

УДК 621.391.8

© 2025 г. В.Г. Бабиков, А.А. Галяев

**АНАЛИТИЧЕСКИЕ МЕТРИЧЕСКИЕ СЕТКИ ДЛЯ ДИАГРАММ  
СТАТИСТИЧЕСКОЙ И СПЕКТРАЛЬНОЙ СЛОЖНОСТЕЙ<sup>1</sup>**

Статья посвящена исследованию диаграмм статистической и спектральной сложности, которые играют важную роль в решении задачи классификации сигналов. Сформулированы и доказаны леммы о построении аналитических метрических сеток для диаграмм статистической и спектральной сложности. Полученные теоретические результаты верифицированы численными экспериментами, которые подтвердили эффективность теоретических оценок.

*Ключевые слова:* информационная энтропия, дискретное преобразование Фурье, периодограмма, дисбаланс, статистическая сложность, спектральная сложность, метрические сетки

**DOI:** 10.31857/S0555292325010036, **EDN:** BVGKFI

**§ 1. Введение**

Множество прикладных задач физики и техники, биологии и медицины связано с обнаружением слабых сигналов, детектируемых на фоне естественного шума. Классические способы решения задачи обнаружения затрагивают в том числе вопросы классификации и используют дополнительную информацию о свойствах сигнала. Так решение задачи оптимальной фильтрации требует знания свойств сигнала: периодичности, полосы частот и др. [1]. Решение задачи различения двух гипотез опирается на лемму Неймана – Пирсона, определяет факт превышения оптимального порога при заданной вероятности ложной тревоги и требует оценки статистических свойств выборочных распределений шума и смеси сигнала и шума [2]. В работе [3] представлена разновидность леммы Неймана – Пирсона, особенность которой состоит в том, что мы не хотим ошибаться, принимая одну гипотезу вместо другой. Решение задачи о разладке требует настройки алгоритма на изменение неизвестных статистических характеристик распределений шума и смеси сигнала и шума, так же как в задаче поиска аномалий [4]. Все перечисленные способы демонстрируют качественную и надежную работу при превышении сигнала над шумом, но при малых отношениях сигнал/шум или в тяжелой помеховой обстановке зачастую дают неверный ответ.

Задача обнаружения сигнала  $s(n)$  традиционно сводится к задаче различения двух гипотез:

$$\begin{cases} \Gamma_0: x(n) = w(n), \\ \Gamma_1: x(n) = s(n) + w(n), \quad n = 1, \dots, N. \end{cases}$$

<sup>1</sup> Работа выполнена при частичной финансовой поддержке гранта Российского научного фонда (номер гранта 23-19-00134) и частично в рамках темы FFGZ-2025-0005 государственного задания ИПУ РАН.

Гипотеза  $\Gamma_0$  связана с принятием решения о приеме только шума, а гипотеза  $\Gamma_1$  – о приеме смеси полезного сигнала и шума, где последовательность  $\{x(n)\}$ ,  $n = 1, \dots, N$ , – спектральный ряд из принятых данных,  $\{s(n)\}$  – полезный сигнал,  $\{w(n)\}$  – аддитивный случайный шум,  $N$  – длина спектрального ряда данных. При этом считаем, что этот ряд получается после применения ДПФ к временному ряду отсчетов.

Случайные величины спектрального ряда  $(x(1), \dots, x(n), \dots, x(N))$  принимают значения  $(x_1, \dots, x_n, \dots, x_N) \in \mathbb{R}^N$ ; те же значения, нормированные на полную энергию,  $(p_1, \dots, p_n, \dots, p_N) \in \mathbb{R}^N$ , так что

$$\forall p_i \in [0, 1], \quad \sum_{i=1}^N p_i = 1. \quad (1)$$

В качестве статистики, отвечающей гипотезе  $\Gamma_0$ , можно использовать различные известные спектральные распределения и на их основе формировать критерий для вторичной обработки.

В работах [5, 6] было установлено, что одним из лучших критериев для задачи обнаружения слабого сигнала в белом шуме является статистическая сложность на основе полной вариации меры со знаком:

$$C_{TV}(p) = -\frac{1}{4 \log_2 N} \left( \sum_{i=1}^N p_i \log_2 p_i \right) \left( \sum_{i=1}^N \left| p_i - \frac{1}{N} \right| \right)^2.$$

Надо отметить, что статистическая сложность  $C_{TV}$  состоит из произведения двух информационных характеристик – информационной энтропии  $S$  и дисбаланса  $D_{TV}$ :

$$\begin{aligned} S(p) &= -\frac{1}{\log_2 N} \left( \sum_{i=1}^N p_i \log_2 p_i \right), \\ D_{TV}(p) &= TV^2(P, 1/N) = \frac{1}{4} \left( \sum_{i=1}^N \left| p_i - \frac{1}{N} \right| \right)^2, \end{aligned} \quad (2)$$

а  $TV(P, 1/N)$  вычисляется относительно равномерного распределения  $N$  отсчетов спектра.

Дополнительно к уже известным информационным характеристикам следует добавить к исследованию свойств так называемый спектральный дисбаланс  $D_S$ , вычисляемый относительно упорядоченного спектра одного окна наблюдения, и спектральную сложность  $C_S$  [7] на его основе:

$$\begin{aligned} D_S(p) &= \frac{1}{4} \left( \sum_{k=1}^N \left| p_k + \frac{1}{N \cdot K_N} \ln \frac{k}{N+1} \right| \right)^2, \\ C_S(p) &= -\frac{1}{4 \log_2 N} \left( \sum_{k=1}^N p_k \log_2 p_k \right) \left( \sum_{k=1}^N \left| p_k + \frac{1}{N \cdot K_N} \ln \frac{k}{N+1} \right| \right)^2, \end{aligned}$$

где  $K_N$  – нормировочный множитель, близкий к единице и зависящий от  $N$ .

При использовании критерия спектральной сложности для построения зависимости качества бинарной классификации AUC ROC от отношения сигнал/шум (SNR) получаются лучшие результаты по сравнению с другими критериями [8, 9]. Более того, критерий спектральной сложности может быть использован как критерий согласия наравне с критерием Фроцини.

В задачах обнаружения требуются пороги, или, другими словами, значения информационных характеристик, которые свидетельствуют о наличии сигнала в шуме. В свою очередь классификация сигналов может быть осуществлена на основании набора значений информационных характеристик. При этом необходимо учитывать, что информационные характеристики чувствительны к типу шума. Следовательно, приступая к решению задачи классификации сигналов, требуется исследовать информационные меры  $C_{TV}$ ,  $C_S$  при приеме разнообразных сигналов и различных сигнально-шумовых смесей [7–10].

Использование информационных характеристик для решения задач статистической физики, классификации медицинских анализов (например, в целях изучения текстуры изображений гистопатологии молочной железы) и др. имеет долгую историю. Уже в 1999 г. в работе [11] для количественного описания сложности двумерных узоров была введена мера сложности, основанная на среднем приросте информации.

В работе [12] поднимаются вопросы, связанные с разработкой общей меры сложности – или закономерности, или структуры – для двумерных систем. Такие задачи возникают, например, при исследовании поверхностей в геологии.

Работа [13] посвящена исследованию возможности классификации текстур посредством двумерных порядковых паттернов, в частности, посредством диаграммы статистической сложности. В [14] авторы подтверждают, что диаграмма статистической сложности является популярным инструментом для различения стохастических сигналов (шума) от детерминированного хаоса (в англоязычной литературе для этого инструмента встречается термин complexity–entropy (CE) plane). Однако авторами работы было обнаружено, что как высокоразмерные детерминированные временные ряды, так и стохастические суррогатные данные (шумы) могут быть расположены в одной и той же области диаграммы, и их представления показывают очень похожее поведение для различных сигнально-шумовых смесей или даже типов сигналов. Поэтому классификация этих данных по их положению в плоскости диаграммы может оказаться сложной или даже вводящей в заблуждение.

Вопрос о сложности различения шума и шумоподобных детерминированных сигналов также подробно освещается в работе [3]. Эта работа посвящена исследованию диаграмм статистической и спектральной сложности. Диаграммы строятся следующим образом: информационная энтропия отмечается по оси абсцисс, статистическая (спектральная) сложность – по оси ординат. В тексте работы диаграммы статистической сложности обозначаются через  $(S, C_{TV})$ , и соответственно, диаграммы спектральной сложности – через  $(S, C_S)$ . В работе даны соответствующие определения, сделаны существенные замечания о чувствительности диаграмм, сформулированы и доказаны леммы об оценке верхней и нижней границ на диаграммах статистической и спектральной сложности для различных сигнально-шумовых смесей. Определены некоторые важные закономерности поведения шумоподобных и слабых сигналов, возможности их обнаружения в условиях белого и голубого шумов.

В перечисленных и других многочисленных работах, посвященных методам двумерного информационного анализа и использующих информационные диаграммы, показана возможность решения задач обнаружения и классификации, в то время как в настоящей статье представлена альтернативная интерпретация этих методов, которая состоит в том, что диаграммы статистической сложности (как и некоторые другие двумерные информационные диаграммы) находятся во взаимно-однозначном соответствии с количеством дискретных компонент спектра сигнала и отношением сигнал/шум. Совершенно неожиданным представляется то, что, рассчитывая значения информационной энтропии и статистической сложности для полученной сигнально-шумовой смеси, одновременно можно определить количество эффективных дискрет спектра и энергетические доли сигнала и шума, но тем не менее это

так – в статье последовательно посредством построения аналитических сеток показано, что диаграммы статистической и спектральной сложностей могут быть сведены к двумерной диаграмме дискретности и отношения сигнал/шум.

Уточним, что в статье представлены аналитические оценки метрических сеток для диаграмм статистической и спектральной сложностей. Инструмент метрических сеток (при определенных условиях) позволяет определить однозначное соответствие между кортежем информационных характеристик  $S$ ,  $C_{TV}$  и кортежем параметров  $d$  (количество дискрет в спектре), SNR (отношение сигнал/шум) [3]. При этом при построении метрических сеток требуется прежде определить, какой шум присутствует в сигнально-шумовой смеси (белый шум, голубой шум, шум с равномерным спектром).

## § 2. Построение метрических сеток на диаграммах информационных характеристик

В целях удобства изложения материала повторим и расширим некоторые важные определения, сформулированные в работе [3].

**Определение 1.** Сигнал, который состоит из  $d$  синусоидальных сигналов равной амплитуды, будем называть  $d$ -сигналом, а отдельные частоты этого сигнала будем называть дискретами. Тогда спектр этого сигнала будет иметь вид

$$A_d = \sum_{i=1}^d A_0 \exp(-j2\pi f_i T), \quad (3)$$

где  $A_0$  – амплитуда единичного отсчета спектра,  $f_i$  – частота  $i$ -го отсчета (Гц),  $T$  – временной интервал сигнала в секундах.

Пример спектра и упорядоченного спектра  $d$ -сигнала приведены на рис. 1.

*Замечание.* Частоты  $d$ -дискрет  $f_i$  и размер окна  $T$ , где окном будем называть временной интервал сигнала, для которого производится дискретное преобразование Фурье (ДПФ), таковы, что выполняется следующее условие:  $\forall f_i: f_i T = r_i$ , где  $r_i \in \mathbb{N}$ ,  $r_i < N/2$ .

Смысл выражения  $r_i \in \mathbb{N}$  заключается в том, что целое количество периодов  $r_i$  соответствующей  $i$ -й  $d$ -дискреты спектра должно укладываться во временное окно  $T$  сигнально-шумовой смеси.

**Определение 2.** Энергетической долей  $q$  сигнала будем называть отношение энергии сигнала к сумме энергий сигнала и шума в сигнально-шумовой смеси, и соответственно,  $1 - q$  означает энергетическую долю шума.

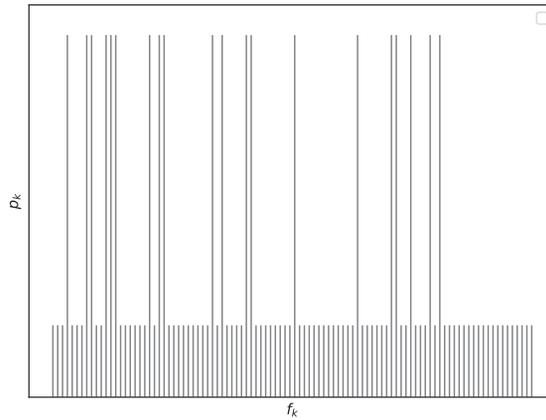
Отношение сигнал/шум определим как

$$\text{SNR} = 10 \lg \left( \frac{q}{1 - q} \right).$$

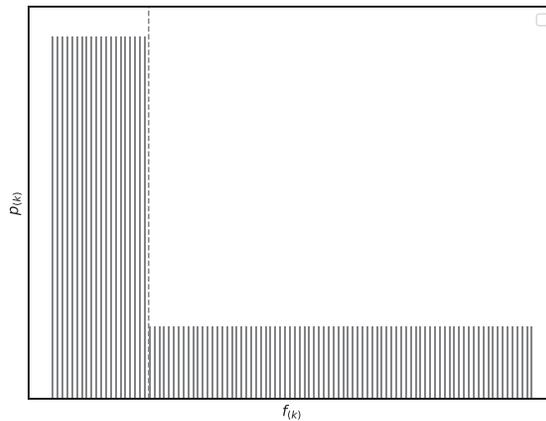
Сигналы при отсутствии шума, когда  $q = 1$ , будем называть “чистыми”.

Наконец, сформулируем определения равномерной метрической сетки, белой и голубой сеток, а также сформулируем и докажем леммы об аналитических оценках метрических сеток для диаграмм статистической и спектральной сложностей.

**Определение 3.** Равномерной сеткой будем называть метрическую двумерную разметку, по которой можно оценивать количество дискрет в спектре сигнала, а также определять отношение сигнал/шум на диаграммах информационной энтропии и статистической или спектральной сложности в условиях равномерного шума



а) Неупорядоченный спектр



б) Упорядоченный спектр

Рис. 1. Энергетический спектр  $d$ -сигнала в равномерном шуме ( $d = 20$ ,  $N = 100$ )

(соответственно, метрическую сетку в условиях белого шума будем называть белой, и т.д. по цвету или типу шума (помехи)).

Далее в статье подробнее раскрывается понятие сетки. В этом контексте сетка – это приближенное взаимно-однозначное отображение кортежа таких информационных характеристик сигнально-шумовой смеси, как информационная энтропия и статистическая (спектральная) сложность, и кортежа таких характеристик, как дискретность (количество дискрет в спектре  $d$ -сигнала) и отношение сигнал/шум.

Прежде чем приступить к анализу возможностей метрических сеток, сформулируем задачи исследования.

**Задача 1** (равномерная сетка для статистической сложности). Оценить зависимости значений  $C_{TV}$  от энтропии  $S$  для различных значений дискретности  $d$  и для различных уровней сигнала в шуме с равномерным спектром ( $q$  – энергетическая доля сигнала в сигнально-шумовой смеси).

**Лемма 1** (равномерная сетка для диаграммы  $C_{TV}$ ). *Статистическая сложность  $C_{TV} = C_{TV}(q, d, N)$  и энтропия  $S = S(q, d, N)$  определяются следующими*

зависимостями от  $q$ ,  $d$  и  $N$ :

$$S(q, d, N) = \frac{1}{\ln N} \left( q \ln \frac{q}{d} + (1 - q) \ln \frac{1 - q}{N - d} \right),$$

$$C_{TV}(q, d, N) = S(q, d, N) \left( q - \frac{d}{N} \right)^2,$$

где  $q$  – энергетическая доля  $d$ -сигнала в сигнально-шумовой смеси (шум с равномерным спектром).

**Доказательство.** Согласно определению, приведенному выше,  $p_i$  – нормированный упорядоченный энергетический спектр (1). Подставляя в выражение для энтропии (2) нормированный упорядоченный энергетический спектр смеси равномерного шума и  $d$ -сигнала (упорядоченный спектр такой смеси выглядит в виде “ступеньки”, см. рис. 1б), получаем

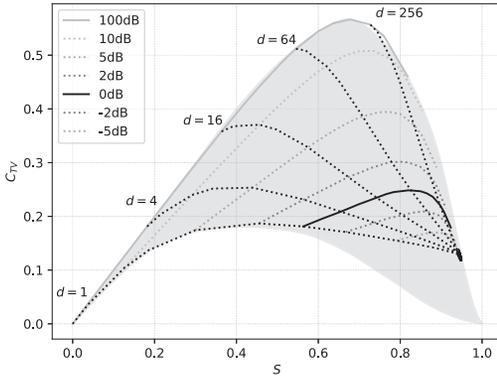
$$\begin{aligned} S(q, d, N) &= \frac{1}{\ln N} \left( \sum_{k=1}^d \frac{q}{d} \ln \frac{q}{d} + \sum_{k=1}^{N-d} \frac{1-q}{N-d} \ln \frac{1-q}{N-d} \right) = \\ &= \frac{1}{\ln N} \left( q \ln \frac{q}{d} + (1-q) \ln \frac{1-q}{N-d} \right), \\ D_{TV} &= \left( \frac{d}{N} - q \right)^2, \quad C_{TV} = S \cdot D_{TV}. \quad \blacktriangle \end{aligned}$$

На рис. 2 на диаграммах информационной сложности ( $S, C_{TV}$ ) и ( $S, C_S$ ) представлено при помощи метрической сетки расположение  $d$ -сигналов ( $-15 \text{ дБ} < \text{SNR} < 100 \text{ дБ}$ ) с количеством дискрет 1, 4, 16, 64 и 256 (тонкие линии, нижняя линия соответствует дискретности  $d = 1$ ) в смеси с белым (рис. 2а,б) и голубым (рис. 2в,г) шумом. Графики построены по результатам численного эксперимента (каждый узел сетки рассчитан как среднее по 100 реализациям спектра сигнально-шумовой смеси). На рисунках показаны также уровни отношения сигнал/шум SNR (тонкие линии серого цвета соответствуют уровням  $-5 \text{ дБ}$ ,  $-2 \text{ дБ}$ ,  $2 \text{ дБ}$ ,  $5 \text{ дБ}$ ,  $10 \text{ дБ}$ ,  $100 \text{ дБ}$ , черная линия соответствует уровню  $0 \text{ дБ}$ ). Закрашенная область – оценки области значений энтропии и статистической и спектральной сложностей, которые были получены в результате численного эксперимента и представлены в работе [3]. Здесь стоит сказать, что границы этих областей были получены посредством имитации сигнально-шумовых смесей  $d$ -сигналов, импульсов и различных шумов (белого, голубого, фиолетового, розового, коричневого).

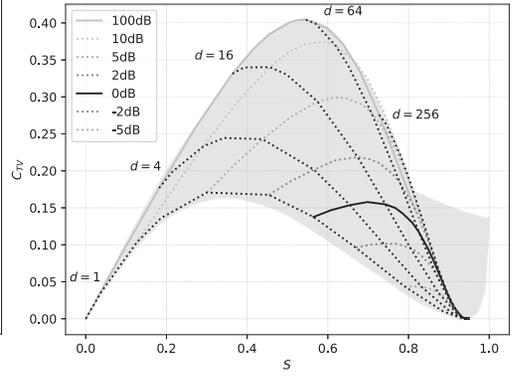
**Задача 2** (белая сетка для статистической сложности). Оценить зависимости значений  $C_{TV}(q, d, N)$  и энтропии  $S(q, d, N)$  для  $d$  дискрет в спектре  $d$ -сигнала и для различных уровней сигнала в белом шуме ( $q$  – энергетическая доля сигнала в сигнально-шумовой смеси).

**Лемма 2** (белая сетка для диаграммы  $C_{TV}$ ). *Приближенные оценки статистической сложности  $C_{TV}(q, d, N)$ , статистического дисбаланса  $D_{TV}(q, d, N)$  и информационной энтропии  $S(q, d, N)$  для смеси  $d$ -сигнала и белого шума задаются следующими зависимостями от  $q$ ,  $d$  и  $N$ :*

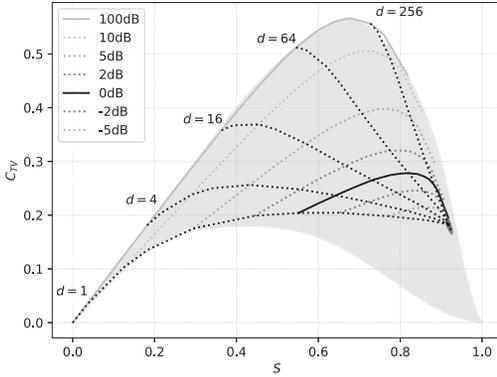
$$S(q, d, N) = S_0(N) \cdot (1 - q) - \frac{\ln \left( \left( \frac{q}{d} \right)^q (1 - q)^{1-q} \right)}{\ln N},$$



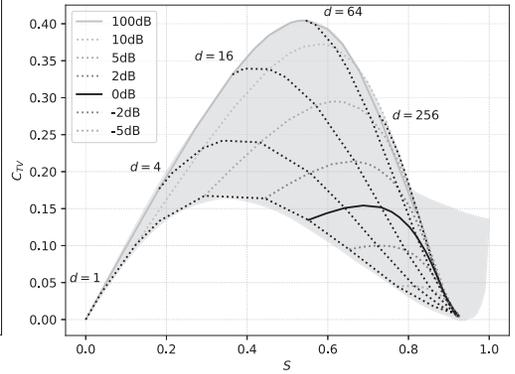
(a)  $C_{TV}$ , белый шум



(б)  $C_S$ , белый шум



(e)  $C_{TV}$ , голубой шум



(e)  $C_S$ , голубой шум

Рис. 2. Метрические сетки на диаграммах статистической и спектральной сложности (численный расчет)

$$D_{TV}(q, d, N) = \frac{1}{4} \left( \alpha(d, N) + 1 + q + (1 - q) \left( -1 + 2 \exp \left( \frac{1}{q - 1} \right) \right) \right)^2,$$

$$C_{TV}(q, d, N) = S(q, d, N) \cdot D_{TV}(q, d, N),$$

где  $N \gtrsim 1000$  – размер спектрального ряда, а  $S_0(N)$  – значение информационной энтропии для белого шума (согласно работе [7] имеем

$$S_0(N) \approx 1 + \frac{\gamma - 1}{\ln N},$$

где  $\gamma$  – константа Эйлера–Маскерони; для  $N = 2048$  получаем  $S_0 \approx 0,9446$ ),  $\alpha(d, N) = -\frac{2d}{N}$ ,  $q$  – энергетическая доля  $d$ -сигнала в сигнально-шумовой смеси.

Доказательство. Воспользуемся результатами, сформулированными в работе [7]. Согласно этой работе среднее значение упорядоченного нормированного энергетического спектра белого шума подчиняется следующему распределению:

$$n_{(k)} = -\frac{1}{N \cdot K_N} \cdot \ln \left( \frac{k}{N + 1} \right).$$

Воспользовавшись этим, получаем следующие выражения для информационной энтропии и статистического дисбаланса (см. также [3]):

$$S(q, d, N) = S_0(N) \cdot (1 - q) - \frac{\ln \left( \left( \frac{q}{d} \right)^q (1 - q)^{1-q} \right)}{\ln N},$$

$$D_{TV}(q, d, N) = \frac{1}{4} \left( 1 + q + \frac{1}{K_N} (1 - q) \left( -1 + B + 2 \exp \left( \frac{K_N}{q - 1} \right) \right) \right)^2,$$

где

$$B = \frac{N}{N + 1} \ln N + \ln \frac{1}{N + 1} \approx -0,00421$$

(для  $N = 2048$ ), а  $K_N$  – нормировочный множитель в распределении частот белого шума, близкий к единице ( $K_N \approx 0,99818$  для  $N = 2048$ ).

Коэффициент  $\alpha(d, N)$  (поправка к дисбалансу) оценен исходя из тех соображений, что при малых  $q$  и при больших  $d$  “ступенька”  $d$ -сигнала становится маленькой (энергия шума в области  $d$ -дискрет становится сравнимой с энергией сигнала) и влияние шума проявляется еще и при оценке энергии  $d$ -дискрет в сигнально-шумовой смеси:  $\alpha(d, N) = -\frac{2d}{N}$ . В силу того, что  $B \approx 0$ , а  $K_N \approx 1$  (для  $N \gtrsim 1000$ ), получаем следующее выражение:

$$D_{TV}(q, d, N) = \frac{1}{4} \left( \alpha(d, N) + 1 + q + (1 - q) \left( -1 + 2 \exp \left( \frac{1}{q - 1} \right) \right) \right)^2,$$

откуда следует, что

$$C_{TV}(q, d, N) = S(q, d, N) \cdot D_{TV}(q, d, N). \quad \blacktriangle$$

**Задача 3** (белая сетка для спектральной сложности). Оценить зависимости значений  $C_S(q, d, N)$  и энтропии  $S(q, d, N)$  для  $d$  дискрет в спектре  $d$ -сигнала и для различных уровней сигнала в белом шуме ( $q$  – энергетическая доля сигнала в сигнально-шумовой смеси).

**Лемма 3** (белая сетка для диаграммы  $C_S$ ). *Приближенные оценки спектральной сложности  $C_S(q, d, N)$ , спектрального дисбаланса  $D_S(q, d, N)$  и информационной энтропии  $S(q, d, N)$  для смеси  $d$ -сигнала и белого шума задаются следующими зависимостями от  $q$ ,  $d$  и  $N$ :*

$$S(q, d, N) = S_0(N) \cdot (1 - q) - \frac{\ln \left( \left( \frac{q}{d} \right)^q (1 - q)^{1-q} \right)}{\ln N},$$

$$D_S(q, d, N) = \left( q + \frac{d(1 + q)}{2N \cdot K_N} \left( 1 + \ln \left( \frac{N + 1}{d} \right) \right) + \beta \right)^2,$$

$$C_S(q, d, N) = S(q, d, N) \cdot D_S(q, d, N),$$

где  $N \gtrsim 1000$  – размер спектрального ряда,  $S_0(N)$  – значение информационной энтропии для белого шума,  $\beta(q, d, N) = -\frac{2}{N}(1 - q)(d - 1)$ ,  $q$  – энергетическая доля  $d$ -сигнала в сигнально-шумовой смеси.

**Доказательство.** Результат для оценки информационной энтропии представлен в лемме 2. В целях оценки спектрального дисбаланса введем следующие

обозначения:

$$\varphi_0(d, N) = - \sum_{k=1}^d \frac{1}{N \cdot K_N} \ln \frac{k}{N+1}, \quad \varphi(q, d, N) = (1-q)\varphi_0(d, N).$$

Тогда выражение для дисбаланса принимает вид

$$D_S(q, d, N) = \frac{1}{4}(|q - \varphi(q, d, N)| + 1 - (1-q) + |\varphi_0(d, N) - \varphi(q, d, N)|)^2,$$

$$\varphi_0(d, N) = - \sum_{k=1}^d \frac{1}{N \cdot K_N} \ln \frac{k}{N+1} = - \frac{1}{N \cdot K_N} (\ln d! - d \ln(N+1)).$$

Далее для приближенного вычисления применяем формулу Стирлинга:

$$\varphi_0(d, N) \approx - \frac{1}{N \cdot K_N} (d \ln d - d - d \ln(N+1)) = \frac{d}{N \cdot K_N} \left( \ln \left( \frac{N+1}{d} \right) + 1 \right),$$

$$D_S(q, d, N) \approx \frac{1}{4} (2q + (1+q)\varphi_0(d, N))^2.$$

Подставляем выражение для  $\varphi_0(d, N)$  в формулу для  $D_S(q, d, N)$ , затем спектральный дисбаланс умножаем на энтропию, далее получаем выражение для спектральной сложности  $C_S(q, d, N)$ :

$$C_S(q, d, N) = S(q, d, N) \cdot D_S(q, d, N),$$

$$D_S(q, d, N) = \left( q + \frac{d(1+q)}{2N \cdot K_N} \left( 1 + \ln \left( \frac{N+1}{d} \right) \right) \right)^2.$$

Дополнительно необходимо иметь ввиду, что к дискретам сигнала добавляются дискреты шума (в доказательстве леммы 2 показан закон распределения упорядоченного нормированного энергетического спектра белого шума), поэтому в целях построения метрической сетки для сигналов с бóльшим количеством дискрет необходимо уточнение оценок энтропии и дисбаланса сигнально-шумовой смеси.

Сравнение численного моделирования и аналитических выражений показывает (см. сравнение расчета и численного эксперимента на рис. 3б), что теоретические формулы для информационной энтропии и спектральной сложности близко оценивают их фактическое поведение. Для повышения качества оценки для большого количества дискрет в сигнале требуется дополнительный поправочный коэффициент  $\beta(q, d, N)$  при расчете дисбаланса  $D_S(q, d, N)$ .

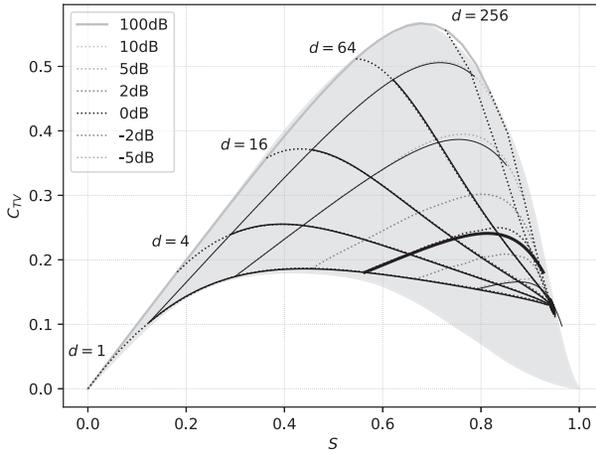
Коэффициент  $\beta(q, d, N)$  (поправка к дисбалансу) оценивается исходя из тех предположений, что при малых  $q$  и при больших  $d$  “ступенька”  $d$ -сигнала становится маленькой (энергия шума в области  $d$ -дискрет становится сравнимой с энергией сигнала) и влияние шума проявляется еще и при оценке энергии  $d$ -дискрет в сигнально-шумовой смеси:

$$\beta(q, d, N) = - \frac{2}{N} (1-q)(d-1).$$

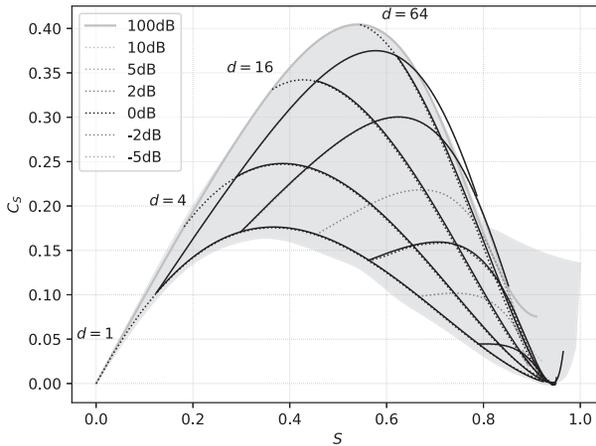
Для оценки дисбаланса окончательно получаем

$$D_S(q, d, N) = \left( q + \frac{d(1+q)}{2N \cdot K_N} \left( 1 + \ln \left( \frac{N+1}{d} \right) \right) + \beta(q, d, N) \right)^2,$$

что завершает доказательство леммы. ▲



(a)  $C_{TV}$



(б)  $C_S$

Рис. 3. Сравнение численного эксперимента (тонкие пунктирные линии) и расчета (сплошные линии, которые расходятся веером из точки с энтропией  $S_0 \approx 0,9446$ ,  $-d$ -сигналы, количество дискрет 1, 4, 16, 64, а также пересекающие их сплошные линии – уровни отношения сигнал/шум, леммы 2, 3)

На рис. 3 представлен сравнительный анализ фактических и расчетных (по леммам 2 и 3) метрических белых сеток ( $N = 2048$ ). Согласно иллюстрациям для соответствующего размера ряда  $N$  диаграмма  $(S, C_{TV})$  позволяет классифицировать сигналы с дискретностью до 256 и отношением сигнал/шум до  $-5$  дБ.

Сформулированные и доказанные в настоящей статье леммы позволяют сформулировать тезис о том, что при определенных условиях (имеется ввиду при определенных уровнях дискретности и отношения сигнал/шум) между кортежем информационных характеристик  $S, C_{TV}$  с одной стороны и кортежем таких параметров, как  $d$  (дискретность сигнала) и SNR (отношение сигнал/шум), существует взаимнооднозначное соответствие. При этом предельные уровни SNR зависят от количества дискрет в сигнале. Так, одиночная дискрета согласно работе [3] может быть обнаружена при уровнях сигнала ниже  $-20$  дБ, в то время как для обнаружения и клас-

сификации сигналов с дискретностью 256 дискрет требуются уровни  $\text{SNR} > -5$  дБ (это справедливо для размера ряда  $N = 2048$ , с увеличением размера ряда пределы меняются и возможности обнаружения и классификации возрастают). В работе [3] представлены оценки границы отношения сигнал/шум, при которых возможна классификация сигналов в зависимости от уровня дискретности  $d$  для размера ряда  $N$  по диаграмме статистической сложности. Выше этой границы существует взаимно-однозначное соответствие между соответствующей точкой на информационной диаграмме  $(S, C_{TV})$  и значениями  $(d, \text{SNR})$ , где  $d$  – количество дискрет в  $d$ -сигнале, а  $\text{SNR}$  – отношение сигнал/шум, выраженное в децибелах.

Получается, что для указанной области для любой сигнально-шумовой смеси (тип шума должен быть определен) имеется возможность однозначно определить количество  $d$  дискрет принимаемого сигнала и уровень отношения сигнал/шум. При этом надо понимать, что количество дискрет является некоторой обобщенной характеристикой спектра, здесь может речь идти о некотором “эффективном” количестве дискрет, которое можно оценить аппроксимацией реального сигнала некоторым  $d$ -сигналом [15]. Сформулируем определение, которое будем использовать в дальнейшем в целях краткости изложения.

**Определение 4.** Произвольную сигнально-шумовую смесь будем называть “белым оригиналом  $d$ -сигнала уровня  $\text{SNR}$  и дискретностью  $d$ ”, если информационные характеристики (информационная энтропия и статистическая сложность) смеси  $d$ -сигнала с количеством дискрет  $d$  и белого шума в такой пропорции, что уровень отношения сигнал/шум определен как  $\text{SNR}$ , равны информационным характеристикам этой сигнально-шумовой смеси.

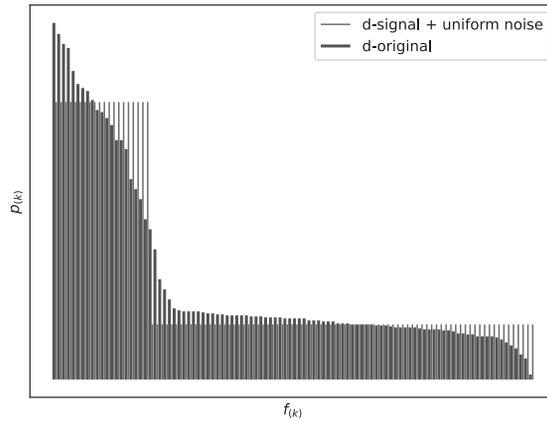
Аналогичные определения можно дать для “голубого оригинала  $d$ -сигнала уровня  $\text{SNR}$ ”, “равномерного оригинала  $d$ -сигнала уровня  $\text{SNR}$ ”. Понятно, что у каждой смеси  $d$ -сигнала и шума бесконечное множество оригиналов. На рис. 4 показаны примеры упорядоченных спектров смеси  $d$ -сигнала и шума с равномерным спектром (импульса) и смеси  $d$ -сигнала и белого шума, а также примеры упорядоченных спектров оригиналов этих сигнально-шумовых смесей.

### § 3. Численный анализ свойств метрических сеток

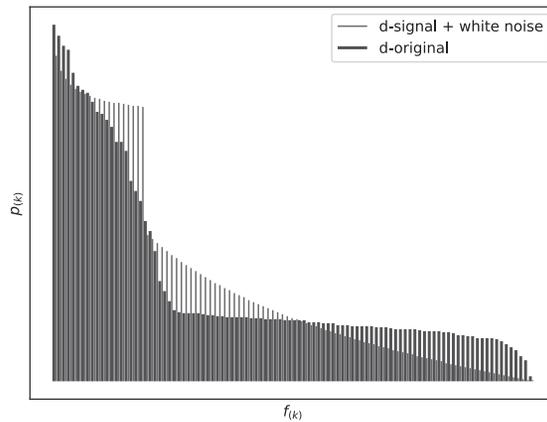
Применим аналитические оценки, полученные в статье, для построения диаграмм на основе метрических сеток при различных шумовых и сигнальных условиях.

На рис. 5, 6 для получения энергетического спектра сигнально-шумовой смеси было применено БПФ (быстрое преобразование Фурье) и представлен анализ сигналов, принадлежащих различным классам. На диаграмме статистической сложности (см. рис. 5а) на фоне метрической сетки представлены сигналы, принадлежащие различным классам. Посредством преобразования диаграммы статистической сложности в диаграмму дискретности на рис. 5б представлены сигналы этих же классов. На диаграмме дискретности дополнительно показаны средние значения (пунктирные вертикальные линии) и области стандартных отклонений по дискретности. Оценки дискретности и отношения сигнал/шум получены по узлам сетки (ближайший узел сетки к результату расчета по отдельному временному окну сигнала (фрейму), в настоящей статье временной отрезок равен одной секунде). Большой разброс в области  $\text{SNR} < -5$  дБ (более темные точки) связан с тем, что такие слабые сигналы с количеством дискрет  $d > 100$  практически невозможно классифицировать при помощи диаграммы сложности на фоне голубого шума – т.е. определить количество дискрет в сигнале при таком малом отношении сигнал/шум и размере спектрального ряда  $N = 2048$ .

На рис. 6 представлен анализ таких же сигналов, но спектры которых получены при помощи периодограммы (метод периодограмм позволяет точнее определить



(a) Упорядоченный спектр смеси  $d$ -сигнала и равномерного шума



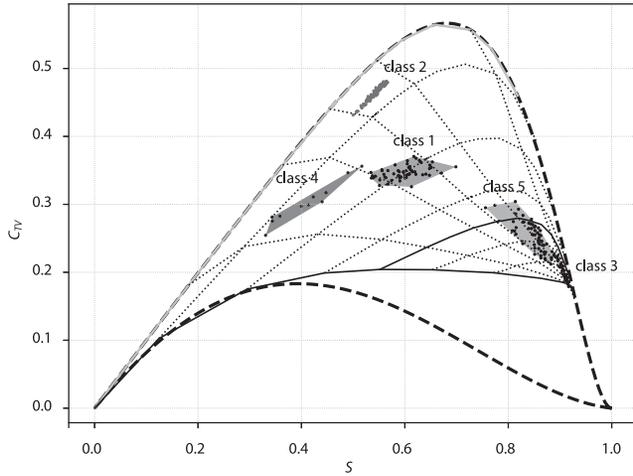
(б) Упорядоченный спектр смеси  $d$ -сигнала и белого шума

Рис. 4. Упорядоченные нормированные энергетические спектры. Сравнение реальной сигнально-шумовой смеси ( $d$ -оригинал) и смеси  $d$ -сигнала с равномерным по спектру шумом (сверху) и смеси  $d$ -сигнала с белым шумом (снизу)

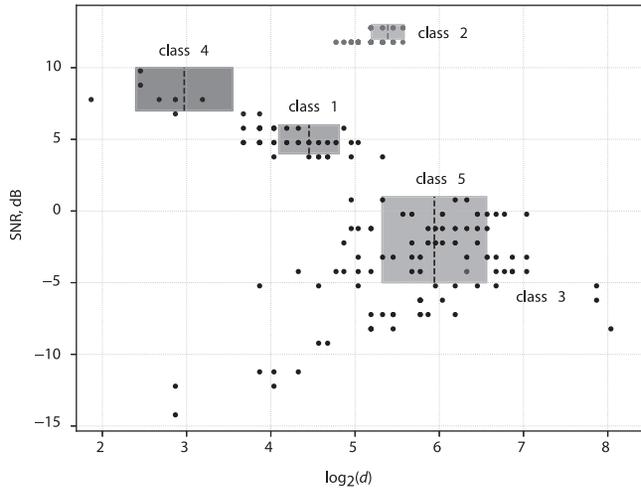
“истинные” частоты дискрет сигнала). Дискретный набор размера  $N = 2048$  получен следующим образом: полный ряд полученных частот ограничивается частотой  $f = 2048$  Гц, далее весь спектр разделяется на дискретные интервалы ( $0 \text{ Гц} < f_0 \leq 1 \text{ Гц}$ ,  $1 \text{ Гц} < f_0 \leq 2 \text{ Гц}$ , ...,  $2047 \text{ Гц} < f_0 \leq 2048 \text{ Гц}$ ), в каждом интервале энергии всех дискрет суммируются, и по полученным значениям формируется ряд соответствующего размера. Такой метод построения спектра позволяет точнее определить уровень дискретности сигнала в сравнении с БПФ.

В целях иллюстрации метода для периодограмм (рис. 6) использована белая сетка, но стоит еще раз отметить, что для корректных оценок уровней дискретности и отношения сигнал/шум метрическая сетка должна соответствовать типу шума (белый, голубой и т.д.)

Для получения метрических сеток на диаграммах статистической и спектральной сложности (в примерах на рис. 2, 3, 6) были сгенерированы сигнально-шумовые смеси  $d$ -сигналов и белого и голубого шумов с частотой дискретизации 4098 Гц, соответственно, длина временного ряда данных (фрейма) для расчета информационных



(а) Диаграмма сложности ( $S, C_{TV}$ )

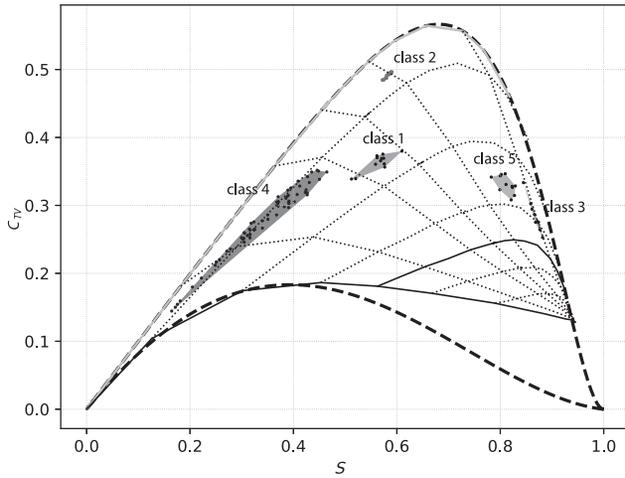


(б) Диаграмма дискретности ( $d, SNR$ )

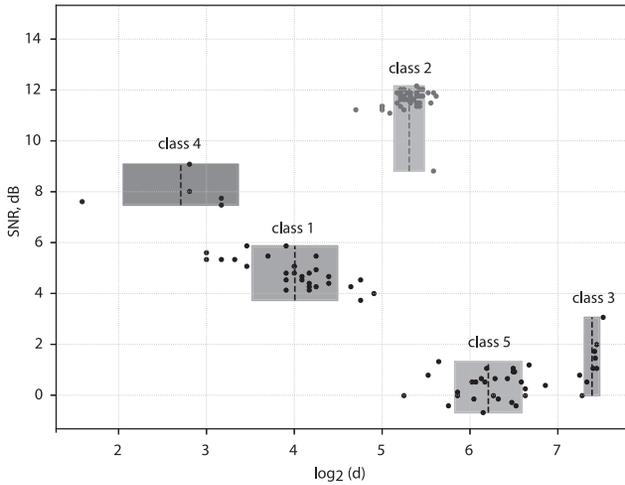
Рис. 5. Классификация сигналов (голубая сетка). Преобразование Фурье

характеристик определяется как  $2N + 2 = 4098$ , размер энергетического спектра равен  $N = 2048$  (убирается нулевая частота). Подробнее об исследовании спектрального ряда и вычислении информационных характеристик см. [7–9].

Для определения каждого узла сетки набиралась статистика на  $Q = 100$  энергетических спектрах численно сгенерированных фреймах сигнально-шумовой смеси размера  $N = 2048$ . Во всех реализациях параметры  $d$ -сигнала оставались одними и теми же, т.е. фиксировались наборы синусоид по количеству и амплитудам со случайными фазами. Аддитивный белый гауссовский шум (как и голубой шум) был получен генератором белого (голубого) шума библиотеки `python-acoustics` с одними и теми же параметрами (в рамках одного набора  $Q$  фреймов). Далее для каждого узла сетки на основании полученных упорядоченных спектральных после-



(а) Диаграмма сложности ( $S, C_{TV}$ )



(б) Диаграмма дискретности ( $d, SNR$ )

Рис. 6. Классификация сигналов (белая сетка). Периодограммы

довательностей рассчитывались средние значения  $S$ ,  $C_{TV}$  и  $C_S$  для смеси шума с  $d$ -сигналом.

Основной количественной характеристикой “зашумленности” сигнала являлось отношение сигнал/шум – SNR (signal-to-noise ratio), которое описывается формулой

$$SNR = 10 \lg \left( \frac{E_{\text{signal}}}{E_{\text{noise}}} \right) = 10 \lg \left( \frac{q}{1 - q} \right),$$

где  $E_{\text{signal}}$ ,  $E_{\text{noise}}$  – полные энергии сигнала и шума соответственно. Таким образом, каждый узел сетки получен для конкретного значения дискретности  $d$  сигнала, а также конкретного значения  $q$  – энергетической доли сигнала в сигнально-шумовой смеси.

## § 4. Заключение

К категории наиболее сложных задач вторичной обработки информации относится задача классификации. Для решения этой задачи необходимо сформировать признаковое пространство. Классификационные критерии, образующие признаковое пространство, должны быть достаточно информативны, чтобы посредством методов машинного обучения (нейронные сети, метод опорных векторов, метод ближайших соседей, генетический алгоритм) ошибки первого и второго рода были минимальными.

В задаче классификации при выбранном пороге (например, по критерию максимального правдоподобия) уменьшить ошибку классификации можно двумя путями: увеличением объема обучающей выборки или использованием признаков, дающих наибольшее расстояние между классами.

В ряде случаев создание больших обучающих выборок является неразрешимой проблемой, поэтому остается только один путь уменьшения вероятности ошибки классификации – это повышение качества признаков.

Описанные в настоящей статье подходы по оценке метрических сеток позволяют из совокупности информационных признаков составить классификационное ядро информативности. Следующие работы будут посвящены новым классификационным признакам, составленным на основе уже известных информационных характеристик: энтропии, статистической и спектральной сложности, статистического и спектрального дисбаланса.

## СПИСОК ЛИТЕРАТУРЫ

1. *Ronald L.A., Duncan W.M.* Signal Analysis: Time, Frequency, Scale, and Structure. Piscataway, NJ: IEEE Press, 2004.
2. *Ширяев А.Н.* Вероятностно-статистические методы в теории принятия решений. М.: МЦНМО, 2020.
3. *Бабиков В.Г., Галяев А.А.* Диаграммы статистической и спектральной сложности // Пробл. передачи информ. 2024. Т. 60. № 2. С. 25–35. <https://doi.org/10.31857/S0555292324020037>
4. *Mehrotra K.G., Mohan C.K., Huang H.-M.* Anomaly Detection Principles and Algorithms. Cham: Springer, 2017. <https://doi.org/10.1007/978-3-319-67526-8>
5. *Галяев А.А., Лысенко П.В., Берлин Л.М.* Статистическая сложность как критерий задачи обнаружения полезного сигнала // Автомат. и телемех. 2023. № 7. С. 121–145. <https://www.mathnet.ru/rus/at16133>
6. *Berlin L.M., Galyaev A.A., Lysenko P.V.* Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments // Sensors. 2023. V. 23. № 4. P. 2133 (17 pp.). <https://doi.org/10.3390/s23042133>
7. *Галяев А.А., Бабиков В.Г., Лысенко П.В., Берлин Л.М.* Новая спектральная мера сложности и ее возможности по обнаружению сигналов в шуме // Докл. РАН. Матем., информ., процессы упр. 2024. Т. 518. С. 80–88. <https://doi.org/10.31857/S2686954324040122>
8. *Galyaev A.A., Babikov V.G., Lysenko P.V., Berlin L.M.* Addition to the Article “A New Spectral Measure of Complexity and Its Capabilities for Detecting Signals in Noise” // Dokl. Math. 2024. V. 110. № 1. P. 369–371. <https://doi.org/10.1134/S1064562424702247>
9. *Галяев А.А., Берлин Л.М., Лысенко П.В., Бабиков В.Г.* Порядковые статистики нормированного спектрального распределения для обнаружения слабых сигналов в белом шуме // Автомат. и телемех. 2024. № 12. С. 49–69. <https://doi.org/10.31857/S0005231024120039>
10. *Бабиков О.В., Лысенко П.В., Берлин Л.М.* Информационные характеристики в задачах обнаружения // Тр. 66-й Всерос. научной конф. МФТИ “Радиотехника и компьютерные технологии” (Москва 2024). М: Физматкнига, 2024. С. 45–50.

11. *Andrienko Yu.A., Brilliantov N.V., Kurths J.* Complexity of Two-Dimensional Patterns // Eur. Phys. J. B. 2012. V. 15. № 3. P. 539–546. <https://doi.org/10.1007/s100510051157>
12. *Feldman D.P., Crutchfield J.P.* Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy // Phys. Rev. E. 2003. V. 67. № 5. P. 051104 (9 pp.). <https://doi.org/10.1103/PhysRevE.67.051104>
13. *Bandt C., Wittfeld K.* Two New Parameters for the Ordinal Analysis of Images // Chaos. 2023. V. 33. № 4. P. 043124 (12 pp.). <https://doi.org/10.1063/5.0136912>
14. *Kottlarz I., Parlitz U.* Ordinal Pattern-Based Complexity Analysis of High-Dimensional Chaotic Time Series // Chaos. 2023. V. 33. № 5. P. 053105 (11 pp.). <https://doi.org/10.1063/5.0147219>
15. *Babikov V.G., Galyaev A.A.* Analytical Representation of Complexity Diagrams // Probl. Inf. Transm. 2025. V. 61. № 1. P. 27–40. <https://doi.org/10.1134/S003294602501003X>

*Бабиков Владимир Георгиевич*  
*Галеев Андрей Алексеевич*  
 Институт проблем управления  
 им. В.А. Трапезникова РАН, Москва  
 babikov@ipu.ru  
 galaev@ipu.ru

Поступила в редакцию  
 11.11.2024  
 После доработки  
 12.03.2025  
 Принята к публикации  
 20.03.2025

## ANALYTICAL METRIC GRIDS FOR STATISTICAL AND SPECTRAL COMPLEXITY DIAGRAMS

© 2025     **V.G. Babikov, A.A. Galyaev**

Trapeznikov Institute of Control Sciences  
 of the Russian Academy of Sciences, Moscow  
 babikov@ipu.ru, galaev@ipu.ru

The paper is devoted to the study of statistical and spectral complexity diagrams, which play an important role in solving the signal classification task. Lemmas on the construction of analytical metric grids for statistical and spectral complexity diagrams are formulated and proved. The obtained theoretical results are verified by numerical experiments, which confirmed the effectiveness of theoretical estimates.

*Keywords:* information entropy, discrete Fourier transform, periodogram, imbalance, statistical complexity, spectral complexity, metric grids

**DOI:** 10.31857/S0555292325010036, **EDN:** BVGKFI

Received 11.11.2024  
 Revised 12.03.2025  
 Accepted 20.03.2025