W,\mathcal{A}}$. Таким образом, $E_{\mathrm{sp}}(R, W, \mathcal{A})$ непрерывна на $(C_{0^+,W,\mathcal{A}}, \infty)$ в силу [35, теорема 6.3.3]. Чтобы продолжить непрерывность на $[C_{0^+,W,\mathcal{A}}, \infty)$, заметим, что $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ не

убывает и непрерывна по R для каждого $\alpha \in (0,1)$. Таким образом, $E_{\mathrm{sp}}(R, W, \mathcal{A})$ не возрастает и полунепрерывна снизу по R. Следовательно, $E_{\mathrm{sp}}(R, W, \mathcal{A})$ непрерывна справа, а значит, и в точке $R = C_{0^+,W,\mathcal{A}}$. ▲

Доказательство следствия 2. По лемме 1 для некоторого $\alpha \in (\eta, 1)$ справедливо неравенство $\frac{1-\alpha}{\alpha}\Big(C_{\alpha,W_{[1,n]},n\varrho} - \ln\frac{M}{L}\Big) \geqslant E_{\mathrm{sp}}\Big(\ln\frac{M}{L}, W_{[1,n]}, n\varrho\Big) - \frac{1}{n}$. Существует $p \in \mathcal{P}(\widetilde{\mathcal{X}}_1^n)$ вида $p = \bigotimes_{t=1}^{n} p_t$, такая что $\mathbf{E}_p[\rho_{[1,n]}] \leqslant n(\widetilde{\varrho} - \delta)$. Положим $\widehat{\varrho} \triangleq \varrho - \frac{3e\varsigma}{n}$. В силу [26, леммы 14 и 28] существует $\widetilde{p} \in \mathcal{P}(\widetilde{\mathcal{X}}_1^n)$ вида $\widetilde{p} = \bigotimes_{t=1}^{n} \widetilde{p}_t$, такая что $I_\alpha\big(\widetilde{p}; \widetilde{W}_{[1,n]}\big) \geqslant C_{\alpha,\widetilde{W}_{[1,n]},n\widehat{\varrho}} - \epsilon$ и $\mathbf{E}_{\widetilde{p}}[\rho_{[1,n]}] \leqslant n\widehat{\varrho}$. Пусть $\widehat{W}_{[1,n]} \colon \widehat{\mathcal{X}}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ – канал-произведение, такой что $\widehat{W}_{[1,n]}(x_1^n) = W_{[1,n]}(x_1^n)$ для всех $x_1^n \in \widehat{\mathcal{X}}_1^n$ и $\widehat{\mathcal{X}}_t = \operatorname{supp}\widetilde{p}_t \cup \operatorname{supp} p_t$ для всех $t \in \{1, \ldots, n\}$. Тогда $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} \geqslant C_{\alpha,\widetilde{W}_{[1,n]},n\widehat{\varrho}} - \epsilon$ и $\widetilde{\varrho} \leqslant \widehat{\varrho}$ по построению. Так как $\widehat{\mathcal{X}}_1^n$ конечно, то в силу [26, лемма 19] существует $\widehat{p} \in \mathcal{P}(\widehat{\mathcal{X}}_1^n)$, такая что $I_\alpha\big(\widehat{p}; \widehat{W}_{[1,n]}\big) = C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$ и $\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}] \leqslant n\widehat{\varrho}$. Кроме того, так как функция стоимости $\rho_{[1,n]}$ аддитивна, то в силу [26, лемма 14] без ограничения общности можно считать, что $\widehat{p}$ имеет вид $\widehat{p} = \bigotimes_{t=1}^{n} \widehat{p}_t$.

Заметим, что $\mathbf{E}_{\widehat{p}_t}\big[\big|\rho_t - \mathbf{E}_{\widehat{p}_t}[\rho_t]\big|^\kappa\big]^{1/\kappa} \leqslant \varsigma$ для всех $\kappa \in \mathbb{R}_+$ и $t \in \{1, \ldots, n\}$, поскольку ρ_t положительно и меньше ς с вероятностью единица относительно $\widehat{p}_t$. Тогда, применяя [25, лемма 19] при $\kappa = \ln n$, получаем

$$\widehat{p}\big(\big|\rho_{[1,n]}(x) - \mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]\big| < 3\varsigma e\big) \geqslant \frac{1}{2\sqrt{n}}.$$

С другой стороны, в силу [26, лемма 29(c)] существует $\widehat{\lambda} = \lambda_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$, для которого $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} = C^{\widehat{\lambda}}_{\alpha,\widehat{W}_{[1,n]}} + \widehat{\lambda}n\widehat{\varrho}$, поскольку по построению $n\widehat{\varrho}$ принадлежит внутренней области допустимых ограничений по стоимости $\widehat{W}_{[1,n]}$. При этом $I^{\widehat{\lambda}}_\alpha\big(\widehat{p}; \widehat{W}_{[1,n]}\big) = C^{\widehat{\lambda}}_{\alpha,\widehat{W}_{[1,n]}}$ в силу [26, лемма 29(d)], так как $I_\alpha\big(\widehat{p}; \widehat{W}_{[1,n]}\big) = C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$ и $\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}] \leqslant n\widehat{\varrho}$. В результате получаем, что $D_\alpha\big(W_{[1,n]}(x_1^n) \,\|\, q_{\alpha,\widehat{p}}\big) - \widehat{\lambda}\rho_{[1,n]}(x_1^n) = C^{\widehat{\lambda}}_{\alpha,\widehat{W}_{[1,n]}}$ для всех $x_1^n \in \operatorname{supp}\widehat{p}$ и $\widehat{\lambda}n\widehat{\varrho} = \widehat{\lambda}\,\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]$. Применяя лемму 6 при $\mathcal{B} = \big\{x_1^n \in \operatorname{supp}\widehat{p} : \big|\rho_{[1,n]}(x_1^n) - \mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]\big| < 3\varsigma e\big\}$, заключаем, что существует (M, L)-код, для которого

$$\ln P_{\mathrm{e}} \leqslant \frac{\alpha-1}{\alpha}\left[C^{\widehat{\lambda}}_{\alpha,\widehat{W}_{[1,n]}} + \inf_{x_1^n \in \mathcal{B}} \widehat{\lambda}\rho_{[1,n]}(x_1^n) - \ln\frac{(M-1)e}{L}\right] + \frac{\ln 4n}{2\alpha} \leqslant$$
$$\leqslant \frac{\alpha-1}{\alpha}\left[C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} - 3\varsigma e\widehat{\lambda} - \ln\frac{(M-1)e}{L}\right] + \frac{\ln 4n}{2\alpha}.$$

Так как $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} + \widehat{\lambda}n(\varrho - \widehat{\varrho}) \geqslant C_{\alpha,\widehat{W}_{[1,n]},n\varrho}$ согласно [26, лемма 27(b)], получаем

$$\ln P_{\mathrm{e}} \leqslant -E_{\mathrm{sp}}\Big(\ln\frac{M}{L}, W_{[1,n]}, \varrho\Big) + \frac{1}{n} + \frac{1-\alpha}{\alpha}\big(6\varsigma e\widehat{\lambda} + 2\epsilon + 1\big) + \frac{\ln 4n}{2\alpha}.$$

Теперь (25) выполнено, поскольку $\frac{1-\alpha}{\alpha}C_{\alpha,W_{[1,n]},n\varrho}$ не возрастает по α согласно [26, лемма 23(b)] при условии, что $\widehat{\lambda} \leqslant \frac{C_{\alpha,W_{[1,n]},n\varrho}}{n\delta}$. Чтобы увидеть, почему это условие выполнено, заметим вначале, что $\widehat{\varrho} \leqslant \varrho$ по определению, и поэтому $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} \leqslant C_{\alpha,\widehat{W}_{[1,n]},n\varrho}$. При этом

$$C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} = \inf_{\lambda\geqslant\widehat{\lambda}} C^{\lambda}_{\alpha,\widehat{W}_{[1,n]}} + \lambda n\widehat{\varrho} > n\delta\widehat{\lambda} + \inf_{\lambda\geqslant\widehat{\lambda}} C^{\lambda}_{\alpha,\widehat{W}_{[1,n]}} + n\lambda(\widehat{\varrho}-\delta) \geqslant$$
$$\geqslant n\delta\widehat{\lambda} + C_{\alpha,\widehat{W}_{[1,n]},n(\widehat{\varrho}-\delta)} \geqslant n\delta\widehat{\lambda}. \quad \blacktriangle$$

Доказательство леммы 8. Согласно (39) и [26, лемма 23(a),(b)] имеем $C_{\alpha,W,\mathcal{A}} \leqslant \widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$. Тогда в силу выражения для $E_{\mathrm{sp}}(R,W,\mathcal{A})$, данного в (11), и определения $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A})$, данного в (40), имеем

$$E_{\mathrm{sp}}(R,W,\mathcal{A}) \leqslant \widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}), \quad \forall R\in\mathbb{R}_{\geqslant 0}. \tag{C.1}$$

Будем оценивать $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}) - E_{\mathrm{sp}}(R,W,\mathcal{A})$ сверху при $R\in[C_{\phi,W,\mathcal{A}},\infty)$. Во-первых,

$$\frac{1-\alpha}{\alpha}\big(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R\big) = \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left(\frac{1-\alpha}{\alpha}\vee\frac{1-\eta}{\eta}\right) C_{\eta,W,\mathcal{A}}\,d\eta - \frac{1-\alpha}{\alpha}R =$$
$$= \frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)\,d\eta +$$
$$+ \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)\,d\eta + \frac{R}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha}\frac{\alpha-\eta}{\eta\alpha}\,d\eta \leqslant$$
$$\leqslant \frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)\,d\eta + \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)\,d\eta + \frac{\epsilon}{1-\epsilon}R. \tag{C.2}$$

Для оценки $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A})$ оценим выражение в (C.2) на двух интервалах значений α по отдельности.

Чтобы оценить (C.2) для $\alpha\in[\phi,1)$, используем тот факт, что $\sup\limits_{\eta\in(0,1)}\dfrac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R) = E_{\mathrm{sp}}(R,W,\mathcal{A})$. Имеем

$$\frac{1-\alpha}{\alpha}\big(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R\big) \leqslant$$
$$\leqslant \frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)d\eta + \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)d\eta + \frac{\epsilon}{1-\epsilon}R \leqslant$$
$$\leqslant \frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} \frac{\eta}{1-\eta}E_{\mathrm{sp}}(R,W,\mathcal{A})d\eta + \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha} E_{\mathrm{sp}}(R,W,\mathcal{A})d\eta + \frac{\epsilon}{1-\epsilon}R \leqslant$$
$$\leqslant E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}\frac{1-\alpha}{\alpha}E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}R.$$

Так как $(1-\alpha)E_{\mathrm{sp}}(R,W,\mathcal{A}) + \alpha R \leqslant (R\vee E_{\mathrm{sp}}(R,W,\mathcal{A}))$ для всех $\alpha\in(0,1)$, то

$$\frac{1-\alpha}{\alpha}\big(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R\big) \leqslant E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}\frac{R\vee E_{\mathrm{sp}}(R,W,\mathcal{A})}{\phi}, \quad \alpha\in[\phi,1). \tag{C.3}$$

Чтобы оценить выражение в (C.2) для $\alpha\in(0,\phi]$, напомним, что $C_{\alpha,W,\mathcal{A}}$ не убывает по α согласно [26, лемма 23(a)]. Таким образом, для любой $R\geqslant C_{\phi,W,\mathcal{A}}$ имеем

$$\frac{1-\alpha}{\alpha}\big(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R\big) \leqslant$$
$$\leqslant \frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int\limits_{\phi}^{\alpha+\epsilon(1-\alpha)} \frac{\eta}{1-\eta} E_{\mathrm{sp}}(R,W,\mathcal{A})\,d\eta\,\mathbb{1}_{\left\{\alpha\in\left[\frac{\phi-\epsilon}{1-\epsilon},\phi\right]\right\}} + \frac{\epsilon}{1-\epsilon}R \leqslant$$
$$\leqslant \frac{1}{\epsilon}\frac{\alpha+\epsilon(1-\alpha)}{\alpha(1-\epsilon)}\int\limits_{\phi}^{\alpha+\epsilon(1-\alpha)} E_{\mathrm{sp}}(R,W,\mathcal{A})\,d\eta\,\mathbb{1}_{\left\{\alpha\in\left[\frac{\phi-\epsilon}{1-\epsilon},\phi\right]\right\}} + \frac{\epsilon}{1-\epsilon}R =$$
$$= \left[\frac{\alpha(1-\epsilon)+2\epsilon-\phi}{\epsilon} - \frac{\phi-\epsilon}{\alpha(1-\epsilon)}\right] E_{\mathrm{sp}}(R,W,\mathcal{A})\mathbb{1}_{\left\{\alpha\in\left[\frac{\phi-\epsilon}{1-\epsilon},\phi\right]\right\}} + \frac{\epsilon}{1-\epsilon}R \leqslant$$
$$\leqslant (1-\phi)E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}\frac{(1-\phi)E_{\mathrm{sp}}(R,W,\mathcal{A}) + \phi R}{\phi}, \quad \alpha\in(0,\phi]. \tag{C.4}$$

Теперь (41) следует из (C.1), (C.3) и (C.4).

С другой стороны, $C_{\alpha,W,\mathcal{A}}$ не убывает, а $\frac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ не возрастает по α согласно [26, лемма 23(a),(b)]. Тогда в силу выражения для $E_{\mathrm{sp}}(R,W,\mathcal{A})$, данного в (11), имеем $E_{\mathrm{sp}}(R,W,\mathcal{A}) \leqslant \frac{1-\phi}{\phi}R$ для всех $R\in[C_{\phi,W,\mathcal{A}},\infty)$. Следовательно,

$$R \vee E_{\mathrm{sp}}(R,W,\mathcal{A}) \leqslant R/\phi, \quad \forall R\in[C_{\phi,W,\mathcal{A}},\infty). \tag{C.5}$$

Теперь (42) следует из (41) и (C.5). ▲

Доказательство теоремы 1. Будем доказывать теорему 1, используя леммы 8 и 9. Мы можем выбирать различные значения κ и ϵ для различных значений n при условии, что выполняются условия лемм 8 и 9.

В силу предположения 1 существуют $K\in[1,\infty)$ и $n_0\in\mathbb{Z}_+$, такие что

$$\max_{t:\,t\leqslant n} C_{1/2,U_{\mathcal{A}}^{(t)}} \leqslant K\ln(n), \quad \forall n\geqslant n_0.$$

Положим $\kappa_n = K\ln(1+n)$. Тогда

$$\gamma_n \leqslant 40(K+1)\ln(1+n), \quad \forall n\geqslant n_0. \tag{C.6}$$

Функция $C_{\alpha,W,\mathcal{A}}$ не убывает по α в силу [26, лемма 23(a)], а $\frac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ не возрастает по α на интервале $(0,1)$ в силу [26, лемма 23(b)]. Таким образом, мы можем оценить $\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$, используя ее определение, данное в (39):

$$\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} = \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[1\vee\left(\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\right] C_{\eta,W,\mathcal{A}}\,d\eta \leqslant$$
$$\leqslant \frac{1}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[1\vee\left(\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\right]\left[1\vee\left(\frac{1-\alpha}{\alpha}\frac{\eta}{1-\eta}\right)\right] C_{\alpha,W,\mathcal{A}}\,d\eta =$$
$$= \frac{C_{\alpha,W,\mathcal{A}}}{\epsilon}\int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[\left(\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\vee\left(\frac{1-\alpha}{\alpha}\frac{\eta}{1-\eta}\right)\right] d\eta \leqslant$$
$$\leqslant \left(1+\frac{\epsilon}{1-\epsilon}\frac{\alpha^2+(1-\alpha)^2}{\alpha(1-\alpha)}\right) C_{\alpha,W,\mathcal{A}}.$$

Тогда при $\epsilon_n = 1/n$ из оценки (C.6) следует, что

$$n\widetilde{C}^{\epsilon}_{\phi,W,\mathcal{A}} + \frac{\gamma_n}{1-\phi} + \ln\frac{8e^3n^{1,5}}{\epsilon} \leqslant$$
$$\leqslant nC_{\phi,W,\mathcal{A}} + \frac{n}{n-1}\frac{C_{\phi,W,\mathcal{A}}}{\phi(1-\phi)} + \frac{40(K+1)\ln(1+n)}{1-\phi} + 3\ln(2en), \quad \forall n \geqslant n_0. \tag{C.7}$$

Таким образом, из условий теоремы 1 вытекают условия леммы 9 при всех достаточно больших n. Значит, для всех достаточно больших n из леммы 9 и оценки (C.6) следует, что

$$P_{\mathrm{e}} \geqslant \left(\frac{(1+n)^{-80(K+1)}}{8e^3n^{2,5}}\right)^{1/\phi} e^{-n\widetilde{E}^{\epsilon_n}_{\mathrm{sp}}\left(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\right)}. \tag{C.8}$$

С другой стороны, из леммы 8, соотношения (31) и монотонности $C_{\alpha,W,\mathcal{A}}$ по α следует, что для достаточно больших n

$$\widetilde{E}^{\epsilon_n}_{\mathrm{sp}}\Big(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\Big) \leqslant E_{\mathrm{sp}}\Big(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\Big) + \frac{C_{\alpha_1,W,\mathcal{A}}}{(n-1)\phi^2}. \tag{C.9}$$

Теперь (32) следует из (C.8) и (C.9). ▲

Доказательство теоремы 2. Будем доказывать теорему 2, используя леммы 8 и 10. Мы можем выбирать различные значения κ и ϵ для различных значений n при условии, что выполняются условия лемм 8 и 10.

В силу предположения 2 существуют $K \in [1,\infty)$ и $n_0 \in \mathbb{Z}_+$, такие что

$$\max_{t:\, t\leqslant n} C_{1/2,W_t,n\varrho} \leqslant K\ln n, \quad \forall n \geqslant n_0. \tag{C.10}$$

Положим $\kappa_n = K\ln(1+n)$. Тогда

$$\gamma_n \leqslant 40(K+1)\ln(1+n), \quad \forall n \geqslant n_0. \tag{C.11}$$

Функция $C_{\alpha,W,\varrho}$ не убывает по α в силу [26, лемма 23(a)], а $\dfrac{1-\alpha}{\alpha}C_{\alpha,W,\varrho}$ не возрастает по α на интервале $(0,1)$ в силу [26, лемма 23(b)]. Таким образом, мы можем оценить $\widetilde{C}^{\epsilon}_{\alpha,W,\varrho}$, используя ее определение, данное в (39):

$$\widetilde{C}^{\epsilon}_{\alpha,W_{[1,n]},n\varrho} \leqslant \left(1 + \frac{\epsilon}{1-\epsilon}\frac{\alpha^2+(1-\alpha)^2}{\alpha(1-\alpha)}\right) C_{\alpha,W_{[1,n]},n\varrho}.$$

Тогда при $\epsilon_n = \dfrac{1}{n}$ из оценок (C.10) и (C.11), а также [26, леммы 23(a),(b), 27(a) и 28] следует, что

$$\widetilde{C}^{\epsilon}_{\alpha,W_{[1,n]},n\varrho} + \frac{\gamma_n}{1-\phi} + \ln\frac{8e^3n^{1,5}}{\epsilon} \leqslant C_{\alpha,W_{[1,n]},n\varrho} + \frac{n}{n-1}\frac{K\ln(n)}{\phi(1-\phi)} +$$
$$+ \frac{40(K+1)\ln(1+n)}{1-\phi} + 3\ln(2en), \quad \forall n \geqslant n_0. \tag{C.12}$$

Таким образом, из условий теоремы 2 вытекают условия леммы 10 при всех достаточно больших n. Значит, для всех достаточно больших n из леммы 10 и оценки (C.11) следует, что

$$P_{\mathrm{e}} \geqslant \left(\frac{(1+n)^{-80(K+1)}}{8e^3n^{2,5}}\right)^{1/\phi} e^{-\widetilde{E}^{\epsilon_n}_{\mathrm{sp}}\left(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\right)}. \tag{C.13}$$

С другой стороны, из леммы 8, соотношения (33) и монотонности $C_{\alpha,W,\varrho}$ по α следует, что для достаточно больших n

$$\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\Big(\ln\frac{M_n}{L_n}, W_{[1,n]}, n\varrho\Big) \leqslant E_{\mathrm{sp}}\Big(\ln\frac{M_n}{L_n}, W_{[1,n]}, n\varrho\Big) + \frac{C_{\alpha_1,W_{[1,n]},n\varrho}}{(n-1)\phi^2}. \tag{C.14}$$

Заметим, что $C_{\alpha,W_{[1,n]},n\varrho}$ не убывает по α согласно [26, лемма 23(a)], а функция $\dfrac{1-\alpha}{\alpha}C_{\alpha,W_{[1,n]},n\varrho}$ не возрастает по α на интервале $(0,1)$ согласно [26, лемма 23(b)]. Таким образом,

$$C_{\alpha_1,W_{[1,n]},n\varrho} \leqslant (\frac{\alpha_1}{1-\alpha_1} \vee 1)C_{1/2,W_{[1,n]},n\varrho}. \tag{C.15}$$

Так как $C_{\alpha,W,\varrho}$ не убывает по ϱ согласно [26, лемма 27(a)], то из соотношения (C.10) и [26, лемма 28] следует, что для всех достаточно больших n

$$C_{1/2,W_{[1,n]},n\varrho} \leqslant Kn\ln n. \tag{C.16}$$

Для достаточно больших n из оценок (C.14), (C.15) и (C.16) вытекает, что

$$\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\Big(\ln\frac{M_n}{L_n}, W_{[1,n]}, n\varrho\Big) \leqslant E_{\mathrm{sp}}\Big(\ln\frac{M_n}{L_n}, W_{[1,n]}, n\varrho\Big) + \frac{2}{\phi^2}\Big(\frac{\alpha_1}{1-\alpha_1} \vee 1\Big)K\ln n. \tag{C.17}$$

Теперь (34) следует из (C.13) и (C.17). ▲

Автор выражает благодарность Фатьме и Мехмету Накибоглу за гостеприимство, без которого эта работа не была бы возможной. Автор также благодарит М. Далаи за указание на неявное утверждение Фано о неподвижной точке в [9], Г. Васкеса-Вилара за указание на работу Полтырева [30] о границе случайного кодирования и за предложения по улучшению статьи, В. Яна за указание на [54, формула (3.63)] и ее отношение к подходу Блейхута, а также рецензента за предложения по улучшению статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Nakiboğlu B.* The Augustin Center and the Sphere Packing Bound for Memoryless Channels // Proc. 2017 IEEE Int. Sympos. on Information Theory (ISIT'2017). Aachen, Germany. June 25–30, 2017. P. 1401–1405.
2. *Elias P.* Coding for Noisy Channels // IRE Conv. Rec. 1955. V. 4. P. 37–46. Reprinted in: Key Papers in the Development of Information Theory. New York: IEEE Press, 1974. P. 48–55.
3. *Elias P.* Coding for Two Noisy Channels // Information Theory: Third London Symposium (London, England, Sept. 1955). London: Butterworth Sci., 1956. P. 61–74.
4. *Добрушин Р.Л.* Асимптотические оценки вероятности ошибки при передаче сообщения по дискретному каналу связи без памяти с симметрической матрицей вероятностей перехода // Теория вероятн. и ее примен. 1962. V. 7. № 3. P. 283–311.
5. *Altuğ Y., Wagner A.B.* On Exact Asymptotics of the Error Probability in Channel Coding: Symmetric Channels, `arXiv:1908.11419 [cs.IT]`, 2019.
6. *Altuğ Y., Wagner A.B.* Refinement of the Sphere Packing Bound for Symmetric Channels // Proc. 49th Annu. Allerton Conf. on Communication, Control, and Computing. Monticello, IL, USA. Sept. 27–30, 2011. P. 30–37.
7. *Shannon C.E., Gallager R.G., Berlekamp E.R.* Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. I // Inform. Control. 1967. V. 10. № 1. P. 65–103.
8. *Gallager R.G.* Information Theory and Reliable Communication. New York: John Wiley & Sons, 1968.
9. *Fano R.M.* Transmission of Information: A Statistical Theory of Communications. New York: M.I.T. Press, 1961.

10. *Арутюнян Е.А.* Оценки экспоненты вероятности ошибки для полунепрерывного канала без памяти // Пробл. передачи информ. 1968. Т. 4. № 4. С. 37–48.
11. *Omura J.K.* A Lower Bounding Method for Channel and Source Coding Probabilities // Inform. Control. 1975. V. 27. № 2. P. 148–177.
12. *Csiszár I., Körner J.* Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge, UK: Cambridge Univ. Press, 2011.
13. *Altuğ Y., Wagner A.B.* Refinement of the Sphere-Packing Bound: Asymmetric Channels // IEEE Trans. Inform. Theory. 2014. V. 60. № 3. P. 1592–1614.
14. *Nakiboğlu B.* A Simple Derivation of the Refined SPB for the Constant Composition Codes // Proc. 2019 IEEE Int. Symp. on Information Theory (ISIT'2019). Paris, France. July 7–12, 2019. P. 2659–2663.
15. *Wyner A.D.* Capacity and Error Exponent for the Direct Detection Photon Channel. II // IEEE Trans. Inform. Theory. 1988. V. 34. № 6. P. 1462–1471.
16. *Бурнашев М.В., Кутоянц Ю.А.* О границе сферической упаковки, пропускной способности и о подобных результатах для пуассоновского канала // Пробл. передачи информ. 1999. Т. 35. № 2. С. 3–22.
17. *Shannon C.E.* Probability of Error for Optimal Codes in a Gaussian Channel // Bell System Tech. J. 1959. V. 38. № 3. P. 611–656.
18. *Ebert P.M.* Error Bounds for Gaussian Noise Channels // Quarterly Progress Rep. № 77. Research Lab. of Electronics, MIT. Cambridge, MA, USA. 1965. P. 292–302. Available at `https://dspace.mit.edu/handle/1721.1/55609`.
19. *Ebert P.M.* Error Bounds for Parallel Communication Channels // Tech. Rep. № 448. Research Lab. of Electronics, MIT. Cambridge, MA, USA. 1966. Available at `https://dspace.mit.edu/handle/1721.1/4295`.
20. *Richters J.S.* Communication over Fading Dispersive Channels. Tech. Rep. № 464. Research Lab. of Electronics, MIT. Cambridge, MA, USA. 1967. Available at `https://dspace.mit.edu/handle/1721.1/4279`.
21. *Lancho Serrano A.* Fundamental Limits of Short-Packet Wireless Communications: Ph.D. Thesis. Departamento de Teoría de la Señal y Comunicaciones, Univ. Carlos III de Madrid, Spain. 2019. Available at `https://e-archivo.uc3m.es/handle/10016/29596`.
22. *Lancho A., Östman J., Durisi G., Koch T., Vazquez-Vilar G.* Saddlepoint Approximations for Short-Packet Wireless Communications // IEEE Trans. Wireless Commun. 2020. V. 19. № 7. P. 4831–4846.
23. *Augustin U.* Error Estimates for Low Rate Codes // Z. Wahrsch. Verw. Gebiete. 1969. V. 14. № 1. P. 61–88.
24. *Augustin U.* Noisy Channels // Habilitation Thesis. Univ. Erlangen-Nürnberg, Germany, 1978. Avaliable at `http://bit.ly/2ID8h7m`.
25. *Nakiboğlu B.* The Sphere Packing Bound via Augustin's Method // IEEE Trans. Inform. Theory. 2019. V. 65. № 2. P. 816–840.
26. *Накибоглу Б.* Августиновская пропускная способность и августиновский центр // Пробл. передачи информ. 2019. Т. 55. № 4. С. 3–51.
27. *Nakiboğlu B.* The Rényi Capacity and Center // IEEE Trans. Inform. Theory. 2019. V. 65. № 2. P. 841–860.
28. *Nakiboğlu B.* A Simple Derivation of the Refined Sphere Packing Bound under Certain Symmetry Hypotheses // Turkish J. Math. 2020. V. 44. № 3. P. 919–948.
29. *Gallager R.G.* A Simple Derivation of the Coding Theorem and Some Applications // IEEE Trans. Inform. Theory. 1965. V. 11. № 1. P. 3–18.
30. *Полтырев Г.Ш.* Границы случайного кодирования для дискретных каналов без памяти // Пробл. передачи информ. 1982. Т. 18. № 1. С. 12–26.
31. *Blahut R.E.* Hypothesis Testing and Information Theory // IEEE Trans. Inform. Theory. 1974. V. 20. № 4. P. 405–417.
32. *Blahut R.E.* Information Bounds of the Fano–Kullback Type // IEEE Trans. Inform. Theory. 1976. V. 22. № 4. P. 410–421.
33. *Blahut R.E.* Principles and Practice of Information Theory. Reading, MA: Addison-Wesley, 1987.

34. *Vazquez-Vilar G., Martinez A., Guillén i Fàbregas A.* A Derivation of the Cost-Constrained Sphere-Packing Exponent // Proc. 2015 IEEE Int. Sympos. on Information Theory (ISIT'2015). Hong Kong, China. June 14–19, 2015. P. 929–933.
35. *Dudley R.M.* Real Analysis and Probability. Cambridge: Cambridge Univ. Press, 2002.
36. *van Erven T., Harremoës P.* Rényi Divergence and Kullback–Leibler Divergence // IEEE Trans. Inform. Theory. 2014. V. 60. № 7. P. 3797–3820.
37. *Rudin W.* Principles of Mathematical Analysis. New York: McGraw-Hill, 1976.
38. *Bogachev V.I.* Measure Theory. Berlin: Springer, 2007.
39. *Elias P.* List Decoding for Noisy Channels // Tech. Rep. № 335. Research Lab. of Electronics, MIT. Cambridge, MA, USA. 1957. Available at `https://dspace.mit.edu/handle/1721.1/4484`.
40. *Dalai M.* Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels // IEEE Trans. Inform. Theory. 2013. V. 59. № 12. P. 8027–8056.
41. *Dalai M.* Some Remarks on Classical and Classical-Quantum Sphere Packing Bounds: Rényi vs. Kullback–Leibler // Entropy. 2017. V. 19. № 7. P. 355 (11 pp.).
42. *Dalai M., Winter A.* Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels // IEEE Trans. Inform. Theory. 2017. V. 63. № 9. P. 5603–5617.
43. *Cheng H.-C., Hsieh M.H.* Moderate Deviation Analysis for Classical-Quantum Channels and Quantum Hypothesis Testing // IEEE Trans. Inform. Theory. 2018. V. 64. № 2. P. 1385–1403.
44. *Cheng H.-C., Hsieh M.H., Tomamichel M.* Quantum Sphere-Packing Bounds with Polynomial Prefactors // IEEE Trans. Inform. Theory. 2019. V. 65. № 5. P. 2872–2898.
45. *Altuğ Y., Wagner A.B.* Refinement of the Random Coding Bound // IEEE Trans. Inform. Theory. 2014. V. 60. № 10. P. 6005–6023.
46. *Scarlett J., Martinez A., Guillén i Fàbregas A.* Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations // IEEE Trans. Inform. Theory. 2014. V. 60. № 5. P. 2647–2666.
47. *Honda J.* Exact Asymptotics for the Random Coding Error Probability // Proc. 2015 IEEE Int. Sympos. on Information Theory (ISIT'2015). Hong Kong, China. June 14–19, 2015. P. 91–95.
48. *Бурнашев М.В., Холево А.С.* О функции надежности квантового канала связи // Пробл. передачи информ. 1998. Т. 34. № 2. С. 3–15.
49. *Telatar E.* Capacity of Multi-antenna Gaussian Channels // Eur. Trans. Telecommun. 1999. V. 10. № 6. P. 585–595.
50. *Vu M.* MISO Capacity with Per-Antenna Power Constraint // IEEE Trans. Commun. 2011. V. 59. № 5. P. 1268–1274.
51. *Cao P.L., Oechtering T.J., Schaefer R.F., Skoglund M.* Optimal Transmit Strategy for MISO Channels with Joint Sum and Per-Antenna Power Constraints // IEEE Trans. Signal Process. 2016. V. 64. № 16. P. 4296–4306.
52. *Loyka S.* The Capacity of Gaussian MIMO Channels under Total and Per-Antenna Power Constraints // IEEE Trans. Commun. 2017. V. 65. № 3. P. 1035–1043.
53. *Cheng H.-C., Nakiboğlu B.* Refined Strong Converse for the Constant Composition Codes, `arXiv:2002.11414 [cs.IT]`, 2020.
54. *Yang W.* Fading Channels: Capacity and Channel Coding Rate in the Finite-Blocklength Regime: Ph.D. Thesis. Communication Systems Group, Dept. of Signals and Systems, Chalmers Univ. of Technology, Gothenburg, Sweden, 2015.

Накибоглу Барыш
Средневосточный технический университет,
Анкара, Турция
`bnakib@metu.edu.tr`

Поступила в редакцию
18.04.2018
После доработки
06.03.2020
Принята к публикации
29.04.2020

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.725

© 2020 г. **Л.А. Бассалыго, В.А. Зиновьев, В.С. Лебедев**

СИММЕТРИЧНЫЕ БЛОК-СХЕМЫ И ОПТИМАЛЬНЫЕ ЭКВИДИСТАНТНЫЕ КОДЫ[1]

Доказывается, что любая симметричная блок-схема (v, k, λ) порождает оптимальные троичный и четверичный равновесные эквидистантные коды, параметры которых n, N, w, d, q однозначно определяются параметрами блок-схемы. Для одного весьма частного случая построены посимвольно равномерные эквидистантные коды минимальной длины.

Ключевые слова: блок-схема, посимвольно равномерный код, эквидистантный код.

DOI: 10.31857/S055529232003002X

1. Назовем q-ичный код длины n мощности N с расстоянием d *посимвольно равномерным*, если в матрице кодовых слов размера $N \times n$ каждый ненулевой символ содержится в каждом столбце матрицы одно и то же число раз независимо от выбора символа и выбора столбца. Всюду в дальнейшем это число полагается равным t, а следовательно, нулевой символ в каждом столбце содержится $m = N - (q-1)t$ раз. Обозначим через $\bar{w}$ средний вес кодовых слов:

$$\bar{w} = \frac{\text{число ненулевых символов всех кодовых слов}}{N}.$$

Лемма. *Если*

$$q\bar{w}^2 > (q-1)(2\bar{w} - d)n,$$

то имеет место следующее неравенство:

$$N \leqslant \frac{(q-1)dn}{q\bar{w}^2 - (q-1)(2\bar{w} - d)n}. \tag{1}$$

Доказательство. Доказательство леммы для равновесных кодов приведено в работе [1], и здесь мы лишь повторим его с использованием среднего веса кодовых слов. Пусть $C = \{c_1, c_2, \ldots, c_N\}$ – код длины n с расстоянием d над q-ичным алфавитом $Q = \{\ell : \ \ell = 0, 1, \ldots, q-1\}$. Рассмотрим матрицу размера $N \times n$ кодовых слов кода C. Обозначим через $w_{\ell j}$ число использования символа ℓ в j-м столбце кодовой матрицы ($\ell = 0, 1, \ldots, q-1$, $j = 1, 2, \ldots, n$). Из определения среднего веса $\bar{w}$ вытекает, что

$$\sum_{j=1}^{n} \sum_{\ell=1}^{q-1} w_{\ell j} = N\bar{w}, \quad \sum_{j=1}^{n} w_{0j} = N(n - \bar{w}). \tag{2}$$

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 19-01-00364).

Так как

$$\sum_{\substack{k,s=1\\k\neq s}}^{N} d(c_k,c_s)=\sum_{j=1}^{n}\sum_{\ell=0}^{q-1} w_{\ell j}(N-w_{\ell j}), \tag{3}$$

то

$$\sum_{j=1}^{n}\sum_{\ell=0}^{q-1} w_{\ell j}(N-w_{\ell j})\geqslant N(N-1)d, \tag{4}$$

причем равенство достигается тогда и только тогда, когда код является эквидистантным. С другой стороны,

$$\sum_{j=1}^{n}\sum_{\ell=0}^{q-1} w_{\ell j}(N-w_{\ell j})=N^2n-\sum_{j=1}^{n}\sum_{\ell=1}^{q-1} w_{\ell j}^2-\sum_{j=1}^{n} w_{0j}^2,$$

и из (2) следует, что

$$\sum_{j=1}^{n}\sum_{\ell=1}^{q-1} w_{\ell j}^2\leqslant\frac{N^2\bar{w}^2}{n(q-1)}\quad\text{и}\quad\sum_{j=1}^{n} w_{0j}^2\leqslant\frac{N^2(n-\bar{w})^2}{n},$$

причем равенство в этих соотношениях достигается тогда и только тогда, когда

$$w_{\ell j}=\frac{N\bar{w}}{n(q-1)},\quad \ell=1,\dots,q-1,\quad j=1,\dots,n,$$

и

$$w_{0j}=\frac{N(n-\bar{w})}{n},\quad j=1,\dots,n.$$

Следовательно,

$$\sum_{j=1}^{n}\sum_{\ell=0}^{q-1} w_{\ell j}(N-w_{\ell j})\leqslant N^2n-\frac{N^2\bar{w}^2}{n(q-1)}+\frac{N^2(n-\bar{w})^2}{n},$$

причем равенство достигается тогда и только тогда, когда код является посимвольно равномерным. Отсюда и из (4) получаем

$$N(q\bar{w}^2-(q-1)(2\bar{w}-d)n)\leqslant(q-1)dn,$$

что и завершает доказательство. ▲

Из доказательства леммы вытекает следующее

У т в е р ж д е н и е 1. *Код достигает границы* (1) *тогда и только тогда, когда он эквидистантен и посимвольно равномерен.*

Обратим внимание, что мы не потребовали равновесности кода, как в стандартной формулировке границы (1), ибо верно следующее

У т в е р ж д е н и е 2. *Если* $m\neq t$, *то эквидистантный посимвольно равномерный код равновесен.*

Д о к а з а т е л ь с т в о. Пусть x_s – слово кода, а w_s – его вес, $s=1,2,\dots,N$. В силу эквидистантности кода

$$\sum_{i=1}^{N} d(x_i,x_s)=(N-1)d. \tag{5}$$

С другой стороны, подсчитывая эту сумму по столбцам, получим

$$\sum_{i=1}^{N} d(x_i, x_s) = (N-t)w_s + (N-m)(n-w_s). \tag{6}$$

Следовательно,

$$w_s(m-t) = (N-1)d - (N-m)n, \tag{7}$$

и если $m \neq t$, то вес w_s кодового слова не зависит от s, что доказывает равновесность кода. ▲

Если же $m = t$, то $(N-1)d = (N-m)n$ и $N = qm$. Отсюда получаем, что

$$N = \frac{qd}{qd-(q-1)n},$$

т.е. имеет место

Следствие 1 [2]. *Если $m = t$, то эквидистантный посимвольно равномерный (n, N, d)-код C лежит на границе Плоткина*

$$N \leqslant \frac{qd}{qd-(q-1)n}. \tag{8}$$

Много семейств кодов, лежащих на границе Плоткина (8), было указано в [2]. Так как сдвиг такого кода на любое слово сохраняет не только эквидистантность кода, но и его посимвольную равномерность, то вычтя из всех кодовых слов фиксированное кодовое слово, придем к коду, все кодовые слова которого за исключением нулевого слова имеют вес, равный расстоянию кода (при этом операция сложения в алфавите определяется по модулю q). Средний вес этого кода равен

$$\bar{w} = \frac{(N-1)d}{N} = \frac{(q-1)n}{q},$$

и согласно утверждению 1 этот код лежит на границе Джонсона (1), которая в данном случае совпадает с границей Плоткина, что нетрудно проверить. Удалив из этого кода нулевое слово, получим равновесный код, лежащий на границе Джонсона (1) (но, естественно, не лежащий на границе Плоткина). С другой стороны, если код лежит на границе Джонсона (и следовательно, эквидистантен) и при этом средний вес кода $\bar{w} = d$ (и следовательно, код равновесен), то присоединяя к этому коду нулевое слово, получим код, лежащий на границе Плоткина.

2. *Блок-схемой* $B(v, k, \lambda)$ называется такое размещение v различных элементов $a_1, \ldots, a_v$ по b блокам $B_1, \ldots, B_b$, при котором каждый блок содержит ровно k различных элементов, каждый элемент принадлежит ровно r блокам, и каждая пара различных элементов a_i, a_j, $i \neq j$, принадлежит ровно λ блокам. Параметры v, k, λ блок-схемы однозначно определяют параметры b и r [3]:

$$b = \frac{\lambda v(v-1)}{k(k-1)}, \quad r = \frac{\lambda(v-1)}{k-1}.$$

Отметим, что $b \geqslant v$ согласно неравенству Фишера [3]. Блок-схема полностью описывается своей матрицей инцидентности $A = [a_{ij}]$, где

$$a_{ij} = \begin{cases} 1, & \text{если } i \in B_j, \\ 0, & \text{если } i \notin B_j, \end{cases} \qquad i = 1, \ldots, v, \quad j = 1, \ldots, b.$$

Утверждение 3. *Любая блок-схема $B(v,k,\lambda)$ порождает троичный равновесный эквидистантный код с параметрами*

$$N=v,\quad n=\binom{b}{2},\quad w=r(b-r),\quad d=(r-\lambda)(2b-3(r-\lambda)) \tag{9}$$

и четверичный равновесный эквидистантный код с параметрами

$$N=v,\quad n=\binom{b}{2},\quad w=\frac{r}{2}(2b-r-1),\quad d=(r-\lambda)(2b-2(r-\lambda)-1). \tag{10}$$

Доказательство. Построение как троичного, так и четверичного кодов происходит по одной и той же весьма простой схеме: выбираются любые два столбца матрицы инцидентности блок-схемы (таковых $\binom{b}{2}$) и заменяются на один столбец q-ичного кода по следующему правилу: для $q=3$

$$(0,0)\to 0,\quad (1,1)\to 0,\quad (0,1)\to 1,\quad (1,0)\to 2,$$

а для $q=4$

$$(0,0)\to 0,\quad (0,1)\to 1,\quad (1,0)\to 2,\quad (1,1)\to 3,$$

так что в результате получаем троичный и четверичный коды. Очевидно, что длина построенных кодов $n=\binom{b}{2}$, а число слов в каждом из них $N=v$. Нетрудно проверить, что число совпадений между любыми двумя кодовыми словами троичного кода равно

$$\binom{b-2(r-\lambda)}{2}+2\binom{r-\lambda}{2},$$

а четверичного –

$$\binom{b-2(r-\lambda)}{2},$$

так как число позиций, в которых два кодовых слова совпадают, не зависит от выбора кодовых слов. Вес кодовых слов при $q=3$ равен $r(b-r)$, а при $q=4$ равен

$$\binom{b}{2}-\binom{b-r}{2}=\frac{r}{2}(2b-r-1).\quad \blacktriangle$$

Так как предложенный в утверждении 3 способ построения троичного и четверичного кодов приводит к равновесным эквидистантным кодам, то естественно задаться вопросом, когда эти коды лежат на обобщенной границе Джонсона (1). Из утверждения 1 вытекает, что для выполнения этого условия коды должны быть посимвольно равномерны. Нетрудно видеть, что для посимвольной равномерности кодов из утверждения 3 необходимо, чтобы в (v,k,λ)-схеме скалярное произведение любых двух столбцов матрицы инцидентности не зависело от выбора столбцов. Если $b>v$, то такие схемы не существуют [3]. Если $b=v$, то такая (v,k,λ)-схема называется *симметричной*, и скалярное произведение любых двух столбцов матрицы инцидентности симметричной схемы равно λ [3].

Поэтому верна

Теорема 1. *Троичный код из утверждения* 3 *лежит на границе Джонсона тогда и только тогда, когда (v,k,λ)-схема симметрична. Четверичный код из*

утверждения 3 *лежит на границе Джонсона тогда и только тогда, когда* (v,k,λ)-*схема симметрична и выполняется равенство* $\lambda = k - \lambda$.

Итак, все симметричные (v,k,λ)-схемы порождают оптимальные троичные равновесные эквидистантные коды с параметрами (9), которые лежат на границе Джонсона (1). Если же симметричная схема является к тому же $(4\lambda-1, 2\lambda, \lambda)$-схемой, то она порождает оптимальный четверичный равновесный эквидистантный код с параметрами (10), также лежащий на границе Джонсона (1). Присоединяя к этому коду нулевое слово, получим четверичный код, лежащий на границе Плоткина. Такие четверичные коды из симметричных схем уже были построены ранее [4].

3. Построенные в предыдущем разделе троичные и четверичные коды, лежащие на границе Джонсона (см. теорему 1), согласно утверждению 1 являются посимвольно равномерными. Параметры (m,t) этих кодов легко вычисляются через параметры симметричных (v,k,λ)-схем. Например, у таких троичных кодов $m = v - 2(k-\lambda)$, $t = k - \lambda$. В [5,6] были приведены семейства q-ичных эквидистантных равновестных кодов, лежащих на границе Джонсона, – эти коды в силу утверждения 1 являются посимвольно равномерными, и их параметры также легко вычисляются. Возникает естественный вопрос: при любых ли значениях параметров q, m, t существует посимвольно равномерный код, лежащий на границе Джонсона при $m \neq t$ и Плоткина при $m = t$? Согласно утверждению 1 и следствию 1 этот вопрос эквивалентен следующему: при любых ли значениях этих параметров существует посимвольно равномерный эквидистантный код? Ответ почти очевиден: при любых значениях параметров q, m, t такой код существует. Действительно, любая матрица размера

$$N \times \frac{N!}{N_0!\,N_1!\ldots N_{q-1}!},$$

состоящая из всех столбцов заданной композиции $N_0 + N_1 + \ldots + N_{q-1} = N$ над q-ичным алфавитом, представляет собой эквидистантный код длины n с расстоянием d, где

$$n = \frac{N!}{N_0!\,N_1!\ldots N_{q-1}!}, \quad d = n - \sum_{i=0}^{q-1} \frac{(N-2)!}{N_0!N_1!\ldots(N_i-2)!\ldots N_{q-1}!},$$

и числом кодовых слов, равным N. Следовательно, это верно и для композиции посимвольно равномерного кода, где $N = m + (q-1)t$ и $N_0 = m$, $N_1 = \ldots = N_{q-1} = t$. Поэтому интерес представляет задача построения эквидистантных кодов минимальной длины при заданных параметрах q, m, t. Но даже при $m = t$, т.е. для кодов, лежащих на границе Плоткина, эта задача далека от завершения, хотя в работах [2,7] такие коды были построены для многих пар параметров (q,t) (в частности, когда q и t являются произвольными степенями любого простого числа p). Тем более интересна такая задача для случая $m \neq t$. Мы рассмотрим первый самый простой случай $m = 3$, $t = 2$, когда число кодовых слов $N = 3 + (q-1)2 = 2q+1$. Начнем с нижней оценки минимальной длины. Так как посимвольно равномерный эквидистантный код при $m \neq t$ равновесен, то нетрудно видеть, что $wN = n(N-m)$. Отсюда и из равенства (7) получаем, что

$$d = \frac{(N-m)(N+m-t)n}{N(N-1)}. \tag{11}$$

Подставляя в это равенство значения $m = 3$, $t = 2$ и учитывая, что $N = 2(q-1)+3 = 2q+1$, получаем

$$d = \frac{(N+1)(N-3)n}{N(N-1)} = \frac{2(q-1)(q+1)n}{q(2q+1)}.$$

Из целочисленности d вытекает следующее легко проверяемое

Утверждение 4. *Пусть n – длина q-ичного эквидистантного посимвольно равномерного кода ($m = 3$, $t = 2$). Тогда*

1. *Если q нечетно и $q \not\equiv 1 \pmod 6$, то $n \geqslant q(2q+1)$;*
2. *Если q нечетно и $q \equiv 1 \pmod 6$, то $n \geqslant q(2q+1)/3$;*
3. *Если q четно и $q \not\equiv 4 \pmod 6$, то $n \geqslant q(2q+1)/2$;*
4. *Если q четно и $q \equiv 4 \pmod 6$, то $n \geqslant q(2q+1)/6$.*

Хотя мы полагаем, что указанные нижние границы минимальной длины достигаются во всех четырех случаях, здесь мы докажем это лишь для первого и третьего. В обоих этих случаях указывается способ построения эквидистантных кодов, минимальная длина которых не превосходит $q(2q+1)$ при любом q и не превосходит $q(2q+1)/2$ при четных q.

Утверждение 5. *Минимальная длина q-ичного эквидистантного посимвольно равномерного кода ($m = 3$, $t = 2$) не превосходит $q(2q+1)$.*

Доказательство. Рассмотрим матрицу размера $(2q+1) \times n$ кодовых слов посимвольно равномерного кода. В любом столбце (длины $2q+1$) этой матрицы каждый ненулевой символ встречается два раза, а нулевой символ – три раза. Перенумеруем строки кодовой матрицы целыми числами удобным для дальнейших описаний образом: первая строка получит номер q, вторая $q-1$ и т.д., последняя строка получит номер $-q$ (т.е. номера строк заполняют целочисленный отрезок $[-q, q]$). Каждый столбец кодовой матрицы индуцирует набор из q непересекающихся подмножеств номеров строк – одно подмножество мощности три, состоящее из номеров строк, содержащих нулевой символ в данном столбце (далее называемое тройкой), остальные $q-1$ подмножеств мощности два, состоящие из номеров строк, содержащих одинаковый ненулевой символ в данном столбце (далее называемые парой). Такой набор, индуцированный j-м столбцом, будем обозначать через V_j, $j = 1, 2, \ldots, n$. Дистанционные свойства кода полностью определяются этими наборами, так как расстояние Хэмминга между двумя кодовыми словами с номерами u и u', $u \neq u'$, равно

$$d(u, u') = n - \left(\text{число пар } \{u, u'\} \text{ в } \bigcup_{j=1}^{n} V_j \right)$$

(каждая тройка в V_j, $j = 1, 2, \ldots, n$, заменена на три соответствующие пары, так что каждый набор V_j содержит $q+2$ пар). Отсюда вытекает, что если каждая пара номеров кодовых слов встречается в $\bigcup\limits_{j=1}^{n} V_j$ одинаковое число раз, то код является эквидистантным. К построению наборов, обладающих этим свойством, мы и переходим. Сначала построим следующие q наборов V_u, $u = 1, 2, \ldots, q$ (назовем их генераторами):

$$V_u = \{u, 0, -u\} \cup \bigcup_{\substack{v=1,\ldots,q \\ v \neq u}} \{v, -v\}.$$

Очевидно, что в $\bigcup\limits_{u=1}^{q} V_u$ каждая пара $\{v, -v\}$ входит q раз, и по одному разу входит каждая пара $\{v, 0\}$ и пара $\{0, -v\}$, $v = 1, 2, \ldots, q$ (напомним, что каждая тройка $\{u, 0, -u\}$ заменяется на три пары $\{u, -u\}$, $\{u, 0\}$, $\{0, -u\}$).

Проделав все циклические сдвиги наших генераторов, получим $n = q(2q+1)$ наборов V_j, $j = 1, 2, \ldots, n$. Докажем теперь, что каждая пара $\{u, u'\}$, $u \neq u'$ (таких пар всего $q(2q+1)$), входит в $\bigcup\limits_{j=1}^{n} V_j$ одинаковое число раз, а именно $q+2$ раз.

Назовем орбитой $O(\{u,u'\})$ пары $\{u,u'\}$ множество пар, полученных из этой пары всеми ее $2q+1$ сдвигами на пары вида x,x:

$$O(\{u,u'\}) = \{\{u+x, u'+x\},\ x \in [-q,q]\}$$

(операция сложения выполняется по модулю $(2q+1)$ в классе вычетов $[-q,q]$). Следующие свойства орбиты очевидны: а) все пары в орбите различны; б) пара $\{v,v'\}$ тогда и только тогда принадлежит орбите $O(\{u,u'\})$, когда $|v-v'| = |u-u'|$. Из свойства б) следует, что орбиты пар $O(\{u,u'\})$ и $O(\{v,v'\})$ либо совпадают, либо не пересекаются. Обозначим через $O(V_1,\ldots,V_q)$ множество орбит (с повторениями), порожденных всеми парами, входящими во все генераторы ($|O(V_1,\ldots,V_q)| = q(q+2)$, так как число генераторов равно q, а число пар в каждом генераторе равно $q+2$). Согласно построению V_j число пар $\{u,u'\}, u \neq u'$, в $\bigcup\limits_{j=1}^{n} V_j$ равно числу орбит из $O(V_1,\ldots,V_q)$, которым принадлежит пара $\{u,u'\}$. Следовательно, для эквидистантности кода достаточно доказать, что каждая пара принадлежит одному и тому же числу таких орбит.

Рассмотрим сначала все пары $\{v,-v\}$, $v=1,\ldots,q$, принадлежащие одному генератору V_u. Нетрудно видеть, что орбиты этих пар не пересекаются, а так как каждая орбита содержит $2q+1$ различных пар, то любая пара $\{u,u'\}$, $u \neq u'$, принадлежит одной такой орбите. Так как все генераторы включают набор пар $\{v,-v\}$, $v=1,\ldots,q$, то любая пара $\{u,u'\}$, $u \neq u'$, принадлежит q таким орбитам. Осталось рассмотреть пары $\{v,0\}$ и $\{0,-v\}$, $v=1,\ldots,q$. Очевидно, что орбиты пар $\{v,0\}$, $v=1,\ldots,q$, не пересекаются, и любая пара $\{u,u'\}$, $u \neq u'$, принадлежит одной такой орбите. Орбиты же пары $\{0,-v\}$ совпадают с орбитами пары $\{v,0\}$, и следовательно, любая пара $\{u,u'\}$, $u \neq u'$, принадлежит $q+2$ орбитам.

Напомним теперь, что наборы V_j были индуцированы столбцами кодовой матрицы, так что обратное преобразование этих наборов в столбцы кодовой матрицы очевидно – каждый набор превращается в столбец длины $2q+1$ кодовой матрицы по следующему правилу: на места, определяемые тройкой набора, ставится нулевой символ, а на места, определяемые парами набора, ставятся одинаковые ненулевые символы (различные пары набора должны содержать различные символы алфавита). Ясно, что такой код является q-ичным посимвольно равномерным ($m=3$, $t=2$) эквидистантным кодом мощности $N = 2q+1$ длины $n = q(2q+1)$ с расстоянием $d = q(2q+1) - q - 2 = 2(q^2-1)$. ▲

Из этого утверждения и п. 1 утверждения 4 вытекает

Следствие 2. *Если q нечетно и $q \not\equiv 1 \pmod 6$, то минимальная длина q-ичного эквидистантного посимвольно равномерного кода ($m=3$, $t=2$) равна $q(2q+1)$.*

В случае четного q имеет место

Утверждение 6. *Если q четно, то минимальная длина q-ичного эквидистантного посимвольно равномерного кода ($m=3$, $t=2$) не превосходит $q(2q+1)/2$.*

Доказательство. Пусть $q = 2s$. Построение наборов V_j весьма похоже на построение наборов при доказательстве утверждения 5, но число генераторов будет в два раза меньше, и поэтому после циклических сдвигов генераторов общее число наборов окажется в два раза меньше, чем в утверждении 5. При этом $s-1$ генераторов входят в число генераторов утверждения 5, а именно

$$V_{2u} = \{2u, 0, -2u\} \cup \bigcup_{\substack{v=1,\ldots,q \\ v \neq 2u}} \{v,-v\}, \quad u = 2,\ldots,s,$$

и лишь один генератор (обозначим его через V_1) строится особым способом:

$$V_1 = \{2s-1, 0, -(2s-1)\}, \{2s, -2s\}, \{s, s-2\}, \{s+1, s-1\},$$
$$\{-(s-2), -(s+1)\}, \{-(s-1), -s\},$$
$$\bigcup_{v=1,\ldots,s-3} \{v, 2s-1-v\}, \bigcup_{v=1,\ldots,s-3} \{-v, -(2s-1-v)\}.$$

Как и в утверждении 5, достаточно доказать, что каждая пара $\{u, u'\}$, $u \neq u'$, принадлежит одному и тому же числу орбит, порожденных всеми парами, входящими в генераторы. Рассмотрим сначала пары $\{2u, 0\}$ и $\{-2u, 0\}$, $u = 2, \ldots, s$, входящие в $s-1$ генераторов. Орбиты этих пар совпадают, и поэтому каждая пара $\{u, u'\}$ такая, что $|u-u'| = 4, 6, \ldots, 2s$, принадлежит этим двум орбитам. Далее рассмотрим орбиты, принадлежащие парам, входящим в генератор V_1. Орбиты пар $\{s, s-2\}$ и $\{s+1, s-1\}$ совпадают, так как $|s-(s-2)| = |s+1-(s-1)| = 2$, и поэтому любая пара $\{u, u'\}$, $|u-u'| = 2$, принадлежит орбитам этих двух пар. Орбиты пар $\{2s, -2s\}$ и $\{-(s-1), -s\}$ совпадают, так как $|2s+2s| = |-(s-1)+s| = 1$, и поэтому любая пара $\{u, u'\}$, $|u-u'| = 1$, принадлежит орбитам этих двух пар. Орбиты пар $\{2s-1, -(2s-1)\}$ и $\{-(s-2), -(s+1)\}$ совпадают, так как $|2s-1+(2s-1)| = |-(s-2)+(s+1)| = 3$, и поэтому любая пара $\{u, u'\}$, $|u-u'| = 3$, принадлежит орбитам этих двух пар. Орбиты пар $\{v, 2s-1-v\}$ и $\{-v, -(2s-1-v)\}$, $v = 0, 1, \ldots, s-3$ (пары $\{0, 2s-1\}$ и $\{0, -(2s-1)\}$ при $v = 0$ получены из тройки $\{2s-1, 0, -(2s-1)\}$), совпадают, и следовательно, любая пара $\{u, u'\}$, $|u-u'| = 5, 7, \ldots, 2s-1$, принадлежит двум орбитам.

Кроме того, в утверждении 5 мы доказали, что любая пара $\{u, u'\}$, $u \neq u'$, принадлежит $q-1$ орбитам тех пар из $s-1$ генераторов, которые мы еще не использовали. Поэтому q-ичный код, полученный из $n = q(2q+1)/2$ наборов V_j способом, указанным в утверждении 5, является эквидистантным кодом мощности $N = 2q+1$ и длины $n = q(2q+1)/2$ с расстоянием

$$d = \frac{q(2q+1)}{2} - \frac{q}{2} - 1 = q^2 - 1. \quad \blacktriangle$$

Из этого утверждения и п. 3 утверждения 4 вытекает

Следствие 3. *Если q четно и $q \not\equiv 4 \pmod 6$, то минимальная длина q-ичного эквидистантного посимвольно равномерного кода $(m = 3, t = 2)$ равна $q(2q+1)/2$.*

Замечание. Соотношение (11) с учетом равенства $N = (q-1)t + m$ и целочисленности d позволяет получать нижние оценки минимальной длины и для произвольных параметров q, m, t. Для конкретных кодов эти нижние оценки оказываются иногда точными (например, они оказываются таковыми для всех троичных кодов из теоремы 1, лежащих на границе Джонсона, при условии, что $(k-\lambda)(2v-3(k-\lambda))$ и $v(v-1)$ взаимно просты).

СПИСОК ЛИТЕРАТУРЫ

1. *Бассалыго Л.А.* Новые верхние границы для кодов, исправляющих ошибки // Пробл. передачи информ. 1965. Т. 1. № 4. С. 41–44.
2. *Семаков Н.В., Зиновьев В.А.* Эквидистантные q-ичные коды с максимальным расстоянием и разрешимые уравновешенные неполные блок-схемы // Пробл. передачи информ. 1968. Т. 4. № 2. С. 3–10.
3. *Beth T., Jungnickel D., Lenz B.* Design Theory. Cambridge: Cambridge Univ. Press, 1999.
4. *Sinha K., Sinha N.* A Class of Optimal Quaternary Codes // Ars Combin. 2010. V. 94. P. 61–64.

5. *Бассалыго Л.А., Зиновьев В.А.* Замечание об уравновешенных неполных блок-схемах, почти разрешимых блок-схемах и q-ичных равновесных кодах // Пробл. передачи информ. 2017. Т. 53. № 1. С. 55–58.
6. *Бассалыго Л.А., Зиновьев В.А., Лебедев В.С.* Об m-квазиразрешимых блок-схемах и q-ичных равновесных кодах // Пробл. передачи информ. 2018. Т. 54. № 3. С. 54–61.
7. *Семаков Н.В., Зиновьев В.А., Зайцев Г.В.* Класс максимальных эквидистантных кодов // Пробл. передачи информ. 1969. Т. 5. № 2. С. 84–87.

Бассалыго Леонид Александрович
Зиновьев Виктор Александрович
Лебедев Владимир Сергеевич
Институт проблем передачи информации
им. А.А. Харкевича РАН
bass@iitp.ru
zinov@iitp.ru
lebedev37@mail.ru

Поступила в редакцию
07.02.2020
После доработки
11.06.2020
Принята к публикации
14.06.2020

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.725 : 512.772.7

© 2020 г. **Н. Патанкер, С.К. Сингх**[1]

О ГЕОМЕТРИЧЕСКИХ КОДАХ ГОППЫ ПО ЭЛЕМЕНТАРНЫМ АБЕЛЕВЫМ p-РАСШИРЕНИЯМ ПОЛЯ $\mathbb{F}_{p^s}(x)$

Пусть p – простое число, а $s > 0$ – натуральное. Рассматриваются одноточечные геометрические коды Гоппы, ассоциированные с элементарными абелевыми p-расширениями поля $\mathbb{F}_{p^s}(x)$. Вычисляется их размерность и точное значение минимального расстояния в нескольких случаях. Эти коды являются частным случаем слабых за́мковых кодов. Также приводится список точных значений второго обобщенного веса Хэмминга этих кодов в нескольких случаях. Получены простые критерии самодвойственности и квазисамодвойственности этих кодов. Кроме того, построены примеры квантовых, сверточных и локально восстанавливаемых кодов по этим функциональным полям.

Ключевые слова: элементарное абелево p-расширение поля $\mathbb{F}_{p^s}(x)$, геометрические коды Гоппы, обобщенный вес Хэмминга.

DOI: 10.31857/S0555292320030031

§ 1. Введение

Пусть $\mathbb{F}_{p^s}$ – конечное поле из p^s элементов характеристики p (где s – натуральное число). Линейным кодом называется $\mathbb{F}_{p^s}$-подпространство пространства $\mathbb{F}_{p^s}^n$ – стандартного n-мерного векторного пространства над $\mathbb{F}_{p^s}$. Такие коды используются для передачи информации.

В [1] Гоппа заметил, что можно использовать дивизоры в поле алгебраических функций для построения класса линейных кодов. В конструкции Гоппы выбирается дивизор G и n рациональных точек (т.е. точек степени 1) алгебраического функционального поля, чтобы получить линейный код длины n. Такие коды называются геометрическими кодами Гоппы. Если G имеет вид rQ для точки Q функционального поля и некоторого целого r, то такие коды называются одноточечными. Одноточечные геометрические коды Гоппы на функциональном поле эрмитовой кривой изучались, например, в [2–5].

Элементарное абелево p-расширение поля рациональных функций $\mathbb{F}_{p^s}(x)$ – это расширение Галуа F поля $\mathbb{F}_{p^s}(x)$, такое что $\mathrm{Gal}(F/\mathbb{F}_{p^s}(x))$ является элементарной абелевой группой порядка p. Примером такого расширения является функциональное поле, ассоциированное с эрмитовой кривой. Свойства элементарных абелевых p-расширений поля $\mathbb{F}_{p^s}(x)$ изучались, в частности, в [2, 6–8]. Другим примером элементарного абелева p-расширения поля $\mathbb{F}_{p^s}(x)$ является функциональное поле $\mathbb{F}_{p^s}(x, y)/\mathbb{F}_{p^s}$, заданное уравнением $A(y) = B(x)$, где $A(T) \in \mathbb{F}_{p^s}[T]$ – сепарабельный аддитивный многочлен степени $q = p^t$ для некоторого t, все корни которого лежат в $\mathbb{F}_{p^s}$, а степень многочлена $B(T) \in \mathbb{F}_{p^s}[T]$ не кратна p. Неособая проективная

[1] Работа выполнена при финансовой поддержке премии ECR/2016/000649 Департамента науки и технологии (DST) правительства Индии.

кривая $\mathcal{X}$, ассоциированная с таким функциональным полем, изучалась в [8], где были получены параметры геометрических кодов Гоппы $C_{\mathcal{L}}(D,G)$ на этом функциональном поле в предположении, что $\deg D \geqslant 4g-2$. В настоящей статье мы изучаем одноточечные геометрические коды Гоппы на элементарном абелевом p-расширении поля $\mathbb{F}_{p^s}(x)$ без этого предположения.

Специальным типом элементарных абелевых p-расширений поля $\mathbb{F}_{p^s}(x)$, рассматриваемых в настоящей статье, являются примеры функциональных полей, ассоциированных со слабыми за́мковыми кривыми[2]. Замковые и слабые замковые кривые представляют интерес для целей теории кодирования. Многие известные коды принадлежат классу замковых и слабых замковых кодов. Замковые и слабые замковые кривые и коды на них изучались, в частности, в [9–12]. В [9] рассматривались одноточечные геометрические коды Гоппы, возникающие из замковых и слабых замковых кривых. Были получены границы на минимальное расстояние и обобщенные веса Хэмминга таких кодов. В [10] была выведена граница d^* типа порядковой границы на минимальное расстояние некоторых замковых кодов (т.е. одноточечных геометрических кодов Гоппы по замковым кривым), в частности, связанных с полугруппами, порожденными двумя элементами, и телескопическими полугруппами. В [11] была вычислена граница d_2^* на второй обобщенный вес Хэмминга для некоторых замковых кодов. В [12] изучались геометрические коды Гоппы, по которым можно построить квантовые коды. Особое внимание было уделено семейству замковых и слабых замковых кодов. В [13] были получены новые квантовые коды с хорошими параметрами, построенные по самоортогональным геометрическим кодам Гоппы на функциональных полях, ассоциированных с широким классом кривых. Наша цель в настоящей статье – определить точное значение второго обобщенного веса Хэмминга одноточечных геометрических кодов Гоппы, построенных по элементарным абелевым расширениям поля $\mathbb{F}_{p^s}(x)$.

Статья имеет следующую структуру. В § 2 напоминаются некоторые результаты о конструкции Гоппы линейных кодов и об обобщенных весах Хэмминга линейных кодов. В § 3 изучаются свойства элементарного абелева p-расширения $F/\mathbb{F}_{p^s}$. В § 4 определяются одноточечные геометрические коды Гоппы на этом функциональном поле и изучаются их параметры. В § 5 приведен список значений второго обобщенного веса Хэмминга этих кодов. В § 6 описываются простые условия самодвойственности и квазисамодвойственности таких кодов. В §§ 7, 8 получены примеры квантовых и сверточных кодов по одноточечным геометрическим кодам Гоппы, построенным в § 4. В § 9 получены локально восстанавливаемые коды по функциональным полям $F/\mathbb{F}_{p^s}$.

§ 2. Предварительные сведения

2.1. Геометрические коды Гоппы. Конструкция Гоппы линейных кодов над $\mathbb{F}_{p^s}$ (см. [2, гл. 2]) состоит в следующем.

Пусть $F'/\mathbb{F}_{p^s}$ – алгебраическое функциональное поле рода g'. Пусть $P_1,\dots,P_n$ – попарно различные точки степени 1 поля $F'/\mathbb{F}_{p^s}$. Положим $D' := P_1+\ldots+P_n$, и пусть G – дивизор поля $F'/\mathbb{F}_{p^s}$, такой что $\operatorname{supp}(G)\cap\operatorname{supp}(D')=\varnothing$. Геометрический код Гоппы $C_{\mathcal{L}}(D',G)$, ассоциированный с D' и G, определяется как

$$C_{\mathcal{L}}(D',G) := \{(x(P_1),\dots,x(P_n)) :\ x\in\mathcal{L}(G)\} \subseteq \mathbb{F}_{p^s}^n.$$

Таким образом, $C_{\mathcal{L}}(D',G)$ является $[n,k,d]$-кодом с параметрами $k=\dim(\mathcal{L}(G))-$ $-\dim(\mathcal{L}(G-D'))$ и $d\geqslant n-\deg(G)$.

[2] Замковые кривые (Castle curves) названы в честь замка (крепости) Ла Мота в Медина-дель-Кампо, Испания, на конференции в которой был впервые представлен доклад, посвященный таким кривым – *прим. ред.*

Еще один код, ассоциированный с дивизорами G и D', определяется с помощью локальных компонент дифференциалов Вейля. Код $C_\Omega(D', G) \subseteq \mathbb{F}_{p^s}^n$ определяется как

$$C_\Omega(D', G) := \{(\omega_{P_1}(1), \ldots, \omega_{P_n}(1)) : \omega \in \Omega_{F'}(G - D')\}.$$

Таким образом, $C_\Omega(D', G)$ является $[n, k, d]$-кодом с параметрами $k = i(G-D')-i(G)$ и $d \geqslant \deg(G) - (2g' - 2)$.

Код $C_\Omega(D', G)$ является двойственным к $C_\mathcal{L}(D', G)$ относительно евклидова скалярного произведения на $\mathbb{F}_{p^s}^n$ т.е. $C_\Omega(D', G) = C_\mathcal{L}(D', G)^\perp$. Пусть η – дифференциал Вейля поля F', такой что $\nu_{P_i}(\eta) = -1$ и $\eta_{P_i}(1) = 1$ для $i = 1, \ldots, n$. Тогда $C_\mathcal{L}(D', G)^\perp = C_\Omega(D', G) = C_\mathcal{L}(D', D' - G + (\eta))$.

2.2. Обобщенные веса Хэмминга линейных кодов. Носителем линейного $[n, k]$-кода C над $\mathbb{F}_{p^s}$ называется множество

$$\operatorname{supp}(C) := \{i : x_i \neq 0 \text{ для некоторого } \boldsymbol{x} = (x_1, \ldots, x_n) \in C\}.$$

Для $1 \leqslant \ell \leqslant k$ назовем *ℓ-м обобщенным весом Хэмминга* кода C величину

$$d_\ell(C) := \min\{|\operatorname{supp}(D)| : D \text{ – линейный подкод } C, \text{ такой что } \dim(D) = \ell\}.$$

В частности, первый обобщенный вес Хэмминга кода C – это обычное минимальное расстояние. *Иерархией весов* кода C называется множество $\{d_1(C), \ldots, d_k(C)\}$ обобщенных весов Хэмминга. Обобщенные веса Хэмминга для линейных кодов были введены в [14, 15] и затем независимо в [16]. Мотивацией к изучению этих весов были некоторые приложения в криптографии.

Некоторые свойства обобщенных весов Хэмминга кода C перечислены в следующих теоремах.

Теорема 1 [16, теорема 1]. *Для линейного $[n, k]$-кода C, где $k > 0$, справедливы неравенства*

$$1 \leqslant d_1(C) < d_2(C) < \ldots < d_k(C) \leqslant n.$$

Пусть $\boldsymbol{H}$ – проверочная матрица кода C, и пусть $\boldsymbol{h}_i$, $1 \leqslant i \leqslant n$, – ее векторы-столбцы. Для $I \subseteq \{1, \ldots, n\}$ обозначим через $\langle \boldsymbol{h}_i : i \in I\rangle$ пространство, порожденное этими векторами. Тогда имеет место следующая

Теорема 2 [16, теорема 2]. *Справедливо равенство*

$$d_\ell(C) = \min\{|I| : |I| - \operatorname{rank}(\langle \boldsymbol{h}_i : i \in I\rangle) \geqslant \ell\}.$$

Для геометрического кода Гоппы $C_\mathcal{L}(D', G)$ его ℓ-й обобщенный вес Хэмминга описывает следующая

Теорема 3 [17, следствие 1]. *Пусть $C = C_\mathcal{L}(D', G)$ – код размерности k, и пусть $a := \dim(\mathcal{L}(G - D')) \geqslant 0$. Тогда для любого ℓ, $1 \leqslant \ell \leqslant k$, справедливы равенства*

$$d_\ell(C) = \min\{\deg(D'') : 0 \leqslant D'' \leqslant D', \dim(\mathcal{L}(G - D' + D'')) \geqslant \ell + a\} =$$
$$= \min\{n - \deg(D'') : 0 \leqslant D'' \leqslant D', \dim(\mathcal{L}(G - D'')) \geqslant \ell + a\}.$$

2.3. Расстояния Фенга – Рао числовых полугрупп. Расстояния Фенга – Рао числовых полугрупп были введены в [18]. Вкратце поясним их в этом пункте.

Пусть A – числовая полугруппа. Если для некоторого множества $G \subseteq A$ любой элемент $x \in A$ можно представить в виде линейной комбинации

$$x = \sum_{y \in G} \lambda_y y,$$

где лишь конечное число коэффициентов $\lambda_y \in \mathbb{N} \cup \{0\}$ отличны от нуля, то говорят, что A порождается множеством G. Хорошо известно, что любая числовая полугруппа конечно порождена. Элемент $x \in A$ называется неприводимым, если из равенства $x = a + b$ для $a, b \in A$ следует, что $ab = 0$. Любое порождающее множество содержит множество всех неприводимых элементов, и это множество неприводимых элементов порождает A. Число неприводимых элементов называется *размерностью вложения* полугруппы A. Пронумеруем элементы полугруппы A в порядке возрастания:

$$A = \{\rho_1 = 0 < \rho_2 < \ldots\}.$$

Для заданных $a, b \in \mathbb{Z}$ будем говорить, что a *делит* b, и записывать это в виде

$$a \leqslant_A b, \quad \text{если} \quad b - a \in A.$$

Это бинарное отношение является отношением порядка.

Через $D(y)$ обозначается множество *делителей* элемента y в A, и для заданного множества $M = \{m_1, \ldots, m_t\} \subseteq A$ положим $D(M) = D(m_1, \ldots, m_t) = \bigcup\limits_{i=1}^{t} D(m_i)$.

Определение 1. Пусть A – числовая полугруппа, т.е. подмоноид в $\mathbb{N}$, такой что $|\mathbb{N} \backslash A| < \infty$ и $0 \in A$. Величина $g := |\mathbb{N} \backslash A|$ называется *родом* полугруппы A. Единственный элемент $c \in A$, такой что $c - 1 \notin A$ и $c + u \in A$ для всех $u \in \mathbb{N}$, называется *кондуктором* полугруппы A. (Классическое) расстояние Фенга – Рао полугруппы A задается функцией

$$\delta_{\mathrm{FR}} \colon A \to \mathbb{N}, \quad x \mapsto \delta_{\mathrm{FR}}(x) := \min\{|D(m_1)| : \ m_1 \geqslant x, \ m_1 \in A\}.$$

Имеется несколько хорошо известных результатов о функции δ_{FR} для произвольной числовой полугруппы A. Одним из важных результатов является такой:

$$\delta_{\mathrm{FR}}(x) \geqslant x + 1 - 2g \quad \text{для всех } x \in A, \text{ таких что } x \geqslant c.$$

Следующий результат дает границу на обобщенные веса Хэмминга некоторых кодов через функцию δ_{FR}.

Теорема 4 [18, теорема 46]. *Пусть* $A = \{0 = \rho_1 < \rho_2 < \ldots < \rho_n < \ldots\}$ – *числовая полугруппа с размерностью вложения* 2. *Тогда*

$$d_\ell(C_t) \geqslant \delta_{\mathrm{FR}}(t+1) + \rho_\ell$$

для $\ell = 1, \ldots, k_t$, *где* C_t – *код из массива кодов, определенного в* [19], *а* k_t – *размерность кода* C_t.

§ 3. Элементарные абелевы p-расширения $F/\mathbb{F}_{p^s}$

Пусть $K := \mathbb{F}_{p^s}$, и пусть q – степень числа p. Предположим, что все корни уравнения $T^q + \mu T = 0$ лежат в поле K (выберем s достаточно большим, так чтобы все корни принадлежали K). Обозначим через $\beta_1, \ldots, \beta_q$ корни многочлена $T^q + \mu T$ в K. Пусть m – натуральное число, взаимно простое с p, такое что $m < p^s$. Выберем m различных элементов $\alpha_1, \ldots, \alpha_m \in K$. Положим $f(x) := \prod\limits_{i=1}^{m} (x - \alpha_i)$.

Рассмотрим функциональное поле F/K, заданное уравнением

$$y^q + \mu y = f(x) \in K[x], \tag{1}$$

где $0 \neq \mu \in K$.

Замечание 1. Имеет место неравенство $\min\{q, m\} \geqslant 2$ за исключением случая, когда $m = 1$ и поэтому F/K – поле рациональных функций.

Функциональное поле, заданное уравнением (1), является элементарным абелевым p-расширением поля $K(x)$, как указано в [2]. Некоторые свойства расширения F/K описывает следующая

Лемма 1 [2, с. 232]. *Справедливы следующие утверждения*:

1. *Расширение F/K имеет род $g = (q-1)(m-1)/2$;*
2. *Полюс $P_\infty \in \mathbb{P}_{K(x)}$ функции x в $K(x)$ имеет единственный прообраз $Q_\infty \in \mathbb{P}_F$ степени 1, причем $e(Q_\infty|P_\infty) = q$;*
3. *Дифференциал dx имеет дивизор*

$$(dx) = (2g-2)Q_\infty = ((q-1)(m-1)-2)Q_\infty;$$

4. *Функция x имеет дивизор полюсов $(x)_\infty = qQ_\infty$, а функция y – дивизор полюсов $(y)_\infty = mQ_\infty$;*
5. *Пусть $r \geqslant 0$. Тогда элементы $x^i y^j$, где*

$$0 \leqslant i, \quad 0 \leqslant j \leqslant q-1, \quad qi + mj \leqslant r,$$

образуют базис пространства $\mathcal{L}(rQ_\infty)$ над K;

6. *Точками F/K степени 1 являются точки P_{α_i,β_j}, где $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant q$.*

Пусть Q – точка степени 1 функционального поля F'/K'. Целое число $\ell \geqslant 0$ называется порядком полюса в точке Q, если существует элемент $z \in F'$, такой что $(z)_\infty = \ell Q$. Пусть $p_1 < p_2 < \ldots$ – последовательность порядков полюсов в точке Q (т.е. p_a – a-й порядок полюса в Q); таким образом, $\dim(\mathcal{L}(p_a Q)) = a$, так что $p_1 = 0$. Полугруппа Вейерштрасса H точки Q – это множество порядков полюсов в Q.

Имеется следующий результат о полугруппе Вейерштрасса H точки Q_∞.

Лемма 2 [2,20]. *Полугруппа Вейерштрасса H точки Q_∞ порождена числами m и q, т.е. $H = \langle q, m\rangle$. Наибольший пробел в точке Q_∞ равен $2g-1$, и H является симметрической числовой полугруппой.*

Пусть $\mathcal{X}'$ – неособая проективная кривая, ассоциированная с полем F/K. Тогда $K(\mathcal{X}') \cong F$. Далее мы покажем, что $\mathcal{X}'$ – слабая замковая кривая.

Кривой с отмеченной точкой над полем $\mathbb{F}_{p^s}$ называется пара $(\mathcal{X}, Q)$, где $\mathcal{X}$ – кривая, определенная над $\mathbb{F}_{p^s}$, а $Q \in \mathcal{X}(\mathbb{F}_{p^s})$ – рациональная точка.

Определение 2 [9, определение 2.1]. Кривая с отмеченной точкой $(\mathcal{X}, Q)$ над $\mathbb{F}_{p^s}$ называется слабой замковой кривой, если

- Полугруппа Вейерштрасса H точки Q симметрическая;
- Существуют морфизм $\varphi\colon \mathcal{X} \to \mathbb{P}^1$, такой что $\mathrm{div}_\infty(\varphi) = hQ$, а также элементы $\gamma_1, \ldots, \gamma_b \in \mathbb{F}_{p^s}$, такие что для всех $i = 1, \ldots, b$ выполнено $\varphi^{-1}(\gamma_i) \subseteq \mathcal{X}(\mathbb{F}_{p^s})$, причем $\#\varphi^{-1}(\gamma_i) = h$.

Лемма 3. $(\mathcal{X}', Q_\infty)$ – *слабая замковая кривая.*

Доказательство. По лемме 2 полугруппа H точки Q_∞ симметрическая. Положим $\varphi := x$, т.е.

$$\varphi\colon \mathcal{X}' \to \mathbb{P}^1, \quad P \mapsto [x(P) : 1].$$

Тогда $(x)_\infty = qQ_\infty$. Для $1 \leqslant i \leqslant m$ имеем

$$\varphi^{-1}(\alpha_i) = \{(\alpha_i, \beta_1), \ldots, (\alpha_i, \beta_q)\} \subseteq \mathcal{X}'(K).$$

(Из [21, теорема 3.1.15] следует, что K-рациональные точки кривой $\mathcal{X}'$ соответствуют точкам степени 1 поля F. Обозначим через P_{α_i} нуль функции $(x-\alpha_i)$ в $K(x)$, т.е. точку степени 1, соответствующую K-рациональной точке α_i на $\mathbb{P}^1$. Тогда согласно [2, с. 232] имеется ровно q точек P_{α_i,β_j}, $1 \leqslant j \leqslant q$, степени 1, лежащих над P_{α_i}, для каждого i, $1 \leqslant i \leqslant m$. Таким образом, P_{α_i,β_j} соответствует (α_i, β_j).) ▲

§ 4. Геометрические коды Гоппы на F/K

В [2] исследовались одноточечные геометрические коды Гоппы на эрмитовом функциональном поле. В [9] были определены одноточечные геометрические коды Гоппы на слабых замковых кривых (т.е. слабые замковые коды) и изучены их параметры. В этом параграфе мы определяем одноточечные геометрические коды Гоппы на F/K, следуя [9], и находим их параметры, используя идеи из [2, 5, 9].

Определение 3. Для $r \in \mathbb{Z}$ положим

$$C_r := C_{\mathcal{L}}(D, rQ_\infty),$$

где

$$D := \sum_{i=1}^{m} \sum_{j=1}^{q} P_{\alpha_i,\beta_j}.$$

Тогда C_r – код длины $N := qm$ над полем K. Для $r < 0$ имеем $\mathcal{L}(rQ_\infty) = \{0\}$, и поэтому $C_r = \{(0, \ldots, 0)\}$. Для $r > N + (2g-2) = 2qm - q - m - 1$ имеем $\dim(C_r) = N$, и поэтому $C_r = K^N$. Таким образом, остается изучить коды C_r для $0 \leqslant r \leqslant 2qm - q - m - 1$.

4.1. Код, двойственный к C_r. В [9, предложение 3.1] были описаны коды, двойственные к слабым замковым кодам. В этом пункте представлено подробное доказательство двойственности для C_r.

Пусть $f(x)$ и D те же, что и в §§ 3, 4. Рассмотрим дифференциал $\eta := \dfrac{f'(x)}{f(x)}\,dx$, тогда $\nu_P(\eta) = -1$ и $\operatorname{res}_P(\eta) = 1$ для всех точек $P \in \operatorname{supp}(D)$. Таким образом, получаем

$$C_{\mathcal{L}}(D, rQ_\infty)^\perp = C_{\mathcal{L}}(D, D + (\eta) - rQ_\infty).$$

Предложение 1. *Код, двойственный к C_r, имеет вид*

$$C_r^\perp = \bar{a} \star C_{2qm-q-m-1-r},$$

где $(\bar{a})^{-1} = \big((f'(x))(P_{\alpha_1,\beta_1}), \ldots, (f'(x))(P_{\alpha_m,\beta_q})\big) \in (K^*)^N$, *а через* $\star$ *обозначено покоординатное произведение в* K^N.

Доказательство. Имеем

$$\begin{aligned}
&C_r^\perp = C_{\mathcal{L}}(D, D + (\eta) - rQ_\infty) = \\
&= C_{\mathcal{L}}(D, D + (f'(x)) - (f(x)) + (dx) - rQ_\infty) = \\
&= C_{\mathcal{L}}(D, D + (f'(x)) - D + qmQ_\infty + (2g-2)Q_\infty - rQ_\infty) = \\
&= C_{\mathcal{L}}(D, (f'(x)) + (qm + 2g - 2 - r)Q_\infty) = \\
&= \bar{a} \star C_{\mathcal{L}}(D, (2qm - q - m - 1 - r)Q_\infty) = \\
&= \bar{a} \star C_{2qm-q-m-1-r}. \quad ▲
\end{aligned}$$

4.2. Параметры кода C_r. В [9, предложения 3.4, 3.6 и 3.8] были найдены размерность и минимальное расстояние кодов C_r для некоторых определенных значений r. В этом пункте мы повторим эти известные результаты и определим параметры кодов C_r для остальных значений r.

Для параметров геометрических кодов Гоппы $C_{\mathcal{L}}(D', G)$ имеется следующий результат.

Теорема 5 [2, теорема 2.2.2 и следствие 2.2.3]. *Код $C_{\mathcal{L}}(D', G)$ является $[n, k, d]$-кодом с параметрами*

$$k = \dim(\mathcal{L}(G)) - \dim(\mathcal{L}(G - D')) \quad и \quad d \geqslant n - \deg(G).$$

Если $\deg(G) < n$, *то* $k = \dim(\mathcal{L}(G))$.

Перейдем к определению параметров кодов C_r.

Рассмотрим множество H порядков полюсов в точке Q_∞ (т.е. полугруппу Вейерштрасса точки Q_∞). Для $b \geqslant 0$ положим

$$H(b) := \{u \in H : \ u \leqslant b\}.$$

Тогда $|H(b)| = \dim(\mathcal{L}(bQ_\infty))$. По лемме 1 имеем

$$H(b) = \{u \leqslant b : \ u = iq + jm, \text{ где } i \geqslant 0 \text{ и } 0 \leqslant j \leqslant q - 1\}.$$

Отсюда

$$|H(b)| = |\{(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0 : \ j \leqslant q - 1 \text{ и } iq + jm \leqslant b\}|.$$

Теорема 6. *Пусть* $0 \leqslant r \leqslant 2qm - q - m - 1$. *Тогда справедливы следующие утверждения*:

1. *Имеет место равенство* $\dim(C_r) = \dim(\mathcal{L}(rQ_\infty)) - \dim(\mathcal{L}((r - qm)Q_\infty))$. *Кроме того*,
   (a) *Для* $0 \leqslant r < qm$ *имеем* $\dim(C_r) = |H(r)|$;
   (b) *Положим* $s := 2qm - q - m - 1 - r$. *Для* $qm \leqslant r \leqslant 2qm - q - m - 1$ *имеем*
   $$\dim(C_r) = qm - |H(s)|;$$
   (c) *Для* $qm - q - m - 1 < r < qm$ *имеем* $\dim(C_r) = r + 1 - \dfrac{(q-1)(m-1)}{2}$.
2. *Минимальное расстояние* $d(C_r)$ *кода* C_r *удовлетворяет неравенству*
   $$d(C_r) \geqslant qm - r.$$
   Если $r = qb$, *где* $0 \leqslant b < m$, *а также если* $r = cm$, *где* $0 \leqslant c < q$, *то* $d(C_r) = qm - r$. *Кроме того, если* $r \geqslant qm - q - m$, *то* C_r *не является МДР-кодом.*

Доказательство. 1. Так как $D \sim qmQ_\infty$, то по теореме 5 имеем

$$\dim(C_r) = \dim(\mathcal{L}(rQ_\infty)) - \dim(\mathcal{L}((r - qm)Q_\infty)),$$

и если $0 \leqslant r < qm$, то $\dim(C_r) = \dim(\mathcal{L}(rQ_\infty)) = |H(r)|$.

Для $qm \leqslant r \leqslant 2qm - q - m - 1$ имеем $0 \leqslant s < qm$. Таким образом, из предложения 1 следует, что

$$\dim C_r = qm - \dim C_r^\perp = qm - \dim C_s = qm - |H(s)|.$$

Если $qm - q - m - 1 < r < qm$ (т.е. $2g - 2 < \deg(rQ_\infty) < N$), то по теореме Римана – Роха имеем

$$\dim(C_r) = \dim(\mathcal{L}(rQ_\infty)) = \deg(rQ_\infty) + 1 - g = r + 1 - \frac{(q-1)(m-1)}{2}.$$

2. Неравенство $d(C_r) \geqslant qm - r$ непосредственно вытекает из теоремы 5. Если $r = qb$, где $0 \leqslant b < m$, выберем b различных элементов из множества $\{\alpha_1, \ldots, \alpha_m\}$. Назовем эти элементы $\gamma_1, \ldots, \gamma_b$. Тогда элемент

$$z_1 := \prod_{j=1}^{b} (x - \gamma_j) \in \mathcal{L}(rQ_\infty)$$

имеет ровно $qb = r$ различных нулей в D. Вес соответствующего слова кода C_r равен $qm - r$. Следовательно, $d(C_r) = qm - r$.

Аналогично, если $r = cm$, где $0 \leqslant c < q$, то выберем c различных элементов из множества $\{\beta_1, \ldots, \beta_q\}$. Назовем их $\tau_1, \ldots, \tau_c$. Тогда элемент

$$z_2 := \prod_{j=1}^{c} (y - \tau_j) \in \mathcal{L}(rQ_\infty)$$

имеет в точности $cm = r$ различных нулей в D. Вес соответствующего слова кода C_r равен $qm - r$. Следовательно, $d(C_r) = qm - r$.

Если $r = qb$ и C_r является МДР-кодом, то из равенства $d(C_r) = qm - \dim(C_r) + 1$ следует $g = 0$, что невозможно. Аналогично для $r = cm$. ▲

В следующей теореме мы определим минимальное расстояние кода C_r для значений $qm \leqslant r \leqslant 2qm - q - m - 1$. Используя идеи из [5] и теорему 2, приходим к следующему результату.

Теорема 7. *Пусть $m > q$. Для $qm \leqslant r \leqslant 2qm - q - m - 1$ имеем $0 \leqslant r^\perp := 2qm - q - m - 1 - r \leqslant qm - q - m - 1$. Пусть $t^\perp \leqslant r^\perp$ – наибольшее целое число, такое что $t^\perp$ является порядком полюса в точке Q_∞, т.е. $t^\perp = aq + bm$, где $0 \leqslant a \leqslant m - 2$ и $0 \leqslant b \leqslant q - 1$. Тогда минимальное расстояние кода C_r имеет вид*

$$d(C_r) = a + 2.$$

Доказательство. Пусть $\boldsymbol{H}$ – проверочная матрица кода C_r. По лемме 1 множество $\{1, x, y, \ldots, x^a, x^{a-1}y, \ldots, y^b\}$ является базисом пространства $\mathcal{L}(t^\perp Q_\infty)$. Выберем элемент $\beta \in K$, такой что $\beta^q + \mu\beta = 0$. Пусть $\boldsymbol{H}_1$ – подматрица матрицы $\boldsymbol{H}$ со столбцами, соответствующими точкам $P_{\alpha_1,\beta}, \ldots, P_{\alpha_{a+2},\beta}$. Используя преобразования строк, представим $\boldsymbol{H}_1$ в следующем виде:

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \ldots & \alpha_{a+2} \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \ldots & \alpha_{a+2}^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha_1^a & \alpha_2^a & \alpha_3^a & \ldots & \alpha_{a+2}^a \\ 0 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Здесь $\operatorname{rank}(\boldsymbol{H}_1) = a + 1$, и $\boldsymbol{H}_1$ имеет $a + 2$ столбцов, так что столбцы матрицы $\boldsymbol{H}_1$ линейно зависимы. Поэтому $d(C_r) \leqslant a + 2$.

С другой стороны, выберем любые $a + 1$ различных столбцов из $\boldsymbol{H}$. Назовем эту матрицу $\boldsymbol{H}_2$. Поскольку каждый столбец матрицы $\boldsymbol{H}$ соответствует точке $P_{\alpha,\beta}$ степени 1, переупорядочим столбцы матрицы $\boldsymbol{H}_2$ в соответствии со значениями α

следующим образом:

$$\begin{array}{llll} P_{\alpha_1,\beta_{1,1}}, & P_{\alpha_1,\beta_{1,2}}, & \ldots, & P_{\alpha_1,\beta_{1,w_1}} \\ P_{\alpha_2,\beta_{2,1}}, & P_{\alpha_2,\beta_{2,2}}, & \ldots, & P_{\alpha_1,\beta_{2,w_2}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & \\ P_{\alpha_\gamma,\beta_{\gamma,1}} & P_{\alpha_\gamma,\beta_{\gamma,2}} & \ldots, & P_{\alpha_\gamma,\beta_{\gamma,w_\gamma}}, \end{array}$$

где α_i попарно различны, $w_1+w_2+\ldots+w_\gamma = a+1$, причем $w_1 \geqslant w_2 \geqslant \ldots \geqslant w_\gamma \geqslant 1$. Для $0 \leqslant j_i \leqslant w_i - 1$, $1 \leqslant i \leqslant \gamma$, элемент $x^{i-1}y^{j_i}$ принадлежит базису $\mathcal{L}(t^\perp Q_\infty)$. Перепишем эти базисные элементы в виде

$$\begin{array}{lllll} 1, & y, & y^2, & \ldots, & y^{w_1-1} \\ x, & xy, & xy^2, & \ldots, & xy^{w_2-1} \\ x^2, & x^2y, & x^2y^2, & \ldots, & x^2y^{w_3-1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & & \\ x^{\gamma-1}, & x^{\gamma-1}y, & x^{\gamma-1}y^2, & \ldots, & x^{\gamma-1}y^{w_\gamma-1}. \end{array}$$

Теперь выделим из $\boldsymbol{H}_2$ подматрицу $\boldsymbol{H'}$ размера $(a+1)\times(a+1)$ таким образом, чтобы каждая строка соответствовала вышеуказанным функциям в заданном порядке. Иначе говоря, $\boldsymbol{H'} = [\boldsymbol{H'}_{i,j}]$, $i,j = 1,2,\ldots,\gamma$, где матрица $\boldsymbol{H'}_{i,j}$ размера $w_i \times w_j$ имеет вид $\boldsymbol{H'}_{i,j} = \alpha_j^{i-1}\boldsymbol{B}_{i,j}$, где

$$\boldsymbol{B}_{i,j} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ \beta_{j,1} & \beta_{j,2} & \beta_{j,3} & \ldots & \beta_{j,w_j} \\ \beta_{j,1}^2 & \beta_{j,2}^2 & \beta_{j,3}^2 & \ldots & \beta_{j,w_j}^2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & & & \\ \beta_{j,1}^{w_i-1} & \beta_{j,2}^{w_i-1} & \beta_{j,3}^{w_i-1} & \ldots & \beta_{j,w_j}^{w_i-1} \end{bmatrix}.$$

Тогда согласно [5, леммы 2 и 3] имеем

$$\det(\boldsymbol{H'}) = \left(\prod_{i=1}^{\gamma}\det(\boldsymbol{B}_{i,i})\right)\left(\prod_{j=2}^{\gamma}\rho_j^{w_j}\right),$$

где

$$\rho_j = \prod_{i=1}^{j-1}(\alpha_j - \alpha_i), \quad j = 2,3,\ldots,\gamma,$$

и любые $a+1$ столбцов матрицы $\boldsymbol{H}$ линейно независимы над K. Следовательно, $d(C_r) \geqslant a+2$. ▲

§ 5. Обобщенные веса Хэмминга кода C_r

В [9, предложения 3.7, 3.8] были получены границы на обобщенные веса Хэмминга кодов C_r с помощью понятий гональности и порядковых границ. Но в общем случае вычисление гональности является трудной задачей. В этом параграфе мы установим точные значения обобщенных весов Хэмминга, в частности, второго обобщенного веса Хэмминга кода C_r в нескольких случаях.

Следуя идеям из [17], получаем следующее утверждение.

Лемма 4. *Пусть $r \leqslant qm$ – порядок полюса в точке Q_∞. Тогда $r = iq + jm$, где $i \geqslant 0$ и $0 \leqslant j \leqslant q-1$. Если либо $i = 0$, либо $j = 0$, то существует дивизор $0 \leqslant D' \leqslant D$, такой что $rQ_\infty \sim D'$.*

Доказательство. Для $i = 0$ и $j = 0$ подходит дивизор $D' = 0$. Далее, если $i = 0$ и $j \neq 0$, то $r = jm$. Выберем j элементов из $\beta_1, \ldots, \beta_q$ и обозначим их через $\tau_1, \ldots, \tau_j$. Положим

$$z := \prod_{t=1}^{j} (y - \tau_t).$$

Тогда $(z) = D' - rQ_\infty$, откуда $D' \sim rQ_\infty$. Для случая $j = 0$ и $i \neq 0$ доказательство аналогично. ▲

Определение 4. Будем говорить, что натуральное число $r \leqslant qm$ обладает свойством $(*)$, если r является порядком полюса в точке Q_∞, $r = iq + jm$ для $i \geqslant 0$, $0 \leqslant j \leqslant q - 1$ и либо $i = 0$, либо $j = 0$.

Теорема 8. *Если для* $1 \leqslant \ell \leqslant \dim(C_r)$ *хотя бы одно из чисел* $r - p_\ell$ *и* $qm - r + p_\ell$ *обладает свойством* $(*)$, *то*

$$d_\ell(C_r) \leqslant qm - r + p_\ell.$$

Доказательство. Если $r - p_\ell$ обладает свойством $(*)$, то по лемме 4 существует дивизор D', такой что $0 \leqslant D' \leqslant D$ и $(r - p_\ell)Q_\infty \sim D'$. Таким образом,

$$\dim(\mathcal{L}(rQ_\infty - D')) = \dim(\mathcal{L}(p_\ell Q_\infty)) = \ell.$$

Тогда из теоремы 3 следует, что

$$d_\ell(C_r) \leqslant qm - r + p_\ell.$$

Если же $qm - r + p_\ell$ обладает свойством $(*)$, то снова найдется дивизор D'', такой что $0 \leqslant D'' \leqslant D$ и $(qm - r + p_\ell)Q_\infty \sim D''$. При этом $qmQ_\infty \sim D$, откуда $D - rQ_\infty + p_\ell Q_\infty \sim D''$. Поэтому $D' := D - D'' \sim (r - p_\ell)Q_\infty$. Следовательно,

$$d_\ell(C_r) \leqslant qm - r + p_\ell. \quad ▲$$

Из теоремы 8 немедленно вытекает

Следствие 1. *Если для* $1 \leqslant r < qm$ *хотя бы одно из чисел* r *и* $qm - r$ *обладает свойством* $(*)$, *то код* C_r *имеет минимальное расстояние* $d(C_r) = qm - r$.

Замечание 2. Для $k = \dim(C_r)$ справедливо $d_k(C_r) = qm$.

По лемме 2 полугруппа Вейерштрасса $H = \langle q, m \rangle$ точки Q_∞ является числовой полугруппой с размерностью вложения 2. Таким образом, из теоремы 4 получаем следующие результаты о втором обобщенном весе Хэмминга кода C_r.

Теорема 9. *Пусть* $m > q$. *Для* $r < qm$ *справедливо неравенство*

$$d_2(C_r) \geqslant qm - r + q.$$

Теорема 10. *Пусть* $m > q$. *Если для* $r < qm - 1$ *хотя бы одно из чисел* $r - q$ *и* $qm - r + q$ *обладает свойством* $(*)$, *то*

$$d_2(C_r) = qm - r + q.$$

Доказательство. Применяя теорему 8, получаем $d_2(C_r) \leqslant qm - r + q$.

С другой стороны, так как $\rho_2 = q$ и код, двойственный к C_r, образует кодовый массив (подробнее см. в [19]), то из теоремы 4 и предложения 1 получаем

$$d_2(C_r) = d_2(C^\perp_{2qm-q-m-1-r}) \geqslant \delta_{\mathrm{FR}}(2qm - q - m - r) + q.$$

Так как $r < qm$, то $2qm - q - m - r \geqslant 2g = qm - q - m + 1$, и поэтому

$$d_2(C_r) \geqslant \delta_{\mathrm{FR}}(2qm - q - m - r) + q \geqslant$$
$$\geqslant 2qm - q - m - r + 1 - (qm - q - m + 1) + q = qm - r + q,$$

что и требовалось. ▲

Теорему 10 можно обобщить на все значения ℓ, $1 \leqslant \ell \leqslant \dim(C_r)$.

Теорема 11. *Для $r < qm - 1$ и $1 \leqslant \ell \leqslant \dim(C_r)$, если хотя бы одно из чисел $r - p_\ell$ и $qm - r + p_\ell$ обладает свойством $(*)$, то*

$$d_\ell(C_r) = qm - r + p_\ell.$$

Доказательство. По теореме 8 имеем

$$d_\ell(C_r) \leqslant qm - r + p_\ell.$$

Обратное неравенство следует из доказательства теоремы 10. ▲

Следующий результат был доказан в [17].

Теорема 12 [17, предложение 4]. *Пусть $C_{\mathcal{L}}(D', G)$ – код размерности k с обильностью $\dim(\mathcal{L}(G - D')) =: a \geqslant 0$. Если существует точка Q степени 1, не входящая в D', и $C_{\mathcal{L}}(D', G - p_{\ell+a}Q) \neq \{0\}$, где p_ℓ – ℓ-й порядок полюса в точке Q, то для любого ℓ, $1 \leqslant \ell \leqslant k$,*

$$d_\ell(C_{\mathcal{L}}(D', G)) \leqslant d_1(C_{\mathcal{L}}(D', G - p_{\ell+a}Q)).$$

Применяя теоремы 9 и 12, получаем следующий результат.

Теорема 13. *Пусть $m > q$. Для $r < qm$, если числа $r - q$ и $qm - r + q$ не обладают свойством $(*)$, то*

$$qm - r + q \leqslant d_2(C_r) \leqslant qm - \overline{(r - q)},$$

где $\overline{(r-q)}$ – наибольший порядок полюса, меньший или равный $(r-q)$ и обладающий свойством $(*)$.

Теорема 14. *Для $1 \leqslant \ell \leqslant \dim C_r$ и $r < qm$, если числа $r - p_\ell$ и $qm - r + p_\ell$ не обладают свойством $(*)$, то*

$$d_\ell(C_r) \leqslant qm - \overline{(r - p_\ell)},$$

где $\overline{(r-p_\ell)}$ – наибольший порядок полюса, меньший или равный $(r - p_\ell)$ и обладающий свойством $(*)$.

Теперь перейдем к вычислению второго обобщенного веса Хэмминга кода C_r в случае $qm \leqslant r$. В доказательствах будем использовать следующий результат из [17].

Теорема 15 [17, предложение 6]. *Пусть $C = C_{\mathcal{L}}(D', G)$ – код размерности k, и пусть $\dim(\mathcal{L}(G - D')) =: a > 0$. Тогда для $1 \leqslant \ell \leqslant k$ справедливо неравенство $d_\ell(C) \leqslant \deg(D'')$ для любого эффективного дивизора $D'' \leqslant D$, такого что $\dim(\mathcal{L}(D'')) > \ell$.*

Теорема 16. *Пусть $m > q$. Если $qm \leqslant r \leqslant 2qm - q - m - 1$, то*

$$d_2(C_r) \leqslant \min\{2q, m\}.$$

Доказательство. Так как $p_3 = \min\{2q, m\}$, то по лемме 4 существует дивизор D', такой что $0 \leqslant D' \leqslant D$ и $p_3 Q_\infty \sim D'$. Тогда $\dim(\mathcal{L}(D')) = \dim(\mathcal{L}(p_3 Q_\infty)) = 3$. Отсюда $d_2(C_r) \leqslant \deg D' = p_3 = \min\{2q, m\}$. ▲

Теорема 17. *Пусть* $m > 2q$, *и пусть* $qm \leqslant r \leqslant 2qm - q - m - 1$. *Тогда* $0 \leqslant r^{\perp} := 2qm - q - m - 1 - r \leqslant qm - q - m - 1$. *Пусть* $t^{\perp} \leqslant r^{\perp}$ – *наибольшее целое число, такое что* $t^{\perp}$ *является порядком полюса в точке* Q_{∞}, *т.е.* $t^{\perp} = aq + bm$, *где* $0 \leqslant a \leqslant m - 3$ *и* $0 \leqslant b \leqslant q - 1$. *Тогда*

$$d_2(C_r) = a + 3.$$

Доказательство. Пусть $\boldsymbol{H}$ – проверочная матрица кода C_r над полем K. Выберем элемент $\beta \in K$, такой что $\beta^q + \mu\beta = 0$. Множество $\{1, x, y, \ldots, x^a, x^{a-1}y, \ldots, y^b\}$ является базисом пространства $\mathcal{L}(t^{\perp}Q_{\infty})$. Пусть $\boldsymbol{H}_1$ – подматрица матрицы $\boldsymbol{H}$ со столбцами, соответствующими точкам $P_{\alpha_1,\beta}, \ldots, P_{\alpha_{a+3},\beta}$ (это возможно, так как $a + 3 \leqslant m$). Применяя преобразования строк, приведем $\boldsymbol{H}_1$ к следующему виду:

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \ldots & \alpha_{a+3} \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \ldots & \alpha_{a+3}^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha_1^a & \alpha_2^a & \alpha_3^a & \ldots & \alpha_{a+3}^a \\ 0 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Здесь $\operatorname{rank}(\boldsymbol{H}_1) = a + 1$, и $\boldsymbol{H}_1$ имеет $a + 3$ столбцов. Поэтому по теореме 2 имеем $d_2(C_r) \leqslant a + 3$.

С другой стороны, из теоремы 1 имеем $d_2(C_r) \geqslant d_1(C_r) + 1 = (a + 2) + 1 = a + 3$, что завершает доказательство. ▲

Пример 1. Для $p = 2$ и $K = \mathbb{F}_4$ рассмотрим функциональное поле $F = K(x, y)$, заданное уравнением

$$y^2 + \omega y = x(x - 1)(x - \omega),$$

где ω – примитивный элемент поля K. Здесь $q = 2$, $m = 3$ и род $g = 1$. Список значений длины, размерности, минимального расстояния и второго обобщенного веса Хэмминга кода C_r приведен в следующей таблице:

r	N	$\dim(C_r)$	$d(C_r)$	$d_2(C_r)$
1	6	1	6	–
2	6	2	4	6
3	6	3	3	$\geqslant 5$ и $\leqslant 6$
4	6	4	2	4
5	6	5	$\geqslant 1$	3
6	6	5	2	$\leqslant 4$

§ 6. Условия квазисамодвойственности и самодвойственности кодов

Линейный код C называется самодвойственным, если $C = C^{\perp}$, где $C^{\perp}$ – код, двойственный к C относительно евклидова скалярного произведения на $\mathbb{F}_{p^s}^n$. Самодвойственные коды образуют важный класс линейных кодов. В этом параграфе приводится простой критерий самодвойственности геометрических кодов Гоппы на F/K.

Имеется следующий результат, доказанный в [22]. Но прежде чем сформулировать его, введем следующее определение для произвольного алгебраического функционального поля F'/K' рода g'.

Определение 5 [22, определение 2.15]. Выберем n точек $P_1, \ldots, P_n$ степени 1 поля F' и положим $D' := P_1 + \ldots + P_n$. Два дивизора G и H называются эквивалентными относительно D', если существует элемент $u \in F'$, такой что $H = G + (u)$ и $u(P_i) = 1$ для всех $i = 1, \ldots, n$.

Предложение 2 [22, следствие 4.15]. *Пусть $n > 2g' + 2$. Пусть G и H – два дивизора одинаковой степени m' в F'. Если $C_{\mathcal{L}}(D', G)$ не равно ни 0, ни $(K')^n$ и при этом $2g' - 1 < m' < n - 1$, то $C_{\mathcal{L}}(D', G) = C_{\mathcal{L}}(D', H)$ тогда и только тогда, когда G и H эквивалентны относительно D'.*

Теорема 18. *Если $qm - q - m + 1 \leqslant r \leqslant qm - 2$, то код C_r квазисамодвойствен тогда и только тогда, когда $r = (2qm - q - m - 1)/2$.*

Доказательство. Если $qm - q - m + 1 \leqslant r \leqslant qm - 2$, то по предложению 2 код C_r квазисамодвойствен тогда и только тогда, когда $r = (2qm - q - m - 1)/2$. ▲

Пусть G – дивизор поля F, такой что $\deg(G) = \dfrac{2qm - q - m - 1}{2}$. Очевидно, что $qm > qm - q - m + 3$ и $qm - q - m < \deg(G) < qm - 1$. Положим $H := D + (\eta) - G$, где D – такой дивизор, как в § 4. Тогда $\deg(G) = \deg(H)$. Условие самодвойственности кода $C_{\mathcal{L}}(D, G)$ дает следующая

Теорема 19. *Код $C_{\mathcal{L}}(D, G)$ самодвойствен тогда и только тогда, когда дивизор $2G$ эквивалентен $(f'(x)) + (2qm - q - m - 1)Q_\infty$ относительно D.*

Доказательство. По предложению 2

$$
\begin{aligned}
&C_{\mathcal{L}}(D, G) = C_{\mathcal{L}}(D, D + (\eta) - G) \Leftrightarrow \\
&\Leftrightarrow G = D + (\eta) - G + (u) \text{ для некоторого } u \in F, \text{ такого что } u(P) = 1 \\
&\quad \text{для каждой точки } P \in \operatorname{supp}(D) \Leftrightarrow \\
&\Leftrightarrow (u) + (\eta) = 2G - D \Leftrightarrow \\
&\Leftrightarrow (u) + (f'(x)) + (dx) - (f(x)) = 2G - D \Leftrightarrow \\
&\Leftrightarrow (u) + (f'(x)) + [(q-1)(m-1) - 2]Q_\infty - D + qmQ_\infty = 2G - D \Leftrightarrow \\
&\Leftrightarrow (f'(x)) = 2G - (2qm - q - m - 1)Q_\infty - (u). \quad ▲
\end{aligned}
$$

Пример 2. Пусть $p = 2$ и $K = \mathbb{F}_4$. Пусть ω – примитивный элемент поля $\mathbb{F}_4$. Рассмотрим поле $F = K(x, y)$, где

$$y^2 + y = x(x - 1)(x - \omega).$$

Тогда все корни многочлена $T^2 + T$ лежат в K. Расширение F/K имеет род $g = 1$. Положим

$$f(x) := x(x - 1)(x - \omega).$$

Пусть P_0, P_1 и P_ω – нули элементов x, $(x - 1)$ и $(x - \omega)$ в $K(x)$ соответственно. Тогда каждая из точек P_0, P_1 и P_ω имеет ровно два прообраза в F. Аналогично, нуль функции $(x - \omega^2)$, который мы обозначим через P_{ω^2}, имеет два прообраза в F, скажем, Q_1 и Q_2. Положим $D := (f(x))_0$, и пусть G – дивизор поля F, эквивалентный $Q_1 + Q_2 + Q_\infty$ относительно D. Тогда код $C_{\mathcal{L}}(D, G)$ самодвойствен. И наоборот, если $C_{\mathcal{L}}(D, G)$ – самодвойственный код, то дивизор $2G$ эквивалентен $2(Q_1 + Q_2 + Q_\infty)$ относительно D.

§ 7. Квантовые коды по одноточечным геометрическим кодам Гоппы на $F/\mathbb{F}_{p^s}$

В этом параграфе строятся квантовые коды на основе одноточечных геометрических кодов Гоппы на $F/\mathbb{F}_{p^s}$. Вначале дадим краткое введение в квантовые коды.

Пусть q_0 – степень простого числа. Через $V_n := (\mathbb{C}^{q_0})^{\otimes n}$ обозначим n-ю тензорную степень q_0-мерного гильбертова пространства $\mathbb{C}^{q_0}$. Квантовым $[[n,k,d]]_{q_0}$-кодом называется q_0^k-мерное векторное подпространство пространства V_n с минимальным расстоянием d. Связь между квантовыми кодами и классическими линейными кодами была установлена в работе [23], после которой было построено много классов квантовых кодов на основе классических кодов, исправляющих ошибки.

Граница Синглтона для квантовых кодов утверждает, что для всякого квантового $[[n,k,d]]_{q_0}$-кода справедливо неравенство $2d \leqslant n-k+2$. Квантовый дефект Синглтона определяется как $\delta^Q := n-k-2d+2 \geqslant 0$, а относительным квантовым дефектом Синглтона называется $\Delta^Q := \delta^Q/n$. Если $\delta^Q = 0$, то код называется квантовым МДС-кодом.

Следующая лемма описывает конструкцию квантовых кодов по классическим линейным кодам.

Лемма 5 [24, лемма 17(a)]. *Пусть C_1 и C_2 – два линейных кода с параметрами $[n,k_i,d_i]_{q_0}$, $i=1,2$, и пусть $C_1 \subset C_2$. Тогда существует $[[n,k_2-k_1,d]]_{q_0}$-код с расстоянием $d = \min\{\mathrm{wt}(c) :\ c \in (C_2 \setminus C_1) \cup (C_1^\perp \setminus C_2^\perp)\}$, где $\mathrm{wt}(c)$ – вес слова c.*

Применим лемму 5 для получения квантовых кодов по одноточечным геометрическим кодам Гоппы, построенным в § 4.

Предложение 3. *Пусть m и q $(= p^{k'})$ такие, как в § 3. Пусть a,b – натуральные числа, такие что $qm-q-m-1 < a < b < qm$. Тогда существует $[[qm, b-a, d]]_{p^s}$-код с расстоянием $d \geqslant \min\{qm-b, a-qm+q+m+1\}$.*

Кроме того, если $m = q+x$ для некоторого $x \in \mathbb{N}$ и при этом $a = cq$ для некоторого натурального c, так что $qm-b \leqslant a-qm+q+m+1$, то относительный квантовый дефект Синглтона $\Delta^Q_{k'}$ получаемых квантовых кодов удовлетворяет условию

$$\lim_{k'\to\infty} \Delta^Q_{k'} = 0.$$

Доказательство. Пусть $C_1 := C_a$ и $C_2 := C_b$ такие, как в § 4. Тогда код C_1 имеет параметры $[qm,\ a+1-\dfrac{(q-1)(m-1)}{2},\ d_1 \geqslant qm-a]_{p^s}$, а C_2 – параметры $[qm,\ b+1-\dfrac{(q-1)(m-1)}{2},\ d_2 \geqslant qm-b]_{p^s}$. При этом $C_1 \subset C_2$. Тогда из леммы 5 следует, что существует $[[qm, b-a, d]]_{p^s}$-код с расстоянием $d \geqslant \min\{qm-b, a-qm+q+m+1\}$.

Далее, если $m = q+x$ и $a = cq$ таковы, что $qm-b \leqslant a-qm+q+m+1$, то

$$\Delta^Q_{k'} = \frac{qm-b+a-2d+2}{qm} \leqslant \frac{qm-b+a-2qm+2b+2}{qm} \leqslant$$

$$\leqslant \frac{a+1}{qm} = \frac{cq+1}{q(q+x)} = \frac{cp^{k'}+1}{p^{k'}(p^{k'}+x)} \to 0 \quad \text{при } k' \to \infty. \quad \blacktriangle$$

Пример 3. Пусть $K = \mathbb{F}_4$, $q=2$, $m=3$, $a=2$ и $b=4$. Тогда a и b удовлетворяют условиям предложения 3, и получаем $[[6,2,\geqslant 2]]_4$-код с наилучшими параметрами согласно таблице в [25]. Аналогично получаются квантовые коды с параметрами $[[6,4,\geqslant 1]]_4$, $[[10,4,\geqslant 2]]_8$, $[[12,4,\geqslant 2]]_9$, $[[14,6,\geqslant 2]]_{16}$ и т.д.

§ 8. Сверточные коды по одноточечным геометрическим кодам Гоппы на $F/\mathbb{F}_{p^s}$

Рассмотрим кольцо многочленов $R = \mathbb{F}_{q_0}[X]$. Сверточным кодом C называется R-подмодуль ранга k модуля R^n. Пусть $G(X) = (g_{ij}(X)) \in (\mathbb{F}_{q_0}[X])^{k\times n}$ – порождающая матрица кода C над $\mathbb{F}_{q_0}[X]$, $\gamma_i = \max\{\deg g_{ij}(X) :\ 1 \leqslant j \leqslant n\}$, $\gamma = \sum\limits_{i=1}^{k} \gamma_i$,

$m' = \max\{\gamma_i : 1 \leqslant i \leqslant k\}$, и пусть d_f – минимальный вес элемента $c \in C$. Тогда говорят, что C имеет длину n, размерность k, степень γ, память m' и свободное расстояние d_f. Если $m' = 1$, то C называется сверточным кодом с единичной памятью и обозначается через $(n, k, \gamma; 1, d_f)_{q_0}$.

Следующая теорема описывает метод построения сверточных кодов по геометрическим кодам Гоппы.

Лемма 6 [26, теорема 3]. *Пусть $F'/\mathbb{F}_{q_0}$ – функциональное поле рода g'. Рассмотрим код $C_\Omega(D', G)$, такой что $2g' - 2 < \deg(G) < n$, где $\deg(G)$ – степень дивизора G. Тогда существует сверточный код с единичной памятью с параметрами $(n, k - \ell, \ell; 1, d_f \geqslant d)_{q_0}$, где $\ell \leqslant k/2$, $k = \deg(G) + 1 - g'$ и $d \geqslant n - \deg(G)$.*

В следующем предложении с помощью леммы 6 строятся сверточные коды с единичной памятью по одноточечным геометрическим кодам Гоппы на $F/\mathbb{F}_{p^s}$. Длина получаемых таким образом сверточных кодов равна qm, где $q = p^{k'}$ и $\text{НОД}(m, q) = 1$ (так что $\max\{q, m\} < p^s$), отличаясь тем самым от кодов из [27].

Предложение 4. *Пусть $qm - q - m - 1 < r < qm$, где r обладает свойством $(*)$ (определение 4). Тогда существует сверточный код с единичной памятью с параметрами $(qm, r + 1 - g - a, a; 1, d_f \geqslant d)_{p^s}$, где $g = (q-1)(m-1)/2$, $a \leqslant (r+1-g)/2$ и $d = qm - r$.*

Доказательство. Рассмотрим код $C_\Omega(D, rQ_\infty)$ на $F/\mathbb{F}_{p^s}$ с дивизором D таким, как в § 4. Поскольку $qm - q - m - 1 < r < qm$, по лемме 6 получаем сверточный код с единичной памятью с параметрами $(qm, r + 1 - g - a, a; 1, d_f \geqslant d)_{p^s}$, где $g = (q-1)(m-1)/2$, $a \leqslant (r+1-g)/2$. Так как r обладает свойством $(*)$, то $d = qm - r$ согласно следствию 1. ▲

Пример 4. Пусть $K = \mathbb{F}_4$, $q = 2$, $m = 3$ и $r = 4$. Тогда получаем сверточный код с единичной памятью с параметрами $(6, 3, 1; 1, \geqslant 2)_4$. Аналогично получаются сверточные коды с единичной памятью с параметрами $(10, 3, 2; 1, \geqslant 4)_8$, $(14, 6, 2; 1, \geqslant 4)_{16}$ и т.д.

§ 9. Локально восстанавливаемые коды на $F/\mathbb{F}_{p^s}$

Код $C \subset \mathbb{F}_{q_0}^n$ называется локально восстанавливаемым (LRC-кодом) с локальностью r_0, если для любого $i \in [n] := \{1, 2, \ldots, n\}$ существует подмножество $A_i \subset [n] \setminus \{i\}$, $|A_i| \leqslant r_0$, и функция φ_i, такие что для любого кодового слова $x \in C$ выполнено $x_i = \varphi_i(\{x_j,\ j \in A_i\})$. LRC-код C длины n, мощности q_0^k и локальности r_0 обозначается через (n, k, r_0). Минимальное расстояние (n, k, r_0)-LRC-кода удовлетворяет неравенству

$$d \leqslant n - k - \left\lceil \frac{k}{r_0} \right\rceil + 2. \tag{2}$$

Коды, лежащие на границе (2), называются оптимальными LRC-кодами. Скорость (n, k, r_0)-LRC-кода удовлетворяет неравенству

$$\frac{k}{n} \leqslant \frac{r_0}{r_0 + 1}. \tag{3}$$

В [28] были построены LRC-коды на алгебраических кривых. Опишем вкратце эту конструкцию; подробнее см. в [28]. Пусть $F_X/\mathbb{F}_{q_0}$ и $F_Y/\mathbb{F}_{q_0}$ – функциональные поля. Пусть $\psi\colon X \to Y$ – рациональное сепарабельное отображение степени $r_0 + 1$ гладких проективных абсолютно неприводимых кривых X и Y, соответствующих полям F_X и F_Y соответственно. Пусть $\psi^*\colon F_Y \to F_X$ – соответствующее отображение функциональных полей. Поскольку ψ сепарабельно, по теореме о примитивном

элементе существует функция $y \in F_X$, такая что $F_X = F_Y(y)$. Эту функцию y можно рассматривать как отображение $y\colon X \to \mathbb{P}^1$, и его степень $\deg(y)$ обозначим через h. Пусть $S = \{P_1, \dots, P_{s_0}\}$ – множество точек степени 1 поля F_Y, а A – положительный дивизор степени $\ell \geqslant 1$, носитель которого не пересекается с S. Для каждого j пусть $\{P_{ij}\}$ – множество точек поля F_X, лежащих над P_j. Будем предполагать, что каждая P_j полностью распадается в F_X. Пусть $\{f_1, \dots, f_{m_0}\}$ – базис линейного пространства $\mathcal{L}(A)$. Пусть V – подпространство F_X размерности $r_0 m_0$, порожденное функциями $\{f_j y^i,\ i = 0, \dots, r_0 - 1,\ j = 1, \dots, m_0\}$. Тогда код $C(A, \psi)$ определяется как образ отображения

$$e\colon V \to \mathbb{F}_{q_0}^{(r_0+1)s_0}, \quad F \mapsto (F(P_{ij}),\ i = 0, \dots, r_0,\ j = 1, \dots, s_0).$$

Теорема 20 [28, теорема 3.1]. *Подпространство $C(A, \psi) \subseteq \mathbb{F}_{q_0}^{(r_0+1)s_0}$ образует линейный (n, k, r_0)-LRC-код с параметрами*

$$n = (r_0 + 1)s_0,$$
$$k = r_0 m_0 \geqslant r_0(\ell - g_Y + 1),$$
$$d \geqslant n - (r_0 + 1)\ell - (r_0 - 1)h$$

при условии, что правая часть неравенства для d является натуральным числом.

В следующем предложении с помощью теоремы 20 строятся локально восстанавливаемые коды.

Предложение 5. *Пусть q, m, $F_X := F$, $F_Y := \mathbb{F}_{p^s}(x)$ и $\psi := \varphi$ такие, как в §3. Положим $S := \{P_{\alpha_1}, \dots, P_{\alpha_m}\}$ и $A := \ell P_\infty$, $\ell \in \mathbb{N}$, и пусть P_∞ – бесконечная точка поля F_Y. Если $2m - \ell q$ – натуральное число, то существует линейный (n, k, r_0)-LRC-код $C(A, \psi)$ с параметрами*

$$n = qm, \quad k \geqslant (q-1)(\ell+1) \quad \text{и} \quad d \geqslant 2m - \ell q.$$

Если $q = 2$, то этот код оптимален с локальностью $r_0 = 1$.

Доказательство. В предыдущих обозначениях имеем $h := m$, $s_0 := m$, $r_0 := q - 1$, $n := qm$ и $g_Y = 0$. Таким образом, по теореме 20 получаем линейный (n, k, r_0)-LRC-код $C(A, \psi)$ с параметрами $n = qm$, $k \geqslant (q-1)(\ell+1)$ и $d \geqslant 2m - \ell q$. При этом

$$d + k + \frac{k}{r_0} \geqslant 2m - \ell q + q(\ell + 1) = n + 2 - (m-1)(q-2).$$

Поэтому при $q = 2$ получаем оптимальный локально восстанавливаемый код с локальностью $r_0 = 1$. ▲

Пример 5. Пусть $K = \mathbb{F}_9$, $q = 3$, $m = 4$ и $\ell = 2$. Тогда получаем локально восстанавливаемый $(12, \geqslant 6, 2)$-код с минимальным расстоянием $d \geqslant 2$. Тогда $d + k + \left\lceil \dfrac{k}{r_0} \right\rceil \geqslant 11$, в то время как для кода, лежащего на границе (2), должно быть $d + k + \left\lceil \dfrac{k}{r_0} \right\rceil = 13$. Поэтому этот локально восстанавливаемый код не оптимален, но разница невелика.

Аналогично получается LRC-код с параметрами $(15, \geqslant 6, 2)$ и минимальным расстоянием $d \geqslant 4$. В этом случае разница с (2) также невелика.

Следствие 2. *Пусть выполнены все условия предложения 5. Если $\ell \geqslant m - 1$, то существует линейный (n, k, r_0)-LRC-код $C(A, \psi)$ с параметрами*

$$n = qm, \quad k \geqslant (q-1)(\ell+1) \quad \text{и} \quad d \geqslant 2m - \ell q.$$

Параметры этого кода лежат на границе (3).

Доказательство. Первое утверждение вытекает из предложения 5. Далее, из (3) получаем

$$\frac{(q-1)}{q} \geqslant \frac{k}{n} \geqslant \frac{(q-1)(\ell+1)}{qm} \geqslant \frac{(q-1)}{q}.$$

Таким образом, полученные коды имеют наибольшую возможную скорость. ▲

§ 10. Заключение

В статье определены одноточечные геометрические коды Гоппы по элементарным абелевым p-расширениям поля $\mathbb{F}_{p^s}(x)$ и вычислены их размерность и точное минимальное расстояние в нескольких случаях. Приведен список точных значений второго обобщенного веса Хэмминга этих кодов в нескольких случаях. Также приведены простые критерии квазисамодвойственности одноточечных геометрических кодов Гоппы и самодвойственности геометрических кодов Гоппы с дивизором G (не обязательно одноточечных). Получены семейства квантовых и сверточных кодов по построенным одноточечным геометрическим кодам Гоппы. Получены также локально восстанавливаемые коды на $F/\mathbb{F}_{p^s}$. Было бы интересно вычислить высшие обобщенные веса Хэмминга этих кодов и других классов слабых замковых кодов.

Авторы выражают глубокую благодарность рецензенту за замечания и предложения, позволившие значительно улучшить качество изложения.

СПИСОК ЛИТЕРАТУРЫ

1. *Гоппа В.Д.* Коды, ассоциированные с дивизорами // Пробл. передачи информ. 1977. Т. 13. № 1. С. 33–39.
2. *Stichtenoth H.* Algebraic Function Fields and Codes. Berlin: Springer-Verlag, 2009.
3. *Stichtenoth H.* A Note on Hermitian Codes over GF(q^2) // IEEE Trans. Inform. Theory. 1988. V. 34. № 5. Part 2. P. 1345–1348.
4. *Tiersma H.J.* Remarks on Codes from Hermitian Curves // IEEE Trans. Inform. Theory. 1987. V. 33. № 4. P. 605–609.
5. *Yang K., Kumar P.V.* On the True Minimum Distance of Hermitian Codes // Coding Theory and Algebraic Geometry (Proc. Int. Workshop held in Luminy, France, June 17–21, 1991). Lect. Notes Math. V. 1518. Berlin: Springer, 1992. P. 99–107.
6. *Garzón Rojas Á., Teherán Herrera A.* Elementary Abelian p-Extensions and Curves with Many Points // Rev. Acad. Colombiana Cienc. Exact. Fís. Natur. 2012. V. 36. № 139. P. 243–252.
7. *García A., Stichtenoth H.* Elementary Abelian p-Extensions of Algebraic Function Fields // Manuscripta Math. 1991. V. 72. № 1. P. 67–79.
8. *Johnsen T., Manshadi S., Monzavi N.* A Determination of the Parameters of a Large Class of Goppa Codes // IEEE Trans. Inform. Theory. 1994. V. 40. № 5. P. 1678–1681.
9. *Munuera C., Sepúlveda A., Torres F.* Castle Curves and Codes // Adv. Math. Commun. 2009. V. 3. № 4. P. 399–408.
10. *Olaya-León W., Munuera C.* On the Minimum Distance of Castle Codes // Finite Fields Appl. 2013. V. 20. P. 55–63.
11. *Olaya-León W., Granados-Pinzón C.* The Second Generalized Hamming Weight of Certain Castle Codes // Des. Codes Cryptogr. 2015. V. 76. № 1. P. 81–87.
12. *Munuera C., Tenório W., Torres F.* Quantum Error-Correcting Codes from Algebraic Geometry Codes of Castle Type // Quantum Inf. Process. 2016. V. 15. № 10. P. 4071–4088.
13. *Hernando F., McGuire G., Monserrat F., Moyano-Fernández J.J.* Quantum Codes from a New Construction of Self-orthogonal Algebraic Geometry Codes // Quantum Inf. Process. 2020. V. 19. № 4. Paper No. 117 (25 pp.).

14. *Helleseth T., Kløve T., Mykkeltveit J.* The Weight Distribution of Irreducible Cyclic Codes with Block Length $n_1((q^l-1)/N)$ // Discrete Math. 1977. V. 18. № 2. P. 179–211.
15. *Kløve T.* The Weight Distribution of Linear Codes over $GF(q^l)$ Having Generator Matrix over $GF(q)$ // Discrete Math. 1978. V. 23. № 2. P. 159–168.
16. *Wei V.K.* Generalized Hamming Weights for Linear Codes // IEEE Trans. Inform. Theory. 1991. V. 37. № 5. P. 1412–1418.
17. *Munuera C.* On the Generalized Hamming Weights of Geometric Goppa Codes // IEEE Trans. Inform. Theory. 1994. V. 40. № 6. P. 2092–2099.
18. *Delgado M., Farrán J.I., García-Sánchez P.A., Llena D.* On the Weight Hierarchy of Codes Coming from Semigroups with Two Generators // IEEE Trans. Inform. Theory. 2014. V. 60. № 1. P. 282–295.
19. *Kirfel C., Pellikaan R.* The Minimum Distance of Codes in an Array Coming from Telescopic Semigroups // IEEE Trans. Inform. Theory. 1995. V. 41. № 6. Part 1. P. 1720–1732.
20. *Geil O., Munuera C., Ruano D., Torres F.* On the Order Bounds for One-Point AG Codes // Adv. Math. Commun. 2011. V. 5. № 3. P. 489–504.
21. *Niederreiter H., Xing C.* Algebraic Geometry in Coding Theory and Cryptography. Princeton, NJ: Princeton Univ. Press, 2009.
22. *Munuera C., Pellikaan R.* Equality of Geometric Goppa Codes and Equivalence of Divisors // J. Pure Appl. Algebra. 1993. V. 90. № 3. P. 229–252.
23. *Calderbank A.R., Rains E.M., Shor P.W., Sloane N.J.A.* Quantum Error Correction via Codes over GF(4) // *IEEE Trans. Inform. Theory.* 1998. V. 44. № 4. P. 1369–1387.
24. *Aly S.A., Klappenecker A., Sarvepalli P.K.* On Quantum and Classical BCH Codes // IEEE Trans. Inform. Theory. 2007. V. 53. № 3. P. 1183–1188.
25. *Grassl M.* Bounds on the Minimum Distance of Linear Codes and Quantum Codes (electronic tables). Available online at `http://www.codetables.de` (accessed on June 6, 2020).
26. *Pereira F.R.F., La Guardia G.G., de Assis F.M.* Classical and Quantum Convolutional Codes Derived from Algebraic Geometry Codes // IEEE Trans. Commun. 2019. V. 67. № 1. P. 73–82.
27. *La Guardia G.G.* On Optimal Constacyclic Codes // Linear Algebra Appl. 2016. V. 496. P. 594–610.
28. *Barg A., Tamo I., Vlăduţ S.* Locally Recoverable Codes on Algebraic Curves // IEEE Trans. Inform. Theory. 2017. V. 63. № 8. P. 4928–4939.

Патанкер Нупур
Сингх Санджай Кумар
Индийский институт науки, образования
и исследований, Бхопал, Индия
`nupurp@iiserb.ac.in`
`sanjayks@iiserb.ac.in`

Поступила в редакцию
12.02.2020
После доработки
15.06.2020
Принята к публикации
30.06.2020

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.17

© 2020 г. **С. Ван, В. Чжан**

ИССЛЕДОВАНИЕ ДРОБНЫХ ПРЕДЕЛЬНЫХ ПОКРЫТЫХ ГРАФОВ

Граф G называется дробным (g, f)-покрытым, если для любого $e \in E(G)$ граф G допускает дробный (g, f)-фактор, покрывающий e. Граф G называется дробным (g, f, n)-предельным покрытым графом, если для любого $S \subseteq V(G)$, такого что $|S| = n$, граф $G - S$ является (g, f)-покрытым графом. Скажем, что дробный (g, f, n)-предельный покрытый граф является дробным (a, b, n)-предельным покрытым графом, если $g(x) = a$ и $f(x) = b$ для любого $x \in V(G)$. Дробные (a, b, n)-предельные покрытые графы были впервые введены и исследованы в работе [1]. В настоящей статье мы исследуем (g, f, n)-предельные покрытые графы и приводим условие их существования в терминах связующего числа. Таким образом, мы улучшили и обобщили предыдущие результаты, полученные в работе [2].

Ключевые слова: граф, связующее число, дробный (g, f)-фактор, дробный (g, f)-покрытый граф, дробный (g, f, n)-предельный покрытый граф.

DOI: 10.31857/S0555292320030043

§ 1. Введение

Многие реальные сети могут быть естественно смоделированы с помощью графов. Вершины графа в этом случае обозначают узлы сети, а ребра графа служат связями между узлами. Можно привести следующие примеры: коммуникационная сеть с вершинами, обозначающими города, и ребрами, обозначающими каналы коммуникации; социальная сеть, в которой вершины представляют собой людей, а ребра соответствуют личному контакту между ними; сеть Интернет с вершинами, являющимися веб-страницами, и ребрами, моделирующими гиперссылки между ними. Далее мы будем использовать термин “граф” вместо “сеть”. Наука о сетях (также известная как анализ сложных сетей) представляет собой развивающуюся область в парадигме больших данных и соответствует анализу сложных реальных и основанных на теоретических моделях сетей с точки зрения теории графов.

Задача отыскания дробного фактора может рассматриваться как более слабая версия хорошо известной задачи о сопоставлении мощностей и применяться в близких областях, таких как проектирование сетей, планирование и комбинаторика многогранников. Например, допустим, что несколько больших пакетов данных могут быть отправлены различным адресатам несколькими каналами в сети связи. Эффективность данной сети можно повысить, если разделить большие пакеты данных на маленькие части. Возможное переназначение пакетов данных можно рассматривать как задачу дробления потока, и оно становится задачей о дробном факторе, когда множества пунктов назначения и отправки в сети не пересекаются (см. [3]).

В модели теории графов удаляются вершины графа, соответствующие поврежденным участкам сети. Ребра оставшегося графа соответствуют связям в получившейся сети. В настоящей статье мы пытаемся найти дробный фактор с помощью специальных ребер в результирующем графе. Очевидно, что задача передачи

данных в коммуникационной сети в заданный момент может быть сведена к задаче о существовании дробного предельного покрытого подграфа в соответствующем графе сети. Условие на связующее число графа часто используется для оценки уязвимости и надежности сети, которые играют ключевую роль в передаче данных и проектировании сетей.

В этой статье мы будем рассматривать только конечные простые графы. Пусть G – граф с множеством вершин $V(G)$ и множеством ребер $E(G)$. Для вершины x из G обозначим через $d_G(x)$ степень вершины x в G, а через $N_G(x)$ – окрестность x в G. Рассмотрим также окрестность множества вершин. Для множества вершин $S \subseteq V(G)$ определим $N_G(S)$ как $N_G(S) = \bigcup\limits_{x \in S} N_G(x)$. Кроме того, зададим $G[S]$ как подграф G, индуцированный S, и будем обозначать граф $G[V(G) \setminus S]$ через $G - S$. Множество вершин S графа G является независимым, если никакие две вершины из S не смежны. Обозначим минимальную степень в графе G через $\delta(G)$. Связующее число $\mathrm{bind}(G)$ определим как минимум отношения $\dfrac{|N_G(S)|}{|S|}$, взятый по всем $S \subseteq V(G)$, удовлетворяющим условиям $S \neq \varnothing$ и $N_G(S) \neq V(G)$.

Пусть $g, f\colon V(G) \to \{0, 1, 2, 3, \ldots\}$ – функции, удовлетворяющие неравенству $g(x) \leqslant f(x)$ для любого $x \in V(G)$. Остовный подграф F графа G называется (g, f)-фактором, если неравенство $g(x) \leqslant d_F(x) \leqslant f(x)$ выполняется для всех $x \in V(G)$. Пусть h – функция, определенная на $E(G)$ и удовлетворяющая условию $h(e) \in [0, 1]$ для всякого $e \in E(G)$. Мы будем называть h дробной (g, f)-индикаторной функцией, если для любого $x \in V(G)$ верно, что $g(x) \leqslant d^h_G(x) \leqslant f(x)$, где

$$d^h_G(x) = \sum_{e \in E(x)} h(e)$$

– дробная степень x в G, где $E(x) = \{e :\ e = xy \in E(G)\}$. Положим

$$E_h = \{e \in E(G) :\ h(e) > 0\}.$$

Если G_h – остовный подграф G с $E(G_h) = E_h$, то назовем G_h дробным (g, f)-фактором G с индикаторной функцией h. Если $h(e) \in \{0, 1\}$ для любого $e \in E(G)$, то дробный (g, f)-фактор является просто (g, f)-фактором. Граф G называется дробным (g, f)-покрытым графом, если для любого $e \in E(G)$ граф G допускает такой дробный (g, f)-фактор G_h, что $h(e) = 1$. Если $g(x) = f(x) = r$ для любого $x \in V(G)$, то дробный (g, f)-покрытый граф назовем дробным r-покрытым графом. Будем говорить, что G – дробный (g, f, n)-предельный покрытый граф, если для любого $S \subseteq V(G)$, такого что $|S| = n$, граф $G - S$ является (g, f)-покрытым. Дробный (g, f, n)-предельный покрытый граф назовем (a, b, n)-предельным покрытым графом, если $g(x) = a$ и $f(x) = b$ для всякого $x \in V(G)$. Дробный (r, r, n)-предельный покрытый граф называется просто дробным (r, n)-предельным покрытым графом.

Напомним, что k-факторы были введены в 1950 г. в работе [4] и впоследствии были исследованы в работах [5–9]. Обобщение данной задачи, (g, f)-факторы, были введены в работе [10] в 1970 г., а в работах [11–16] были даны достаточные (необходимые) условия их существования в терминах степеней, окрестностей, числа стабильности, связующего числа и т.д. В 1988 г. в работе [17] было впервые введено понятие (g, f)-покрытых графов и получена их характеризация. Дробные (g, f)-факторы были введены в работе [18] в 1985 г. и изучались в работах [19–25]. В 2002 г. в работе [26] были впервые исследованы (g, f)-покрытые графы. Наконец, определение дробных (a, b, n)-предельных покрытых графов было дано в 2019 г. [1], а в [1,27,28] были получены достаточные условия существования данных графов, выраженные в терминах степеней вершин, чисел независимости и связности. Некоторые другие результаты о факторах и дробных факторах в графах представлены в работах [29–33]. В на-

стоящей статье мы изучаем дробные (g, f, n)-предельные покрытые графы и даем условие существования дробных (g, f, n)-предельных покрытых графов, выраженное в терминах связующего числа. Этот результат представляет следующая

Теорема 1. *Рассмотрим целые числа a, b, μ, n, такие что $\mu \geqslant 0$, $n \geqslant 0$, $a \geqslant 1$, $a + \mu \geqslant 2$ и $b - \mu \geqslant \max\{a, 2\}$. Пусть G – граф порядка p, такой что*

$$p \geqslant \frac{(a+b-1)(a+b-2)+1}{a+\mu} + \frac{(a+\mu)n}{a+\mu-1}.$$

Кроме того, пусть $g, f\colon V(G) \to \{0, 1, 2, 3, \ldots\}$ – функции, удовлетворяющие условию $a \leqslant g(x) \leqslant f(x) - \mu \leqslant b - \mu$ для всякого $x \in V(G)$. Если

$$\operatorname{bind}(G) > \frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b},$$

то G является дробным (g, f, n)-предельным покрытым графом.

При $p \to +\infty$ можно доказать, что

$$\frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b} \to \frac{a+b-1}{a+\mu}$$

и

$$\frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b} < \frac{a+b}{a+\mu},$$

где n – неотрицательное конечное число. В этом случае из теоремы 1 следует, что G является дробным (g, f, n)-предельным покрытым графом, если только

$$\operatorname{bind}(G) \geqslant \frac{a+b}{a+\mu}.$$

Если $n = \dfrac{p}{2}$, то из теоремы 1 вытекает, что G – дробный (g, f, n)-предельный покрытый граф при условии

$$\operatorname{bind}(G) > \frac{2(a+b-1)(p-1)}{(a+\mu)p-2(a+b)}.$$

В этом случае при $p \to +\infty$

$$\frac{2(a+b-1)(p-1)}{(a+\mu)p-2(a+b)} \to \frac{2(a+b-1)}{a+\mu}$$

и выполняется неравенство

$$\frac{2(a+b-1)(p-1)}{(a+\mu)p-2(a+b)} < \frac{2a+2b-1}{a+\mu}.$$

Тогда из теоремы 1 следует, что G является дробным (g, f, n)-предельным покрытым графом, если

$$\operatorname{bind}(G) \geqslant \frac{2a+2b-1}{a+\mu}.$$

Положим $\mu = 0$ в теореме 1. Тогда мы получим следующее следствие о (g, f, n)-предельных покрытых графах.

Следствие 1. *Рассмотрим целые числа a, b, n, такие что $n \geqslant 0$ и $b \geqslant a \geqslant 2$, и граф G порядка p, такой что*

$$p \geqslant \frac{(a+b-1)(a+b-2)+1}{a} + \frac{an}{a-1}.$$

Пусть $g, f\colon V(G) \to \{0, 1, 2, 3, \ldots\}$ – функции, удовлетворяющие неравенствам $a \leqslant g(x) \leqslant f(x) \leqslant b$ для любого $x \in V(G)$. Если

$$\operatorname{bind}(G) > \frac{(a+b-1)(p-1)}{a(p-n)-a-b},$$

то G – дробный (g, f, n)-предельный покрытый граф.

Пусть $g(x) = f(x)$ для любого $x \in V(G)$ в теореме 1. Тогда можно вывести следующее следствие относительно (f, n)-предельного покрытого графа.

Следствие 2. *Рассмотрим целые числа a, b, n, такие что $n \geqslant 0$ и $b \geqslant a \geqslant 2$, и граф G порядка p, такой что*

$$p \geqslant \frac{(a+b-1)(a+b-2)+1}{a} + \frac{an}{a-1}.$$

Пусть $f\colon V(G) \to \{0, 1, 2, 3, \ldots\}$ – функция, удовлетворяющая условию $a \leqslant f(x) \leqslant b$ для всякого $x \in V(G)$. Если

$$\operatorname{bind}(G) > \frac{(a+b-1)(p-1)}{a(p-n)-a-b},$$

то G – дробный (f, n)-предельный покрытый граф.

Положим $a = b = r$ и $n = 0$ в теореме 1. Тогда $f(x) = r$ для любого $x \in V(G)$. Мы получаем следующее следствие о дробных r-покрытых графах.

Следствие 3. *Рассмотрим целое число r, такое что $r \geqslant 2$, и граф G порядка p, такой что $p \geqslant 4r - 6 + \frac{3}{r}$. Если*

$$\operatorname{bind}(G) > \frac{(2r-1)(p-1)}{rp-2r},$$

то G – дробный r-покрытый граф.

Из следствия 3 мы сразу получаем следующую теорему, которая была доказана Юанем и Хао в [2]. Следовательно, основной результат данной статьи является улучшением и расширением предыдущего результата Юаня и Хао.

Теорема 2 [2]. *Пусть $r \geqslant 2$ – целое число, а G – граф порядка p, такой что $p \geqslant 4r + 2$. Предположим, что $\delta(G) \geqslant r + 1$. Если G удовлетворяет условию*

$$\operatorname{bind}(G) > \frac{(2r-1)(p-1)}{rp-2r},$$

то G – дробный r-покрытый граф.

§ 2. Доказательство теоремы 1

Доказательство теоремы 1 в значительной степени опирается на следующую теорему, представляющую собой критерий того, что граф является дробным (g, f)-покрытым.

Теорема 3 [26]. *Рассмотрим граф G. Пусть $g, f\colon V(G) \to \{0,1,2,3,\ldots\}$ – такие функции, что $g(x) \leqslant f(x)$ для всякого $x \in V(G)$. Тогда G является дробным (g,f)-покрытым, если и только если*

$$\theta_G(X,Y) = f(X) + d_{G-X}(Y) - g(Y) \geqslant \varepsilon(X)$$

для всех $X \subseteq V(G)$ и $Y = \{x :\ x \in V(G) \setminus X,\ d_{G-X}(x) \leqslant g(x)\}$, где $\varepsilon(X)$ определено следующим образом:

$$\varepsilon(X) = \begin{cases} 2, & \text{если } X \text{ не является независимым множеством,} \\ 1, & \text{если } X \text{ является независимым множеством и есть ребро,} \\ & \text{соединяющее } X \text{ и } V(G) \setminus (X \cup Y), \text{ или есть ребро } e = xy, \\ & \text{соединяющее } X \text{ и } Y, \text{ и при этом } d_{G-X}(y) = g(y) \text{ для } y \in Y, \\ 0 & \text{в противном случае.} \end{cases}$$

Доказательство теоремы 1. Для любого $N \subseteq V(G)$, такого что $|N| = n$, мы пишем $H = G - N$. Чтобы доказать теорему 1, достаточно показать, что H – дробный (g,f)-покрытый граф. Предположим от противного, что H не является дробным (g,f)-покрытым. Тогда из теоремы 3 получаем соотношение

$$\theta_H(X,Y) = f(X) + d_{H-X}(Y) - g(Y) \leqslant \varepsilon(X) - 1 \tag{1}$$

для некоторого подмножества $X \subseteq V(H)$, где $Y = \{x :\ x \in V(H) \setminus X,\ d_{H-X}(x) \leqslant g(x)\}$.

Если $Y = \varnothing$, то из (1) и условия $\varepsilon(X) \leqslant |X|$ мы получаем, что $\varepsilon(X) - 1 \geqslant \theta_H(X,\varnothing) = f(X) \geqslant (a+\mu)|X| \geqslant |X| \geqslant \varepsilon(X)$, что приводит к противоречию. Следовательно, $Y \neq \varnothing$. Таким образом, можно определить

$$h = \min\{d_{H-X}(x) :\ x \in Y\}.$$

С учетом определения Y выполняется соотношение $0 \leqslant h \leqslant b - \mu$.

Чтобы убедиться в справедливости теоремы 1, мы отыщем противоречие, рассматривая следующие два случая.

Случай 1: $h = 0$.

Пусть $Q = \{x : x \in Y, d_{H-X}(x) = 0\}$. Ясно, что $Q \neq \varnothing$ и $N_G(V(G)\setminus(X\cup N))\cap Q = \varnothing$, т.е. $|N_G(V(G)\setminus(X\cup N))| \leqslant p - |Q|$. С учетом определения $\mathrm{bind}(G)$ и неравенства

$$\mathrm{bind}(G) > \frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b}$$

получаем следующее соотношение:

$$\frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b} < \mathrm{bind}(G) \leqslant \frac{|N_G(V(G)\setminus(X\cup N))|}{|V(G)\setminus(X\cup N)|} \leqslant \frac{p-|Q|}{p-n-|X|}. \tag{2}$$

Из (2) и неравенства $p \geqslant \dfrac{(a+b-1)(a+b-2)+1}{a+\mu} + \dfrac{(a+\mu)n}{a+\mu-1}$ следует, что

$$\begin{aligned} &(a+b-1)(p-1)|X| > \\ &> (a+b-1)(p-1)(p-n) - ((a+\mu)(p-n)-a-b)(p-|Q|) > \\ &> (b-\mu-1)(p-1)(p-n) + p - 1 + (p-1)|Q|, \end{aligned}$$

откуда

$$|X| > \frac{(b-\mu-1)(p-n)+|Q|+1}{a+b-1}. \tag{3}$$

С другой стороны, из (1) и неравенств $\varepsilon(X) \leqslant 2$ и $|X| + |Y| + n \leqslant p$ следует цепочка соотношений

$$1 \geqslant \varepsilon(X) - 1 \geqslant \theta_H(X,Y) = f(X) + d_{H-X}(Y) - g(Y) \geqslant$$
$$\geqslant (a+\mu)|X| + |Y| - |Q| - (b-\mu)|Y| = (a+\mu)|X| - (b-\mu-1)|Y| - |Q| \geqslant$$
$$\geqslant (a+\mu)|X| - (b-\mu-1)(p-n-|X|) - |Q| =$$
$$= (a+b-1)|X| - (b-\mu-1)(p-n) - |Q|,$$

из которой вытекает неравенство

$$|X| \leqslant \frac{(b-\mu-1)(p-n) + |Q| + 1}{a+b-1},$$

противоречащее (3).

Случай 2: $1 \leqslant h \leqslant b-\mu$.

Существует $x_1 \in Y$ с условием $d_{H-X}(x_1) = h$. Положим $W = (V(H) \setminus X) \setminus N_{H-X}(x_1)$. Очевидно, что $x_1 \in W$ и $x_1 \notin N_G(W)$. Следовательно, $W \neq \varnothing$ и $N_G(W) \neq V(G)$. Заметим, что $|V(H)| = |V(G)| - n = p - n$. Объединяя эти соотношения с определением $\mathrm{bind}(G)$ и неравенством $\mathrm{bind}(G) > \frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b}$, получаем, что

$$\frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b} < \mathrm{bind}(G) \leqslant \frac{|N_G(W)|}{|W|} \leqslant \frac{p-1}{p-n-|X|-h},$$

откуда следует оценка

$$|X| > \frac{(b-\mu-1)(p-n)+a+b}{a+b-1} - h. \tag{4}$$

Согласно (4) и неравенству $|X| + |Y| + n \leqslant p$ имеем

$$\theta_H(X,Y) = f(X) + d_{H-X}(Y) - g(Y) \geqslant (a+\mu)|X| + h|Y| - (b-\mu)|Y| =$$
$$= (a+\mu)|X| - (b-\mu-h)|Y| \geqslant (a+\mu)|X| - (b-\mu-h)(p-n-|X|) =$$
$$= (a+b-h)|X| - (b-\mu-h)(p-n) >$$
$$> (a+b-h)\left(\frac{(b-\mu-1)(p-n)+a+b}{a+b-1} - h\right) - (b-\mu-h)(p-n),$$

т.е.

$$\theta_H(X,Y) > (a+b-h)\left(\frac{(b-\mu-1)(p-n)+a+b}{a+b-1} - h\right) - (b-\mu-h)(p-n). \tag{5}$$

Если $h = 1$, то из (5) и $\varepsilon(X) \leqslant 2$ получаем

$$\theta_H(X,Y) > (a+b-1)\left(\frac{(b-\mu-1)(p-n)+a+b}{a+b-1} - 1\right) - (b-\mu-1)(p-n) =$$
$$= 1 \geqslant \varepsilon(X) - 1,$$

что противоречит (1). Поэтому далее мы рассматриваем только $2 \leqslant h \leqslant b-\mu$.

Пусть $\varphi(h) = (a+b-h)\Big(\dfrac{(b-\mu-1)(p-n)+a+b}{a+b-1} - h\Big) - (b-\mu-h)(p-n)$. Тогда

$$\varphi'(h) = -\frac{(b-\mu-1)(p-n)+a+b}{a+b-1} + h - a - b + h + p - n =$$
$$= \frac{(a+\mu)(p-n)-a-b}{a+b-1} - a - b + 2h \geqslant$$
$$\geqslant \frac{(a+b-1)(a+b-2)+1-a-b}{a+b-1} - a - b + 2h = 2h - 3 \geqslant 1,$$

поскольку $p \geqslant \dfrac{(a+b-1)(a+b-2)+1}{a+\mu} + \dfrac{(a+\mu)n}{a+\mu-1}$ и $2 \leqslant h \leqslant b-\mu$. Отсюда следует, что $\varphi(h)$ достигает своего минимального значения при $h = 2$, т.е. $\varphi(h) \geqslant \varphi(2) = \dfrac{(a+\mu)(p-n)-(a+b-2)^2}{a+b-1}$. Объединяя полученное соотношение с (5), неравенствами $\varepsilon(X) \leqslant 2$ и $p \geqslant \dfrac{(a+b-1)(a+b-2)+1}{a+\mu} + \dfrac{(a+\mu)n}{a+\mu-1}$, получаем оценку

$$\theta_H(X,Y) > \varphi(h) \geqslant \varphi(2) = \frac{(a+\mu)(p-n)-(a+b-2)^2}{a+b-1} \geqslant 1 \geqslant \varepsilon(X) - 1,$$

которая противоречит (1). Таким образом, теорема 1 доказана. ▲

§ 3. Точность теоремы 1

Теперь мы покажем, что оценка $\mathrm{bind}(G) > \dfrac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b}$ в теореме 1 наилучшая возможная. Будем обозначать $G = mR$, если граф G состоит из m ($m \geqslant 2$) непересекающихся копий графа R. Соединение графов $G = A \vee B$ задается соотношениями $V(G) = V(A) \cup V(B)$ и $E(G) = E(A) \cup E(B) \cup \{uv : u \in V(A),\ v \in V(B)\}$.

Пусть $b = a + \mu$, $g(x) \equiv b - \mu$ и $f(x) \equiv a + \mu$, а β – достаточно большое положительное целое число, для которого $\dfrac{(a+\mu)\beta - 1}{2(b-\mu-1)}$ является целым. Обозначим $\gamma = \dfrac{(a+\mu)\beta-1}{2(b-\mu-1)}$. Построим граф $G = K_{\beta+n} \vee (\gamma K_2)$ порядка p, где K_t обозначает полный граф порядка t. Легко видеть, что

$$\mathrm{bind}(G) = \frac{|N(V(\gamma K_2) \setminus \{w\})|}{|V(\gamma K_2) \setminus \{w\}|} = \frac{p-1}{2\gamma - 1} = \frac{(a+b-1)(p-1)}{(a+b-1)(2\gamma-1)} =$$
$$= \frac{(a+b-1)(p-1)}{(a+\mu)(p-n)-a-b},$$

где $w \in V(\gamma K_2)$. Положим $H = G - N$ для $N \subseteq V(K_{\beta+n})$, такого что $|N| = n$. Пусть $X = V(K_{\beta+n}) \setminus N$ и $Y = V(\gamma K_2)$. Заметим, что X не является независимым множеством. Следовательно, $\varepsilon(X) = 2$. Таким образом, мы заключаем, что

$$\theta_H(X,Y) = f(X) + d_{H-X}(Y) - g(Y) = (a+\mu)|X| + \frac{(a+\mu)\beta-1}{b-\mu-1} - (b-\mu)|Y| =$$
$$= (a+\mu)\beta + \frac{(a+\mu)\beta-1}{b-\mu-1} - (b-\mu)\frac{(a+\mu)\beta-1}{b-\mu-1} = 1 < 2 = \varepsilon(X).$$

С учетом теоремы 3 граф H не является дробным (g,f)-покрытым, поэтому граф G не является дробным (g,f,n)-предельным покрытым.

Авторы благодарят рецензентов за ценные комментарии и полезные предложения.

СПИСОК ЛИТЕРАТУРЫ

1. *Zhou S., Xu Y., Sun Z.* Degree Conditions for Fractional (a, b, k)-Critical Covered Graphs // Inform. Process. Lett. 2019. V. 152. Article 105838 (5 pp.).
2. *Yuan Y., Hao R.-X.* Neighborhood Union Conditions for Fractional $[a, b]$-Covered Graphs // Bull. Malays. Math. Sci. Soc. 2020. V. 43. № 1. P. 157–167.
3. *Liu G., Zhang L.* Characterizations of Maximum Fractional (g, f)-Factors of Graphs // Discrete Appl. Math. 2008. V. 156. № 12. P. 2293–2299.
4. *Belck H.-B.* Reguläre Faktoren von Graphen // J. Reine Angew. Math. 1950. V. 188. P. 228–252.
5. *Enomoto H., Jackson B., Katerinis P., Satio A.* Toughness and the Existence of k-Factors // J. Graph Theory. 1957. V. 9. № 1. P. 87–95.
6. *Katerinis P., Woodall D.R.* Binding Numbers of Graphs and the Existence of k-Factors // Quart. J. Math. Oxford Ser. (2). 1987. V. 38. № 150. P. 221–228.
7. *Plummer M.D., Saito A.* Toughness, Binding Number and Restricted Matching Extension in a Graph // Discrete Math. 2017. V. 340. № 11. P. 2665–2672.
8. *Kano M., Tokushige N.* Binding Numbers and f-Factors of Graphs // J. Combin. Theory Ser. B. 1992. V. 54. № 2. P. 213–221.
9. *Cymer R., Kano M.* Generalizations of Marriage Theorem for Degree Factors // Graphs Combin. 2016. V. 32. № 6. P. 2315–2322.
10. *Lovász L.* Subgraphs with Prescribed Valencies // J. Combin. Theory. 1970. V. 8. P. 391–416.
11. *Egawa Y., Kano M.* Sufficient Conditions for Graphs to Have (g, f)-Factors // Discrete Math. 1996. V. 151. № 1–3. P. 87–90.
12. *Matsuda H.* A Neighborhood Condition for Graphs to Have $[a, b]$-Factors // Discrete Math. 2000. V. 224. № 1–3. P. 289–292.
13. *Kouider M., Ouatiki S.* Sufficient Condition for the Existence of an Even $[a, b]$-Factor in Graph // Graphs Combin. 2013. V. 29. № 4. P. 1051–1057.
14. *Zhou S., Zhang T., Xu Z.* Subgraphs with Orthogonal Factorizations in Graphs // Discrete Appl. Math. 2020 (in press). Available online at `https://doi.org/10.1016/j.dam.2019.12.011`.
15. *Zhou S., Sun Z.* Binding Number Conditions for $P_{\geqslant 2}$-Factor and $P_{\geqslant 3}$-Factor Uniform Graphs // Discrete Math. 2020. V. 343. № 3. Article 111715 (6 pp.).
16. *Zhou S.Z., Sun Z.R.* Some Existence Theorems on Path Factors with Given Properties in Graphs // Acta Math. Sin. (Engl. Ser.). 2020. V. 36. № 8. P. 917–928.
17. *Liu G.Z.* On (g, f)-Covered Graphs // Acta Math. Sci. (English Ed.). 1988. V. 8. № 2. P. 181–184.
18. *Anstee R.P.* An Algorithmic Proof of Tutte's f-Factor Theorem // J. Algorithms. 1985. V. 6. № 1. P. 112–131.
19. *Liu G., Zhang L.* Toughness and the Existence of Fractional k-Factors of Graphs // Discrete Math. 2008. V. 308. № 9. P. 1741–1748.
20. *Katerinis P.* Fractional ℓ-Factors in Regular Graphs // Australas. J. Combin. 2019. V. 73. Part 3. P. 432–439.
21. *Zhou S., Sun Z., Ye H.* A Toughness Condition for Fractional (k, m)-Deleted Graphs // Inform. Process. Lett. 2013. V. 113. № 8. P. 255–259.
22. *Gao W., Wang W.F., Guirao J.L.G.* The Extension Degree Conditions for Fractional Factor // Acta Math. Sin. (Engl. Ser.). 2020. V. 36. № 3. P. 305–317.
23. *Zhou S., Xu L., Xu Z.* Remarks on Fractional ID-k-Factor-Critical Graphs // Acta Math. Appl. Sin. Engl. Ser. 2019. V. 35. № 2. P. 458–464.
24. *Yuan Y., Hao R.-X.* Toughness Condition for the Existence of All Fractional (a, b, k)-Critical Graphs // Discrete Math. 2019. V. 342. № 8. P. 2308–2314.

25. *Lu H., Yu Q.* General Fractional f-Factor Numbers of Graphs // Appl. Math. Lett. 2011. V. 24. № 4. P. 519–523.
26. *Li Z., Yan G., Zhang X.* On Fractional (g, f)-Covered Graphs // OR Trans. (in Chinese). 2002. V. 6. № 4. P. 65–68.
27. *Zhou S., Wu J., Liu H.* Independence Number and Connectivity for Fractional (a, b, k)-Critical Covered Graphs, `arXiv:1909.01070 [math.CO]`, 2019.
28. *Lv X.* A Degree Condition for Fractional (g, f, n)-Critical Covered Graphs // AIMS Math. 2020. V. 5. № 2. P. 872–878.
29. *Zhou S.* Remarks on Orthogonal Factorizations of Digraphs // Int. J. Comput. Math. 2014. V. 91. № 10. P. 2109–2117.
30. *Zhou S., Sun Z., Xu Z.* A Result on r-Orthogonal Factorizations in Digraphs // European J. Combin. 2017. V. 65. P. 15–23.
31. *Zhou S., Sun Z., Liu H.* Sun Toughness and $P_{\geqslant 3}$-Factors in Graphs // Contrib. Discrete Math. 2019. V. 14. № 1. P. 167–174.
32. *Zhou S., Yang F., Xu L.* Two Sufficient Conditions for the Existence of Path Factors in Graphs // Sci. Iran. D: Comput. Sci. Eng. Electr. Eng. 2019. V. 26. № 6. P. 3510–3514.
33. *Zhou S.* Remarks on Path Factors in Graphs // RAIRO – Oper. Res. (forthcoming article). 2019. Available online at `https://doi.org/10.1051/ro/2019111`.

Ван Суфан
Школа государственного управления,
Научно-технологический университет Цзянсу,
Чжэньцзян, провинция Цзянсу, КНР
`wangsufangjust@163.com`
Чжан Вэй
Колледж Оуцзян, Университет Вэньчжоу,
Вэньчжоу, провинция Чжэцзян, КНР

Поступила в редакцию
11.11.2019
После доработки
13.05.2020
Принята к публикации
12.06.2020

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 3**

УДК 621.391.1 : 519.713 : 517.977.5

© 2020 г. **А.В. Колногоров**

ГАУССОВСКИЙ ДВУРУКИЙ БАНДИТ: ПРЕДЕЛЬНОЕ ОПИСАНИЕ[1]

Для гауссовского двурукого бандита, который возникает при анализе пакетной обработки данных, изучается предельное поведение минимаксного риска, если горизонт управления N неограниченно растет. Минимаксный риск ищется как байесовский, вычисленный относительно наихудшего априорного распределения. Показано, что наиболее высокие требования к управлению предъявляются в области "близких" распределений, где математические ожидания доходов различаются на величину порядка $N^{-1/2}$. В области "близких" распределений получены рекуррентное интегро-разностное уравнение для нахождения байесовского риска относительно наихудшего априорного распределения в инвариантной форме с горизонтом управления, равным единице, и дифференциальное уравнение в частных производных второго порядка в предельном случае. Результаты позволяют оценить качество пакетной обработки. Например, минимаксный риск, соответствующий пакетной обработке данных, разбитых на 50 пакетов, может быть лишь на 2% выше своего предельного значения, если число пакетов неограниченно растет. В случае бернуллиевского двурукого бандита показано, что оптимальная обработка данных по одному не является более эффективной, чем пакетная, если N неограниченно растет.

Ключевые слова: гауссовский двурукий бандит, минимаксный и байесовский подходы, пакетная обработка, асимптотическая минимаксная теорема.

DOI: 10.31857/S0555292320030055

§ 1. Введение

Статья продолжает работу [1], в которой дан обзор некоторых других подходов к рассматриваемой задаче и библиография. Опишем кратко результаты [1]. Гауссовский двурукий бандит – это управляемый случайный процесс ξ_n, $n = 1, \ldots, N$, значения которого интерпретируются как доходы, зависят только от выбираемых в текущие моменты времени действий y_n и имеют нормальные распределения с плотностями $f_{D_\ell}(x \,|\, m_\ell)$, если $y_n = \ell$ $(\ell = 1, 2)$, где $f_D(x \,|\, m) = (2\pi D)^{-1/2} \exp\{-(x-m)^2/(2D)\}$. Предполагается, что D_1, D_2 – априори известные дисперсии, а m_1, m_2 – неизвестные математические ожидания. Такой двурукий бандит описывается параметром $\theta = (m_1, m_2)$. Допустимое множество параметров имеет вид $\Theta = \{\theta : \ |m_1 - m_2| \leqslant \leqslant 2C\}$, где $0 < C < \infty$.

Гауссовский двурукий бандит возникает, если одинаковые действия применяются к пакетам данных, а для управления используются суммарные доходы в пакетах. В силу центральной предельной теоремы для широкого класса процессов распределения суммарных доходов в пакетах данных близки к нормальным, если размеры пакетов достаточно велики. При обработке начальных пакетов могут быть получены

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 20-01-00062).

оценки дисперсий одношаговых доходов D_1, D_2. Их можно использовать при дальнейшей обработке, так как минимаксный риск мало меняется при малом изменении дисперсий (см. лемму 8 этой статьи). Важное свойство пакетной обработки, установленное в [1], состоит в том, что она практически не увеличивает минимаксный риск, если число пакетов, на которые разбиты данные, достаточно велико.

Управляющая стратегия σ в момент времени $n+1$ определяет выбор действия y_{n+1} в зависимости от известной предыстории (X_1, n_1, X_2, n_2), где n_1, n_2 – текущие полные количества применений обоих действий $(n_1+n_2=n)$, X_1, X_2 – соответствующие полные доходы. Таким образом $\sigma_\ell(X_1, n_1, X_2, n_2) = \Pr(y_{n+1} = \ell \,|\, X_1, n_1, X_2, n_2)$, $\ell = 1, 2$.

Для формулировки цели управления определим функцию потерь

$$L_N(\sigma, \theta) = N(m_1 \vee m_2) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N} \xi_n\right), \tag{1.1}$$

которая описывает математическое ожидание потерь полного дохода относительно максимально возможной величины вследствие неполноты информации. Здесь $(m_1 \vee m_2)$ – максимум из m_1, m_2, а $\mathbf{E}_{\sigma,\theta}$ обозначает знак математического ожидания, вычисленного по мере, порожденной стратегией σ и параметром θ. Обозначим через $\lambda(\theta)$ априорную плотность распределения параметра на множестве Θ. По функции потерь определяются минимаксный и байесовский риски

$$R_N^M(\Theta) = \inf_{\{\sigma\}} \sup_{\Theta} L_N(\sigma, \theta), \tag{1.2}$$

$$R_N^B(\lambda) = \inf_{\{\sigma\}} \int_{\Theta} L_N(\sigma, \theta)\lambda(\theta)\, d\theta, \tag{1.3}$$

соответствующие оптимальные стратегии называются минимаксной и байесовской стратегиями. Преимущество минимаксного подхода состоит в его независимости от априорного распределения, для выбора которого, как правило, нет ясных критериев. Однако прямых методов нахождения минимаксных стратегии и риска не существует. С другой стороны, байесовский подход дает возможность написать рекуррентное уравнение, позволяющее найти байесовские стратегию и риск методом динамического программирования. Объединяет минимаксный и байесовский подходы основная теорема теории игр, согласно которой при широких предположениях выполнено равенство

$$R_N^M(\Theta) = \sup_{\lambda} R_N^B(\lambda), \tag{1.4}$$

т.е. минимаксный риск (1.2) равен байесовскому риску (1.3), вычисленному относительно наихудшего априорного распределения, на котором байесовский риск максимален, а минимаксная стратегия совпадает с соответствующей байесовской. Именно этот подход для нахождения минимаксных стратегии и риска рассмотрен в [1]. Отметим, что другой подход к пакетной обработке, предполагающий использование небольшого числа пакетов возрастающего объема, в близких постановках независимо рассмотрен в [2, 3], где установлены, по существу, одинаковые оценки минимаксного риска в этом случае.

В данной статье изучаются предельные свойства минимаксных стратегии и риска, если число обрабатываемых пакетов неограниченно растет. Установлено, что при $N \to \infty$ существует предел для нормированного минимаксного риска $N^{-1/2} R_N^M(\Theta)$, что позволяет оценить качество пакетной обработки. Например, в случае разбиения данных на 50 пакетов нормированный минимаксный риск может быть лишь на 2% выше своего предельного значения. Особая роль в исследовании отводится обла-

сти "близких" распределений, удовлетворяющих условию $|m_1 - m_2| \leqslant 2cN^{-1/2}$, где $c > 0$ – достаточно большое фиксированное число. Наибольшие потери достигаются именно в этой области и, соответственно, в ней же предъявляются наиболее высокие требования к управлению.

Структура статьи следующая. В § 2 обсуждается роль "близких" распределений и предлагается класс стратегий, позволяющих отделить не "близкие" распределения за счет определения для них лучшего действия на коротком начальном этапе управления. В § 3 уравнение для вычисления байесовских стратегии и риска относительно класса распределений, к которому принадлежит наихудшее, дается в инвариантной форме с горизонтом управления, равным единице. Такое представление справедливо именно в области "близких" распределений. В § 4 показано, что если количество обрабатываемых пакетов неограниченно растет, то существует непрерывный предел у решения этого уравнения, а само инвариантное рекуррентное уравнение переходит в дифференциальное уравнение в частных производных второго порядка. Строгое доказательство этого результата требует некоторых дополнительных оценок гладкости предельного решения, которые пока не получены. Но численные эксперименты, представленные в § 5, подтверждают близость решений рекуррентного и дифференциального уравнений. Наконец, в § 6 показано, что в случае бернуллиевского двурукого бандита обработка данных по одному не позволяет уменьшить величину минимаксного риска в сравнении с пакетной обработкой, если количество данных неограниченно растет, поскольку предельные описания обработок пакетами и по одному даются одинаковыми дифференциальными уравнениями. Заключение содержится в § 7.

§ 2. "Близкие" распределения и их роль в управлении

Под "близкими" будем понимать распределения, определяемые множеством параметров $\Theta_{cN^{-1/2}} = \{\theta : |m_1 - m_2| \leqslant 2cN^{-1/2}\}$, где $c > 0$ – достаточно большая фиксированная константа. Ниже устанавливается, что при больших N для анализа максимальных потерь достаточно ограничиться только "близкими" распределениями. В частности, приводится известный результат о том, что максимальные потери, имеющие порядок $N^{1/2}$, достигаются именно в области "близких" распределений. Далее предложен класс стратегий, способных отделить "удаленные" распределения за счет определения для них на коротком начальном этапе лучшего действия и применения его затем до конца управления. Нормированный минимаксный риск $N^{-1/2}R_N^M(\Theta_C)$, обеспечиваемый этими стратегиями на множестве $\Theta_C = \{\theta : |m_1 - m_2| \leqslant 2C\}$, путем выбора подходящих параметров стратегии может быть сделан при больших N сколь угодно близким к нормированному риску $N^{-1/2}R_N^M(\Theta_{cN^{-1/2}})$ для некоторого $c > 0$.

Начнем с оценок максимальных потерь. Справедлива

Теорема 1. *В области "близких" распределений порядок максимальных потерь не ниже* $N^{1/2}$. *Если же известно, что* $|m_1 - m_2| \geqslant \delta > 0$, *то порядок максимальных потерь не превосходит* $\ln(N)$.

Доказательство. Для оценки снизу потерь в области "близких" распределений рассмотрим поведение гауссовского двурукого бандита, характеризуемого дисперсиями D_1, D_2 и более узким множеством параметров $\{\theta_1 = (D_1^{1/2}m, -D_2^{1/2}m), \theta_2 = (-D_1^{1/2}m, D_2^{1/2}m)\}$. Управление будем осуществлять на основе наблюдений процесса z_n, $n = 1, \ldots, N$, где $z_n = \xi_n D_1^{-1/2}$, если $y_n = 1$, и $z_n = -\xi_n D_2^{-1/2}$, если $y_n = 2$. Легко видеть, что при любой стратегии управления $\{z_n\}$ – независимые нормально распределенные случайные величины с единичной дисперсией и математическим ожиданием, равным m, если $\theta = \theta_1$, и $-m$, если $\theta = \theta_2$. Положим $Z_n = \sum\limits_{i=1}^{n} z_i$. То-

гда согласно лемме Неймана – Пирсона на n-м шаге принимаются гипотезы $m > 0$ и $m < 0$, если $Z_n > 0$ и $Z_n < 0$ соответственно.

Пусть для определенности $\theta = \theta_1$, причем $m = xN^{-1/2}$, $x > 0$. Обозначим через $\Phi(x) = \int\limits_{-\infty}^{x} f_1(t\,|\,0)\,dt$ функцию стандартного нормального распределения, и пусть $\widetilde{D}^{1/2} = 0{,}5(D_1^{1/2} + D_2^{1/2})$. Тогда одношаговые потери на n-м шаге равны

$$2m\widetilde{D}^{1/2}\Pr(Z_n < 0) = (\widetilde{D}/N)^{1/2}2x\Phi(-xN^{-1/2}n/n^{1/2}) = (\widetilde{D}N)^{1/2}2x\Phi(-xt^{1/2})\Delta t,$$

где $t = n/N$, $\Delta t = N^{-1}$. А потери на всем горизонте управления равны

$$(\widetilde{D}N)^{1/2}\left(2x\sum_{n=1}^{N}\Phi(-xt^{1/2})\Delta t\right) \sim (\widetilde{D}N)^{1/2}\left(2x\int\limits_{0}^{1}\Phi(-xt^{1/2})\,dt\right).$$

Максимальное значение выражения в скобках достигается при $x \approx 1{,}22$ и приблизительно равно 0,53. Поэтому порядок максимальных потерь в области "близких" распределений не ниже $N^{1/2}$.

Предположим теперь, что $|m_1 - m_2| \geqslant \delta > 0$, и рассмотрим стратегию, которая сначала применяет каждое из действий по $k = \varkappa\ln(N)$ раз, получая за это полные доходы X_1 и X_2, а затем оставшиеся $N - 2k$ раз применяет только то действие, которому соответствует больший из доходов X_1, X_2. Отметим, что $X_1 - X_2$ имеет нормальное распределение с параметрами $\mathbf{E}(X_1 - X_2) = k(m_1 - m_2)$, $\mathbf{D}(X_1 - X_2) = 2kD$, где $D = 0{,}5(D_1 + D_2)$. Если $m_1 > m_2$, то потери равны

$$L_N(m_1, m_2) = (m_1 - m_2)\,(k + \Pr(X_1 - X_2 < 0)(N - 2k)) =$$
$$= (m_1 - m_2)\left(k + \Phi\left(-(m_1 - m_2)(0{,}5k/D)^{1/2}\right)(N - 2k)\right).$$

С использованием оценки (см. [4, с. 43])

$$\Phi(-x) < (2\pi)^{-1/2}x^{-1}\exp(-x^2/2) < \exp(-x^2/2) \quad \text{при } x > 1 \tag{2.1}$$

получаем, что

$$L_N(m_1, m_2) < (m_1 - m_2)\left(\varkappa\ln(N) + N^{1-0{,}25(m_1-m_2)^2\varkappa D^{-1}}\right) =$$
$$= (m_1 - m_2)\varkappa\ln(N) + \alpha_N,$$

где $\alpha_N \to 0$ при $N \to \infty$, если $0{,}25\varkappa\delta^2 D^{-1} > 1$. ▲

Отметим, что в случае равных дисперсий порядок оценки $N^{1/2}$ дан ранее в [5,6], а оценки $\ln(N)$ – в [7]. Логарифмическая оценка порядка роста максимальных потерь справедлива и для ряда других стратегий (см., например, [8,9]).

Установим, что выбором подходящей стратегии можно перевести управление в область "близких" распределений. Для некоторой стратегии σ, действующей на горизонте управления $[2n_0 + 1, N]$, где $2n_0 < N$, определим стратегию σ^*, действующую на всем горизонте $[1, N]$. Эта стратегия на начальном этапе управления $[1, 2n_0]$ применяет действия по очереди до тех пор, пока абсолютная разность полных доходов за их применение не превысит величины порога $\widetilde{\alpha}N^{1/2}$ или не закончится начальный этап управления ($\widetilde{\alpha} = \alpha D^{1/2}$, $D = 0{,}5(D_1 + D_2)$, $\alpha > 0$). Если величина порога превышена на начальном этапе в момент времени $2k_0$ ($2k_0 \leqslant 2n_0$), то на заключительном этапе $[2k_0 + 1, N]$ применяется только то действие, которому соответствует бóльшая величина полного дохода в момент времени $2k_0$. Если же до конца начального этапа порог не был достигнут, то на заключительном этапе управления $[2n_0 + 1, N]$ применяется стратегия управления σ.

Покажем, что выбором подходящих n_0 и α эта стратегия позволяет перевести управление в область "близких" распределений. Нам потребуются оценки, представленные ниже в леммах 1–3.

Лемма 1. *Пусть* $m_1 - m_2 = b$, $0 < b < \infty$, *величины порогов равны* $\pm a$, $0 < a < \infty$. *Тогда вероятность достижения порога в точке* $-a$ *не превосходит величины*

$$P_e(a,b) \leqslant \frac{e^{-ab/D}}{1+e^{-ab/D}}. \tag{2.2}$$

Доказательство. Обозначим через $z_k = (\xi_{2k-1} - \xi_{2k})$ текущие разности доходов при применении пороговой стратегии, распределенные с плотностями $(4\pi D)^{-1/2}\exp(-(z_k-b)^2/(4D))$, $k = 1,\dots,n_0$. Следуя [5], обозначим $u_k = \sum\limits_{i=1}^{k} z_i$ и введем подмножества

$$A_{1,k} = \{|u_j| < a \text{ при } 1 \leqslant j < k \text{ и } u_k \geqslant a\},$$
$$A_{2,k} = \{|u_j| < a \text{ при } 1 \leqslant j < k \text{ и } u_k \leqslant -a\}, \quad B = \{|u_j| < a \text{ при } 1 \leqslant j \leqslant n_0\}.$$

Принадлежность процесса подмножествам $A_{1,k}$, $A_{2,k}$ означает, что пороги a и $-a$ впервые были достигнуты в момент времени $2k$, попадание в подмножество B означает, что ни один из этих порогов не был достигнут до конца начального этапа. Ясно, что

$$\Pr(A_{1,k}) = \int\limits_{A_{1,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{(z_i-b)^2}{4D}} dz^k =$$
$$= \int\limits_{A_{1,k}} (4\pi D)^{-k/2} e^{\frac{bu_k}{2D}} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k \geqslant e^{\frac{ab}{2D}} \int\limits_{A_{1,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k. \tag{2.3}$$

Здесь $dz^k = dz_1 \dots dz_k$. Аналогично,

$$\Pr(A_{2,k}) =$$
$$= \int\limits_{A_{2,k}} (4\pi D)^{-k/2} e^{\frac{bu_k}{2D}} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k \leqslant e^{-\frac{ab}{2D}} \int\limits_{A_{2,k}} (4\pi D)^{-k/2} \prod_{i=1}^{k} e^{-\frac{z_i^2+b^2}{4D}} dz^k. \tag{2.4}$$

Так как подмножества $A_{1,k}$, $A_{2,k}$ переходят одно в другое при преобразовании координат $z_i \leftarrow -z_i$, $i = 1,\dots,k$, то значения интегралов в (2.3), (2.4) одинаковы, и следовательно,

$$\frac{\Pr(A_{2,k})}{\Pr(A_{1,k})} \leqslant e^{-ab/D}. \tag{2.5}$$

Обозначим через $A_1 = \bigcup\limits_{k=1}^{n_0} A_{1,k}$, $A_2 = \bigcup\limits_{k=1}^{n_0} A_{2,k}$ события, состоящие в том, что были достигнуты пороги a и $-a$ соответственно. Так как все события $A_{1,k}$, $A_{2,k}$, $k = 1,\dots,n_0$, попарно несовместны, то из (2.5) следует, что

$$\frac{\Pr(A_2)}{\Pr(A_1)} \leqslant e^{-ab/D}.$$

Так как $P_e(a,b) = \Pr(A_2)$ и $\Pr(A_1) + \Pr(A_2) = 1 - \Pr(B) < 1$, то формула (2.2) справедлива. ▲

Лемма 2. *Пусть* $m_1 - m_2 = 2\beta(D/N)^{1/2}$, $0 < c < \infty$, $\Pr(\overline{B}) = 1 - \Pr(B)$. *Справедливы оценки*

$$L_N(\sigma^*, \theta) \leqslant 2c \Pr(\overline{B})(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \Pr(B) \tag{2.6}$$

при $|\beta| \leqslant c$ *и*

$$L_N(\sigma^*, \theta) \leqslant \left(\alpha + \max_{\beta \geqslant c}(2\beta e^{-2\alpha\beta}) + o(1)\right)(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \Pr(B) \tag{2.7}$$

при $|\beta| \geqslant c$.

Доказательство. Оценка (2.6) следует из того, что применение одного и того же действия на оставшемся горизонте управления после достижения одного из порогов, которое происходит с вероятностью $\Pr(\overline{B})$, не может дать потери, бóльшие чем $2c(DN)^{1/2}$. При этом применение стратегии σ на заключительном этапе происходит с вероятностью $\Pr(B)$.

Докажем оценку (2.7). Положим $a = \widetilde{\alpha}N^{1/2}$, $b = 2\widetilde{\beta}N^{-1/2}$, где $\widetilde{\beta} = \beta D^{1/2}$ и без ограничения общности будем считать, что $\beta > 0$. Через $k^* = \min(k : |u_k| \geqslant \widetilde{\alpha}N^{1/2})$ обозначим момент достижения одного из порогов и положим $k_0 = \min(k^*, n_0)$. Тогда ожидаемые потери допускают оценку

$$L_N(\sigma^*, \theta) \leqslant L_N^{(1)}(\alpha, \beta) + L_N^{(2)}(\alpha, \beta) + L_{N-2n_0}(\sigma, \theta) \Pr(B), \tag{2.8}$$

где $L_N^{(1)}(\alpha, \beta), L_N^{(2)}(\alpha, \beta)$ возникают в случае достижения одного из порогов и равны

$$\begin{aligned} &L_N^{(1)}(\alpha, \beta) = \\ &= \mathbf{E}\left(\sum_{n=1}^{2k_0}(m_1 - \xi_n)\right) = 2\,\mathbf{E}\left(\sum_{k=1}^{k_0}(m_1 - \xi_{2k-1})\right) + \mathbf{E}\left(\sum_{k=1}^{k_0}(\xi_{2k-1} - \xi_{2k})\right), \\ &L_N^{(2)}(\alpha, \beta) = \mathbf{E}\left(\sum_{n=2k_0+1}^{N}(m_1 - \xi_n)\right). \end{aligned} \tag{2.9}$$

Так как $\sum\limits_{k=1}^{n}(m_1 - \xi_{2k-1})$, $n = 1, 2, \ldots,$ – мартингал, а k_0 – момент остановки, то

$$\mathbf{E}\left(\sum_{k=1}^{k_0}(m_1 - \xi_{2k-1})\right) = 0,$$

поэтому из первого равенства в (2.9) следует, что

$$L_N^{(1)}(\alpha, \beta) = \mathbf{E}\left(\sum_{k=1}^{k_0}(\xi_{2k-1} - \xi_{2k})\right) \leqslant (\widetilde{\alpha} + o(1))N^{1/2}, \tag{2.10}$$

где $o(1)$ вызвано превышением порога на последнем шаге. Из (2.2) и второго равенства в (2.9) следует оценка

$$L_N^{(2)}(\alpha, \beta) \leqslant (2\widetilde{\beta})N^{-1/2}\,\mathbf{E}(N - 2k_0)P_e(a, b) \leqslant \max_{\beta \geqslant c}(2\beta e^{-2\alpha\beta})(DN)^{1/2}. \tag{2.11}$$

Из (2.8), (2.10), (2.11) следует справедливость оценки (2.7). ▲

Сделаем несколько замечаний. Пороговая стратегия, действующая на всем горизонте управления, т.е. при $2n_0 = N$, впервые предложена в [5,6] для бернуллиевского

двурукого бандита. Лемма 1 использует идею, предложенную в [5], остальные представленные здесь оценки оригинальны.

Оценка (2.7) при $2n_0 = N$ и $c = 0$ соответствует тому, что пороговая стратегия применяется ко всем двуруким бандитам на всем горизонте управления. В этом случае максимальные потери не превосходят $\Big(\alpha + \max\limits_{\beta \geqslant 0}(2\beta e^{-2\alpha\beta}) + o(1)\Big)(DN)^{1/2} \leqslant$ $\leqslant (\alpha + (\alpha e)^{-1} + o(1))(DN)^{1/2}$ и достигаются при $2\alpha\beta = 1$, т.е. в области "близких" распределений. При $\alpha = e^{-1/2}$ максимальные потери асимптотически не превосходят величины $2e^{-1/2}(DN)^{1/2} \approx 1{,}213(DN)^{1/2}$. В [5, 6] дана более точная оценка множителя 0,752 (при $D = 0{,}25$), но представленные здесь выкладки значительно короче.

Наконец отметим, что если c достаточно велико, то $\max\limits_{\beta \geqslant c}(2\beta e^{-2\alpha\beta}) = 2ce^{-2\alpha c}$.

Для дальнейшего нам потребуется оценка вероятности $\Pr(\overline{B})$, где $\overline{B}$ означает дополнение события B. Здесь также используется идея из [5].

Лемма 3. *Пусть* $a = \alpha(DN)^{1/2}$, $\alpha > 0$, *и* $m_1 - m_2 = 2\beta(D/N)^{1/2}$, $|\beta| \leqslant c < \infty$. *Пусть оценка вероятности* $\Pr(\overline{B})$ *выполняется на момент времени* $2tN$. *Тогда при* $N \to \infty$ *справедлива оценка*

$$\Pr(\overline{B}) \leqslant \Phi\left(\frac{-2\beta t - \alpha}{(2t)^{1/2}}\right)(1 + e^{2\alpha\beta}) + \Phi\left(\frac{2\beta t - \alpha}{(2t)^{1/2}}\right)(1 + e^{-2\alpha\beta}). \tag{2.12}$$

Доказательство. Последовательность случайных величин

$$s_t = N^{-1/2}\sum_{\nu=1}^{tN}(\xi_{2\nu-1} - \xi_{2\nu}), \quad \text{где } t = nN^{-1},\ n = 1, 2, \ldots,$$

описывает одномерное случайное блуждание с дисперсией шага $2DN^{-1}$ и сносом $2\widetilde{\beta}N^{-1}$, где $\widetilde{\beta} = \beta D^{1/2}$. Рассмотрим случайное блуждание, начинающееся из точки $x = 0$ с поглощающим экраном в точке $x = -\widetilde{\alpha}$, где $\widetilde{\alpha} = \alpha D^{1/2}$. При $N \to \infty$ плотность распределения s_t слабо сходится к плотности $f(x, t)$, $0 \leqslant t \leqslant 0{,}5$, $x \geqslant -\widetilde{\alpha}$, удовлетворяющей уравнению Фоккера – Планка – Колмогорова

$$f'_t = -2\widetilde{\beta}f'_x + Df''_{xx}$$

с начальным условием $f(x, 0) = \delta(x)$ и граничным условием $f(-\widetilde{\alpha}, t) = 0$, $0 \leqslant t \leqslant 0{,}5$, где $\delta(x)$ – дельта-функция Дирака. Решение этого уравнения имеет вид

$$f(x,t) = \frac{1}{(4\pi Dt)^{1/2}}\exp\left(-\frac{(x - 2\widetilde{\beta}t)^2}{4Dt}\right) - \frac{e^{-2\widetilde{\alpha}\widetilde{\beta}/D}}{(4\pi Dt)^{1/2}}\exp\left(-\frac{(x + 2\widetilde{\alpha} - 2\widetilde{\beta}t)^2}{4Dt}\right).$$

Поэтому вероятность того, что случайное блуждание не было поглощено до момента времени t, есть

$$P_1(\alpha,\beta) = \int\limits_{-\widetilde{\alpha}}^{\infty} f(x,t)\,dx = \Phi\left(\frac{2\widetilde{\beta}t + \widetilde{\alpha}}{(2Dt)^{1/2}}\right) - e^{-2\widetilde{\alpha}\widetilde{\beta}/D}\Phi\left(\frac{2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right).$$

Так как $\Phi(x) + \Phi(-x) = 1$ при всех x, то вероятность $P_2(\alpha, \beta)$ поглощения до момента времени t равна

$$P_2(\alpha,\beta) = 1 - P_1(\alpha,\beta) = \Phi\left(\frac{-2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right) + e^{-2\widetilde{\alpha}\widetilde{\beta}/D}\Phi\left(\frac{2\widetilde{\beta}t - \widetilde{\alpha}}{(2Dt)^{1/2}}\right).$$

Ясно, что для случайного блуждания, начинающегося из точки $x = 0$, с двумя поглощающими экранами в точках $x = -\widetilde{\alpha}$ и $x = \widetilde{\alpha}$ вероятность поглощения в точке $x = -\widetilde{\alpha}$ не превосходит $P_2(\alpha, \beta)$. В силу симметрии вероятность поглощения в точке $x = \widetilde{\alpha}$ не превосходит $P_2(\alpha, -\beta)$. Поэтому $\Pr(\overline{B}) \leqslant P_2(\alpha, \beta) + P_2(\alpha, -\beta)$, откуда следует (2.12). ▲

Теорема 2. *Положим $n_0 = \varepsilon_0 N$ и выберем параметры стратегии σ^* следующим образом. По заданным малым $\alpha > 0$ и $\delta > 0$ определим достаточно большое $c > 0$ так, чтобы $\max\limits_{\beta \geqslant c} 2\beta e^{-2\alpha\beta} = 2ce^{-2\alpha c} = \delta$. После этого определим ε_0 из условия $2c\varepsilon_0 = 0{,}1\alpha$. Тогда при $N \to \infty$ справедливы оценки*

$$L_N(\sigma^*, \theta) \leqslant (2c)^{-1}\delta^2 (DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \tag{2.13}$$

при $|\beta| \leqslant c$,

$$L_N(\sigma^*, \theta) \leqslant (\alpha + \delta + o(1))\,(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta) \tag{2.14}$$

при $c \leqslant |\beta| \leqslant 20c$ и

$$L_N(\sigma^*, \theta) \leqslant \left(\alpha + 20\delta^{20}(2c)^{-19} + 20\delta^{5/2}(2c)^{-3/2} + o(1)\right)(DN)^{1/2} \tag{2.15}$$

при $|\beta| \geqslant 20c$.

Доказательство. Без ограничения общности считаем, что $\beta > 0$. Докажем оценку (2.13). Рассмотрим оценку (2.6), в которой $\Pr(\overline{B})$ оценивается из (2.12) при $t = \varepsilon_0$. Отметим, что $\exp(-(\alpha + 2\beta\varepsilon_0)^2/(4\varepsilon_0))e^{2\alpha\beta} = \exp(-(\alpha - 2\beta\varepsilon_0)^2/(4\varepsilon_0))$, причем при выбранных параметрах $(\alpha - 2\beta\varepsilon_0)(2\varepsilon_0)^{-1/2} \gg 1$. С учетом (2.1), (2.12) отсюда следует, что $\Pr(\overline{B}) < \exp(-(\alpha - 2\beta\varepsilon_0)^2/(4\varepsilon_0))$. Отметим, что $\alpha - 2\beta\varepsilon_0 > 0{,}9\alpha$ и $\alpha = 20c\varepsilon_0$. Поэтому $\Pr(\overline{B}) < \exp(-0{,}81\alpha^2/(4\varepsilon_0)) = \exp(-0{,}81 \times 5\alpha c) < \exp(-4\alpha c)$. Из (2.6) следует, что в этом случае $L_N(\sigma^*, \theta) \leqslant 2ce^{-4\alpha c}(DN)^{1/2} + L_{N-2n_0}(\sigma, \theta)$. Так как $e^{-2\alpha c} = (2c)^{-1}\delta$, то оценка (2.14) выполнена.

Пусть теперь $c \leqslant \beta \leqslant 20c$. Оценка (2.14) следует из (2.7), так как $\Pr(B) \leqslant 1$.

Чтобы установить (2.15), используем (2.7) при $\beta \geqslant 20c$, при этом $\max\limits_{\beta \geqslant 20c} 2\beta e^{-2\alpha\beta} = 40ce^{-40\alpha c} = 20\delta(2c)^{-19}\delta^{19}$. Требуется дополнительно оценить $L_{N-2n_0}(\sigma, \theta)\Pr(B)$. При выбранных параметрах $\Pr(B) < \Phi(-(2\beta\varepsilon_0 - \alpha)/(2\varepsilon_0)^{1/2})$, причем $\beta\varepsilon_0 > \alpha$. С учетом (2.1) получаем оценку $\Pr(B) < \exp\left(-(2\beta\varepsilon_0 - \alpha)^2/(4\varepsilon_0)\right) < \exp\left(-\beta^2\varepsilon_0/4\right) < e^{-\alpha\beta/4}$. Так как всегда верно $L_{N-2n_0}(\sigma, \theta) \leqslant 2\beta(DN)^{1/2}$, то $L_{N-2n_0}(\sigma, \theta)\Pr(B) \leqslant (DN)^{1/2} \max\limits_{\beta \geqslant 20c}(2\beta e^{-\alpha\beta/4})$. При выбранных параметрах имеем $\max\limits_{\beta \geqslant 20c}(2\beta e^{-\alpha\beta/4}) = 40ce^{-5\alpha c} = 20\delta(2c)^{-3/2}\delta^{3/2}$. ▲

Следствие 1. *Обозначим $R_N^{*M}(\Theta) = \inf\limits_{\{\sigma^*\}} \sup\limits_{\Theta} L_N(\sigma, \theta)$. Ясно, что $R_N^M(\Theta_C) \leqslant R_N^{*M}(\Theta_C)$. Из теоремы следует, что для любого $\varepsilon > 0$ с помощью выбора подходящих α, c и ε_0 можно при больших N обеспечить выполнение неравенства $R_N^{*M}(\Theta_C) \leqslant R_{N-2n_0}^M(\Theta_{cN^{-1/2}}) + \varepsilon(DN)^{1/2}$. Так как всегда верно $R_{N-2n_0}^M(\Theta_{cN^{-1/2}}) \leqslant R_N^M(\Theta_{cN^{-1/2}})$, а $R_N^M(\Theta_{cN^{-1/2}}) \leqslant R_N^M(\Theta_C)$ при $cN^{-1/2} < C$, то из теоремы следует, что*

$$\limsup_{N\to\infty} (DN)^{-1/2} \left| R_N^M\left(\Theta_{cN^{-1/2}}\right) - R_N^M\left(\Theta_C\right)\right| \leqslant \varepsilon.$$

Поэтому в следующих параграфах анализ ведется в области "близких" распределений. Приведем примеры выбора значений α, c, ε_0. Пусть $\alpha = \delta = 0{,}01$. Тогда $2c \approx 1170$, $\varepsilon_0 \approx 0{,}85 \times 10^{-6}$. Если $\alpha = \delta = 0{,}001$, то $2c \approx 16700$, $\varepsilon_0 \approx 0{,}6 \times 10^{-8}$.

Это говорит о том, что оценки (2.13)–(2.15) носят скорее теоретический характер. Практически приемлемые параметры можно выбрать с помощью математического моделирования, один такой пример приведен в [10].

Таким образом, при больших N имеет место парадоксальная ситуация: наиболее высокие требования к управлению надо обеспечить там, где, казалось бы, оно не требуется совсем, т.е. в области "близких" распределений. Покажем теперь, что использование пакетной обработки позволяет выйти из области "близких" распределений. Рассмотрим двурукий бандит с доходами ζ_t, $t = 1, \dots, T$, характеризуемыми математическими ожиданиями m_1, m_2 и дисперсиями D_1, D_2 в случае применения первого и второго действий, причем $m_1 - m_2 = xT^{-1/2}$. Пусть $T = NM$. Будем применять одинаковые действия к пакетам из M данных, для управления используем значения процесса $\xi_n = M^{-1/2} \sum\limits_{t=(n-1)M+1}^{nM} \zeta_t$, $n = 1, \dots, N$. Тогда $m'_\ell = \mathbf{E}(\xi_n \,|\, y_n = \ell) = M^{1/2} m_\ell$, $D'_\ell = \mathbf{D}(\xi_n \,|\, y_n = \ell) = D_\ell$, $\ell = 1, 2$, и следовательно, $m'_1 - m'_2 = xN^{-1/2}$. Если число пакетов и полное число данных связаны соотношением $N \ll T$, то распределения доходов в пакетах уже не будут "близкими" относительно T. При этом в [1] установлено, что пакетная обработка практически не приводит к увеличению минимаксного риска, если количество пакетов достаточно велико.

§ 3. Инвариантное описание

Далее будем рассматривать управление в области "близких" распределений и на всем горизонте $[1, N]$. Получим рекуррентное уравнение для нахождения байесовских стратегии и риска относительно класса априорных распределений, к которому принадлежит наихудшее, в инвариантной форме с горизонтом управления, равным единице. Для вычислений удобно поменять параметризацию следующим образом: $m_1 = m + v$, $m_2 = m - v$, тогда множество "близких" распределений принимает вид $\Theta_{cN^{-1/2}} = \{\theta : |v| \leqslant cN^{-1/2}\}$. В [1] установлено, что асимптотически наихудшая априорная плотность распределения может быть выбрана в виде

$$\nu_a(m, v) = \varkappa_a(m)\rho(v), \tag{3.1}$$

где $\varkappa_a(m) = (2a)^{-1}$ – плотность равномерного распределения на отрезке $|m| \leqslant a$, причем $a \to \infty$. Плотность $\rho(v)$, соответствующая наихудшему распределению вида (3.1), может быть найдена из условия (1.4). В случае $D_1 = D_2$ плотность $\rho(v)$ может быть выбрана симметричной, т.е. $\rho(v) = \rho(-v)$.

Рассмотрим стратегии пакетной обработки, которые сначала применяют действия поровну к $2M_0$ данным, а затем управляют оптимально, применяя действия к пакетам из M данных. Пусть n_1, n_2 – полные количества применений обоих действий к моменту времени $n = n_1 + n_2$, а X_1, X_2 – полные доходы за применение обоих действий. Обозначим $n^*_\ell = n_\ell D_\ell$, $n'_\ell = n_\ell / D_\ell$, $M^*_\ell = MD_\ell$, $M'_\ell = M/D_\ell$, $\ell = 1, 2$, а также $U = (X_1 n_2 - X_2 n_1)/n'$, где $n' = n'_1 + n'_2$. В [1] получено рекуррентное уравнение (см. [1, теорема 2, формулы (3.8)–(3.11)]), позволяющее найти байесовские стратегию и риск, соответствующие априорной плотности вида (3.1) с помощью рекуррентного вычисления рисков $R_M(U, n_1, n_2)$. Для получения рекуррентного уравнения в инвариантной форме с горизонтом управления, равным единице, положим

$$\begin{aligned}
&C = cN^{-1/2}, \quad w = N^{1/2}v, \quad \varrho(w) = N^{-1/2}\rho(v), \quad u = UN^{-1/2},\\
&t_\ell = n_\ell N^{-1}, \quad t^*_\ell = n^*_\ell N^{-1}, \quad t'_\ell = n'_\ell N^{-1}, \quad t = t_1 + t_2, \quad t' = t'_1 + t'_2,\\
&r_\varepsilon(u, t_1, t_2) = N^{-1/2} R_M(U, n_1, n_2), \quad r^{(\ell)}_\varepsilon(u, t_1, t_2) = N^{-1/2} R^{(\ell)}_M(U, n_1, n_2),\\
&\varepsilon_0 = M_0 N^{-1}, \quad \varepsilon = MN^{-1}, \quad \varepsilon^*_\ell = M^*_\ell N^{-1}, \quad \varepsilon'_\ell = M'_\ell N^{-1}, \quad \ell = 1, 2.
\end{aligned} \tag{3.2}$$

Кроме того, обозначим $f_D(u) = f_D(u\,|\,0)$, $D_g^2 = D_1 D_2$, $D_h^{-1} = 0{,}5(D_1^{-1} + D_2^{-1})$. Справедлива

Теорема 3. *Пусть априорная плотность распределения имеет вид* (3.1) *и* $a \to \infty$. *Байесовские стратегия и риск могут быть найдены в результате решения рекуррентного уравнения динамического программирования*

$$r_\varepsilon(u, t_1, t_2) = \min_{\ell=1,2} r_\varepsilon^{(\ell)}(u, t_1, t_2), \tag{3.3}$$

где $r_\varepsilon^{(1)}(u, t_1, t_2) = r_\varepsilon^{(2)}(u, t_1, t_2) = 0$ *при* $t_1 + t_2 = 1$, *и далее*

$$\begin{aligned} r_\varepsilon^{(1)}(u, t_1, t_2) &= \varepsilon g^{(1)}(u, t_1, t_2) + r_\varepsilon(u, t_1 + \varepsilon, t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u), \\ r_\varepsilon^{(2)}(u, t_1, t_2) &= \varepsilon g^{(2)}(u, t_1, t_2) + r_\varepsilon(u, t_1, t_2 + \varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1} (t'+\varepsilon_2')^{-1}}(u) \end{aligned} \tag{3.4}$$

при $2\varepsilon_0 \leqslant t_1 + t_2 < 1$. *Здесь* $r(u) * f(u) = \int\limits_{-\infty}^{\infty} r(u - x) f(x)\, dx$ *означает свертку функций,*

$$\begin{aligned} g^{(1)}(u, t_1, t_2) &= \int\limits_{-c}^{0} 2|w| g(w; u, t_1, t_2) \varrho(w)\, dw, \\ g^{(2)}(u, t_1, t_2) &= \int\limits_{0}^{c} 2w g(w; u, t_1, t_2) \varrho(w)\, dw, \\ g(w; u, t_1, t_2) &= \exp\left(2D_g^{-2}(uw - w^2 t_1 t_2 (t')^{-1})\right). \end{aligned} \tag{3.5}$$

Байесовская стратегия при $t \leqslant 2\varepsilon_0$ $(n \leqslant 2M_0)$ *применяет действия по очереди. При* $t > 2\varepsilon_0$ $(n > 2M_0)$ *на промежутке времени* $(t, t + \varepsilon]$ (*иначе говоря,* $tN < n \leqslant tN + M$) *применяется то действие, которому соответствует меньшая величина* $r_\varepsilon^{(\ell)}(u, t_1, t_2)$; *при равенстве* $r_\varepsilon^{(1)}(u, t_1, t_2) = r_\varepsilon^{(2)}(u, t_1, t_2)$ *выбор может быть произвольным. Тогда байесовский риск, соответствующий* (3.1), *вычисляется по формуле*

$$N^{-1/2} R_N^B(\rho(v)) = l(\varrho(w)) + \int\limits_{-\infty}^{\infty} f_{0,5\varepsilon_0 D_g^2 D_h}(u) r_\varepsilon(u, \varepsilon_0, \varepsilon_0)\, du, \tag{3.6}$$

где

$$l(\varrho(w)) = \varepsilon_0 \int\limits_{-c}^{c} 2|w| \varrho(w)\, dw.$$

Доказательство. Теорема доказывается путем выполнения замены переменных (3.2) в уравнениях (3.8)–(3.11) из [1]. ▲

Отметим, что при различных горизонтах управления N уравнение в инвариантной форме может оказаться одинаковым, так как оно определяется не горизонтом управления и размерами начального и последующих пакетов M_0, M, а относительными размерами пакетов ε_0, ε.

Из теоремы 3 следует, что оптимальная стратегия всегда выбирает либо первое, либо второе действие. Важным свойством стратегии является ее пороговый характер. Положим $\Delta r_\varepsilon(u, t_1, t_2) = r_\varepsilon^{(1)}(u, t_1, t_2) - r_\varepsilon^{(2)}(u, t_1, t_2)$. Ясно, что критериями выбора первого и второго действий являются условия $\Delta r_\varepsilon(u, t_1, t_2) < 0$ и $\Delta r_\varepsilon(u, t_1, t_2) > 0$. Обозначим $G^+(\cdot) = \max(G(\cdot), 0)$, $G^-(\cdot) = \max(-G(\cdot), 0)$.

Теорема 4. *Справедливо представление* $\Delta r_\varepsilon(u,t_1,t_2) = \varepsilon G_\varepsilon(u,t_1,t_2)$, *где* $G_\varepsilon(u,t_1,t_2) = G_\varepsilon^{(1)}(u,t_1,t_2) - G_\varepsilon^{(2)}(u,t_1,t_2)$. *При этом* $G_\varepsilon^{(\ell)}(u,t_1,t_2) = g^{(\ell)}(u,t_1,t_2)$ *при* $t_1+t_2 = 1-\varepsilon$ *и далее* $G_\varepsilon^{(1)}(u,t_1,t_2) = G_\varepsilon^{+}(u,t_1,t_2+\varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u)$, $G_\varepsilon^{(2)}(u,t_1,t_2) = G_\varepsilon^{-}(u,t_1+\varepsilon,t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u)$ *при* $t_1+t_2 < 1-\varepsilon$, $\ell = 1,2$. *Поэтому* $\Delta r_\varepsilon(u,t_1,t_2)$ *монотонно убывает по* u *так, что* $\Delta r_\varepsilon(-\infty,t_1,t_2) = +\infty$, $\Delta r_\varepsilon(+\infty,t_1,t_2) = -\infty$. *Следовательно, существуют пороги* $\{T_\varepsilon(t_1,t_2)\}$, *такие что* $\Delta r_\varepsilon(u,t_1,t_2) < 0$ *при* $u > T_\varepsilon(t_1,t_2)$, $\Delta r_\varepsilon(u,t_1,t_2) > 0$ *при* $u < T_\varepsilon(t_1,t_2)$.

В случае $D_1 = D_2 = 1$ эквивалентная теорема доказана в [11] (см. [11, теорема 2]). Доказательство легко переносится на рассматриваемый случай.

Замечание 1. Из (3.3)–(3.5) следует, что все риски $\{r_\varepsilon(u,t_1,t_2)\}$ непрерывны по u, следовательно, все $\{r_\varepsilon^{(\ell)}(u,t_1,t_2)\}$ – бесконечно дифференцируемы по u. Из теоремы 4 дополнительно следует, что первые производные $\{(r_\varepsilon)'_u(u,t_1,t_2)\}$ имеют ровно по одному скачку порядка ε.

Наконец, дадим инвариантное рекуррентное уравнение для вычисления потерь. Ограничимся стратегиями σ, которые к первым $2M_0$ данным применяют оба действия равное количество раз, а затем одинаковые действия применяют к пакетам из M данных в зависимости от текущей предыстории U, n_1, n_2, так как именно к этому классу принадлежит оптимальная стратегия.

Теорема 5. *Пусть* $\sigma_\ell(u,t_1,t_2)$ *обозначает стратегию, т.е.* $\sigma_\ell(u,t_1,t_2) = \Pr(y_{(t,t+\varepsilon]} = \ell \,|\, u,t_1,t_2)$, $\ell = 1,2$. *В условиях теоремы* 3 *ожидаемые потери вычисляются в соответствии с уравнением*

$$l_\varepsilon(u,t_1,t_2) = \sigma_1(u,t_1,t_2) l_\varepsilon^{(1)}(u,t_1,t_2) + \sigma_2(u,t_1,t_2) l_\varepsilon^{(2)}(u,t_1,t_2), \tag{3.7}$$

где $l_\varepsilon^{(1)}(u,t_1,t_2) = l_\varepsilon^{(2)}(u,t_1,t_2) = 0$ *при* $t_1+t_2 = 1$,

$$\begin{aligned} l_\varepsilon^{(1)}(u,t_1,t_2) &= \varepsilon g^{(1)}(u,t_1,t_2) + l_\varepsilon(u,t_1+\varepsilon,t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u), \\ l_\varepsilon^{(2)}(u,t_1,t_2) &= \varepsilon g^{(2)}(u,t_1,t_2) + l_\varepsilon(u,t_1,t_2+\varepsilon) * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u), \end{aligned} \tag{3.8}$$

если $t_1+t_2 < 1$, $t_1 \geqslant \varepsilon_0$ *и* $t_2 \geqslant \varepsilon_0$. *Ожидаемые потери вычисляются в соответствии с формулой*

$$N^{-1/2} L_N(\sigma;\rho(v)) = l(\varrho(w)) + \int\limits_{-\infty}^{\infty} f_{0,5\varepsilon_0 D_g^2 D_h}(u) l_\varepsilon(u,\varepsilon_0,\varepsilon_0)\, du. \tag{3.9}$$

Доказательство. Теорема доказывается путем выполнения замены переменных (3.2) в уравнениях (3.15)–(3.17) из [1]. Дополнительно выполняются замены $l_\varepsilon(u,t_1,t_2) = N^{-1/2} L_M(U,n_1,n_2)$, $l_\varepsilon^{(\ell)}(u,t_1,t_2) = N^{-1/2} L_M^{(\ell)}(U,n_1,n_2)$, $\ell = 1,2$. ▲

§ 4. Предельное описание

В этом параграфе устанавливаются условия Липшица для риска $r_\varepsilon(u,t_1,t_2)$ по аргументам u,t_1,t_2 и его предельное описание дифференциальным уравнением в частных производных второго порядка. Кроме того, устанавливается упомянутый в § 1 результат о том, что малое изменение дисперсий приводит к малому изменению потерь. Далее без ограничения общности считаем, что $\max(D_1,D_2) = 1$, $\min(D_1,D_2) = \alpha > 0$ (см. [1, теорема 4]).

Положим $c_1 = c\alpha^{-1}$, $k(u) = \min(k^{(1)}(u), k^{(2)}(u))$, где

$$k^{(1)}(u) = \int\limits_{-c}^{0} 2|w| \exp\left(2D_g^{-2}uw\right) \varrho(w)\, dw, \quad k^{(2)}(u) = \int\limits_{0}^{c} 2w \exp\left(2D_g^{-2}uw\right) \varrho(w)\, dw.$$

Очевидно, что $k(u)$ – ограниченная функция, причем $k(u) \to +0$ при $u \to \pm\infty$, а $g^{(\ell)}(u, t_1, t_2) \leqslant k^{(\ell)}(u)$ при всех допустимых u, t_1, t_2 и $\ell = 1, 2$.

Лемма 4. *Справедливы оценки*

$$\begin{aligned} &k(u) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant e^{2cc_1\varepsilon} k(u), \\ &k(u) * f_{\varepsilon_2^* t_1^2 (t')^{-1} (t'+\varepsilon_2')^{-1}}(u) \leqslant e^{2cc_1\varepsilon} k(u). \end{aligned} \tag{4.1}$$

Доказательство. Непосредственно проверяется равенство $\exp(Cu) * f_\varepsilon(u) = \exp(Cu) \times \exp(0{,}5C^2\varepsilon)$, где C – функция, не зависящая от u. Проверим первую оценку в (4.1). Действительно,

$$\begin{aligned} &k^{(\ell)}(u) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant \exp\left(0{,}5 \cdot 4 (D_1 D_2)^{-2} c^2 \varepsilon D_1 D_2^2\right) k^{(\ell)}(u) \leqslant \\ &\leqslant e^{2cc_1\varepsilon} k^{(\ell)}(u). \end{aligned} \tag{4.2}$$

Далее, так как $k(u) \leqslant k^{(\ell)}(u)$, то $k(u) * f_\varepsilon(u) \leqslant k^{(\ell)}(u) * f_\varepsilon(u)$, $\ell = 1, 2$, и следовательно, $k(u) * f_\varepsilon(u) \leqslant \min\left(k^{(1)}(u) * f_\varepsilon(u), k^{(2)}(u) * f_\varepsilon(u)\right)$. Поэтому из (4.2) следует первая оценка в (4.1). Вторая оценка в (4.1) проверяется аналогично. ▲

Установим равномерные ограниченность и непрерывность $r_\varepsilon(u, t_1, t_2)$.

Лемма 5. *Справедлива оценка*

$$r_\varepsilon(u, t_1, t_2) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k(u). \tag{4.3}$$

Доказательство проводится по индукции. При $t = 1 - \varepsilon$ имеем $r^{(\ell)}(u, t_1, t_2) \leqslant \varepsilon k^{(\ell)}(u)$, $\ell = 1, 2$, и следовательно, оценка справедлива. Пусть она справедлива при $t_1 + t_2 = t + \varepsilon$. С учетом (3.4), предположения индукции и (4.1) получаем

$$\begin{aligned} &r_\varepsilon^{(1)}(u, t_1, t_2) \leqslant \varepsilon k^{(1)}(u) + r_\varepsilon(u, t_1 + \varepsilon, t_2) * f_{\varepsilon_1^* t_2^2 (t')^{-1} (t'+\varepsilon_1')^{-1}}(u) \leqslant \\ &\leqslant \varepsilon k^{(1)}(u) + (1 - t - \varepsilon) e^{2(1-t-2\varepsilon)cc_1} e^{2\varepsilon cc_1} k(u) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k^{(1)}(u). \end{aligned}$$

Точно так же $r_\varepsilon^{(2)}(u, t_1, t_2) \leqslant (1 - t) e^{2(1-t-\varepsilon)cc_1} k^{(2)}(u)$. Отсюда следует (4.3). ▲

Лемма 6. *Справедлива оценка*

$$|(r_\varepsilon)'_u(u, t_1, t_2)| \leqslant 2c_1 (1 - t) e^{2(1-t-\varepsilon)cc_1} k(u). \tag{4.4}$$

Доказательство. Отметим, что справедливо представление

$$r_\varepsilon(u, t_1, t_2) = \sigma_1(u, t_1, t_2) r_\varepsilon^{(1)}(u, t_1 + \varepsilon, t_2) + \sigma_2(u, t_1, t_2) r_\varepsilon^{(2)}(u, t_1, t_2 + \varepsilon), \tag{4.5}$$

где $\sigma_\ell(u, t_1, t_2)$ – байесовская стратегия, которая является кусочно-постоянной и принимает только значения 0 и 1, причем ее нужно считать неизменной при дифференцировании (4.5) по u. Поэтому

$$\begin{aligned} &|(r_\varepsilon)'_u(u, t_1, t_2)| \leqslant \\ &\leqslant \sigma_1(u, t_1, t_2) |(r_\varepsilon^{(1)})'_u(u, t_1 + \varepsilon, t_2)| + \sigma_2(u, t_1, t_2) |(r_\varepsilon^{(2)})'_u(u, t_1, t_2 + \varepsilon)|. \end{aligned} \tag{4.6}$$

Кроме того, из (3.4) следуют оценки

$$
\begin{aligned}
&|(r_\varepsilon^{(1)})'_u(u,t_1,t_2)| \leqslant \\
&\leqslant \varepsilon|(g^{(1)})'_u(u,t_1,t_2)| + |(r_\varepsilon)'_u(u,t_1+\varepsilon,t_2)| * f_{\varepsilon_1^* t_2^2 (t')^{-1}(t'+\varepsilon_1')^{-1}}(u), \\
&|(r_\varepsilon^{(2)})'_u(u,t_1,t_2)| \leqslant \\
&\leqslant \varepsilon|(g^{(2)})'_u(u,t_1,t_2)| + |(r_\varepsilon)'_u(u,t_1,t_2+\varepsilon)| * f_{\varepsilon_2^* t_1^2 (t')^{-1}(t'+\varepsilon_2')^{-1}}(u).
\end{aligned}
\tag{4.7}
$$

Отметим, что $|(g^{(\ell)})'_u(u,t_1,t_2)| \leqslant 2cD_g^{-2}k^{(\ell)}(u) = 2c_1k^{(\ell)}(u)$ при всех допустимых u,t_1,t_2 и $\ell = 1,2$. Дальнейшие рассуждения выполняются по индукции аналогично приведенным в лемме 5 с учетом (4.6), (4.7) и того, что $\sigma_\ell(u,t_1,t_2)$ – стратегия, обеспечивающая нахождение $r(u,t_1,t_2)$. ▲

Установим условия Липшица для $r_\varepsilon(u,t_1,t_2)$ по t_1,t_2. Ограничимся оценками по t_1.

Лемма 7. *Пусть* $\delta = K\varepsilon$, $\widetilde{\varepsilon}_i = \varepsilon_1^* t_2^2 (t' + (K-i)\varepsilon_1')^{-1}(t' + (K-i+1)\varepsilon_1')^{-1}$, $\widetilde{\delta}_i = \sum\limits_{j=1}^{i} \widetilde{\varepsilon}_j$. *Тогда справедливо неравенство*

$$
|r_\varepsilon(u,t_1,t_2) - r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)| \leqslant \delta k^{(1)}(u) e^{2cc_1\delta}. \tag{4.8}
$$

Доказательство. Обозначим через $r_\varepsilon^{(1,K)}(u,t_1,t_2)$ риск, соответствующий тому, что сначала K раз применялось первое действие, а затем управление выполнялось оптимально. Тогда при $i = 1,\ldots,K$ выполнено равенство

$$
\begin{aligned}
&r_\varepsilon^{(1,i)}(u,t_1+(K-i)\varepsilon,t_2) = \\
&= \varepsilon g^{(1)}(u,t_1+(K-i)\varepsilon,t_2) + r_\varepsilon^{(1,i-1)}(u,t_1+(K-i+1)\varepsilon,t_2) * f_{\widetilde{\varepsilon}_i}(u),
\end{aligned}
$$

где $r_\varepsilon^{(1,0)}(\cdot) = r_\varepsilon(\cdot)$. Так как $f_{\widetilde{\varepsilon}_i}(u) * f_{\widetilde{\varepsilon}_j}(u) = f_{\widetilde{\varepsilon}_i+\widetilde{\varepsilon}_j}(u)$, а $\widetilde{\delta}_i \leqslant D_1\delta(t_2/t')^2$, то отсюда следует, что

$$
\begin{aligned}
&r_\varepsilon^{(1,K)}(u,t_1,t_2) \leqslant \varepsilon\sum_{i=1}^{K} g^{(1)}(u,t_1+(K-i)\varepsilon,t_2) * f_{\widetilde{\delta}_i}(u) + r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u) \leqslant \\
&\leqslant \delta k^{(1)}(u) e^{2cc_1\delta} + r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u).
\end{aligned}
$$

Здесь оценка первого слагаемого аналогична оценкам леммы 4. Отметим, что $r_\varepsilon(u,t_1,t_2) \leqslant r_\varepsilon^{(1,K)}(u,t_1,t_2)$. С другой стороны, справедлива оценка $r_\varepsilon(u,t_1,t_2) \geqslant \geqslant r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)$, которая следует из того, что байесовский риск на меньшем горизонте управления $[t+\delta,1]$ при наличии дополнительной информации, обусловленной K-кратным применением первого действия, всегда не превосходит байесовского риска на большем горизонте управления $[t,1]$. Отсюда следует (4.8). ▲

Следствие 2. *Так как* $|(r_\varepsilon)'_u(\cdot)| \leqslant k = \max\limits_u 2c_1e^{2(1-\varepsilon)cc_1}k(u)$ *равномерно по* u, *то* $|r_\varepsilon(u,t_1+\delta,t_2) - r_\varepsilon(u-x,t_1+\delta,t_2)| \leqslant k|x|$, *поэтому*

$$
|r_\varepsilon(u,t_1+\delta,t_2) - r_\varepsilon(u,t_1+\delta,t_2) * f_{\widetilde{\delta}_K}(u)| \leqslant k\int\limits_{-\infty}^{\infty} |x| f_{\widetilde{\delta}_K}(x)\,dx = k(2/\pi)^{1/2}\delta_K^{1/2}. \tag{4.9}
$$

Из (4.8), (4.9) *следует, что при малых* δ *разность* $|r_\varepsilon(u,t_1,t_2) - r_\varepsilon(u,t_1+\delta,t_2)|$ *равномерно ограничена величиной порядка* $\delta^{1/2}$. *С использованием замечания* 1 *нетрудно*

установить, что $|r_\varepsilon(u,t_1,t_2)-r_\varepsilon(u,t_1+\delta,t_2)|$ *ограничена величиной порядка* δ, *однако соответствующая равномерная оценка пока не получена.*

Для дальнейшего сделаем несколько замечаний. Во-первых, для наихудшей априорной плотности $\varrho(w)$ найдется такое $w_0>0$, что $\varrho(w)=0$ при $|w|<w_0$. Это следует из того, что соответствующая байесовская стратегия является уравнивающей, т.е. обеспечивает одинаковые потери, равные минимаксному риску, для всех w, на которых $\varrho(w)\neq 0$, а потери, соответствующие w, не превосходят $2|w|$. Поэтому $k(u)\leqslant 2c\exp(-\alpha^{-1}w_0|u|)$, и в частности, $k(u)\leqslant\varepsilon|\delta|$ при $|u|\geqslant U_1=$ $=\alpha w_0^{-1}\left(\ln(2c)-\ln(\varepsilon)-\ln|\delta|\right)$.

Во-вторых, если для вычисления потерь использовать одинаковую стратегию, но различные функции $g^{(\ell)}(u,t_1,t_2)$ и $\widehat{g}^{(\ell)}(u,t_1,t_2)$, удовлетворяющие условию $|g^{(\ell)}(u,t_1,t_2)-\widehat{g}^{(\ell)}(u,t_1,t_2)|\leqslant\varepsilon|\delta|$, то для соответствующих им потерь справедлива оценка $|l(u,t_1,t_2)-\widehat{l}(u,t_1,t_2)|\leqslant|\delta|(1-t)$, которая устанавливается по индукции.

В-третьих, обозначим через U_2 максимальное абсолютное значение порога переключения стратегии σ. Ясно, что достаточно определить $g^{(1)}(u,t_1,t_2)$ при $u\geqslant -U_2$, а $g^{(2)}(u,t_1,t_2)$ при $u\leqslant U_2$. Обозначим $U_3=\max(U_1,U_2)$. Из сделанных замечаний следует, что если положить $\widehat{g}^{(\ell)}(u,t_1,t_2)=g^{(\ell)}(u,t_1,t_2)$ при $|u|\leqslant U_3$ и $\widehat{g}^{(\ell)}(u,t_1,t_2)=$ $=0$ при $|u|>U_3$, то для соответствующих потерь будет выполнена оценка $|l(u,t_1,t_2)-\widehat{l}(u,t_1,t_2)|\leqslant|\delta|(1-t)$.

Установим теперь, что малые изменения дисперсий приводят к малым изменениям потерь. Достаточно рассмотреть изменения D_1 и потери $\widehat{l}_\varepsilon(u,t_1,t_2)$. Справедлива

Лемма 8. *Пусть* D_2 *фиксирована, а* D_1 *может меняться. Обозначим через* $\widehat{l}_\varepsilon(D_1;\sigma;\cdot)$ *потери, в которых явно указана зависимость от* D_1 *и* σ *и использованы функции* $\{\widehat{g}^{(\ell)}(u,t_1,t_2)\}$. *Тогда для достаточно малых* $|\delta|$ *справедлива оценка*

$$|\widehat{l}_\varepsilon(D_1;\sigma;u,t_1,t_2)-\widehat{l}_\varepsilon(D_1+\delta;\sigma;u,t_1,t_2)|\leqslant 3|\delta|\alpha^{-2}(U_3c+c^2)k(u), \tag{4.10}$$

где $\sigma=\sigma_\ell(D_1;u,t_1,t_2)$ – *байесовская стратегия, соответствующая* D_1,u,t_1,t_2, *а* $U_3=\max(U_1,U_2)$ *и при фиксированном* ε *имеет порядок* $\ln|\delta|$.

Доказательство. Так как $|uw-w^2t_1t_2(t')^{-1}|\leqslant(U_3c+c^2)$ при $|u|\leqslant U_3$, а $D_1^{-1}-$ $-(D_1+\delta)^{-1}=D_1^{-2}(\delta+o(\delta))$, то

$$\begin{aligned}
&|g^{(\ell)}(D_1,u,t_1,t_2)-g^{(\ell)}(D_1+\delta,u,t_1,t_2)|\leqslant\\
&\leqslant g^{(\ell)}(D_1,u,t_1,t_2)\times\left(\exp\left\{2D_2^{-1}(U_3c+c^2)\left|D_1^{-1}-(D_1+\delta)^{-1}\right|\right\}-1\right)\leqslant\\
&\leqslant g^{(\ell)}(D_1,u,t_1,t_2)\left(\exp\left\{2(|\delta|+o(\delta))\alpha^{-2}(U_3c+c^2)\right\}-1\right)
\end{aligned}$$

при $|u|\leqslant U_3$ и для достаточно малых $|\delta|$

$$|\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)-\widehat{g}^{(\ell)}(D_1+\delta;u,t_1,t_2)|\leqslant 3|\delta|\alpha^{-2}(U_3c+c^2)g^{(\ell)}(D_1;u,t_1,t_2).$$

Обозначим

$$\begin{aligned}
&\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)=|\widehat{l}_\varepsilon(D_1,\sigma;u,t_1,t_2)-\widehat{l}_\varepsilon(D_1+\delta;\sigma;u,t_1,t_2)|,\\
&\Delta\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)=|\widehat{g}^{(\ell)}(D_1;u,t_1,t_2)-\widehat{g}^{(\ell)}(D_1+\delta;u,t_1,t_2)|.
\end{aligned}$$

Так как $\sigma=\sigma_\ell(D_1;u,t_1,t_2)$ – оптимальная стратегия, соответствующая D_1,u,t_1,t_2, то

$$\begin{aligned}
&\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)\leqslant\sigma_1(D_1;u,t_1,t_2)\Delta\widehat{l}_\varepsilon^{(1)}(D_1;u,t_1,t_2)+\\
&+\sigma_2(D_1;u,t_1,t_2)\Delta\widehat{l}_\varepsilon^{(2)}(D_1;u,t_1,t_2),
\end{aligned}$$

где

$$\Delta\widehat{l}_\varepsilon^{(1)}(D_1;u,t_1,t_2)\leqslant\varepsilon\Delta\widehat{g}^{(1)}(D_1;u,t_1,t_2)+\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)*f_{\varepsilon_1^*t_2^2(t')^{-1}(t'+\varepsilon_1')^{-1}}(u),$$
$$\Delta\widehat{l}_\varepsilon^{(2)}(D_1;u,t_1,t_2)\leqslant\varepsilon\Delta\widehat{g}^{(2)}(D_1;u,t_1,t_2)+\Delta\widehat{l}_\varepsilon(D_1;u,t_1,t_2)*f_{\varepsilon_2^*t_1^2(t')^{-1}(t'+\varepsilon_2')^{-1}}(u).$$

Далее оценки выполняются по индукции аналогично приведенным в леммах 5, 6. ▲

Установим существование непрерывного предела $r_\varepsilon(u,t_1,t_2)$ при $\varepsilon\to 0$.

Теорема 6. *При всех u,t_1,t_2, при которых определено решение уравнения* (3.3)–(3.5), *существует непрерывный предел $r(u,t_1,t_2)=\lim\limits_{\varepsilon\to+0}r_\varepsilon(u,t_1,t_2)$, который можно доопределить по непрерывности на все допустимые u,t_1,t_2 ($t_1\geqslant\varepsilon_0$, $t_2\geqslant\varepsilon_0$, $t_1+t_2\leqslant 1$). Этот предел равномерно ограничен, имеет равномерно ограниченную производную по u и равномерно непрерывен по t_1,t_2 с константами, указанными в* (4.3), (4.4), (4.8) *и* (4.9) *соответственно. При $\rho_N(v)=N^{-1/2}\varrho(w)$, $w=N^{1/2}v$ и $\varepsilon_0\to+0$ для байесовского риска* (3.6) *справедлива асимптотическая оценка*

$$\lim_{N\to\infty}N^{-1/2}R_N^B(\rho_N(v))=\lim_{\varepsilon_0\to 0}r(\varrho;0,\varepsilon_0,\varepsilon_0),\tag{4.11}$$

где в обозначении $r(\varrho;0,\varepsilon_0,\varepsilon_0)$ явно указана зависимость от $\varrho(w)$.

Доказательство. Пусть N – исходный горизонт управления. Тогда решение уравнения (3.3)–(3.5) определено при $t=2\varepsilon_0+n\varepsilon$, где $\varepsilon_0=M_0/N$ и $\varepsilon=M/N$ – относительные размеры начального и последующего пакетов. Если горизонт управления N фиксирован, то уменьшение величины ε соответствует тому, что действия можно менять чаще. Ясно, что при этом байесовские риски не возрастают. Поэтому $r_\varepsilon(u,t_1,t_2)$ является неотрицательной неубывающей функцией ε, и следовательно, имеет предел при $\varepsilon\to 0$. Непрерывность предела устанавливается предельным переходом в (4.3), (4.4), (4.8) и (4.9). Формула (4.11) следует из (3.6) и непрерывности $r(\varrho;u,t_1,t_2)$. ▲

Приведем дифференциальное уравнение в частных производных второго порядка для вычисления $r(u,t_1,t_2)$. Строгое обоснование этого вывода требует равномерных оценок производных $r_\varepsilon(u,t_1,t_2)$, поэтому ограничимся нестрогими рассуждениями, дополнив их результатами численного моделирования в § 5. Если $r_\varepsilon(u,\cdot\,)$ – достаточно гладкая функция u, то $r_\varepsilon(u-x,\cdot\,)=r_\varepsilon(u,\cdot\,)-(r_\varepsilon)'_u(u,\cdot\,)x+0{,}5(r_\varepsilon)''_{uu}(u,\cdot\,)x^2+o(x^2)$. Так как

$$\int\limits_{-\infty}^{\infty}f_{\widetilde{\varepsilon}}(x)\,dx=1,\quad\int\limits_{-\infty}^{\infty}xf_{\widetilde{\varepsilon}}(x)\,dx=0,\quad\int\limits_{-\infty}^{\infty}x^2f_{\widetilde{\varepsilon}}(x)\,dx=\widetilde{\varepsilon},$$

то из (3.4) при малых ε следуют уравнения

$$r_\varepsilon^{(1)}(u,t_1,t_2)=\varepsilon g^{(1)}(u,t_1,t_2)+r_\varepsilon(u,t_1+\varepsilon,t_2)+$$
$$+0{,}5\varepsilon_1^*t_2^2(t')^{-1}(t'+\varepsilon_1')^{-1}(r_\varepsilon)''_{uu}(u,t_1+\varepsilon,t_2)+o(\varepsilon),$$
$$r_\varepsilon^{(2)}(u,t_1,t_2)=\varepsilon g^{(2)}(u,t_1,t_2)+r_\varepsilon(u,t_1,t_2+\varepsilon)+$$
$$+0{,}5\varepsilon_2^*t_1^2(t')^{-1}(t'+\varepsilon_2')^{-1}(r_\varepsilon)''_{uu}(u,t_1,t_2+\varepsilon)+o(\varepsilon).$$

Дополняя их уравнением (3.3), записанным в виде

$$\min_{\ell=1,2}\Big(r_\varepsilon^{(\ell)}(u,t_1,t_2)-r_\varepsilon(u,t_1,t_2)\Big)=0,$$

в пределе при $\varepsilon \to 0$ получаем уравнение для $r = r(u, t_1, t_2)$

$$\min_{\ell=1,2}\left(\frac{\partial r}{\partial t_\ell} + \frac{D_\ell t_\ell^2}{2(t')^2} \times \frac{\partial^2 r}{\partial u^2} + g^{(\ell)}(u, t_1, t_2)\right) = 0, \tag{4.12}$$

где $\overline{\ell} = 3 - \ell$, с начальным условием

$$r(u, t_1, t_2) = 0 \quad \text{при } t_1 + t_2 = 1. \tag{4.13}$$

При этом байесовская стратегия предписывает выбирать ℓ-е действие, если меньше ℓ-е выражение в левой части (4.12). Для численного определения $r(u, t_1, t_2)$ в соответствии с (4.12) следует использовать разностное уравнение

$$\begin{aligned}
&r(u, t_1, t_2) = \min\bigl(r^{(1)}(u, t_1, t_2), r^{(2)}(u, t_1, t_2)\bigr),\\
&r^{(1)}(u, t_1, t_2) = r(u, t_1 + \Delta t, t_2) +\\
&+ \Delta t\bigl(g^{(1)}(u, t_1, t_2) + 0{,}5 D_1 (t_2/t')^2 D^2 r(u, t_1 + \Delta t, t_2)\bigr),\\
&r^{(2)}(u, t_1, t_2) = r(u, t_1, t_2 + \Delta t) +\\
&+ \Delta t\bigl(g^{(2)}(u, t_1, t_2) + 0{,}5 D_2 (t_1/t')^2 D^2 r(u, t_1, t_2 + \Delta t)\bigr),
\end{aligned} \tag{4.14}$$

где $D^2 r(u, \cdot) = \Delta u^{-2}\,(r(u + \Delta u, \cdot) - 2r(u, \cdot) + r(u - \Delta u, \cdot))$, с начальным условием (4.13). Аналогично для вычисления потерь, соответствующих стратегии σ в предельном случае, следует использовать разностное уравнение

$$\begin{aligned}
&l(u, t_1, t_2) = \sigma_1(u, t_1, t_2) l^{(1)}(u, t_1, t_2) + \sigma_2(u, t_1, t_2) l^{(2)}(u, t_1, t_2),\\
&l^{(1)}(u, t_1, t_2) = l(u, t_1 + \Delta t, t_2) +\\
&+ \Delta t\bigl(g^{(1)}(u, t_1, t_2) + 0{,}5 D_1 (t_2/t')^2 D^2 l(u, t_1 + \Delta t, t_2)\bigr),\\
&l^{(2)}(u, t_1, t_2) = l(u, t_1, t_2 + \Delta t) +\\
&+ \Delta t\bigl(g^{(2)}(u, t_1, t_2) + 0{,}5 D_2 (t_1/t')^2 D^2 l(u, t_1, t_2 + \Delta t)\bigr),
\end{aligned} \tag{4.15}$$

где $D^2 l(u, \cdot) = \Delta u^{-2}\,(l(u + \Delta u, \cdot) - 2l(u, \cdot) + l(u - \Delta u, \cdot))$, с начальным условием $l(u, t_1, t_2) = 0$ при $t_1 + t_2 = 1$.

§ 5. Численные эксперименты

На рис. 1 представлены результаты вычисления рисков $\{r_\varepsilon(u, t_1, t_2)\}$ как минимумов из $\{r_\varepsilon^{(1)}(u, t_1, t_2), r_\varepsilon^{(2)}(u, t_1, t_2)\}$ для $(t_1, t_2) = (0{,}4; 0{,}4), (0{,}2; 0{,}2), (0{,}1; 0{,}1)$. Жирными линиями представлены риски $\{r_\varepsilon^{(1)}(u, t_1, t_2)\}$, тонкими – риски $\{r_\varepsilon^{(2)}(u, t_1, t_2)\}$. При этом меньшим значениям $t_1 + t_2$ соответствуют бóльшие риски $r_\varepsilon^{(1)}(u, t_1, t_2)$, $r_\varepsilon^{(2)}(u, t_1, t_2)$.

Рис. 2 демонстрирует непрерывность рисков по u и их близость при малых изменениях t_2. Жирными линиями представлены четыре группы рисков $\{r(u, t_1, t_2)\}$, полученных в результате решения разностного уравнения (4.14), причем чем меньше $t_1 + t_2$, тем выше соответствующая кривая. Снизу вверх четыре пары кривых представлены для $(t_1, t_2) = (0{,}4; 0{,}44), (0{,}4; 0{,}4)$, $(t_1, t_2) = (0{,}2; 0{,}24), (0{,}2; 0{,}2)$, $(t_1, t_2) = (0{,}1; 0{,}14), (0{,}1; 0{,}1)$, $(t_1, t_2) = (0{,}001; 0{,}001), (0{,}001; 0{,}041)$. Последнюю можно рассматривать как предельную пару кривых $(t_1, t_2) = (\varepsilon_0; \varepsilon_0), (\varepsilon_0; \varepsilon_0 + 0{,}04)$ при $\varepsilon_0 \to +0$, так как кривые, вычисленные при $\varepsilon_0 = 0{,}001; 0{,}002; 0{,}004$, практически совпадают. Для первых трех пар также были вычислены риски $\{r_\varepsilon(u, t_1, t_2)\}$ с использованием уравнения (3.3)–(3.4), они представлены тонкими линиями и совсем немного превышают $\{r(u, t_1, t_2)\}$. Отметим, что кривые на рис. 1 и 2 вычислены при

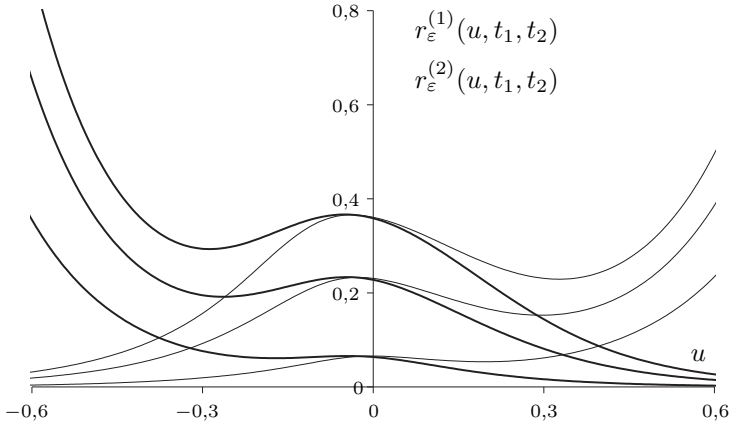

Рис. 1. Риски $r_\varepsilon^{(1)}(u,t_1,t_2)$, $r_\varepsilon^{(2)}(u,t_1,t_2)$ при $(t_1,t_2)=(0{,}4;0{,}4),(0{,}2;0{,}2),(0{,}1;0{,}1)$

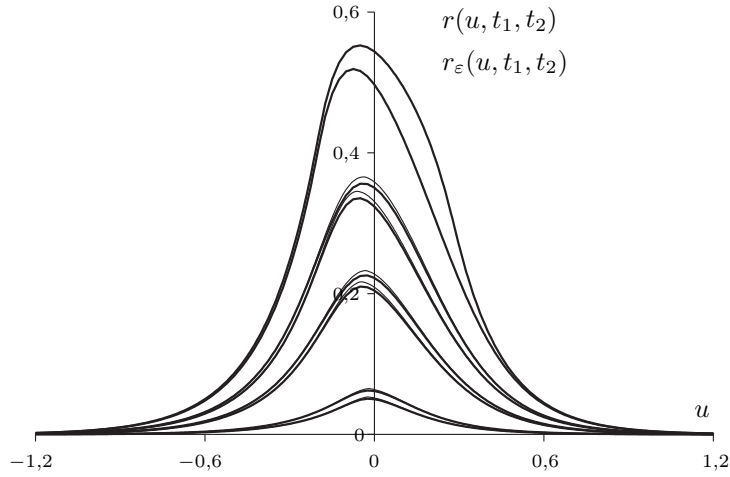

Рис. 2. Риски $r(u,t_1,t_2)$, $r_\varepsilon(u,t_1,t_2)$

$D_1=1$, $D_2=0{,}5$, плотность $\varrho(w)$ сосредоточена в точках $d_1=1{,}49$ и $d_2=-1{,}53$ с вероятностями 0,54 и 0,46 (это распределение близко к наихудшему при выбранных параметрах), $\varepsilon=0{,}02$, $\Delta t=2000^{-1}$. Вычисления выполнялись в области $|u|\leqslant 2{,}5$, шаг Δu равен 0,005 при выполнении численного интегрирования и 0,025 при решении уравнения (4.14). Отметим также, что выбор области и шага численного интегрирования влияет только на точность результата, в то время как при численном решении уравнения (4.14) для обеспечения устойчивости требуется выполнение условия $\Delta t/\Delta u^2<1$ (см., например, [12]).

На рис. 3 представлены риски и потери, вычисленные при $D_1=D_2=1$, если $d_1=-d_2=d$, $\varrho(w)=0{,}5\,(\delta(w-d_1)+\delta(w-d_2))$ и $\delta(\cdot)$ – дельта-функция Дирака. Если наихудшее априорное распределение сосредоточено в двух точках, то оно соответствует d, при котором байесовский риск достигает максимума. Жирные линии *1* и *2* получены в результате решения уравнения (3.3)–(3.5), причем линия *2* характеризует значение байесовского риска без потерь на начальном этапе, равных $2d\varepsilon_0$. Максимум байесовского риска $r_\varepsilon(d)$ приблизительно равен 0,652 при $d\approx 1{,}6$. Жирные линии *3* и *4* соответствуют потерям $l_\varepsilon(d)$ (полным и без потерь на начальном

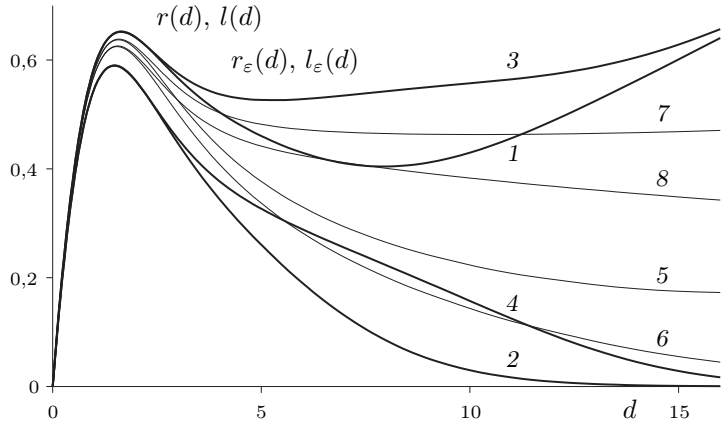

Рис. 3. Риски и потери как функции априорного распределения

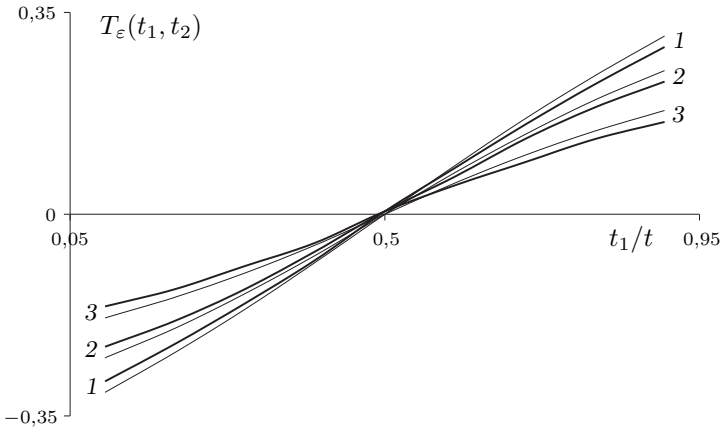

Рис. 4. Значения порогов как функции от t_1/t

этапе), которые вычислены в соответствии с уравнением (3.7)–(3.8) для найденной байесовской стратегии. Они также имеют единственный максимум при $d \approx 1{,}6$. Здесь $\varepsilon = \varepsilon_0 = 0{,}02$, $\Delta u = 0{,}005$, $|u| \leqslant 2{,}5$, т.е. кривые соответствуют пакетной обработке данных, разбитых на 50 пакетов.

Тонкие линии *5* и *6* соответствуют байесовским рискам $r(d)$ (полному и без начальных потерь), найденным в результате решения уравнения (4.14). Максимум байесовского риска при $d \approx 1{,}6$ приблизительно равен 0,637, т.е. меньше, чем для пакетной обработки, приблизительно на 2%. Тонкие линии *7* и *8* соответствуют потерям $l(d)$ (полным и без потерь на начальном этапе), полученным для найденной стратегии в результате решения уравнения (4.15). Здесь $\Delta t = 2000^{-1}$, $\varepsilon_0 = 0{,}004$, $\Delta u = 0{,}025$, $|u| \leqslant 2{,}5$.

Наконец, на рис. 4 представлены значения порогов байесовской стратегии, определенные численными методами при решении уравнения (3.3)–(3.5). Пары линий *1*, *2*, *3* соответствуют значениям $t = 0{,}2; 0{,}6; 0{,}8$, причем жирные линии получены при $\varepsilon = 0{,}02$, а тонкие при $\varepsilon = 0{,}01$. Здесь $D_1 = D_2 = 1$, $d_1 = -d_2 = 1{,}6$, $\varrho(w) = 0{,}5\,(\delta(w - d_1) + \delta(w - d_2))$. Эти результаты показывают, что можно ожидать сходимости стратегии при $\varepsilon \to +0$.

§ 6. Обработка данных по одному. Бернуллиевский двурукий бандит

В этом параграфе показано, что для бернуллиевского двурукого бандита максимальные потери являются такими же, как и для гауссовского, если $N \to \infty$. Это следует из того, что в этом случае уравнение для вычисления байесовского риска совпадает с (4.12), а минимаксный риск не меньше любого байесовского. Доходы бернуллиевского двурукого бандита ξ_n, $n = 1, 2, \ldots, N$, зависят от текущих выбираемых действий y_n следующим образом:

$$\Pr(\xi_n = 1 \,|\, y_n = \ell) = p_\ell, \quad \Pr(\xi_n = 0 \,|\, y_n = \ell) = q_\ell, \quad \ell = 1, 2.$$

Бернуллиевский двурукий бандит описывается векторным параметром $\theta = (p_1, p_2)$. Допустимое множество параметров Θ пока считаем произвольным.

Стратегия управления σ в момент времени $n+1$ описывает выбор действия в зависимости от предыстории (X_1, n_1, X_2, n_2), где n_1, n_2 – текущие полные количества применений обоих действий ($n_1 + n_2 = n$), X_1, X_2 – соответствующие полные доходы. Таким образом, $\sigma_\ell(X_1, n_1, X_2, n_2) = \Pr(y_{n+1} = \ell \,|\, X_1, n_1, X_2, n_2)$, $\ell = 1, 2$. Будем предполагать, что в начале управления стратегия применяет оба действия по n_0 раз. Если $n_0 \ll N$, то это практически не влияет на полные потери. Функция потерь имеет вид

$$L_N(\sigma, \theta) = N(p_1 \vee p_2) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N} \xi_n\right),$$

а байесовский риск относительно априорной плотности распределения $\lambda(p_1, p_2)$ равен

$$R_N^B(\lambda) = \min_{\{\sigma\}} \iint\limits_{\Theta} L_N(\sigma, \theta)\, \lambda(p_1, p_2)\, dp_1\, dp_2. \tag{6.1}$$

Запишем стандартное уравнение динамического программирования для вычисления байесовского риска (6.1). Апостериорная плотность распределения, соответствующая предыстории процесса (X_1, n_1, X_2, n_2), равна

$$\lambda(p_1, p_2 \,|\, X_1, n_1, X_2, n_2) = \frac{B(X_1, n_1 \,|\, p_1) B(X_2, n_2 \,|\, p_2) \lambda(p_1, p_2)}{P(X_1, n_1, X_2, n_2)}, \tag{6.2}$$

$$\text{где} \quad P(X_1, n_1, X_2, n_2) = \iint\limits_{\Theta} B(X_1, n_1 \,|\, p_1) B(X_2, n_2 \,|\, p_2)\, \lambda(p_1, p_2)\, dp_1\, dp_2 \tag{6.3}$$

и

$$B(X, n \,|\, p) = \binom{n}{X} p^X (1-p)^{n-X}.$$

Если дополнительно положить $B(X, n \,|\, p) = 1$ при $n = 0$, $X = 0$, то формула (6.2) сохранится и при $n_1 = 0$ и/или $n_2 = 0$.

Обозначим через $R_{N-n}^B(X_1, n_1, X_2, n_2)$ байесовский риск на последних $(N - n)$ шагах, вычисленный относительно апостериорной плотности распределения (6.2). В дальнейшем будем для сокращения записи обозначать через $\_\ell$ предысторию X_ℓ, n_ℓ, если она одинакова во всем выражении. Для нахождения байесовского риска (6.1) следует решать стандартное уравнение динамического программирования

$$R_{N-n}^B(X_1, n_1, X_2, n_2) = \min\bigl(R_{N-n}^{B(1)}(X_1, n_1, X_2, n_2), R_{N-n}^{B(2)}(X_1, n_1, X_2, n_2)\bigr), \tag{6.4}$$

где $R^{B(1)}_{N-n}(X_1,n_1,X_2,n_2)=R^{B(2)}_{N-n}(X_1,n_1,X_2,n_2)=0$ при $n=N$, и далее

$$
\begin{aligned}
&R^{B(1)}_{N-n}(X_1,n_1,\__2)=\iint\limits_{\Theta}\lambda(p_1,p_2\,|\,X_1,n_1,\__2)\times\\
&\times\Big((p_2-p_1)^+ + \mathbf{E}^{(1)}_x\,R^{B}_{N-n-1}(X_1+x,n_1+1,\__2)\Big)\,dp_1\,dp_2,\\
&R^{B(2)}_{N-n}(\__1,X_2,n_2)=\iint\limits_{\Theta}\lambda(p_1,p_2\,|\,\__1,X_2,n_2)\times\\
&\times\Big((p_1-p_2)^+ + \mathbf{E}^{(2)}_x\,R^{B}_{N-n-1}(\__1,X_2+x,n_2+1)\Big)\,dp_1\,dp_2
\end{aligned}
\tag{6.5}
$$

при $2n_0\leqslant n<N$. Здесь $\mathbf{E}^{(\ell)}_x\,R(x)=q_\ell R(0)+p_\ell R(1)$, $q_\ell=1-p_\ell$.

В уравнениях (6.4), (6.5) риск $R^{B(\ell)}_{N-n}(\cdot)$ равен математическому ожиданию полных потерь на оставшемся горизонте управления длины $N-n$, если сначала было применено ℓ-е действие, а затем управление осуществлялось оптимально ($\ell=1,2$). Байесовская стратегия предписывает выбирать действие, соответствующее меньшей текущей величине $R^{B(\ell)}_{N-n}(\cdot)$, $\ell=1,2$; в случае их равенства выбор может быть произвольным. Байесовский риск (6.1) равен

$$
\begin{aligned}
&R^{B}_{N}(\lambda)=n_0\iint\limits_{\Theta}|p_1-p_2|\lambda(p_1,p_2)\,dp_1\,dp_2+\\
&+\sum_{X_1=0}^{n_0}\sum_{X_2=0}^{n_0}R^{B}_{N-2n_0}(X_1,n_0,X_2,n_0)P(X_1,n_0,X_2,n_0).
\end{aligned}
\tag{6.6}
$$

Обозначим

$$
\begin{aligned}
&\widetilde{R}(X_1,n_1,X_2,n_2)=R^{B}_{N-n}(X_1,n_1,X_2,n_2)P(X_1,n_1,X_2,n_2),\\
&\widetilde{R}^{(\ell)}(X_1,n_1,X_2,n_2)=R^{B(\ell)}_{N-n}(X_1,n_1,X_2,n_2)P(X_1,n_1,X_2,n_2),\quad \ell=1,2,
\end{aligned}
$$

где $P(X_1,n_1,X_2,n_2)$ определена в (6.3). Справедлива

Т е о р е м а 7. *Рассмотрим рекуррентное уравнение*

$$
\widetilde{R}(X_1,n_1,X_2,n_2)=\min\big(\widetilde{R}^{(1)}(X_1,n_1,X_2,n_2),\widetilde{R}^{(2)}(X_1,n_1,X_2,n_2)\big),
\tag{6.7}
$$

где $\widetilde{R}^{(1)}(X_1,n_1,X_2,n_2)=\widetilde{R}^{(2)}(X_1,n_1,X_2,n_2)=0$ *при* $n=N$, *и далее*

$$
\begin{aligned}
&\widetilde{R}^{(1)}(X_1,n_1,\__2)=\widetilde{g}^{(1)}(X_1,n_1,\__2)+\sum_{x=0}^{1}\widetilde{R}(X_1+x,n_1+1,\__2)h(X_1,n_1,x),\\
&\widetilde{R}^{(2)}(\__1,X_2,n_2)=\widetilde{g}^{(2)}(\__1,X_2,n_2)+\sum_{x=0}^{1}\widetilde{R}(\__1,X_2+x,n_2+1)h(X_2,n_2,x)
\end{aligned}
\tag{6.8}
$$

при $2n_0\leqslant n<N$. *Здесь*

$$
\begin{aligned}
&\widetilde{g}^{(1)}(X_1,n_1,X_2,n_2)=\iint\limits_{\Theta}(p_2-p_1)^+B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\lambda(p_1,p_2)\,dp_1\,dp_2,\\
&\widetilde{g}^{(2)}(X_1,n_1,X_2,n_2)=\iint\limits_{\Theta}(p_1-p_2)^+B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\lambda(p_1,p_2)\,dp_1\,dp_2
\end{aligned}
\tag{6.9}
$$

и

$$h(X,n,0)=\frac{n+1-X}{n+1},\quad h(X,n,1)=\frac{X+1}{n+1}. \tag{6.10}$$

Байесовская стратегия предписывает выбирать действие, которое соответствует меньшей текущей величине $\widetilde{R}^{(\ell)}(\cdot)$, $\ell=1,2$; *в случае их равенства выбор может быть произвольным. Байесовский риск* (6.1) *равен*

$$R_N^B(\lambda)=n_0\iint\limits_{\Theta}|p_1-p_2|\lambda(p_1,p_2)\,dp_1\,dp_2+\sum_{X_1=0}^{n_0}\sum_{X_2=0}^{n_0}\widetilde{R}(X_1,n_0,X_2,n_0). \tag{6.11}$$

Доказательство. Формулы (6.7)–(6.9), (6.11) следуют из (6.4)–(6.6). В проверке нуждается (6.10). Достаточно рассмотреть $h(X_1,n_1,x)$. Действительно,

$$\begin{aligned}h(X_1,n_1,x)&=\frac{\iint\limits_{\Theta}p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}{P(X_1+x,n_1+1,X_2,n_2)}=\\&=\frac{\iint\limits_{\Theta}p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}{\iint\limits_{\Theta}B(X_1+x,n_1+1\,|\,p_1)B(X_2,n_2\,|\,p_2)\,\lambda(p_1,p_2)\,dp_1\,dp_2}=\\&=\frac{p_1^xq_1^{1-x}B(X_1,n_1\,|\,p_1)}{B(X_1+x,n_1+1\,|\,p_1)}=\frac{\binom{n_1}{X_1}}{\binom{n_1+1}{X_1+x}}.\end{aligned}$$

Непосредственно проверяется, что это соответствует выражению в (6.10). ▲

Выберем некоторое $0<p<1$ и рассмотрим следующую замену переменных: $p_1=p+(\mu+w)N^{-1/2}$, $p_2=p+(\mu-w)N^{-1/2}$, $t=nN^{-1}$, $X_\ell=n_\ell p+x_\ell N^{1/2}$, $t_\ell=n_\ell N^{-1}$, $\ell=1,2$. В качестве множества параметров выберем следующее: $\Theta=$ $=\{(p+(\mu+w)N^{-1/2},p+(\mu-w)N^{-1/2}):|w|\leqslant c,\,|\mu|\leqslant a_N\}$, где $c>0$ – достаточно большое фиксированное число, $a_N=N^\alpha$, $0<\alpha<1/2$, и N достаточно велико. Геометрически Θ – это узкая полоса с центром в точке (p,p), параллельная главной диагонали квадрата $\{(p_1,p_2):\ p_\ell\in[0,1],\ \ell=1,2\}$. Отношение длин меньшей и большей сторон этой полосы стремится к нулю с ростом N, так же как и сами длины обеих сторон.

Далее удобно изменить параметризацию и от (p_1,p_2) перейти к (μ,w). Априорную плотность распределения выберем в виде $N\varkappa_a(\mu)\varrho(w)$, где $\varkappa_a(\mu)$ – плотность равномерного распределения при $|\mu|\leqslant a_N$, а $\varrho(w)$ – произвольная плотность при $|w|\leqslant c$. Отметим, что близкая к данной априорная плотность распределения выбрана для оценки минимаксного риска снизу в [13].

Положим $D_1=D_2=pq$, где $q=1-p$, $n_\ell^*=D_\ell n_\ell$, $n'_\ell=n_\ell/D_\ell$, $t_\ell^*=n_\ell^*N^{-1}$, $t'_\ell=n'_\ell N^{-1}$, $t'=t'_1+t'_2$, $\varepsilon_0=n_0N^{-1}$, $\varepsilon=N^{-1}$, $\delta=N^{-1/2}$ и $\widetilde{R}(X_1,n_1,X_2,n_2)=$ $=N^{-1/2}\widetilde{r}(x_1,t_1,x_2,t_2)$. Будем писать $x_N\sim y_N$, если $\lim\limits_{N\to\infty}x_N/y_N=1$, и $x_N\lesssim y_N$, если $\lim\limits_{N\to\infty}x_N/y_N\leqslant 1$. Если n_0 достаточно велико, то при $n_\ell>n_0$ справедливы оценки

$$\begin{aligned}&B(X_1,n_1\,|\,p_1)\sim N^{-1/2}f_{t_1^*}(x_1\,|\,(\mu+w)t_1),\\&B(X_2,n_2\,|\,p_2)\sim N^{-1/2}f_{t_2^*}(x_2\,|\,(\mu-w)t_2).\end{aligned} \tag{6.12}$$

Действительно, в силу локальной предельной теоремы имеем $B(X_\ell, n_\ell \,|\, p_\ell) \sim$ $\sim f_{n_\ell^*}(X_\ell \,|\, n_\ell p_\ell) \sim f_{n_\ell^*}(X_\ell - n_\ell p \,|\, n_\ell p_\ell - n_\ell p)$, откуда с учетом сделанной замены переменных следует (6.12).

Пусть последовательность $\{b_N\}$ такова, что $b_N \to +\infty$, $b_N/a_N \to 0$ при $N \to \infty$, и пусть $y = (\overline{x}_1 t_1' + \overline{x}_2 t_2')/t'$, $z = x_1 t_2 - x_2 t_1$, где $\overline{x}_\ell = x_\ell / t_\ell$. Справедливы оценки

$$\begin{aligned}
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) \sim N^{-3/2}(2a_N)^{-1}\widetilde{g}^{(\ell)}(z, t_1, t_2) \quad \text{при } |y| \leqslant a_N - b_N,\\
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) \lesssim N^{-3/2}(2a_N)^{-1}\widetilde{g}^{(\ell)}(z, t_1, t_2)\\
&\text{при } a_N - b_N < |y| \leqslant a_N + b_N,\\
&\widetilde{g}^{(\ell)}(X_1, n_1, X_2, n_2) = N^{-3/2}(2a_N)^{-1} o(e^{-\gamma(|y|-a_N)^2})\\
&\text{при } |y| > a_N + b_N,\ \gamma > 0,
\end{aligned} \tag{6.13}$$

где

$$\begin{aligned}
\widetilde{g}^{(1)}(z, t_1, t_2) &= \int\limits_{-c}^{0} 2|w| f_{t_1^* t_2^* t'}(z - 2wt_1t_2)\varrho(w)\,dw,\\
\widetilde{g}^{(2)}(z, t_1, t_2) &= \int\limits_{0}^{c} 2|w| f_{t_1^* t_2^* t'}(z - 2wt_1t_2)\varrho(w)\,dw.
\end{aligned} \tag{6.14}$$

Достаточно проверить (6.13), (6.14) при $\ell = 1$. Из (6.9), (6.12) следует, что

$$\begin{aligned}
&\widetilde{g}^{(1)}(X_1, n_1, X_2, n_2) \sim N^{-3/2}(2a_N)^{-1} \times\\
&\times \int\limits_{-c}^{0} \left(\int\limits_{-a_N}^{a_N} f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^*}(x_2 \,|\, (\mu - w)t_2)\,d\mu \right) 2|w|\varrho(w)\,dw.
\end{aligned}$$

Отсюда (6.13), (6.14) легко следуют, если для оценки внутреннего интеграла использовать непосредственно проверяемое равенство

$$\begin{aligned}
&f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^s}(x_2 \,|\, (\mu - w)t_2) =\\
&= f_{(t')^{-1}}\left(y + (t')^{-1}(t_2' - t_1')w - \mu\right) f_{t_1^* t_2^* t'}\left(z - 2wt_1t_2\right).
\end{aligned}$$

Аналогично с учетом (6.3), (6.12) выполняется аппроксимация

$$\begin{aligned}
&P(X_1, n_1, X_2, n_2) \sim\\
&\sim N^{-1}(2a_N)^{-1} \int\limits_{-c}^{c} \left(\int\limits_{-a_N}^{a_N} f_{t_1^*}(x_1 \,|\, (\mu + w)t_1) f_{t_2^*}(x_2 \,|\, (\mu - w)t_2)\,d\mu \right) \varrho(w)\,dw.
\end{aligned}$$

Так как $R^B_{N-n}(\cdot) \leqslant 2cN^{1/2}$, то с учетом определения $\widetilde{r}(x_1, t_1, x_2, t_2)$ отсюда можно показать, что

$$\begin{aligned}
&\widetilde{r}(x_1, t_1, x_2, t_2) = (2a_N)^{-1}O(1) \quad \text{при } |y| \leqslant a_N + b_N,\\
&\widetilde{r}(x_1, t_1, x_2, t_2) = (2a_N)^{-1}o(e^{-\gamma(|y|-a_N)^2}) \quad \text{при } |y| > a_N + b_N,\ \gamma > 0.
\end{aligned} \tag{6.15}$$

Покажем, что при $|y| \leqslant a_N - b_N$ и $N \to \infty$ из (6.7)–(6.11) следует дифференциальное уравнение в частных производных второго порядка

$$\min_{\ell=1,2}\left(\widetilde{r}_{t_\ell}' + t_\ell^{-1}\widetilde{r} + zt_\ell^{-1}\widetilde{r}_z' + 0{,}5D_\ell t_\ell^2 \widetilde{r}_{zz}'' + \widetilde{g}^{(\ell)}(z, t_1, t_2)\right) = 0 \tag{6.16}$$

с начальным условием

$$\widetilde{r}(z,t_1,t_2)=0 \quad \text{при } t_1+t_2=1. \tag{6.17}$$

При этом байесовский риск равен

$$R_N^B(\lambda)\sim N^{1/2}\left(\varepsilon_0\int\limits_{-c}^{c}2|w|\varrho(w)\,dw+\int\limits_{-\infty}^{\infty}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dz\right). \tag{6.18}$$

Замечание 2. В нашем случае D_1 и D_2 одинаковы, однако для гауссовского двурукого бандита с различными дисперсиями D_1 и D_2 дифференциальное уравнение и байесовский риск записываются именно в виде (6.16)–(6.18), если за основу взять формулы (3.2)–(3.5) из [1].

Чтобы записать уравнения (6.8) с использованием переменных x_1,t_1,x_2,t_2, заметим, что паре переменных (X_ℓ,n_ℓ) соответствует пара (x_ℓ,t_ℓ) по определению, а парам $(X_\ell,n_\ell+1)$ и $(X_\ell+1,n_\ell+1)$ соответствуют пары $(x_\ell-p\delta,t_\ell+\varepsilon)$ и $(x_\ell+q\delta,t_\ell+\varepsilon)$. Например, для $(X_\ell,n_\ell+1)$ имеем $x_\ell\leftarrow(X_\ell-(n_\ell+1)p)N^{-1/2}=x_\ell-p\delta$, $t_\ell\leftarrow$ $\leftarrow(n_\ell+1)N^{-1}=t_\ell+\varepsilon$. Далее, с учетом (6.10)

$$h(X_\ell,n_\ell,0)=\frac{n_\ell+1-n_\ell p-x_\ell N^{1/2}}{n_\ell+1}=q+pt_\ell^{-1}\varepsilon-x_\ell t_\ell^{-1}\delta+o(\varepsilon),$$
$$h(X_\ell,n_\ell,1)=\frac{n_\ell p+x_\ell N^{1/2}+1}{n_\ell+1}=p+qt_\ell^{-1}\varepsilon+x_\ell t_\ell^{-1}\delta+o(\varepsilon).$$

При $\ell=1$ уравнение (6.8) примет вид

$$\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=\varepsilon(2a_N)^{-1}\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(x_1-p\delta,t_1+\varepsilon,x_2,t_2)\times$$
$$\times(q+pt_1^{-1}\varepsilon-x_1t_1^{-1}\delta)+\widetilde{r}(x_1+q\delta,t_1+\varepsilon,x_2,n_2)(p+qt_1^{-1}\varepsilon+x_1t_1^{-1}\delta)+o(\varepsilon).$$

Предполагая, что $\widetilde{r}(x_1,\cdot\,)$ является достаточно гладкой функцией, и выполняя замены $\widetilde{r}(x_1-p\delta,\cdot\,)=\widetilde{r}(x_1,\cdot\,)-p\delta\widetilde{r}'_{x_1}(x_1,\cdot\,)+0{,}5p^2\varepsilon\widetilde{r}''_{x_1x_1}(x_1,\cdot\,)+o(\varepsilon)$ и $\widetilde{r}(x_1+q\delta,\cdot\,)=$ $=\widetilde{r}(x_1,\cdot\,)+q\delta\widetilde{r}'_{x_1}(x_1,\cdot\,)+0{,}5q^2\varepsilon\widetilde{r}''_{x_1x_1}(x_1,\cdot\,)+o(\varepsilon)$, получаем

$$\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=\varepsilon(2a_N)^{-1}\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(x_1,t_1+\varepsilon,x_2,t_2)(1+\varepsilon t_1^{-1})+$$
$$+\varepsilon\widetilde{r}'_{x_1}(x_1,t_1+\varepsilon,x_2,t_2)x_1t_1^{-1}+\varepsilon 0{,}5D_1\widetilde{r}''_{x_1x_1}(x_1,t_1+\varepsilon,x_2,t_2)+o(\varepsilon), \tag{6.19}$$

Положим $\widetilde{r}^{(1)}(x_1,t_1,x_2,t_2)=(2a_N)^{-1}\widetilde{r}^{(1)}(z,t_1,t_2)$, $\widetilde{r}(x_1,t_1,x_2,t_2)=(2a_N)^{-1}\widetilde{r}(z,t_1,t_2)$. Заметим, что $\widetilde{r}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}\widetilde{r}(z-x_2\varepsilon,t_1+\varepsilon,t_2)=(2a_N)^{-1}(\widetilde{r}(z,t_1+\varepsilon,t_2)-$ $-x_2\varepsilon\widetilde{r}'_z(z,t_1+\varepsilon,t_2)+o(\varepsilon))$, $\widetilde{r}'_{x_1}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}t_2\widetilde{r}'_z(z-x_2\varepsilon,t_1+\varepsilon,t_2)$, $\widetilde{r}''_{x_1x_1}(x_1,t_1+\varepsilon,x_2,t_2)=(2a_N)^{-1}t_2^2\widetilde{r}''_{zz}(z-x_2\varepsilon,t_1+\varepsilon,t_2)$. Поэтому (6.19) принимает вид

$$\widetilde{r}^{(1)}(z,t_1,t_2)=\varepsilon\widetilde{g}^{(1)}(z,t_1,t_2)+\widetilde{r}(z,t_1+\varepsilon,t_2)(1+\varepsilon t_1^{-1})+$$
$$+\varepsilon\widetilde{r}'_z(z,t_1+\varepsilon,t_2)zt_1^{-1}+\varepsilon 0{,}5D_1t_2^2\widetilde{r}''_{zz}(z,t_1+\varepsilon,t_2)+o(\varepsilon). \tag{6.20}$$

Аналогично при $\ell=2$ имеем

$$\widetilde{r}^{(2)}(z,t_1,t_2)=\varepsilon\widetilde{g}^{(2)}(z,t_1,t_2)+\widetilde{r}(z,t_1,t_2+\varepsilon)(1+\varepsilon t_2^{-1})+$$
$$+\varepsilon\widetilde{r}'_z(z,t_1,t_2+\varepsilon)zt_2^{-1}+\varepsilon 0{,}5D_2t_1^2\widetilde{r}''_{zz}(z,t_1,t_2+\varepsilon)+o(\varepsilon). \tag{6.21}$$

Дополняя (6.20), (6.21) уравнением $\min\limits_{\ell=1,2}(\widetilde{r}(z,t_1,t_2)-\widetilde{r}^{(\ell)}(z,t_1,t_2))=0$, соответствующим уравнению (6.7), в пределе при $\varepsilon\to0$ получаем (6.16). Далее, с учетом выбран-

ных множества Θ и априорной плотности распределения байесовский риск (6.11) равен

$$R_N^B(\lambda) = N^{1/2}\left(\varepsilon_0 \int\limits_{-c}^{c} 2|w|\varrho(w)\,dw + \sum_{x_1=-\varepsilon_0 pN^{1/2}}^{\varepsilon_0 qN^{1/2}} \sum_{x_2=-\varepsilon_0 pN^{1/2}}^{\varepsilon_0 qN^{1/2}} \widetilde{r}(x_1,\varepsilon_0,x_2,\varepsilon_0)\delta^2\right),$$

где δ характеризует шаг по x_1, x_2, поэтому при $\varepsilon \to 0$

$$R_N^B(\lambda) \sim N^{1/2}\left(\varepsilon_0 \int\limits_{-c}^{c} 2|w|\varrho(w)\,dw + \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \widetilde{r}(x_1,\varepsilon_0,x_2,\varepsilon_0)\,dx_1\,dx_2\right). \tag{6.22}$$

Снова преобразуем переменные: $y = (t_1'\overline{x}_1 + t_2'\overline{x}_2)t'^{-1}$, $z = x_1t_2 - x_2t_1$, где $t_1 = t_2 = \varepsilon_0$. Якобиан преобразования равен $|\partial(x_1,x_2)/\partial(y,z)| = 1$. С учетом (6.15) двойной интеграл в (6.22) равен

$$(2a_N)^{-1}\int\limits_{-\infty}^{\infty}\left(\int\limits_{-(a_N-b_N)}^{a_N-b_N}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dy\right)dz \sim \left(\int\limits_{-\infty}^{\infty}\widetilde{r}(z,\varepsilon_0,\varepsilon_0)\,dz\right).$$

Отсюда и из (6.22) следует (6.18). Наконец, справедлива

Теорема 8. *Пусть $\widetilde{r}(z,t_1,t_2)$ удовлетворяет дифференциальному уравнению* (6.16) *с начальным условием* (6.17). *Положим $u = z/t'$, $\widetilde{r}(z,t_1,t_2) = r(z/t',t_1,t_2)f_{t_1^*t_2^*t'}(z)$. Тогда $r(u,t_1,t_2)$ удовлетворяет дифференциальному уравнению* (4.12) *с начальным условием* (4.13).

Доказательство. Покажем сначала, что справедливо представление вида $\widetilde{r}(z,t_1,t_2) = \mathrm{r}(z,t_1,t_2)f_{t_1^*t_2^*t'}(z)$, где функция $\mathrm{r}(z,t_1,t_2)$ удовлетворяет уравнению

$$\min_{\ell=1,2}\left(\mathrm{r}'_{t_\ell} + \frac{z}{D_\ell t'}\cdot\mathrm{r}'_z + \frac{D_\ell t_\ell^2}{2}\cdot\mathrm{r}''_{zz} + \mathrm{g}^{(\ell)}(z,t_1,t_2)\right) = 0, \tag{6.23}$$

где $\mathrm{g}^{(\ell)}(z,t_1,t_2) = \widetilde{g}^{(\ell)}(z,t_1,t_2)f^{-1}_{t_1^*t_2^*t'}(z)$, а $\{\widetilde{g}^{(\ell)}(z,t_1,t_2)\}$ определены в (6.14). Для сокращения выкладок опускаем ниже зависимость функций $f, \mathrm{r}, \widetilde{r}, \mathrm{g}, \widetilde{g}$ от z, t_1, t_2. При этом f'_{t_1}, f'_z, f''_{zz} означают соответствующие частные производные. Справедливы равенства

$$f'_{t_1} = \frac{t_1 + D_1t'}{2t_1^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)f, \quad f'_z = -\frac{z}{t_1^*t_2^*t'}f, \quad f''_{zz} = \frac{1}{t_1^*t_2^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)f.$$

Полагая $\widetilde{r} = \mathrm{r}f$ и подставляя в (6.16) при $\ell = 1$, получаем

$$\begin{aligned}
&(\mathrm{r}'_{t_1}f + \mathrm{r}f'_{t_1}) + \frac{\mathrm{r}f}{t_1} + \frac{z}{t_1}(\mathrm{r}'_zf + \mathrm{r}f'_z) + D_1\frac{t_2^2}{2}(\mathrm{r}''_{zz}f + 2\mathrm{r}'_zf'_z + \mathrm{r}f''_{zz}) + \mathrm{g}^{(1)}f = \\
&= \left(\mathrm{r}'_{t_1} + \frac{t_1 + D_1t'}{2t_1^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)\mathrm{r}\right)f + \frac{\mathrm{r}}{t_1}f + \frac{z}{t_1}\cdot\left(\mathrm{r}'_z - \frac{z}{t_1^*t_2^*t'}\cdot\mathrm{r}\right)f + \\
&+ D_1\frac{t_2^2}{2}\cdot\left(\mathrm{r}''_{zz} - \frac{2z}{t_1^*t_2^*t'}\mathrm{r}'_z + \frac{1}{t_1^*t_2^*t'}\left(\frac{z^2}{t_1^*t_2^*t'} - 1\right)\mathrm{r}\right)f + \mathrm{g}^{(1)}f = \\
&= \left(\mathrm{r}'_{t_1} + \frac{z}{D_1t'}\cdot\mathrm{r}'_z + \frac{D_1t_2^2}{2}\cdot\mathrm{r}''_{zz} + \mathrm{g}^{(1)}\right)f.
\end{aligned}$$

Поэтому выражение, соответствующее $\ell = 1$ в (6.16), обращается в нуль тогда и только тогда, когда обращается в нуль выражение, соответствующее $\ell = 1$ в (6.23). Далее положим $r(u, t_1, t_2) = \mathrm{r}(ut', t_1, t_2)$. Тогда $\mathrm{g}^{(\ell)}(ut', t_1, t_2) = g^{(\ell)}(u, t_1, t_2)$, $r'_{t_1}(u, t_1, t_2) = \mathrm{r}'_z(ut', t_1, t_2)D_1^{-1}u + \mathrm{r}'_{t_1}(ut', t_1, t_2)$, $r'_u(u, t_1, t_2) = \mathrm{r}'_z(ut', t_1, t_2)t'$, $r''_{uu}(u, t_1, t_2) = \mathrm{r}''_{zz}(ut', t_1, t_2)(t')^2$. Поэтому

$$\mathrm{r}'_{t_1}(ut', t_1, t_2) + \frac{z}{D_1 t'} \cdot \mathrm{r}'_z(ut', t_1, t_2) + \frac{D_1 t_2^2}{2} \cdot \mathrm{r}''_{zz}(ut', t_1, t_2) + \mathrm{g}^{(\ell)}(ut', t_1, t_2) =$$
$$= r'_{t_1}(u, t_1, t_2) + \frac{D_1 t_2^2}{2(t')^2} \cdot r''_{uu}(u, t_1, t_2) + g^{(\ell)}(u, t_1, t_2).$$

Поэтому выражение, соответствующее $\ell = 1$ в (6.23), обращается в нуль тогда и только тогда, когда обращается в нуль выражение, соответствующее $\ell = 1$ в (4.12). Для выражения, соответствующего $\ell = 2$, проверка выполняется аналогично. Поэтому уравнение (6.16) с начальным условием (6.17) эквивалентно уравнению (4.12) с начальным условием (4.13). ▲

Наконец, подставляя $\widetilde{r}(z, t_1, t_2) = r(u, t_1, t_2) f_{t_1^* t_2^* t'}(ut')$, $z = ut'$ при $t_1 = t_2 = \varepsilon_0$ в (6.18) и предполагая, что $r(u, t_1, t_2)$ является непрерывной функцией u, t_1, t_2, получаем, что байесовский риск при $N \to \infty$ и $\varepsilon_0 \to +0$ определяется формулой (4.11).

§ 7. Заключение

Для гауссовского двурукого бандита, характеризующего пакетную обработку данных, получено асимптотическое описание байесовского риска относительно наихудшего априорного распределения, т.е. эти результаты описывают пакетную обработку больших данных. В силу основной теоремы теории игр это позволяет находить минимаксные стратегию и риск, которые определяются полным количеством данных N и дисперсиями одношаговых доходов D_1, D_2. Поскольку при пакетной обработке одинаковые действия применяются к группам данных, можно ожидать, что эти результаты останутся справедливыми для широкого класса процессов, у которых одношаговые доходы подчиняются центральной предельной теореме. При этом дисперсии доходов можно считать известными, так как их можно оценить на коротком начальном отрезке управления, а минимаксный риск мало меняется при малом изменении дисперсий. Наибольшие потери достигаются в области "близких" распределений, для которых математические ожидания доходов различаются на величину порядка $N^{-1/2}$. Кроме того, оказалось, что в случае бернуллиевского двурукого бандита максимальные потери при оптимальной обработке больших данных по одному не могут быть сделаны меньше потерь, соответствующих оптимальной пакетной обработке. Можно ожидать, что этот результат сохранится и для широкого класса процессов с другими распределениями одношаговых доходов.

Автор выражает глубокую признательность рецензенту за внимание к статье и полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Колногоров А.В.* Гауссовский двурукий бандит и оптимизация групповой обработки данных // Пробл. передачи информ. 2018. Т. 54. № 1. С. 93–111.
2. *Perchet V., Rigollet P., Chassang S., Snowberg E.* Batched Bandit Problems // Ann. Statist. 2016. V. 44. № 2. P. 660–681.
3. *Колногоров А.В.* Задача о двуруком бандите для систем с параллельной обработкой данных // Пробл. передачи информ. 2012. Т. 48. № 1. С. 83–95.
4. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. М.: Наука, 1987.

5. *Vogel W.* A Sequential Design for the Two-Armed Bandit // Ann. Math. Statist. 1960. V. 31. № 2. P. 430–443.
6. *Vogel W.* An Asymptotic Minimax Theorem for the Two-Armed Bandit Problem // Ann. Math. Statist. 1960. V. 31. № 2. P. 444–451.
7. *Lai T.L., Levin B., Robbins H., Siegmund D.* Sequential Medical Trials // Proc. Natl. Acad. Sci. USA. 1980. V. 77. № 6. P. 3135–3138.
8. *Lai T.L., Robbins H.* Asymptotically Efficient Adaptive Allocation Rules // Adv. in Appl. Math. 1985. V. 6. № 1. P. 4–22.
9. *Kaufmann E.* On Bayesian Index Policies for Sequential Resource Allocation // Ann. Statist. 2018. V. 46. № 2. P. 842–865.
10. *Колногоров А.В.* Робастное параллельное управление в случайной среде (задача о двуруком бандите) // АиТ. 2012. № 4. С. 114–130.
11. *Колногоров А.В.* К предельному описанию робастного параллельного управления в случайной среде // АиТ. 2015. № 7. С. 111–126.
12. *Самарский А.А.* Теория разностных схем. М.: Наука, 1989.
13. *Bather J.A.* The Minimax Risk for the Two-Armed Bandit Problem // Mathematical Learning Models — Theory and Algorithms. Lect. Notes Statist. V. 20. New York: Springer-Verlag, 1983. P. 1–11.

Колногоров Александр Валерианович
Новгородский государственный университет
им. Ярослава Мудрого, кафедра
прикладной математики и информатики
kolnogorov53@mail.ru

Поступила в редакцию
06.04.2020
После доработки
02.06.2020
Принята к публикации
02.06.2020

Адрес редакции: 127051, Москва, Б. Каретный пер., 19, стр. 1, тел. (495) 650-47-39

Оригинал-макет подготовил *Д.Ю. Ногин*
по контракту с ООО «ИКЦ «АКАДЕМКНИГА»

Москва
ООО «ИКЦ «АКАДЕМКНИГА»