

S. V. Kozyrev

Steklov Mathematical Institute

p -Adic numbers and complex systems

September 7–12, 2015, Belgrade

International Conference p -ADICS.2015

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

Wavelets — hierarchical function representations.

Clustering — trees,
Multiclustering — buildings.

Spin glasses — Parisi matrix, p -adic parameters.

Protein structure, DNA packing.

Protein dynamics — p -adic diffusion.

Genetic code — 2-adic plane.

Deep learning — hierarchical models in machine learning.

Wavelets

Basis of wavelets in $L^2(\mathbb{R})$.

Standard parameterization — by translations and dilations.

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

The function $\psi(x)$ is called a wavelet. The first example is the Haar wavelet (difference of two characteristic functions)

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

The pair (j, n) of indices of wavelets actually is a parameter on a tree. To see this it is easier to consider wavelets on a half-line $x \geq 0$, when $n \geq 0$.

Tree of balls

Balls in ultrametric space can be considered as vertices of a tree (the tree of balls). Balls are vertices, edges connect balls embedded without intermediary balls.

p -Adic numbers \mathbb{Q}_p — balls are in one to one correspondence with

$$\bigcup_{j \in \mathbb{Z}} \mathbb{Q}_p / p^j \mathbb{Z}_p.$$

Any ball has a form

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

a is a negative integer, \mathbb{Z}_p is the ring of p -adic integers (unit ball). Here n can be considered as a parameter in $\mathbb{Q}_p / \mathbb{Z}_p$.

The Monna map

p -Adic parametrization of positive integers (one to one map)

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

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

Small p -adic distances map to small real distances.

Applying this construction (for $p = 2$) to the set of indices (j, n) of wavelet coefficients on positive half-line we get:

wavelet coefficients are vertices in 2-adic tree of balls.

Already real wavelets are hierarchical.

p -Adic wavelets

Basis of wavelets in $L^2(\mathbb{Q}_p)$.

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

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

Example:

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

where $\Omega(x)$ is a characteristic function of $[0, 1]$ (thus $\Omega(|x|_p)$ is a characteristic function of the unit ball \mathbb{Z}_p), and χ is the character

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

S.V. Kozyrev, Wavelet theory as p -adic spectral analysis, Izvestiya: Mathematics, 2002, 66 no 2, 367–376.

Clustering

Clustering is method of hierarchical classification of data.

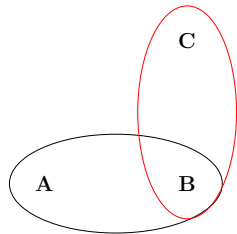
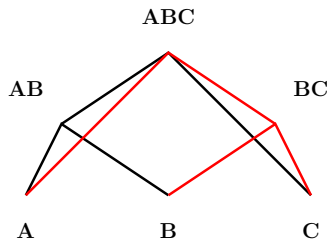
Data is labeled by a hierarchical system (tree, or dendrogram) of clusters. Typical approach (k -means clustering, nearest neighbor clustering) — clusters are generated using some metric in the data.

Can be used for unsupervised learning — extracting of information from unlabeled data.

Multiclustering — several systems of clusters on the same data. In particular, when we have a family of metrics on the data, different metrics generate different cluster trees.

In a typical situation this generates a network of clusters with cycles — cycles are generated when clusters with respect to different metrics coincide as sets.

p -Adic case: cluster networks are related to affine Bruhat-tits buildings.



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.

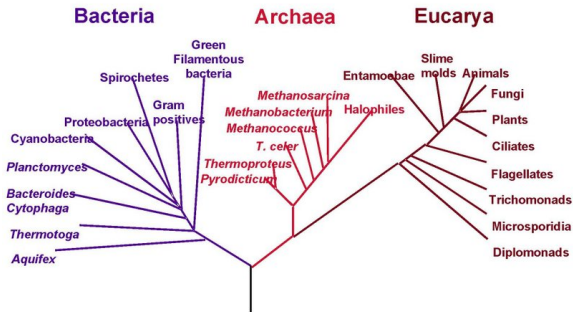
S.Albeverio, S.V.Kozyrev, Clustering by hypergraphs and dimensionality of cluster systems, *p-Adic Numbers, Ultrametric Analysis and Applications*. 2012. V.4. No.3. P.167–178.
arXiv:1204.5952

S.V.Kozyrev, Cluster networks and Bruhat–Tits buildings, *Theoretical and Mathematical Physics*. 2014. V.180. No.2. P.959–967. *arXiv:1404.6960*

Clustering in life sciences: the tree of life

Carl von Linne, Systema Naturae, 1735

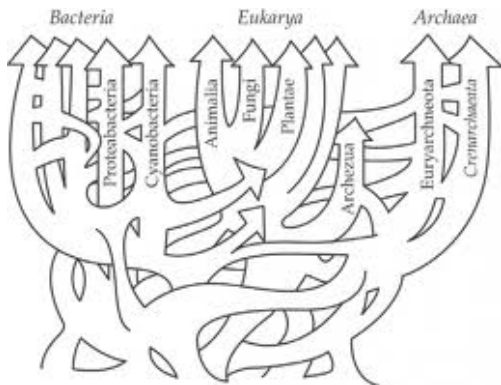
Phylogenetic Tree of Life



Ribosomal phylogenetic tree

Carl Woese, 1977, 1985

Phylogenetic network



Contains cycles — relation to multiclustering.

Spin glasses

Replica symmetry breaking — hierarchical Parisi matrix (Q_{ab}),
built by iterative procedure

$$\begin{pmatrix} 0 & q_1 \\ q_1 & 0 \end{pmatrix}, \quad \left(\begin{array}{cc|cc} 0 & q_1 & q_2 & q_2 \\ q_1 & 0 & q_2 & q_2 \\ \hline q_2 & q_2 & 0 & q_1 \\ q_2 & q_2 & q_1 & 0 \end{array} \right),$$

$$\left(\begin{array}{cccc|cccc} 0 & q_1 & q_2 & q_2 & q_3 & q_3 & q_3 & q_3 \\ q_1 & 0 & q_2 & q_2 & q_3 & q_3 & q_3 & q_3 \\ q_2 & q_2 & 0 & q_1 & q_3 & q_3 & q_3 & q_3 \\ q_2 & q_2 & q_1 & 0 & q_3 & q_3 & q_3 & q_3 \\ \hline q_3 & q_3 & q_3 & q_3 & 0 & q_1 & q_2 & q_2 \\ q_3 & q_3 & q_3 & q_3 & q_1 & 0 & q_2 & q_2 \\ q_3 & q_3 & q_3 & q_3 & q_2 & q_2 & 0 & q_1 \\ q_3 & q_3 & q_3 & q_3 & q_2 & q_2 & q_1 & 0 \end{array} \right),$$

$q_i > 0$ are real (and positive) parameters.

Monna map — reshuffling of rows and columns of the Parisi matrix $p^N \times p^N$ (above $p = 2$ and $N = 1, 2, 3$)

$$I : \{1, \dots, p^N\} \rightarrow p^{-N}\mathbb{Z}/\mathbb{Z},$$

$$I^{-1} : \sum_{j=-N}^{-1} x_j p^j \mapsto 1 + p^{-1} \sum_{j=-N}^{-1} x_j p^{-j}.$$

Matrix elements Q_{ab} of the Parisi matrix is a function of p -adic distance between $I(a)$ and $I(b)$:

$$Q_{ab} = q(|I(a) - I(b)|_p),$$

where $q(p^k) = q_k$, $q(0) = 0$, $k = 1, \dots, N$.

$p^{-N}\mathbb{Z}/\mathbb{Z}$ is a group of fractions $\{\sum_{j=-N}^{-1} x_j p^j\}$, $x_j = 0, \dots, p-1$ with the addition modulo 1.

V.A.Avetisov, A.H.Bikulov, S.V.Kozyrev, Application of p-adic analysis to models of spontaneous breaking of replica symmetry, *J. Phys. A: Math. Gen.* 1999. V.32. N.50. P.8785–8791, [arXiv:cond-mat/9904360](https://arxiv.org/abs/cond-mat/9904360)

G. Parisi, N. Sourlas, p-Adic numbers and replica symmetry breaking, *Eur. Phys. J. B*, 2000, V.14, P.535–542.

A.Yu.Khrennikov, S.V.Kozyrev, Replica symmetry breaking related to a general ultrametric space I: replica matrices and functionals. *Physica A: Statistical Mechanics and its Applications*. 2006. V.359. P.222-240. [arXiv:cond-mat/0603685](https://arxiv.org/abs/cond-mat/0603685)

A.Yu.Khrennikov, S.V.Kozyrev, Replica symmetry breaking related to a general ultrametric space II: RSB solutions and the $n \rightarrow 0$ limit. *Physica A: Statistical Mechanics and its Applications*. 2006. V.359. P.241-266. [arXiv:cond-mat/0603687](https://arxiv.org/abs/cond-mat/0603687)

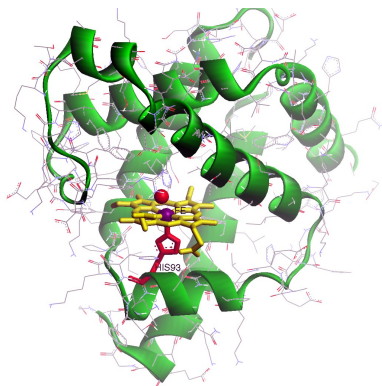
A.Yu.Khrennikov, S.V.Kozyrev, Replica symmetry breaking related to a general ultrametric space III: The case of general measure // Physica A: Statistical Mechanics and its Applications. 2007. V.378. N.2. P.283-298. arXiv:cond-mat/0603694

D.M. Carlucci, C. De Dominicis, On The Replica Fourier Transform, Comptes Rendus Ac.Sci. Ser.IIB Mech.Phys. Chem.Astr., 325 (1997) P.527, arXiv:cond-mat/9709200

C. De Dominicis, D.M. Carlucci, T. Temesvari, Replica Fourier Transforms on Ultrametric Trees, and Block-Diagonalizing Multi-Replica Matrices, Journal de Physique I (France) 7 (1997) P.105-115, arXiv:cond-mat/9703132

Proteins

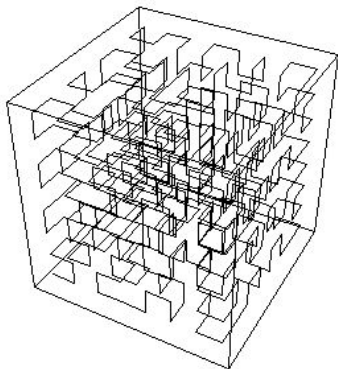
Protein is a peptide chain (chain of amino acids) folded in a compact globule (native state)



Myoglobin

Hierarchical structure of protein globules.

Protein globules are analogous to Peano curves — space-filling curves with hierarchical structure.



DNA packing is also dense and hierarchical

A.Yu. Grosberg, S.K. Nechaev, E.I. Shakhnovich, The role of topological constraints in the kinetics of collapse of macromolecules. J Phys. France 1988; 49:2095–2100.

Dynamics on energy landscapes and protein dynamics

Diffusion in a potential

$$\frac{\partial}{\partial t} f(x, t) = \Delta f(x, t) + \beta \nabla f(x, t) \cdot \nabla U(x) + \beta f(x, t) \Delta U(x)$$

$f(x, t)$ – distribution function

U – potential, $\beta = 1/kT$ – inverse temperature.

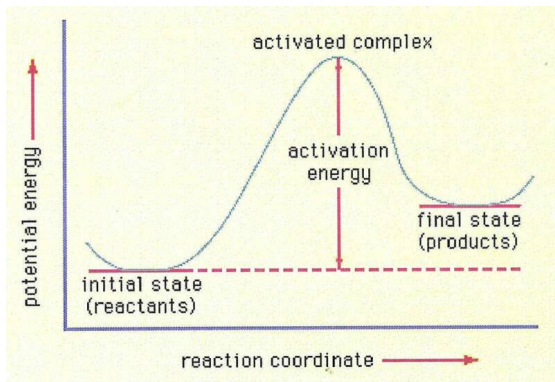
Approximation by the Arrhenius transitions between the local energy minima

$$\frac{d}{dt} f(a, t) = \sum_b (Q_{ab} f(b, t) - Q_{ba} f(a, t)).$$

the Arrhenius formula — the transition rate is proportional to

$$\exp(-\beta\Delta E), \quad \Delta E = E_1 - E_0$$

ΔE — activation barrier.



Complex energy landscapes — many local minima.

Three local minima – two transition states.

Hierarchy of transition states.

Example: local minima A , B , C ,

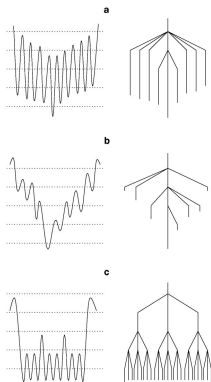
transition state with energy E_1 between A and B ,

transition state with energy E_2 between (A, B) and C , $E_1 < E_2$

Hierarchical matrix of transition energies

$$\left(\begin{array}{cc|c} 0 & E_1 & E_2 \\ E_1 & 0 & E_2 \\ \hline E_2 & E_2 & 0 \end{array} \right).$$

General case — *disconnectivity graph* of local minima and transition states. Hierarchy of basins (branches of the tree) — interbasin kinetics.



O. M. Becker, M. Karplus, The Topology of Multidimensional Protein Energy Surfaces: Theory and Application to Peptide Structure and Kinetics, J. Chem.Phys., 1997, V.106, P.1495–1517.

Example: p -Adic diffusion equation

$$\frac{\partial}{\partial t} f(x, t) + D_x^\alpha f(x, t) = 0$$

with the Vladimirov fractional operator

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

α is proportional to inverse temperature $\alpha = \beta k$.

— describes the relaxation of a protein.

x — conformation parameter.

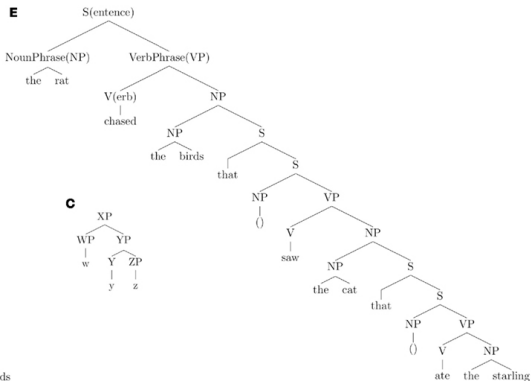
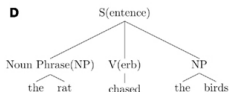
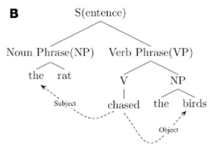
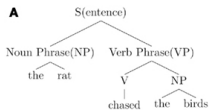
V.A.Avetisov, A.H.Bikulov, S.V.Kozyrev, V.A.Osipov, p -Adic Models of Ultrametric Diffusion Constrained by Hierarchical Energy Landscapes. J. Phys. A: Math. Gen. 2002. V.35. N.2. P.177–189, arXiv:cond-mat/0106506

Genetic code — 2-dimensional 2-adic parametrization describing the degeneracy of the code

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

Humanitarian sciences: Syntax, music, etc.

Hierarchical syntax markup — we speak in a hierarchical way



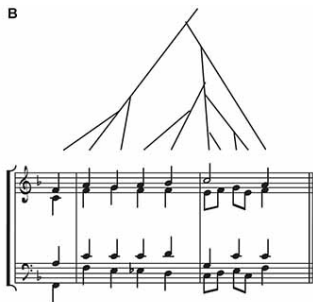
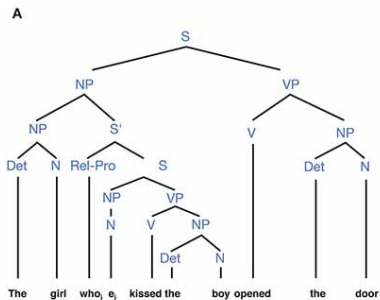
Music — sounds (notes) and musical phrases have lengths equal to degrees of two;

musical phrases have hierarchical 2-adic structure (phrase is a combination of smaller phrases);

repetition of phrases (with small modification) — 2-adic local constancy or 2-adic Lipschitz condition.

Analogy with syntax markup

A.D. Patel, *Language, music, syntax and the brain*. *Nature Neuroscience*. (2003). 6, 674–681.



Christus, der ist mein Leben 1st phrase (J.S. Bach)

F. Lerdahl, *Tonal Pitch Space* (2001, OUP)

Ivelisse Robles

Possible applications to learning

In machine learning, generative models are used for modeling of data. To reduce computations it is important to use simple generative models.

Kolmogorov complexity —
the length of the shortest description of the object.

In particular, for natural number N
the Kolmogorov complexity is estimated from above by $10 \lg N$
(the possible number of digits in the expansion over degrees of 10).

This positional notation for natural numbers is hierarchical.
Analogous approach can be applied for function approximations
using wavelets (i.e. compression), etc.

Hierarchy is a general method for controlling complexity.

Deep learning — hierarchical multilayer neural networks. Recent achievements in machine learning are related to hierarchical function representations.

Summary

Wavelets — hierarchical family of functions
(tree of balls, the Monna map)

Clustering and multiclustering — trees and buildings

Spin glasses — the Parisi matrix, p -adic parameters

Protein structure, DNA packing

Protein dynamics — p -adic diffusion

Genetic code — 2-adic plane

Hierarchy, Kolmogorov complexity, deep learning.