

УДК 51-76, 515.124

МНОГОМЕРНАЯ p -АДИЧЕСКАЯ МЕТРИКА
И ГЕНЕТИЧЕСКИЙ КОДС. В. Козырев¹, А. Ю. Хренников²¹ Математический институт им. В. А. Стеклова РАН,
119991, Москва, ул. Губкина, 8.² Университет им. Карла Линнея, Векшо, Швеция.

E-mails: kozyrev@mi.ras.ru, andrei.khrennikov@lnu.se

Обсуждается класс метрик на многомерных p -адических пространствах. Для рассматриваемых метрик набор шаров отличается от набора шаров для стандартной метрики, более того, можно рассматривать разные наборы шаров, зависящие от положения и масштаба. В качестве примера приложения вводимых метрических пространств рассматривается параметризация генетического кода. Показывается, что вырождение генетического кода описывается метрическим пространством из рассмотренного класса, то есть отображение генетического кода постоянно на шарах относительно метрики из введённого класса.

Ключевые слова: генетический код, ультраметрика.

1. Конструкция метрики. Стандартная многомерная ультраметрика на \mathbb{Q}_p^d имеет вид

$$d(x, y) = \max_{i=1, \dots, d} (|x_i - y_i|_p), \quad x, y \in \mathbb{Q}_p^d, \quad x = (x_1, \dots, x_d), \quad y = (y_1, \dots, y_d). \quad (1)$$

В работе [1] изучалась деформированная многомерная метрика на \mathbb{Q}_p^d вида

$$d_{q_1, \dots, q_d}(x, y) = \max_{i=1, \dots, d} (q_i |x_i - y_i|_p), \quad p^{-1} < q_i \leq 1$$

и строились системы всплесков, отвечающие такой метрике (а также некоторым обобщениям такой метрики, получаемым изометрическими линейными преобразованиями в \mathbb{Q}_p^d). Стандартная метрика имеет вид $d = d_{1, \dots, 1}$, q_i — веса деформации.

В настоящей работе обобщается класс метрик введением зависимости весов q_i от положения и масштаба. А именно, рассматривается дерево шаров $\mathcal{T}(\mathbb{Q}_p^d, d)$. Вершины этого дерева являются шарами в \mathbb{Q}_p^d относительно стандартной метрики d , вершины соединяются рёбрами, если для соответствующих шаров один из шаров является максимальным подшаром во втором шаре. На дереве $\mathcal{T}(\mathbb{Q}_p^d, d)$ есть естественный частичный порядок, отвечающий вложению шаров.

Рассмотрим теперь набор весов $\{q_i\}$, $i = 1, 2, \dots, d$, $q_i = q_i(I)$, $p^{-1} < q_i \leq 1$, являющийся вектор-функцией на дереве $\mathcal{T}(\mathbb{Q}_p^d, d)$ (дереве шаров относительно стандартной метрики).

Введём метрику s на \mathbb{Q}_p^d , имеющую вид

$$s(x, y) = \max_{i=1, \dots, d} (q_i(\sup(x, y)) |x_i - y_i|_p), \quad p^{-1} < q_i(J) \leq 1, \quad J \in \mathcal{T}(\mathbb{Q}_p^d, d). \quad (2)$$

Сергей Владимирович Козырев (д.ф.-м.н.), ведущий научный сотрудник, отдел математической физики. Андрей Юрьевич Хренников (д.ф.-м.н.), директор центра, центр математического моделирования, отделение математики.

Здесь $\text{sup}(x, y)$ есть наименьший шар (относительно стандартной метрики d), содержащий точки x и y .

Рассмотрим также более сложный пример метрики, получаемый из (2) применением зависящей от шара изотметрии. А именно, рассмотрим для шаров \mathbb{Q}_p^d относительно стандартной метрики (1) набор mod p -аффинных изометрий $\{U\}$ (см. [2]), то есть взаимно однозначных отображений шара в этот же шар, являющихся изометриями относительно (1) и удовлетворяющих условию mod p -аффинности (индуцированная рассматриваемой изометрией перестановка на множестве максимальных подшаров является аффинным отображением).

Введём метрику вида

$$q(x, y) = s(Ux, Uy), \quad U = U(\text{sup}(x, y)).$$

Таким образом,

$$q(x, y) = \max_{i=1, \dots, d} (q_i(\text{sup}(x, y)) |U(\text{sup}(x, y))(x_i - y_i)|_p).$$

Здесь q_i удовлетворяют неравенствам, приведённым в (2), U есть описанный выше набор mod p -аффинных изометрий. Для определения метрики q фактически достаточно знать об изометриях U только соответствующие перестановки на множестве максимальных подшаров в $\text{sup}(x, y)$.

Введённые выше метрики можно рассматривать как ультраметрический аналог римановой метрики в \mathbb{R}^d . Для введения римановой метрики необходимо ввести метрику (квадратичную форму) в касательном пространстве к каждой точке в \mathbb{R}^d . В ультраметрическом случае строится метрика введением метрики в наборах максимальных подшаров для каждого шара относительно стандартной метрики.

2. 2-Адическая параметризация генетического кода. Генетический код (способ кодирования белков на нуклеиновых кислотах), см. для обзора подходов к описанию генетического кода [3], имеет триплетную структуру: каждая аминокислота кодируется тройкой расположенных рядом нуклеотидов $C_1 C_2 C_3$ (кодоном).

Обсудим подход работы [4] по 2-адической параметризации генетического кода. 2-Мерная 2-адическая параметризация множества кодонов вводится нижеследующим образом. Такой подход также использовался в работах [5, 6]. Другая p -адическая параметризация генетического кода была введена в [7]. Применение p -адических динамических систем к описанию генетического кода обсуждалось в [8].

1. Множество нуклеотидов параметризуется парами 0,1 следующим образом (такая параметризация отражает химическую природу нуклеотидов):

$$\begin{array}{|c|c|} \hline A & G \\ \hline U & C \\ \hline \end{array} = \begin{array}{|c|c|} \hline 00 & 01 \\ \hline 10 & 11 \\ \hline \end{array}. \quad (\text{A})$$

2. Рассмотрим следующий порядок нуклеотидов в кодоне:

$$2 > 1 > 3. \quad (\text{B})$$

3. Построим 2-адическую плоскость — квадрат 8×8 с координатами (x, y) :

$$x = (x_0x_1x_2) = x_0 + 2x_1 + 4x_2, \quad x_i = 0, 1;$$

$$y = (y_0y_1y_2) = y_0 + 2y_1 + 4y_2, \quad y_i = 0, 1.$$

Такая 2-адическая плоскость может рассматриваться как группа $\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ (со сложением по модулю 8) с 2-мерной 2-адической метрикой

$$d(C_1C_2C_3, C'_1C'_2C'_3) = \max(|x - x'|_2, |y - y'|_2), \quad (3)$$

$$(x, y) = \rho(C_1C_2C_3), \quad (x', y') = \rho(C'_1C'_2C'_3).$$

Расстояние принимает значения 1, 1/2, 1/4.

Рассмотрим отображение ρ , переводящее кодон в точку 2-адической плоскости

$$\rho : C_1C_2C_3 \mapsto (x, y) = (x_0x_1x_2, y_0y_1y_2)$$

по следующим правилам: C_2 определяет пару (x_0, y_0) , C_1 определяет пару (x_1, y_1) , C_3 определяет пару (x_2, y_2) .

Нуклеотид определяет пару цифр по правилу (A), масштаб (2-адическая норма) пары задаётся правилом (B).

Рассмотрим 2-мерный вариант отображения Монна (p -адической замены переменной) следующего вида. Занумеруем строки и столбцы 2-адической плоскости по следующему правилу:

$$\eta : \mathbb{Z}/8\mathbb{Z} \rightarrow \{1, 2, 3, 4, 5, 6, 7, 8\}, \quad \eta : x \mapsto \tilde{x}, \quad y \mapsto \tilde{y};$$

$$\eta : x_0 + 2x_1 + 4x_2 \mapsto 1 + 4x_0 + 2x_1 + x_2;$$

$$\eta : y_0 + 2y_1 + 4y_2 \mapsto 1 + 4y_0 + 2y_1 + y_2.$$

Эквивалентно такое отображение может быть введено следующим образом:

$$\eta : 0, 4, 2, 6, 1, 5, 3, 7 \mapsto 1, 2, 3, 4, 5, 6, 7, 8.$$

Отметим, что метрика на множестве кодонов была введена с использованием отображения ρ , но без использования отображения η . Получаем 2-адическую плоскость кодонов вида

| | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| AAA | AAG | GAA | GAG | AGA | AGG | GGA | GGG |
| AAU | AAC | GAU | GAC | AGU | AGC | GGU | GGC |
| UAA | UAG | CAA | CAG | UGA | UGG | CGA | CGG |
| UAU | UAC | CAU | CAC | UGU | UGC | CGU | CGC |
| AUA | AUG | GUA | GUG | ACA | ACG | GCA | GCG |
| AUU | AUC | GUU | GUC | ACU | ACC | GCU | GCC |
| UUA | UUG | CUA | CUG | UCA | UCG | CCA | CCG |
| UUU | UUC | CUU | CUC | UCU | UCC | CCU | CCC |

Применяя к 2-адической плоскости кодонов митохондриальный генетический код, получим таблицу аминокислот на 2-адической плоскости следующего вида:

| | | | |
|---------------------------------|---------------------------------|---------------------------------|-----|
| $\frac{\text{Lys}}{\text{Asn}}$ | $\frac{\text{Glu}}{\text{Asp}}$ | $\frac{\text{Ter}}{\text{Ser}}$ | Gly |
| $\frac{\text{Ter}}{\text{Tyr}}$ | $\frac{\text{Gln}}{\text{His}}$ | $\frac{\text{Trp}}{\text{Cys}}$ | Arg |
| $\frac{\text{Met}}{\text{Ile}}$ | Val | Thr | Ala |
| $\frac{\text{Leu}}{\text{Phe}}$ | Leu | Ser | Pro |

Таким образом, вырождение генетического кода описывается локальным постоянством отображения 2-адического аргумента на малых расстояниях. В частности, в вышеприведенной таблице прообраз Lys есть половина шара диаметра $1/4$ относительно стандартной 2-мерной метрики (3), прообраз Pro есть целый такой шар. В следующем пункте обсуждается вопрос о том, как выбирать метрику, чтобы вырождение описывалось шарами (а не половинками шаров).

3. Метрическое пространство генетического кода. На пространстве кодонов (2-адической плоскости) можно рассмотреть двумерную 2-адическую деформированную метрику вида

$$d_{1,q}(x, y) = \max(|x_1 - y_1|_2, q|x_2 - y_2|_2), \quad 1/2 < q < 1.$$

Напомним, что на таблице кодонов оси координат направлены следующим образом: x_1 — сверху вниз, x_2 — слева направо.

Теперь прообраз Lys есть шар диаметром $q/4$, прообраз Pro есть шар диаметром $1/4$. Относительно метрики $d_{1,q}$ области вырождения кода являются шарами. Тем не менее хотелось бы иметь возможность более эффективно подгонять параметры метрики под свойства кода.

Для этого рассмотрим следующую метрику из семейства (2) на $\mathbb{Z}/8\mathbb{Z}^2$, отвечающую параметрам на шарах относительно стандартной метрики $d_{1,1}$:

$$(1,1), \quad \begin{array}{|c|c|} \hline (q,1) & (q,1) \\ \hline (q,1) & (q,1) \\ \hline \end{array}, \quad \begin{array}{|c|c|c|c|} \hline (1,q) & (1,q) & (1,q) & (1,1) \\ \hline (1,q) & (1,q) & (1,q) & (1,1) \\ \hline (1,q) & (1,1) & (1,1) & (1,1) \\ \hline (1,q) & (1,1) & (1,1) & (1,1) \\ \hline \end{array}, \quad 1/2 < q < 1.$$

Здесь показаны параметры метрики s на шарах относительно $d_{1,1}$ диаметра 1, $1/2$, $1/4$ соответственно. Параметры на шарах диаметра $1/4$ отвечают за вырождение (описывают области вырождения генетического кода). Параметры на шарах диаметра $1/2$ отвечают за характер вырождения (в частности, в двух подшарах диаметра $q/2$ в правом верхнем шаре диаметра $1/2$ вырождение генетического кода носит разный характер, на 2 и 4 кодона соответственно).

Работа частично поддержана грантом «Математическое моделирование» университета Карла Линнея (Векшо, Швеция). Один из авторов (С. Козырев) был частично поддержан грантами DFG Project 436 RUS 113/809/0-1 и DFG Project 436 RUS 113/951, грантами РФФИ 08-01-00727-а,

09–01–12161–офи-м и 11–01–00828–а, грантом Президента Российской Федерации на поддержку научных школ НШ-7675.2010.1 и программой Отделения математики РАН «Современные проблемы теоретической математики».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Albeverio S., Kozyrev S. V.* Multidimensional p -adic wavelets for the deformed metric // *p-Adic Numbers, Ultrametric Analysis, and Applications*, 2010. Vol. 2, no. 4. Pp. 265–277.
2. *Albeverio S., Kozyrev S. V.* Pseudodifferential p -adic vector fields and pseudodifferentiation of a composite p -adic function // *p-Adic Numbers, Ultrametric Analysis, and Applications*, 2010. Vol. 2, no. 1. Pp. 21–34.
3. *Koonin E. V., Novozhilov A. S.* Origin and Evolution of the Genetic Code: The Universal Enigma // *IUBMB Life*, 2009, February. Vol. 61(2). Pp. 99–111, arXiv: [0807.4749](https://arxiv.org/abs/0807.4749).
4. *Khrennikov A. Yu., Kozyrev S. V.* Genetic code on the dyadic plane // *Physica A*, 2007. Vol. 381. Pp. 265–272, arXiv: [q-bio/0701007](https://arxiv.org/abs/q-bio/0701007).
5. *Khrennikov A. Yu., Kozyrev S. V.* 2-Adic clustering of the PAM matrix // *Journal of Theoretical Biology*, 2009. Vol. 261. Pp. 396–406, arXiv: [0903.0137](https://arxiv.org/abs/0903.0137).
6. *Козырев С. В., Хренников А. Ю.* 2-Адические числа в генетике и симметрия Румера // *Докл. Акад. Наук*, 2010. Т. 430, № 1. С. 32–34; англ. пер. *Kozyrev S. V., Khrennikov A. Yu.* 2-Adic numbers in genetics and Rumer’s symmetry // *Dokl. Math.*, 2010. Vol. 81, no. 1. Pp. 128–130.
7. *Dragovich B., Dragovich A.* A p -Adic Model of DNA Sequence and Genetic Code // *p-Adic Numbers, Ultrametric Analysis and Applications*, 2009. Vol. 1, no. 1. Pp. 34–41, arXiv: [q-bio/0607018v1](https://arxiv.org/abs/q-bio/0607018v1).
8. *Khrennikov A. Yu.* Gene expression from polynomial dynamics in the 2-adic information space // *Chaos, Solitons, and Fractals*, 2009. Vol. 42, no. 1. Pp. 341–347.

Поступила в редакцию 20/X/2010;
в окончательном варианте — 15/II/2011.

MSC: 54E35

MULTIDIMENSIONAL p -ADIC METRIC AND GENETIC CODE

*S. V. Kozyrev*¹, *A. Yu. Khrennikov*²

¹ Steklov Mathematical Institute, Russian Academy of Sciences,
8, Gubkina st., Moscow, 119991, Russia.

² Linnæus University,
82, Kalmar, Växjö, SE-391, Sweden.

E-mails: kozyrev@mi.ras.ru, andrei.khrennikov@lnu.se

We discuss the family of metrics in multimentional p -adic spaces. For a metric under consideration the set of balls differs from the set of balls for the standard multidimensional metric. Moreover, we it is possible to consider the metrics for which the form of the sets of balls depends on the scale and position. As the example of the introduced metric spaces we study the 2-adic parametrization of the genetic code. We show that the degeneracy of the genetic code is described by the metric space from the considered family, i.e. the map of the genetic code is constant for the balls with respect to the metric from the introduced class.

Key words: *genetic code, ultrametric.*

Original article submitted 20/X/2010;
revision submitted 15/II/2011.

Sergey V. Kozyrev (Dr. Sci. (Phys. & Math.)), Leading Researcher, Dept. of Mathematical Physics. *Andrei Yu. Khrennikov* (Dr. Sci. (Phys. & Math.)), Director of the Centre, International Center for Mathematical Modeling, Dept. of Mathematics.