NATIONAL RESEARCH UNIVERSITY

ЛАБОРАТОРИЯ «МНОГОМЕРНАЯ АППРОКСИМАЦИЯ И ПРИЛОЖЕНИЯ»

School-conference
"Approximation and Data Analysis 2019"

Nizhny Novgorod, Russia
September 30 – October 4, 2019

BOOK OF ABSTRACTS

Deep Learning Appliance for Sensitive Data

Alexey Aleshin

EPAM Systems
https://www.epam.com/

GDPR rules apply a lot of restrictions on sharing data between companies, this problem creates a lot of work to identify personally identifiyng information (PII) and its anonymization. During the presentation we will show on examples of the problems and restrictions of natural language algorithm for PII detection and variational autoencoder for data anonymization.

Representing Systems of Dilations and Translations of a Function in Symmetric Spaces

Astashkin S.V.

Samara National Research University
astash56@mail.ru

Let $1 \leqslant p < \infty$, $f \in L_p = L_p[0,1]$. According to the result by Filippov and Oswald proved in [1] the obvious necessary condition $\int_0^1 f(t)\,dt \neq 0$ assures that the sequence $\{f_{k,i}\}$ of dyadic dilations and translations of a function $f \in L_p$ defined by

$$f_{k,i}(t) = \begin{cases} f(2^k t - i), & t \in [\frac{i}{2^k}, \frac{i+1}{2^k}], \\ 0, & elsewhere, \end{cases}$$

$$i = 0, \ldots, 2^k - 1, \qquad k = 0, 1, \ldots,$$

is a representing system in the space L_p. This means that for every function $x \in L_p$ there is a sequence of coefficients $\{\xi_{k,i}\}$ such that $x = \sum_{k=0}^{\infty} \sum_{i=0}^{2^k-1} \xi_{k,i} f_{k,i}$ with convergence in L_p.

Let X be an arbitrary separable symmetric space on $[0,1]$. Basing on a combination of the frame approach with the notion and properties of the multiplicator space $\mathcal{M}(X)$ of X with respect to the tensor product, we will discuss the problem when the sequence of dyadic dilations and translations of a function $f \in X$ is a representing system in the space X. The main result (see [2]) reads that this holds whenever $\int_0^1 f(t)\,dt \neq 0$ and $f \in \mathcal{M}(X)$. Moreover, the condition $f \in \mathcal{M}(X)$ turns out to be sharp in a certain sense. In particular, a decreasing nonnegative function f, $f \neq 0$, from a Lorentz space Λ_φ generates an absolutely representing system of dyadic dilations and translations in Λ_φ if and only if $f \in \mathcal{M}(\Lambda_\varphi)$.

References

[1] *Filippov V.I., Oswald P.* Representation in L_p by series of translates and dilates of one function // J. Approx. Theory. 1995. V. 82. P. 15–29.

[2] *Astashkin S.V., Terekhin P.A.* Representation properties of systems of dilations and translations in symmetric spaces (submitted).

Approximation and Cover Properties of Learning Networks

Andrew Barron

Department of Statistics and Data Science, Yale University, USA
andrew.barron@yale.edu

For functions of bounded variation V with respect to a class of activation functions, we recall how an m term approximation achieves error bounded by $V/m^{1/2}$. Implications are given for the metric entropy and the statistical risk of single hidden-layer neural networks, and other ridge function expansions. This sets the stage for discussion of related complexity properties for deep networks.

Complexity and Risk Properties of Deep Nets

Andrew Barron

Department of Statistics and Data Science, Yale University, USA
andrew.barron@yale.edu

Suitable extension of the notion of variation is developed for deep networks, based on the sum of the absolute values of the products of weights along all paths through the network. The Rademacher and Gaussian complexities of these networks are discussed and shown to be bounded by expressions of the order $V\left(nlogd\right)^{1/2}$ where V is the variation, n is the sample size, and d is the input dimension. Also discussed is whether these complexities can be arranged to be independent of the number of layers of the network. The associated statistical risk (mean squared generalization error) of the least squares estimator among the deep nets is shown to be of order $V\left(\frac{log(d)}{n}\right)^{1/2}$. These results show that favorable risk properties hold for extremely large numbers of input variables as long as the sample size n is large compared to the log of the number of variables.

The Role of Information Theory in Deep Net Analysis

Andrew Barron

Department of Statistics and Data Science, Yale University, USA
andrew.barron@yale.edu

Information theory arises in learning network methodology and analysis in three ways that we discuss. First in the formulation of description-length methods (such as complexity penalized least squares) for estimation of the network and its associated resolvability bound on statistical risk. Second in the determination of optimal rates of function estimation for these classes of functions. Third in the demonstration of the relationship between Gaussian complexity and metric entropy.

Greedy Algorithm with Respect to Assymmetric Dictionary may Diverge

Borodin P.A.

Moscow State University
pborodin@inbox.ru

Let H be a Hilbert space. A subset D of the unit sphere $S(H) = \{s \in H : |s| = 1\}$ is called a *dictionary* if $D = H$. For any dictionary $D \subset S(H)$ and any $x_0 \in H$, the *pure greedy algorithm* (PGA) generates a sequence x_n defined inductively by

$$x_{n+1} = x_n - \langle x_n, g_{n+1}\rangle g_{n+1}, \qquad n = 0, 1, 2, \dots, \tag{1}$$

where the element $g_{n+1} \in D$ is such that

$$\langle x_n, g_{n+1}\rangle = \max\{\langle x_n, g\rangle : g \in D\}.$$

The existence of the above maximum is an additional condition on D.

It is well known that PGA converges for any symmetric dictionary [1]. In general case the condition necessary for convergence is the positive totality of D: linear combinations of elements of D with positive coefficients should be dense in H. Till recently it has been unknown whether this condition is sufficient for convergence. In [2], a special recursive greedy algorithm was invented, which converges for any positively total dictionary and any starting element x_0, though it does not provide an expansion of x_0 into a series of elements of D with positive coefficients.

We present an example of positively total dictionary and a starting element, for which PGA diverges.

References

[1] *V. Temlyakov.* Greedy Approximation, Cambridge, 2011.

[2] *E.D. Livshits.* On n-Term Approximation with Positive Coefficients // Math. Notes, 82:3 (2007), 332–340.

Latent Convolutional Models for Image Restoration

Evgeny Burnaev

Center for Computational and Data-Intensive Science and Engineering, Skoltech Moscow
E.Burnaev@skoltech.ru

Abstract: We present a new latent model of natural images that can be learned on large-scale datasets. The learning process provides a latent embedding for every image in the training dataset, as well as a deep convolutional network that maps the latent space to the image space. After training, the new model provides a strong and universal image prior for a variety of image restoration tasks such as large-hole inpainting, superresolution, and colorization. To model high-resolution natural images, our approach uses latent spaces of very high dimensionality (one to two orders of magnitude higher than previous latent image models). To tackle this high dimensionality, we use latent spaces with a special manifold structure (convolutional manifolds) parameterized by a ConvNet of a certain architecture. In the experiments, we compare the learned latent models with latent models learned by autoencoders, advanced variants of generative adversarial networks, and a strong baseline system using simpler parameterization of the latent space. Our model outperforms the competing approaches over a range of restoration tasks. We illustrate how these models can be used for perceptual depth superresolution.

About Multipoint Schur Criteria

Viktor Buslaev

Steklov Mathematical Institute, Moscow
buslaev@mi-ras.ru

The classical Schur criterion gives an answer to the question whether the function f, given by its power series in ponit 0, is a Schur function, i.e. a function holomorphic in the unit circle $D = \{|z| < 1\}$ and taking values in it modulo not exceeding 1. The answer is given in terms of determinants introduced by Schur, constructed in a special way by coefficients f_k, $k = 0, 1, \ldots$ of power series of f, and the proof relies on a Schur algorithm allowing to represent the Schur function as a continuous fraction of a special kind called a Schur continuous fraction. The report will show that Schur determinants numerically coincide (up to some Prime factor) with Hankel determinants constructed by coefficients f_k, $k = 0, 1, \ldots$ of power series of f and the coefficients of the associated series with the center of the expansion at infinity.

Sparse-grid Polynomial Interpolation Approximation and Integration for Parametric and Stochastic Elliptic PDEs with Lognormal Inputs

Dinh Dũng

Vietnam National University, Hanoi
dinhzung@gmail.com

One of basic problems in Uncertainty Quantification are approximation and numerical integration for parametric and stochastic PDEs. Since the number of parametric variables may be very large or even infinite, they are treated as high-dimensional or infinite-dimensional approximation problems.

Let $D \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Consider the parametric diffusion elliptic equation

$$-\operatorname{div}(a(\mathbf{y})\nabla u(\mathbf{y})) = f \quad \text{in} \quad D, \quad u(\mathbf{y})|_{\partial D} = 0, \tag{1}$$

for a given fixed right-hand side $f \in V'$ and spatially variable scalar diffusion coefficient $a(\mathbf{y})$ parametrized by $\mathbf{y} \in \mathbb{R}^\infty$, where $V := H_0^1(D)$ is the energy space and $V' = H^{-1}(D)$ the conjungate space of V. We consider the so-called lognormal case when the parametrized diffusion coefficient $a(\mathbf{y})$ is of the form

$$a(\mathbf{y}) = \exp(b(\mathbf{y})), \quad b(\mathbf{y}) = \sum_{j=1}^{\infty} y_j\psi_j, \quad \mathbf{y} = (y_j)_{j=1}^{\infty} \in \mathbb{R}^\infty, \tag{2}$$

where the y_j are i.i.d. standard Gaussian random variables and $\psi_j \in L_\infty(D)$.

By combining a certain approximation property in the spatial domain D, and weighted ℓ_2-summability of the Hermite polynomial expansion coefficients in the parametric domain $\mathbb{R}^\infty$, obtained in [1] and [2], we investigate linear non-adaptive methods of fully discrete polynomial interpolation approximation as well as fully discrete weighted quadrature methods of integration for parametric and stochastic elliptic PDEs (1) with random lognormal inputs ([2]). We explicitly construct such methods and prove corresponding convergence rates of the approximations by them. The linear non-adaptive methods of fully discrete polynomial interpolation approximation are sparse-grid collocation methods which are certain sums taken over finite nested Smolyak-type indices sets $G(\xi)$ parametrized by positive number ξ, of mixed tensor products of dyadic scale successive differences of spatial approximations of particular solvers, and of successive differences of their parametric Lagrange interpolating polynomials. Moreover, they generate in a natural way fully discrete weighted quadrature formulas for integration of the solution to parametric and stochastic elliptic PDEs and its linear functionals, and the error of the corrsponding integration can be estimated via the error in the Bochner space $L_1(\mathbb{R}^\infty, V, \gamma)$ norm of the generating methods where γ is the Gaussian probability measure on $\mathbb{R}^\infty$. Our analysis leads to auxiliary convergence rates in ξ of these approximations when ξ going to ∞. For a given $n \in \mathbb{N}$, we choose ξ_n so that the cardinality of $G(\xi_n)$ which in some sense characterizes computation complexity, does not exceed n, and hence obtain the convergence rates in increasing n, of the fully discrete polynomial approximation and weighted integration. We also briefly consider problems of non-fully discrete polynomial interpolation approximation and integration. For details see [3].

References

[1] *Bachmayr M., Cohen A., DeVore A. and Migliorati G.* Sparse polynomial approximation of parametric elliptic PDEs // Part II: lognormal coefficients, ESAIM Math. Model. Numer. Anal. 2017. V. 51. 341-363.

[2] *Bachmayr M., Cohen A., Dũng D. and Schwab C.* Fully discrete approximation of parametric and stochastic elliptic PDEs // SIAM J. Numer. Anal. 2017. V. 55. 2151-2186.

[3] *Dũng D.* Sparse-grid polynomial interpolation approximation and integration for parametric and stochastic elliptic PDEs with lognormal inputs // arXiv preprint 2019. arXiv:1904.06502v4 [math.NA].

Logan Type Uncertainty Principle for Bandlimited Functions

Gorbachev D.V.

Tula State University
dvgmail@mail.ru

We study the uncertainty principles related to the generalized Logan problem in $\mathbb{R}^d$. Our main result provides the complete solution of the following problem: for a fixed $m \in \mathbb{Z}_+$, find

$$\sup\{|x|\colon (-1)^m f(x) > 0\} \cdot \sup\{|x|\colon x \in \operatorname{supp} \widehat{f}\,\} \to \inf,$$

where the infimum is taken over all nontrivial positive definite bandlimited functions such that

$$\int_{\mathbb{R}^d} |x|^{2k} f(x)\,dx = 0 \;\; for \;\; k = 0, \dots, m-1 \;\; if \;\; m \geqslant 1.$$

We also obtain the uncertainty principle for bandlimited functions related to the recent result by Bourgain, Clozel, and Kahane.

This is a joint research with V. Ivanov (Tula State University, Russia) and S. Tikhonov (CRM, ICREA, Barcelona).

The work of D. Gorbachev and V. Ivanov is supported by the Russian Science Foundation under grant 18-11-00199 and performed in Tula State University. S. Tikhonov was partially supported by MTM 2017-87409-P, 2017 SGR 358, and the CERCA Programme of the Generalitat de Catalunya.

References

[1] *Gorbachev D.V., Ivanov V.I., and Tikhonov S.Yu.* Uncertainty principles for eventually constant sign bandlimited functions // arXiv:1904.11328.

Predictive Maintenance

Denis Grachev

EPAM Systems
https://www.epam.com/

The research covers two common problems of predictive maintenance: lack markup data and forecast justification for clogging of the evaporator. We have proposed a technique to set target value and models accuracy by searching constant integral values and a way to summarize the factor importance in dynamics.

Projectors onto Spaces of Chebyshevian Splines

Karen Keryan

Yerevan State University,
American University of Armenia,
karenkeryan@ysu.am, kkeryan@aua.am

The talk is devoted to a generalization of Shadrin's theorem. We prove that the L^∞-norm of the orthoprojector on Chebyshevian splines are bounded by a constant, which is independent of the mesh and depends on the underlying extended complete Chebyshev system. It is also proved that the Chebyshevian B-splines are perturbations of polynomial B-splines.

References

[1] *A. Shadrin.* The L^∞-norm of the L^2-spline projector is bounded independently of the knot sequence: a proof of de Boor's conjecture. Acta Math., 187(1):59-137, 2001.

Finite Sums of Ridge Functions on Convex Sets

Kuleshov A.A.

Lomonosov Moscow State University
kuleshov.a.a@yandex.ru

We assume that $n \geqslant 2$ and $E \subset \mathbb{R}^n$ is a set. A closed convex set E with a non-empty interior is called a *convex body*. A *ridge function* on E is a function of the form $\varphi(\mathbf{a} \cdot \mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n) \in E$, $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $\mathbf{a} \cdot \mathbf{x} = \sum_{j=1}^n a_j x_j$ and φ is a real-valued function defined on $\Delta(\mathbf{a}) = \{\mathbf{a} \cdot \mathbf{x} : \mathbf{x} \in E\}$. On a set E, consider a sum of ridge functions

$$f(\mathbf{x}) = \sum_{i=1}^{m} \varphi_i(\mathbf{a}^i \cdot \mathbf{x}).$$

Let E be a convex body. We study the smoothness properties of the functions φ_i under certain assumptions on smoothness of f. We prove that the continuity of f implies that every φ_i belongs to the VMO space on every compact interval of its domain. Also, we prove that for the existence of finite limits of the functions φ_i at the corresponding boundary points of their domains, it suffices to assume the Dini condition on the modulus of continuity of f at some boundary point of E. Also we prove that the obtained (Dini) condition is sharp. Then we extend some of the results on the $C^k(E)$ classes for $k \geqslant 1$. We prove that for measurable functions φ_i the general implication $f \in C^k(E) \Rightarrow \varphi_i \in C^k(\Delta(\mathbf{a}^i))$ $(i = 1, \ldots, m)$ holds iff the boundary of E is *smooth*.

References

[1] *S.V. Konyagin, A.A. Kuleshov.* On the continuity of finite sums of ridge functions // Math. Notes, 2015, v. 98, pp. 336-338.

[2] *S.V. Konyagin, A.A. Kuleshov.* On some properties of finite sums of ridge functions defined on convex subsets of ^{n} // Proc. Steklov Inst. Math., 2016, v. 293, pp. 186 - 193.

[3] *A.A. Kuleshov.* On some properties of smooth sums of ridge functions // Proc. Steklov Inst. Math., 2016, v. 294, pp. 89 - 94.

[4] *A.A. Kuleshov.* Continuous sums of ridge functions on a convex body and the class VMO // Math. Notes, 2017, v. 102, pp. 799-805.

[5] *S.V. Konyagin, A.A. Kuleshov, V.E. Maiorov.* Some problems in the theory of ridge functions // Proc. Steklov Inst. Math., 2018, v. 301, pp. 144-169.

[6] *A.A. Kuleshov.* Continuous sums of ridge functions on a convex body with Dini condition on moduli of continuity at boundary points // Analysis Mathematica, 2019, v. 45, iss. 2, pp. 335–345.

Some Estimates of Submatrix Norms

Limonova I.V.

Moscow State University
limonova_ irina@rambler.ru

Estimates of the operator norms of the submatrices of a given matrix A find applications in various areas of mathematics, for instance, in computational methods and discretization. One of the first results of this topic was obtained in [1]. There the estimates for the classical operator norm of submatrices are a concequence of the corresponding estimates for $(2,1)$-norms. The case of $(2,1)$-norm was studied in [2] in details. It contains the partial answer to the Srivastava's question arosen in his research blog.[1] Now we consider conditions on a matrix A with unit operator (p,q)-norm ensuring the existence of a partition of this matrix into two submatrices with (p,q)-norms close to $1/2^{1/q}$.

References

[1] *B. S. Kashin.* Some properties of matrices of bounded operators from the space l_2^n into l_2^m. Izv. Akad. Nauk. Armyan. SSR Ser. Mat. 15(5), 379-394 (1980)

[2] *B. S. Kashin, I. V. Limonova.* Decomposing a Matrix into two Submatrices with Extremally Small (2,1)-norm. Math. Notes, 106:1 (2019), 63-70.

[1] https://math.berkeley.edu/ nikhil/courses/270/open.pdf

Some Problems of Optimal Recovery of Linear Operators

Magaril-Il'yaev G.G.

Faculty of Mechanics and Mathematics, Moscow State University
magaril@mech.math.msu.su

Osipenko K.Yu.

Faculty of Mechanics and Mathematics, Moscow State University
kosipenko@yahoo.com

In the report we will talk about the optimal recovery of functions from their inexactly given spectrum and recovery of solutions of differential equations from inaccurate initial data. In the first case, as an illustration, we consider the optimal recovery of functions from inaccurate information about their Fourier coefficients. Next, for a special one-parameter semi-group of operators, we consider the optimal recovery of the operator at a given value of the parameter from inaccurate information about the values of other parameters. We construct a family of optimal methods. As a consequence we find a family of optimal methods in the problem of optimal recovery for the solution of the heat equation on $\mathbb{R}^d$

$$\frac{\partial u}{\partial t} = \Delta u, \quad u(0,\cdot) = f(\cdot),$$

at the time instant t from their approximate measurements at time instants $t_1 < \ldots < t_n$. We also consider the problem of optimal recovery of the solution for the Dirichlet problem in the half-space

$$\Delta w\,(x,y) = 0,\ (x,y) \in \mathbb{R}^d \times \mathbb{R},\ y > 0,\ w(\cdot,0) = f(\cdot),$$

which is to recover the solution on the hyperplane $y = Y$ from its inaccurate measurements on the hyperplanes $y = y_i$, $i = 1,\ldots,n$.

References

[1] *Magaril-Il'yaev G.G., Osipenko K.Yu.* On best harmonic synthesis of periodic functions // J. Math. Sci. 2015. V. 209. N 1. P. 115-129.

[2] *Magaril-Il'yaev G.G., Osipenko K.Yu.* Optimal recovery of the solution of the heat equation from inaccurate data // Sb. Math. 2009. V. 200. N 5. P. 665-682.

Synthetic Data in Deep Learning

Sergey Nikolenko

Laboratory of Mathematical Logic of the
St.-Petersburg Department of the Steklov Mathematical Institute
sergey@logic.pdmi.ras.ru

Abstract: Many major problems of modern AI come down to data: either lack of data or, also very often, lack of labeled data. Synthetic data is an important approach to solving the data problem by either producing artificial data from scratch or using advanced augmentation techniques to produce novel and diverse training examples. In the talk, I will introduce the notion of synthetic data and various approaches to making and using it. In particular, we will discuss domain adaptation, a set of techniques designed to make a model trained on one domain of data, the source domain, work well on a different, target domain. This is a natural fit for synthetic data: in almost all applications, we would like to train the model in the source domain of synthetic data but then apply the results in the target domain of real data. We will survey DA approaches for synthetic-to-real adaptation, concentrating on deep learning models. We will see the gaze estimation story from Apple's Refiner and beyond, DA techniques for learning to drive, GAN-based DA for medical imaging, and much more. Expect a lot of GANs and loss functions!

Human Pose Estimation

Daniil Osokin

Internet of Things group, Intel A/O
https://intel.ru/

Human pose estimation task aims at predicting coordinates of a person's keypoints: shoulders, elbows, knees, ankles, etc. It is useful in different domains, such as action recognition, sports, augmented reality. We will discuss major approaches for human pose estimation based on deep learning. Also it will be shown, how specific pose estimation method can be adopted for the real-time inference on CPU.

Convergence of Stochastic Subspace Correction Methods Fault Tolerance

Peter Oswald

Dresden / INS, Bonn University
agp.oswald@gmail.com

We present convergence results for the expectation $\mathbb{E}(\|e^{(m)}\|^2)$ of the squared error in stochastic subspace correction schemes and their accelerated versions to solve symmetric coercive variational problems. As a motivating application we discuss their potential for achieving fault tolerance in an unreliable compute network. We employ an overlapping domain decomposition algorithm for PDE discretizations to discuss the latter aspect.
This is joint work with Michael Griebel (INS, University Bonn).

References

[1] *Griebel, M., Oswald, P.* Stochastic subspace correction methods and fault tolerance. Math. Comp. (in print); arXiv:1807.11315; Univ. Bonn, INS Preprint No. 1809.

Stochastic Gradient Descent

Peter Richtarik

King Abdullah University of Science and Technology and University of Edinburgh
peter.richtarik@kaust.edu.sa

In this series of talks I will present a new unified analysis of SGD for regularized smooth strongly convex optimization, and through special cases shed light on notions such as importance sampling, minibacthing, variance reduction and gradient compression.

Measurement Matrices of Integers with Small Elements

K.S. Ryutin

Lomonosov Moscow State University,
High–dimensional approximation and applications laboratory
kriutin@yahoo.com

We plan to discuss some problems around the integer sparse recovery problem raised in [1]. We consider the equation $y = Ax$, with A — some integer–valued $m \times d$- matrix (we call it the measurement matrix) and x an unknown s–sparse integer vector (the cardinality of its support is s).

The main questions are: is it possible to construct the measurement $m \times d$–matrix A such that any s–sparse vector x can be recovered if one knows Ax? How small can be the elements of such a matrix? What about the complexity of the recovery algorithm?

Very interesting results on the first 2 questions were obtained in [1],[2],[4]. The algorithm for the recovery was given in [3] with estimates for its complexity. This algorithm works (and is specially designed) for the explicit measurement matrix from [4]. Let us mention one of the results from the recent paper [5]

Proposition. *For large N and any $s \leqslant c_1 N/\log N$ there exists a boolean $(c_2 s \log N) \times N$ measurement matrix such that any s–sparse vector from $\mathbb{Z}^N$ can be recovered.*

The matrix is explicitly given and the recovery algorithm is of a polynomial complexity.

This work was supported be the Russian Federation Government grant 14.W03.31.0031.

References

[1] *L. Fukshansky, D. Needell, B. Sudakov.* An algebraic perspective on integer sparse recovery. // Appl. Mathem. and Comp. (2019). – V. 340. – P. 31–42.

[2] *S. V. Konyagin.* On the recovery of an integer vector from linear measurements. Math. Notes, (2018), V. 104 N. 6, P. 859–865.

[3] *K.S. Ryutin.* The algorithm for the recovery of integer vector via linear measurements. https://arxiv.org/abs/1905.02687

[4] *S.V. Konyagin, B. Sudakov.* An extremal problem for integer sparse recovery. https://arxiv.org/abs/1904.08661.

[5] *К.С. Рютин.* Целочисленные матрицы измерения с малыми элементами, обеспечивающие восстановление векторов. submitted to Math. Notes. (2019)

Fashion AI: One-Shot Clothing Detection

Alexey Sidnev

Huawei Research Center, Nizhny Novgorod, Russia,
and Lobachevsky State University of Nizhny Novgorod
https://www.huawei.com/ru/

Fashion analysis is quite a complex task. Clothing detection is the first technological challenge to be solved for almost any fashion analysis task. Recently presented one-shot approach CenterNet outperforms most of the current detectors and demonstrates outstanding speed-accuracy trade-off. We consider CenterNet architecture in details along with a brief review of deep learning based object detection methods. Also we present our results on DeepFashion2 Challenge and show that CenterNet is able to achieve the state-of-the-art accuracy for bounding box detection and landmark detection tasks.

Approximation by Quasi-Projection Operators and Fourier Multipliers

Skopina M.
Saint Petersburg, Russia
skopina@MS1167.spb.edu
(joint results with Yu. Kolomoitsev and A. Krivoshein)

Quasi-projection operators with a matrix dilation M are defined by

$$Q_j(f,\varphi,\widetilde{\varphi}) = \sum_{k\in\mathbb{Z}^d} \langle f, \widetilde{\varphi}_{jk}\rangle \varphi_{jk},$$

where φ is a function, $\widetilde{\varphi}$ is a tempered distribution,

$$\psi_{jk}(x) := |\det M|^{j/2}\psi(M^j x + k), \quad j\in\mathbb{Z}, k\in\mathbb{Z}^d,$$

and $\langle f, \widetilde{\varphi}_{jk}\rangle$ has meaning in some sense. If $d=1$, $\widetilde{\varphi}$ is the Dirac delta-function and $\varphi = \mathrm{sinc}$, then $Q_j(f,\varphi,\widetilde{\varphi})$ is the classical sampling operator. We consider different classes of such operators and study their approximation properties.

Error estimates in L_p-norm, $2 \leqslant p \leqslant \infty$, are provided for a large class of functions $\varphi \in L_p(\mathbb{R}^d)$ and for $\widetilde{\varphi} \in \mathcal{S}'_N$, where $\mathcal{S}'_N$ is the set of tempered distribution whose Fourier transform $\widehat{\widetilde{\varphi}}$ is a function on $\mathbb{R}^d$ such that $|\widehat{\widetilde{\varphi}}(\xi)| \leqslant C_{\widetilde{\varphi}}(1+|\xi|)^N$. Under the Strang-Fix conditions of order s for φ, so-called weak compatibility condition of order s for φ and $\widetilde{\varphi}$, and enough decay of $\widehat{\varphi}$, the estimates are given in terms of the Fourier transform of f. Approximation order depends on the smoothness of f and on s. Under additional assumption on $\widetilde{\varphi}$, these results are improved in several directions for two classes of functions φ. Namely, the estimates are obtained for all $p \geqslant 1$, for a wider class of functions f, and given in terms of the moduli of smoothness and best approximations of f. The first class consists of functions φ decaying faster than $|x|^{-d-\epsilon}$, $\epsilon > 0$. The second class consists of band-limited functions φ such that the function $\widehat{\varphi}$ and the derivatives of order s of the functions $\widehat{\varphi}\overline{\widehat{\widetilde{\varphi}}}$, $\widehat{\varphi}(\cdot + l)$, $l \in \mathbb{R}^d$, $l \neq 0$, restricted to a set $[-\delta,\delta]^d$, $\delta \in (0, 1/2]$, are the Fourier multipliers in $L_p(\mathbb{R}^d)$.

Uniform Recovery Guarantees for Least Squares Approximation

Ullrich T.

Technical University of Chemnitz, Germany
tino.ullrich@mathematik.tu-chemnitz.de

We consider the problem of learning multivariate functions belonging to a certain reproducing kernel Hilbert space from n given samples. Our focus will be on uniform recovery. In fact, we propose a least squares algorithm for which we manage to control the worst-case recovery error for the whole class. This algorithm uses n function samples at randomly drawn nodes as input and projects onto a subspace spanned by an orthonormal system. Examples could be the multivariate Fourier system or (bi-)orthogonal wavelets. The random nodes are drawn once for the whole class (comparable to the draw of an RIP matrix which works well for all sparse signals). Our recovery guarantees involve explicit constants and preasymptotics and are valid with high probability. The same algorithm may be used for scattered data approximation. We will also comment on the practical potential of the method by showing some numerical experiments. This is joint work with L. Kämmerer and T. Volkmer (TU Chemnitz).

Chebyshev Polynomials and Best Rank-one Approximation Ratio

André Uschmajew
MPI MiS Leipzig

We establish a new extremal property of the classical Chebyshev polynomials in the context of the theory of rank-one approximations of tensors. We also give some necessary conditions for a tensor to be a minimizer of the ratio of spectral and Frobenius norms. This is joint work with Andrei Agrachev and Khazhgali Kozhasov.

Machine Learning in Electrocardiogram Diagnosis

Nikolai Zolotykh

Laboratory of advanced for high-dimensional data analysis,
Nizhny Novgorod Lobachevski State University
nikolai.zolotykh@gmail.com

The talk will focus on different approaches to electrocardiogram (ECG) analysis: from rule-based methods to deep learning. In particular, we will consider two approaches to the ECG segmentation - one of them exploits wavelet transforms (rule-based method) and another uses UNET-like neural networks - and different architectures of neural networks for ECG classification. We will discuss the advantages and disadvantages of different approaches and see which method is best used depending on the size of the training set and on other characteristics.

Contents

1 *Aleshin A.* Deep Learning Appliance for Sensitive Data

2 *Astashkin S.* Representing Systems of Dilations and Translations of a Function in Symmetric Spaces

4 *Barron A.* Approximation and Cover Properties of Learning Networks

5 *Barron A.* Complexity and Risk Properties of Deep Nets

6 *Barron A.* The Role of Information Theory in Deep Net Analysis

7 *Borodin P.* Greedy Algorithm with Respect to Assymmetric Dictionary may Diverge

8 *Burnaev E.* Latent Convolutional Models for Image Restoration

9 *Buslaev V.* About Multipoint Schur Criteria

10 *Dinh Dũng* Sparse-grid Polynomial Interpolation Approximation and Integrationfor Parametric and Stochastic Elliptic PDEs with Lognormal Inputs

13 *Gorbachev D.* Logan Type Uncertainty Principle for Bandlimited Functions

14 *Grachev D.* Predictive Maintenance

15 *Keryan K.* Projectors onto Spaces of Chebyshevian Splines

16 *Kuleshov A.* Finite Sums of Ridge Functions on Convex Sets

18 *Limonova I.* Some Estimates of Submatrix Norms

19 *Magaril-Il'yaev G.* Some Problems of Optimal Recovery of Linear Operators

21 *Nikolenko S.* Synthetic Data in Deep Learning

22 *Osokin D.* Human Pose Estimation

23 *Oswald P.* Convergence of Stochastic Subspace Correction Methods Fault Tolerance

24 *Richtarik P.* Stochastic Gradient Descent

25 *Ryutin K.* Measurement Matrices of Integers with Small Elements

27 *Sidnev A.* Fashion AI: One-Shot Clothing Detection

28 *Skopina M.* Approximation by Quasi-Projection Operators and Fourier Multipliers

29 *Ullrich T.* Uniform Recovery Guarantees for Least Squares Approximation

30 *Uschmajew A.* Chebyshev Polynomialsand Best Rank-One Approximation Ratio

31 *Zolotykh N.* Machine Learning in Electrocardiogram Diagnosis