Sergei Kozyrev
Steklov Mathematical Institute, Russia

p-Adic numbers and complex systems

Sixth International Conference
on *p*-adic Mathematical Physics and its Applications
CINVESTAV, Mexico City, October 23–27, 2017

Hierarchical methods for complex systems
Hierarchy — trees, buildings, wavelets,
ultrametric analysis, *p*-adic numbers.

Multidimensional hierarchy in data analysis:

Clustering — trees,
Multiclustering — networks, Bruhat–Tits buildings.

Genetic code — 2-adic plane.

Other applications of hierarchy:
Spin glasses, protein dynamics,
DNA packaging (chromatin structure).

Field of p-**Adic Numbers** $\mathbb{Q}_p$ — completion of rationals with respect to p-adic norm. p-Adic numbers — series

$$x = \sum_{l=a}^{\infty} x_l p^l, \quad x_l \in \{0, 1, \ldots, p-1\}.$$

Ultrametric $d(\cdot,\cdot)$: strong triangle inequality

$$d(x,y) \leq \max(d(x,z), d(z,y)), \qquad \forall x, y, z \in X.$$

Properties of non–Archimedean geometry:
all triangles are isosceles;
balls either are disjoint or contain one another;
any point in a ball is its center.
Tree of balls in (locally compact) ultrametric space — vertices are balls (balls of non zero diameter or isolated points), edges — pairs (ball, maximal subball).

Example: balls in $\mathbb{Q}_p$

$$p^j\left(n+\mathbb{Z}_p\right), \qquad n=\sum_{l=a}^{-1} n_l p^l, \quad n_l \in \{0,1,\ldots,p-1\},$$

j is integer,
a negative integer,
$\mathbb{Z}_p$ – p-adic integers (ball $|x|_p \leq 1$),
n — parameter on $\mathbb{Q}_p/\mathbb{Z}_p$ – enumerate unit balls.
$p^j n$ – enumerate balls of non zero diameter.

Tree of balls in $\mathbb{Q}_p$: any ball contains p maximal subballs

$$\mathbb{Z}_p=\bigcup_{m=0}^{p-1}\left(m+p\mathbb{Z}_p\right).$$

Wavelet basis in $L^2(\mathbb{R})$.
Parameterized by translations and dilations (of a fixed function or a fixed finite set of functions)

$$\psi_{jn}(x) = 2^{j/2}\psi\left(2^j x - n\right), \qquad x \in \mathbb{R}, \quad j, n \in \mathbb{Z}.$$

Function $\psi(x)$ – wavelet. Example – the Haar wavelet (1909) (difference of two characteristic functions)

$$\psi(x) = \chi_{[0,1/2)}(x) - \chi_{[1/2,1]}(x).$$

General wavelet bases, multiresolution analysis,
S.Mallat, Y.Meyer, I.Daubechies, see
I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.

Theorem: 1) **Basis of p-adic wavelets** in $L^2(\mathbb{Q}_p)$:

$$\psi_{k;jn}(x) = p^{j/2}\psi_k\left(p^{-j}x - n\right),$$

$$x \in \mathbb{Q}_p, \quad j \in \mathbb{Z}, \quad n \in \mathbb{Q}_p/\mathbb{Z}_p, \quad k \in \{1, \dots, p-1\}.$$

$$\psi_k(x) = \psi(kx), \qquad \psi(x) = \chi(p^{-1}x)\Omega(|x|_p),$$

here $\Omega(x)$ is a characteristic function of $[0,1]$
(i.e. $\Omega(|x|_p)$ is a characteristic function of $\mathbb{Z}_p$),
χ is the character

$$\chi(x) = e^{2\pi i\{x\}}, \qquad \{x\} = \sum_{l=a}^{-1} x_l p^l, \quad x = \sum_{l=a}^{\infty} x_l p^l.$$

2) p-Adic wavelets — eigenvectors of the Vladimirov operator of fractional differentiation

$$D^{\alpha}\psi_{k;\,jn}=p^{\alpha(1-j)}\psi_{k;\,jn}.$$

$$D^{\alpha}f(x)=\Gamma_p^{-1}(-\alpha)\int_{\mathbb{Q}_p}\frac{f(x)-f(y)}{|x-y|_p^{1+\alpha}}d\mu(y),\quad \alpha>0,$$

$$\Gamma_p(-\alpha)=\frac{p^{\alpha}-1}{1-p^{-1-\alpha}}.$$

S.V. Kozyrev, Wavelet theory as p-adic spectral analysis, Izvestiya: Mathematics, 2002, **66** no 2, 367–376, arXiv:math-ph/0012019

Review paper:
S.V. Kozyrev, A.Yu. Khrennikov, V.M. Shelkovich, p-Adic Wavelets and Their Applications, Proceedings of the Steklov Institute of Mathematics, 2014, **285**, 157–196.

Ultrametric wavelets.

Locally compact complete ultrametric space X with Borel measure μ, space $L^2(X,\mu)$.

Ultrametric wavelets related to a ball I in X (of non-zero diameter): W_I — space of locally constant μ–mean zero functions supported in ball I, constant on maximal subballs in I. This space is finite dimensional.

Let us choose any orthonormal basis $\{\psi_{Ij}\}$ in W_I and take union of such bases over balls I. This is a basis in $L^2(X,\mu)$ called the basis of ultrametric wavelets.

Spectra of general ultrametric PDO. Pseudodifferential operator on (locally compact) ultrametric space X

$$Tf(x) = \int_X t(\sup(x,y))(f(x) - f(y))d\mu(y).$$

$\sup(x,y)$ is the minimal ball in X containing points x, y.

Eigenvectors of T – utrametric wavelets $\{\psi_{Ij}\}$ with eigenvalues

$$T\psi_{Ij} = \lambda_I \psi_{Ij},$$

$$\lambda_I = t(I)\mu(I) + \sum_{J>I} t(J)(\mu(J) - \mu(J,I)),$$

(J,I) – maximal subball in J containing I. The operator is well defined if the series above converges absolutely.

Relation between real and p-adic wavelets – **Monna map**

$$\mathbb{Q}_p \to \mathbb{R}_+,$$

$$\mathbb{Q}_p/\mathbb{Z}_p \to \mathbb{Z}_+,$$

$$\sum_{l=a}^{\infty} x_l p^l \mapsto \sum_{l=a}^{\infty} x_l p^{-l-1}.$$

Small p-adic distances map to small real distances. One to one almost everywhere and measure preserving.

$p = 2$: Haar basis on $\mathbb{R}_+$ maps to 2-adic wavelet basis (hierarchy in real wavelet basis).

Monna map – p-adic parametrization of the Parisi matrix in theory of spin glasses.

Wavelet frames as orbits

Action of the affine group

$$f(x) \mapsto |a|_p^{-\frac{1}{2}} f\left(\frac{x-b}{a}\right).$$

Orbit of wavelet $\psi(x) = \chi(p^{-1}x)\Omega(|x|_p)$ is the set of products of wavelets $\psi_{k;jn}$ and p-roots of one.

Orbit gives the parametrization of the wavelet basis. Generalization — frames of wavelets as orbits of mean zero test functions (locally constant compactly supported).

Frame $\{e_n\}$ in the Hilbert space $\mathcal{H}$ is a set of vectors in $\mathcal{H}$:
$\exists A, B > 0$: $\forall g \in \mathcal{H}$

$$A\|g\|^2 \leq \sum_n |\langle g, e_n\rangle|^2 \leq B\|g\|^2.$$

Metrics in multidimensional p-adic spaces

Norm in $\mathbb{Q}_p^d$

$$N_{q_1,\dots,q_d}(z) = \max_{i=1,\dots,d}(q_i|z_i|_p), \qquad q_i > 0.$$

More general norm: A-rotation of $N_{q_1,\dots,q_d}$, A is a matrix in $\mathrm{Gl}_d(\mathbb{Q}_p))$

$$N^A_{q_1,\dots,q_d}(z) = N_{q_1,\dots,q_d}(Az).$$

Metric is defined by the norm

$$s(x,y) = N(x-y).$$

Example: Norm $N_{q_1,\dots,q_d}$, $p^{-1} < q_i \leq 1$,
$\mathbb{Z}_p^d$ and $p\mathbb{Z}_p^d$ are balls,
if not all q_i are equal there exist intermediary balls between $\mathbb{Z}_p^d$ and $p\mathbb{Z}_p^d$.

Multidimensional wavelet bases as orbits. Scheme:
1) Metric in $\mathbb{Q}_p^d$, tree $\mathcal{T}$ of balls with respect to this metric.
2) Group of automorphisms of the tree $\mathcal{T}$ of balls (linear transformations and translations).
3) Wavelet bases and frames as orbits of this group.

Example 1: norm

$$\|z\|_p = \max_{i=1,\dots,d}(|z_i|_p).$$

Group of automorphisms of the tree of balls — translations; homogeneous dilations; norm preserving linear transformations (matrices with elements in $\mathbb{Z}_p$ and $|\det(\cdot)|_p = 1$.)

Wavelets — products of characters of $\mathbb{Q}_p^d$ and characteristic functions of balls, character makes a single oscillation on a ball.

Theorem Set of functions $\{\psi_{k;jn}\}$ is an orthonormal basis in $L^2(\mathbb{Q}_p^d)$:

$$\psi_k(x) = \chi(p^{-1}k \cdot x)\Omega(\|x\|_p), \qquad x \in \mathbb{Q}_p^d, \quad k \cdot x = \sum_{l=1}^{d} k_l x_l,$$

$$k = (k_1, \dots, k_d), \qquad k_l = 0, \dots, p-1,$$

at least one of k_l is non zero;

$$\psi_{k;jn}(x) = p^{-\frac{dj}{2}}\psi_k(p^j x - n), \qquad x \in \mathbb{Q}_p^d, \quad j \in \mathbb{Z}, \quad n \in \mathbb{Q}_p^d/\mathbb{Z}_p^d,$$

$$n = \left(n^{(1)}, \dots, n^{(d)}\right), \quad n^{(l)} = \sum_{i=\beta_l}^{-1} n_i^{(l)} p^i,$$

$$n_i^{(l)} = 0, \dots, p-1, \quad \beta_l \in \mathbb{Z}_-.$$

Example 2: norm

$$\|z\| = \max_{i=1,\dots,d}(q_i|z_i|_p).$$

$$p^{-1} < q_1 < q_2 < \dots < q_d \le 1.$$

Sequence of embedded balls in between of $p\mathbb{Z}_p^d$ and $\mathbb{Z}_p^d$

$$\mathbb{Z}_p \times \dots \mathbb{Z}_p \times p\mathbb{Z}_p \dots p\mathbb{Z}_p$$

Matrix dilation — maps $\|\cdot\|$-ball centered in zero to a maximal subball

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ p & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Wavelets with matrix dilation

$$\Psi_k(x) = \chi\left(k \cdot A^{-1}x\right)\Omega(\|x\|),$$

$$k = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The orthonormal **wavelet basis with matrix dilations** is constructed by translations and matrix dilations of wavelets Ψ_k:

$$\Psi_{k;jn}(x) = p^{-\frac{j}{2}}\Psi_k(A^j x - n), \qquad j \in \mathbb{Z}, \quad n \in \mathbb{Q}_p^d/\mathbb{Z}_p^d.$$

Remark Defined above dilation maps a ball centered in zero to a maximal subball centered in zero and this property should be satisfied for all pairs (ball, maximal subball). Such maps exist not for all norms from the described above family (only for sufficiently symmetric). For less symmetric norms one could obtain several matrix dilation maps, each of these dilations can be applied only to some subset of pairs (ball, maximal subball).

Conjugated space to $\mathbb{Q}_p^d$ — space of linear functionals acting by scalar product $k \cdot x$. Conjugated norm on conjugated space.
The unit ball with respect to the norm $\| \cdot \|$ in $\mathbb{Q}_p^d$;

$$B_1 = \{x \in \mathbb{Q}_p^d : \|x\| \leq 1\},$$

the sequence of intermediary balls

$$B_1 \supset B_2 \supset \cdots \supset B_d \supset B_{d+1} = pB_1$$

with the diameters

$$1 \geq q_1 > q_2 > \ldots q_d > p^{-1}.$$

Then we have the sequence of conjugated balls

$$B_1^* \subset B_2^* \subset \dots \subset B_d^* \subset B_{d+1}^* = p^{-1} B_1^*,$$

$$B_l^* = \{k \in \mathbb{Q}_p^d : |k \cdot x|_p \leq 1, \quad \forall x \in B_l, \quad k \cdot x = \sum_{i=1}^{d} k_i x_i\}.$$

We take the diameters of B_l^* equal to

$$1 \leq q_1^{-1} < q_2^{-1} < \dots q_d^{-1} < p.$$

This defines a norm in conjugated space (norms of other balls are defined by linearity).

Lemma If A is a $\|\cdot\|$-dilation then transpose A^* is a dilation with respect to the conjugated norm.

Spectra of PDO over $\mathbb{Q}_p^d$.

Generalization of the Vladimirov fractional derivation operator

$$D^{\alpha} f(x) = F^{-1}\left(\|k\|^{\alpha} F[f]\right)(x),$$

where F is the Fourier transform, $\|\cdot\|$ is the conjugated norm in the conjugated space (with parameter k).

Theorem Wavelets with matrix dilations are eigenvectors of the above operator

$$D^{\alpha}\Psi_{k;jn}(x) = \|A^{*(j-1)}k\|^{\alpha}\Psi_{k;jn}(x).$$

Quincunx basis ($p = 2$) in $L^2(\mathbb{Q}_2^2)$.
Norm

$$q(x) = s(Ux), \quad U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

$$s(x) = \max(q|x_1|_p, |x_2|_p), \quad p^{-1} < q < 1.$$

The **quincunx matrix** is a q-dilation

$$Q = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Theorem The set of functions $\psi_{jn}(x) = p^{-\frac{j}{2}}\psi(Q^j x - n)$, $j \in \mathbb{Z}$, $n \in \mathbb{Q}_2^2/\mathbb{Z}_2^2$ is an orthonormal basis in $L^2(\mathbb{Q}_2^2)$ (the 2-adic quincunx basis).

2-dimensional 2-adic metric in applications:
genetic code on the 2-adic plane.

Codons (triples of nucleotides) encode amino acids.
64 codons, 20 amino acids – the code is degenerate.
Nucleotides are enumerated by pairs of digits $(0, 1)$.
Set of codons (64 codons) can be naturally enumerated by 2-adic plane 8×8 with metric

$$d_{1,q}((x, y), (x', y')) = \max(|x - x'|_2, q|y - y'|_2), \quad 1/2 < q < 1,$$

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

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

Application to this table of codons of mitochondrial genetic code gives parametrization of the code:

close in metric $d_{1,q}$ codons map to the same amino acid,
less close codons map to similar amino acids (hydrophobic amino acids are clustered with respect to this metric):

$\frac{\text{Lys}}{\text{Asn}}$	$\frac{\text{Glu}}{\text{Asp}}$	$\frac{\text{Ter}}{\text{Ser}}$	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

Clustering – hierarchical classification of data.

Data are marked by hierarchic system (partially ordered tree) of clusters. Typically clusters are constructed using metric on data.

Multiclustering — several system of clusters on the same data. Several metrics, different metrics generate different trees of clusters. Clusters networks — cycles are created by gluing together different cluster trees. Clusters for different metrics which coincide as sets are identified.

A.Strehl, J.Ghosh, C.Cardie, Cluster ensembles — a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 2002. 3. P.583–617.

p-Adic case: cluster networks are related to Bruhat–Tits buildings.

Example: Nearest neighbor clustering
(M, ρ) – metric space.

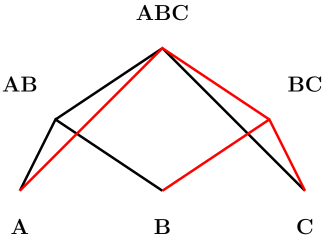
$a = x_0, x_1, \ldots, x_{n-1}, x_n = b$ — ε-chain in (M, ρ).

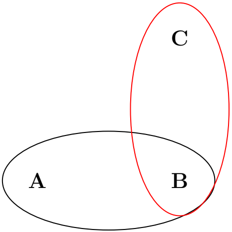
Chain distance between a and b:
$d(a, b) = \inf(\varepsilon$: there exists ε-chain between a, $b)$.

All properties of ultrametric except for non–degeneracy (non–coinciding points can have zero chain distance).
Chain distance defines ultrametric on the set of of equivalence classes of points in M, where a, b are equivalent if $d(a, b) = 0$.

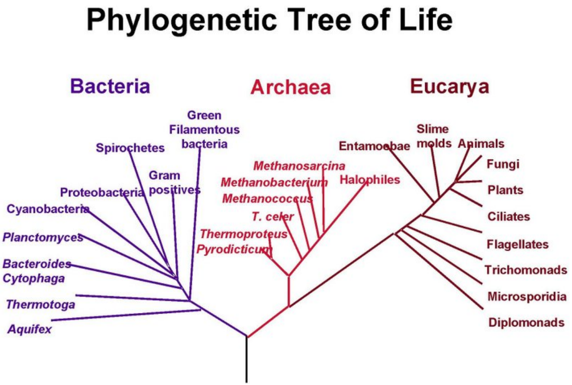
Cluster — ball with respect to chain distance.
Clustering — covering of (M, ρ) by clusters.

Multiclustering — several metrics on M, several cluster trees.
Network of clusters — identify clusters which coincide as sets and glue together different trees of clusters.


ABC
AB
BC
A
B
C

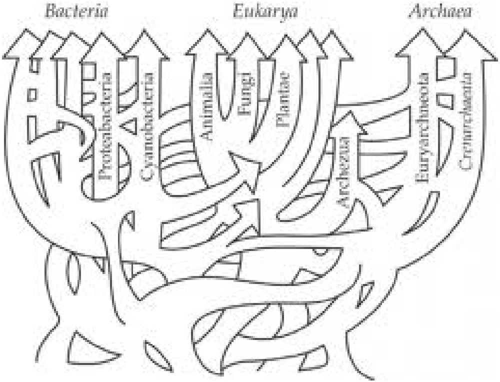

C
A
B

Clustering in life sciences: tree of life
Carl von Linne, Systema Naturae, 1735



Ribosomal tree of life
Carl Woese, 1977, 1985

Plylogenetic network



E.V.Koonin, The Logic of Chance, The Nature and Origin of Biological Evolution, Pearson Education, FT Press Science, 2012

Cycles — relation to multiclustering?

Multiclustering in p-adic spaces $\mathbb{Q}_p^d$: relation to Bruhat–Tits buildings (simplicial complexes of equivalence classes of lattices).

Affine Bruhat–Tits building. Vertices are equivalence classes of lattices. A lattice in $\mathbb{Q}_p^d$ is an open compact $\mathbb{Z}_p$-module in $\mathbb{Q}_p^d$. Any lattice can be put in the form

$$\oplus_{i=1}^{d}\mathbb{Z}_p e_i,$$

where $\{e_i\}$ is a basis in $\mathbb{Q}_p^d$.
Two lattices are equivalent if one is a scalar multiple of the other.
Two lattices L_1 and L_2 are adjacent (connected by an edge) if some representatives from equivalence classes L_1 and L_2 satisfy

$$pL_1 \subset L_2 \subset L_1.$$

$k-1$-Simplices are defined as equivalence classes of k adjacent lattices, i.e. the chains

$$pL_k \subset L_1 \subset L_2 \subset \cdots \subset L_k.$$

Here $1 \leq k \leq d$. Dimension – length of the above sequence.

Relation to multiclustering: Balls centered in zero in $\mathbb{Q}_p^d$ (with respect to some of discussed earlier metrics) — lattices (open compact $\mathbb{Z}_p$-modules).
Simplicial complex of balls with respect to multiclustering with the family of metrics — balls (not necessarily centered in zero).
Restriction to centered in zero balls and factorization by dilations by degrees of p gives the affine building.

General family of metrics on data: one can define a structure of simplicial complex and corresponding notion of dimension (using cycles in cluster networks).
– Analogs of buildings in general data analysis.

Summary

p-Adic wavelets — tree of balls, group of automorphisms of this tree, wavelet basis — system of coherent states for this group. Wavelets are eigenvectors of pseudodifferential operators.

Multidimensional hierarchy.

Example: — 2-adic plane of genetic code.

Data analysis — clustering, multiclustering, cluster networks, relation to affine buildings, dimension for general cluster systems.