

Applied Functional Analysis

Feng Dai (University of Alberta)

Ronald DeVore (Texas A & M University)

Vladimir Temlyakov (University of South Carolina)

Sergey Tikhonov (ICREA and Centre de Recerca Matemàtica - Barcelona)

August 28, 2022—September 2, 2022

The hybrid workshop on applied functional analysis attracted 51 participants, 7 of which participated in-person at the Banff International Research Station. A total of 28 talks (including 26 forty-five minute talks, and 2 twenty-five minute talks) were presented by a wide variety of speakers, ranging from graduate students, postdocs, early career academics, to senior academics. The presentations covered a wide range of topics covering foundational aspects of sampling discretization, applied harmonic analysis, sparse approximation theory, numerical and functional analysis.

1. OVERVIEW OF THE FIELD AND RECENT DEVELOPMENTS

Many multi-parameter real-world problems are modeled on spaces of multivariate functions in n variables, where n may be very large. Since these problems can almost never be solved analytically, one is interested in suitable model assumptions and good approximate solutions within a reasonable computing time. The multivariate approximation theory in the classical setting has close connections with other areas of mathematics and has many applications in numerical computations. However, classical methods do not work for really high dimensions. High-dimensional approximation is a hot rapidly developing area of mathematics and numerical analysis where researchers try to understand which new approaches may work. A promising contemporary approach is based on the concept of sparsity and nonlinear m -term approximation. The last decade has seen great successes in studying nonlinear approximation which was motivated by numerous applications. The fundamental question of nonlinear approximation is how to devise good constructive methods (algorithms) of nonlinear approximation. This problem has two levels of nonlinearity. The first level of nonlinearity is m -term approximation (sparse approximation) with regard to bases. In this problem one can use the unique function expansion with regard to a given basis to build an approximant. Nonlinearity enters by

looking for m -term approximants with terms (i.e. basis elements in approximant) allowed to depend on a given function. Since the elements of the basis used in the m -term approximation are allowed to depend on the function being approximated, this type of approximation is very efficient. On the second level of nonlinearity, we replace a basis by a more general system which is not necessarily minimal (for example, redundant system, dictionary). This setting is much more complicated than the first one (bases case), however, there is a solid justification of importance of redundant systems in both theoretical questions and in practical applications. Recent results have established that greedy type algorithms are suitable methods of nonlinear approximation in both m -term approximation with regard to bases and m -term approximation with regard to redundant systems. It turns out that there is one fundamental principal that allows us to build good algorithms both for arbitrary redundant systems and for very simple well structured bases like the Haar basis. This principal is the use of a greedy step in searching for a new element to be added to a given m -term approximant. By a greedy step, we mean one which maximizes a certain functional determined by information from the previous steps of the algorithm. We obtain different types of greedy algorithms by varying the above mentioned functional and also by using different ways of constructing (choosing coefficients of the linear combination) the m -term approximant from the already found m elements of the dictionary. We payed a special attention to this new promising area of research and invited several experts on greedy approximation to speak at the workshop.

It was understood in the beginning of the 20th century that smoothness properties of a univariate function determine the rate of approximation of this function by polynomials (trigonometric in the periodic case and algebraic in the non-periodic case). A fundamental question is: What is a natural multivariate analog of univariate smoothness classes? Different function classes were considered in the multivariate case: isotropic and unisotropic Sobolev and Besov classes, classes of functions with bounded mixed derivative and others. The next fundamental question is: How to approximate functions from these classes? Kolmogorov introduced the concept of the n -width of a function class. This concept is very useful in answering the above question. The Kolmogorov n -width is a solution to an optimization problem where we optimize over n -dimensional linear subspaces. This concept allows us to understand which n -dimensional linear subspace is the best for approximating a given class of functions. The rates of decay of the Kolmogorov n -width are known for the univariate smoothness classes. In some cases even exact values of it are known. The problem of the rates of decay of the Kolmogorov n -width for

the classes of multivariate functions with bounded mixed derivative is still an open problem. We note that the function classes with bounded mixed derivative are not only interesting and challenging object for approximation theory but they are important in numerical computations. This topic was discussed in detail at the workshop.

Numerical integration is one more challenging multivariate problem where approximation theory methods are very useful. For a given function class F we want to find m points x^1, \dots, x^m in D such that $\sum_{j=1}^m \frac{1}{m} f(x^j)$ approximates well the integral $\int_D f d\mu$, where μ is the normalized Lebesgue measure on D . Classical geometric discrepancy theory is concerned with different versions of the following questions: What is the most uniform way to distribute finitely many points in various geometric settings (in particular, in high dimensions)? In other words, how well can one approximate continuous objects by discrete ones? And how large are the errors that inevitably arise in such approximations? Quite naturally this topic is directly related to approximation theory, more specifically - numerical integration: well-distributed point sets provide good cubature formulas and different notions of discrepancy yield integration error estimates for various function classes. This immediately brings up the connection between discrepancy and functional analysis and the function space theory. Classical discrepancy theory provides constructions of point sets that are good for numerical integration of characteristic functions of parallelepipeds of the form $P = \prod_{j=1}^d [a_j, b_j]$. The typical error bound is of the form $m^{-1}(\log m)^{d-1}$. Note that a regular grid for $m = n^d$ provides an error of the order $m^{-1/d}$. The above mentioned results of discrepancy theory are closely related to numerical integration of functions with bounded mixed derivative (the case of the first mixed derivative). Sparse grids play an important role in numerical integration of functions with bounded mixed derivative. In the case of r th bounded mixed derivative they provide an error of the order $m^{-r}(\log m)^{(d-1)(r+1)}$. Also, they provide the recovery error in the sampling problem of the same order. Note again that the regular grid from above provides an error of the order $m^{-r/d}$. The error bound $m^{-r}(\log m)^{(d-1)(r+1)}$ is reasonably good for moderate dimensions d , say, $d \leq 40$. It turns out that there are practical computational problems with moderate dimensions where sparse grids work well. Sparse grids techniques have applications in quantum mechanics, numerical solutions of stochastic PDEs, data mining, finance. This topic, including numerical integration on the sphere was also discussed in detail at the workshop.

2. PRESENTATION HIGHLIGHTS

Many of the recent developments highlighted in the last section were presented at the workshop. Details are given below.

1. Alexander Litvak: The minimal dispersion in the unit cube.

We improve known upper bounds for the minimal dispersion of a point set in the unit cube. Our bounds are sharp up to logarithmic factors. The talk is partially based on a joint work with G. Livshyts.

2. D. Leviatan: Coconvex approximation of periodic functions.

Let \tilde{C} be the space of continuous 2π -periodic functions f , endowed with the uniform norm $\|f\| := \max_{x \in \mathbb{R}} |f(x)|$, and denote by $\omega_k(f, t)$, the k -th modulus of smoothness of f . Denote by \tilde{C}^r , the subspace of r times continuously differentiable functions $f \in \tilde{C}$, and let \mathbb{T}_n , be the set of trigonometric polynomials T_n of degree $\leq n$ (that is, of order $\leq 2n + 1$). Given a set $Y_s := \{y_i\}_{i=1}^{2s}$, of $2s$ points, $s \geq 1$, such that $-\pi \leq y_1 < y_2 < \dots < y_{2s} < \pi$, and a function $f \in \tilde{C}^r$, $r \geq 3$, that changes convexity exactly at the points Y_s , namely, the points Y_s are all the inflection points of f . We wish to approximate f by trigonometric polynomials which are coconvex with it, that is, satisfy

$$f''(x)T_n''(x) \geq 0, \quad x \in \mathbb{R}.$$

We prove, in particular, that if $r \geq 3$, then for every $k, s \geq 1$, there exists a sequence $\{T_n\}_{n=N}^{\infty}$, $N = N(r, k, Y_s)$, of trigonometric polynomials $T_n \in \mathbb{T}_n$, coconvex with f , such that

$$\|f - T_n\| \leq \frac{c(r, k, s)}{n^r} \omega_k(f^{(r)}, 1/n).$$

It is known that one may not take N independent of Y_s .

3. Speaker: Bin Han: Generalized Hermite subdivision schemes and spline wavelets on intervals

Hermite and Birkhoff interpolation is a classical topic in approximation theory and is useful in CAGD and numerical PDEs, due to their connections to spline theory and wavelets. In this talk, we shall introduce the notation of generalized Hermite subdivision schemes, characterize their convergence and smoothness, and then discuss their connections with spline multiwavelets having the interpolation properties. We provide some examples of generalized Hermite subdivision schemes having the Hermite/Birkhoff interpolation and spline properties. As an application, we first construct spline multiwavelets on the real line from some generalized Hermite subdivision schemes, adapt them to the unit interval through a recent general method for adapting any wavelets to bounded intervals, and then illustrate their application to the cavity problem of the Helmholtz equation. Some are joint work with M. Michelle.

4. Yuri Malykhin: Widths and rigidity.

We will consider Kolmogorov widths of finite systems of functions. It is known that any orthonormal system $\{f_1, \dots, f_N\}$ is *rigid* in L_2 , i.e. it can't be approximated by linear spaces of dimension essentially smaller than N . This is not always true in weaker metrics, e.g. the first N Walsh functions can be $o(1)$ -approximated by linear spaces of dimension $o(N)$ in L_p , $p < 2$. We will give some sufficient conditions for rigidity in that norms. Also we will discuss the connections between widths and the notion of matrix rigidity and give some related positive results on approximation of Fourier and Walsh systems.

5. Ben Adcock: Is Monte Carlo a bad sampling strategy for learning smooth functions in high dimensions?

This talk concerns the approximation of smooth, high-dimensional functions on bounded hypercubes from limited samples using polynomials. This task lies at the heart of many applications in computational science and engineering – notably, those arising from parametric modelling and computational uncertainty quantification. It is common to use Monte Carlo sampling in such applications, so as not to succumb to the curse of dimensionality. However, it is well known that such a strategy is theoretically suboptimal. Specifically, there are many polynomial spaces of dimension n for which the sample complexity scales log-quadratically, i.e., like $c \cdot n^2 \cdot \log(n)$ as $n \rightarrow \infty$. This well-documented phenomenon has led to a concerted effort over the last decade to design improved, in fact, near-optimal strategies, whose sample complexities scale log-linearly, or even linearly in n .

Paradoxically, in this talk we demonstrate that Monte Carlo is actually a perfectly good strategy in high dimensions, despite this apparent suboptimality. We first document this phenomenon empirically via several numerical examples. Next, we present a theoretical analysis that resolves this seeming contradiction for the class of $(\mathbf{b}, \varepsilon)$ -holomorphic functions of infinitely-many variables. We show that there is a least-squares approximation based on m Monte Carlo samples whose error decays algebraically fast in $m/\log(m)$, with a rate that is the same as that of the best n -term polynomial approximation. This result is non-constructive, since it assumes knowledge of a suitable polynomial subspace (depending on \mathbf{b}) in which to compute the approximation. Hence, we then present a constructive scheme based on compressed sensing that achieves the same rate, subject to a slightly stronger assumption on \mathbf{b} and a larger polylogarithmic factor. This scheme is practical, and numerically performs as well as or better than well-known adaptive least-squares schemes.

Finally, while most of our results concern polynomials, we also demonstrate that the same results can be achieved via deep neural networks with standard training procedures. Overall, our findings demonstrate that Monte Carlo sampling is eminently suitable for smooth function approximation tasks on bounded domains when the dimension is sufficiently high. Hence, the benefits of state-of-the-art improved sampling strategies seem to be generically limited to lower-dimensional settings.

This is joint work with Simone Brugiapaglia, Juan M. Cardenas, Nick Dexter and Sebastian Moraga.

References

- B. Adcock & S. Brugiapaglia. *Is Monte Carlo a bad sampling strategy for learning smooth functions in high dimensions?* Preprint (2022).
- B. Adcock, S. Brugiapaglia, N. Dexter & S. Moraga. *On efficient algorithms for computing near-best polynomial approximations to high-dimensional, Hilbert-valued functions from limited samples.* arXiv:2203.13908 (2022).
- B. Adcock, S. Brugiapaglia, & C. G. Webster, *Sparse Polynomial Approximation of High-Dimensional Functions*, SIAM, Philadelphia, PA (2022).
- B. Adcock & J. M. Cardenas, *Near-optimal sampling strategies for multivariate function approximation on general domains*, SIAM J. Math. Data Sci., 2(3):607–630 (2020).

6. Martin Buhmann (Janin Jäger): Strict Positive Definiteness of Convolutional and Axially Symmetric Kernels.

We study new sufficient conditions of strict positive definiteness for generalisations of radial basis and other kernels on multi-dimensional spheres which are no longer radially symmetric but possess specific coefficient structures. The results use the series expansion of the kernel in spherical harmonics. The kernels either have a convolutional form or are axially symmetric with respect to one axis.

References

Martin Buhmann & Janin Jäger [2022] "Strictly positive definite convolutional and axial-symmetric kernels on d -dimensional spheres", *J. of Fourier Analysis and Applica.*

Martin Buhmann & Janin Jäger [2021] "Strictly positive definite kernels on the 2 -sphere: from radial symmetry to eigenvalue block structure". *IMA J. Numerical Analysis*, 1-26.

Martin Buhmann & Janin Jäger [2020] "Pólya type criteria for conditional strict positive definiteness of functions on spheres". *Journal of Approximation Theory* 257.

7. Mario Ullrich (Matthieu Dolbeault, David Krieg): On optimal L_2 -approximation with function values.

Let $F \subset L_2$ be a class of complex-valued functions on a set D , such that, for all $x \in D$, point evaluation $f \mapsto f(x)$ is a continuous linear functional. We study the L_2 -approximation of functions from F and want to compare the power of function values with the power of arbitrary linear information. To be precise, the *sampling number* $g_n(F)$ is the minimal worst-case error (in F) that can be achieved with n function values, whereas the *approximation number* (or Kolmogorov width) $d_n(F)$ is the minimal worst-case error that can be achieved with n pieces of arbitrary linear information (like derivative values or Fourier coefficients).

Here, we report on recent developments in this problem and, in particular, explain how the individual contributions from [1,2,3,4] lead to the following statement:

There is a universal constant $c \in \mathbb{N}$ such that the sampling numbers of the unit ball F of every separable reproducing kernel Hilbert space are bounded by

$$g_{cn}(F) \leq \sqrt{\frac{1}{n} \sum_{k \geq n} d_k(F)^2},$$

We also obtain similar upper bounds for more general classes F , and provide examples where our bounds are attained up to a constant. For example, if we assume that $d_n(F) \asymp n^{-\alpha} (\log n)^\beta$ for some $\alpha > 1/2$ and $\beta \in \mathbb{R}$, then we obtain

$$g_n(F) \asymp d_n(F),$$

showing that function values are (up to constants) as powerful as arbitrary linear information.

The results rely on the solution to the Kadison-Singer problem, which we extend to the subsampling of a sum of infinite rank-one matrices.

Bibliography:

- 1 M. Dolbeault, D. Krieg and M. Ullrich, A sharp upper bound for sampling numbers in L_2 , preprint.
- 2 D. Krieg and M. Ullrich, Function values are enough for L_2 -approximation, *Found. Comput. Math.* **21** (2021), 1141–1151.
- 3 D. Krieg and M. Ullrich, Function values are enough for L_2 -approximation: Part II, *J. Complexity* **66** (2021).
- 4 N. Nagel, M. Schäfer and T. Ullrich, A new upper bound for sampling numbers, *Found. Comput. Math.* **21** (2021).

8. Janin Jäger (Martin Buhmann, Jean Carlo): Strict positive definiteness: From compact Riemannian manifolds to the sphere.

Isotropic positive definite functions are used in approximation theory and are for example applied in geostatistics and physiology. They are also of importance in statistics where they occur as correlation functions of homogeneous random fields on spheres. We study a class of functions applicable for interpolation of arbitrary scattered data on \mathbb{M}^d by linear combination of shifts of an isotropic basis function φ , where \mathbb{M}^d is a compact Riemannian manifold.

A class of functions for which the resulting interpolation problem is uniquely solvable for any distinct point set $\Xi \subset \mathbb{M}^d$ and arbitrary d is the class of strictly positive definite kernels $SPD(\mathbb{M}^d)$. For kernels possessing a certain series expansion in eigenfunctions of the Laplace-Beltrami operator on \mathbb{M}^d , we derive a characterisation of this class. First, for general compact Riemannian manifolds, then for homogeneous manifolds and finally for two-point homogeneous manifolds and the sphere (see [3]).

For the special case of \mathbb{S}^{d-1} , the results extend the characterisation for radial kernels from [1]. For this case, we derive conditions showing that non-radial kernels can be strictly positive definite while possessing significantly less positive coefficients, in the given expansion, compared to radially symmetric kernels (see [2]).

REFERENCES

- [1] Chen, D., Menegatto, V. A., & Sun, X. (2003). A Necessary and Sufficient Condition for Strictly Positive Definite Functions on Spheres. *Proceedings of the AMS*, 131(9), 2733–2740.
- [2] Guella, J. C. & Jäger, J. (2022). Strictly positive definite non-isotropic kernels on two-point homogeneous manifolds: The asymptotic approach, *Arxiv-eprints*, arXiv:2205.07396
- [3] Buhmann, M., & Jäger, J. (2022). Strict positive definiteness of convolutional and axially symmetric kernels on d-dimensional spheres. *Journal of Fourier Analysis and Applications*, 28(3), 1-25.

9. Gideon Schechtman: The problem of dimension reduction of finite sets in normed spaces.

Given a normed space X , $C > 1$ and $n \in \mathbb{N}$ we denote by $k_n^C(X)$ the smallest k such that every $S \subset X$ with $|S| = n$ admits an embedding into a k -dimensional subspace of X which distort the mutual distances in S by at most a factor of C . We shall survey the little that is known about estimating $k_n^C(X)$ for different spaces X . The latest is a result of Assaf Naor, Gilles Pisier and myself: Let S_1 denote the Schatten–von Neumann trace class, i.e., the Banach space of all compact operators $T : \ell_2 \rightarrow \ell_2$ whose trace class norm $\|T\|_{S_1} = \sum_{j=1}^{\infty} \sigma_j(T)$ is finite, where $\{\sigma_j(T)\}_{j=1}^{\infty}$ are the singular values of T . We prove that for each $C > 1$, $k_n^C(S_1)$ has a lower bound which is a positive power of n . This extends a result of Brikmann and Charikar (2003) who proved a similar result with ℓ_1 replacing S_1 . It stands in sharp contrast with the Johnson–Lindenstrauss lemma (1984) which says that the situation in ℓ_2 is very different.

10. A. Kroo: Weierstrass type approximation problem for multivariate homogeneous polynomials.

By the celebrated Weierstrass approximation theorem continuous functions on compact sets admit uniform polynomial approximation. The similar question of density of homogeneous multivariate polynomials has been actively investigated in the past 10-15 years. In this talk we will give a survey of the main developments related to this problem.

11. Yuan Xu: Approximation and analysis in localized homogeneous space.

We consider approximation and localized frames on conic domains. Our approach is based on the orthogonal structure that admits a closed-form formula for its reproducing kernels, akin to the addition formula of spherical harmonics. Such a formula leads to highly localized kernels that serve as the foundation for analysis in such domains. The results will be presented in a general framework encompassing well-studied domains such as the unit sphere and the unit ball.

12. Akram Aldroubi: Dynamical sampling: Source term recovery.

Consider the IVP in a separable Hilbert space : where , , is the time derivative of , and is an initial condition. The goal is to recover from the measurements , where is a countable subset of , is an unknown, but slowly varying background source, and is an additive noise.

13. Boris Kashin: On some problems joint for function theory and theoretical computer science.

14. V. Temlyakov: Sampling discretization of the uniform norm.

Discretization of the uniform norm of functions from a given finite dimensional subspace of continuous functions will be discussed. We will pay special attention to the case of trigonometric polynomials with frequencies from an arbitrary finite set with fixed cardinality. We will discuss the fact that for any N -dimensional subspace of the space of continuous functions it is sufficient to use e^{CN} sample points for an accurate upper bound for the uniform norm. Previous known results show that one cannot improve on the exponential growth of the number of sampling points for a good discretization theorem in the uniform norm. Also, we will present a general result, which connects the upper bound on the number of sampling points in the discretization theorem for the uniform norm with the best m -term bilinear approximation of the Dirichlet kernel associated with the given subspace.

We illustrate the application of our technique on the example of trigonometric polynomials.

The talk is based on a joint work with B. Kashin and S. Konyagin.

15. A. Prymak: Optimal polynomial meshes exist on any multivariate convex domain.

We show that every convex body Ω in \mathbb{R}^d possesses optimal polynomial meshes, which confirms a conjecture by A. Kroo. Namely, there exists a sequence $\{Y_n\}_{n \geq 1}$ of finite subsets of Ω such that the cardinality of Y_n is at most $C_1 n^d$ while for any polynomial P of total degree $\leq n$ in d variables $\|P\|_{\Omega} \leq C_2 \|P\|_{Y_n}$, where $\|P\|_X := \sup\{|P(x)| : x \in X\}$ and C_1, C_2 are positive constants depending only on Ω . This is a joint work with Feng Dai.

16. Ding-Xuan Zhou: Approximation Theory of Structured Deep Neural Networks.

Deep learning has been widely applied and brought breakthroughs in speech recognition, computer vision, natural language processing, and many other domains. The involved deep neural network architectures and computational issues have been well studied in machine learning. But there lacks a theoretical foundation for understanding the modelling, approximation or generalization ability of deep learning models with network architectures. One family of structured neural networks is deep convolutional neural networks (CNNs) with convolutional structures. The convolutional architecture gives essential differences between deep CNNs and fully-connected neural networks, and the classical approximation theory for fully-connected networks developed around 30 years ago does not apply. This talk describes an approximation theory of deep CNNs and related structured deep neural networks.

17. Dmitriy Bilyk: Discrete minimizers of energy integrals.

It is quite natural to expect that minimization of pairwise interaction energies leads to uniform distributions, at least for "nice" kernels. However, the opposite effect occurs in many interesting examples, especially for attractive-repulsive energies or when the repulsion is very weak: minimizing measures are discrete (or at least are very non-uniform, e.g. supported on "thin"

or lower-dimensional sets). We shall discuss some results related to this curious phenomenon and its relation to analysis, signal processing, discrete geometry etc.

18. Egor Kosov: New bounds in the problem of sampling discretization of L^p norms.

In the talk we consider the problem of L^p norms discretization by evaluating a function at a certain finite set of points for functions from a given finite dimensional subspace. We mostly consider the cases $p = 1$ and $p = 2$ and present some new upper bounds for the number of points sufficient for such a discretization. In addition, we discuss some general ideas used to obtain these bounds.

19. Javad Mashreghi: A non-recoverable signal.

Given an analytic function f on the open unit disc \mathbb{D} , or an integrable function on its boundary $\mathbb{T} = \partial\mathbb{D}$, our first attempt to approximate f is via its partial Taylor sums $s_n(f) = \sum_{k=0}^n \hat{f}(k)z^k$ or partial Fourier sums $s_n(f) = \sum_{k=-n}^n \hat{f}(k)e^{ikt}$. If this first direct approach fails, we exploit several well-developed summation methods, e.g., Abel, Borel, Cesàro, Hausdorff, Hölder, Lindelöf, Nörlund, etc., to come up with an appropriate combination of the partial sums which converges to the original function. More explicitly, we consider the weighted sums $\sigma_n(f) = \sum_{k=0}^n w_{nk}\hat{f}(k)z^k$ or $\sigma_n(f) = \sum_{k=-n}^n w_{nk}\hat{f}(k)e^{ikt}$.

While, in many cases, this procedure is a success story, we may naturally wonder if for each space an appropriate summability method via partial Taylor or Fourier sums can be always designed. We show that, unfortunately, this is not always feasible. We construct a Hilbert space \mathcal{H} of analytic functions on \mathbb{D} with the following properties: 1) analytic polynomials are dense in \mathcal{H} , 2) odd polynomials are not dense in the subspace of odd functions in \mathcal{H} . Hence, in particular, there is an $f \in \mathcal{H}$ such that with no lower-triangular summability method one can recover f from its partial Taylor sums $s_n(f)$. At the same token, there is a Hilbert space of integrable functions on \mathbb{T} such that 1) trigonometric polynomials are dense in \mathcal{H} , 2) odd trigonometric polynomials are not dense in the subspace of odd functions in \mathcal{H} . Hence, as an outcome, there is a signal $f \in \mathcal{H}$ such that with no lower-triangular summability method, we can recover f from its partial Fourier sums $s_n(f)$.

20. Qi Ye: Machine Learning in Banach Spaces: A Black-box or White-box Method?

In this talk, we study the whole theory of regularized learning for generalized data in Banach spaces including representer theorems, approximation theorems, and convergence theorems. Specially, we combine the data-driven and model-driven methods to study the new algorithms and theorems of the regularized learning. Usually the data-driven and model-driven methods are used to analyze the black-box and white-box models, respectively. With the same thought of the Tai Chi diagram, we use the discrete local information of the black-box and white-box models to construct the global approximate solutions by the regularized learning. Our original ideas are inspired by the eastern philosophy such as the golden mean. The work of the regularized learning for generalized data provides another road to study the algorithms of machine learning including

- the interpretability in approximation theory,
- the nonconvexity and nonsmoothness in optimization theory,
- the generalization and overfitting in regularization theory

Moreover, based on the theory of the regularized learning, we will construct the composite algorithms combining support vector machines, artificial neural networks, and decision trees for our current research projects of the big data analytics in education and medicine.

21. Felipe Goncalves: Bandlimited extremal functions in higher dimensions.

We will talk about some of the challenging obstructions in constructing higher dimensional bandlimited functions with sign constraints in physical space and support constraints in frequency space. We will give several applications of such “magic” functions in several contexts.

22. Aleh (Oleg) Asipchuk: Construction of exponential Riesz bases on split intervals.

Let Ω be a union of intervals of total length L . It is well known that exponential bases exist on $L^2(\Omega)$ but explicit expressions for such bases are only known in special cases. In this work, we construct exponential Riesz bases on $L^2(\Omega)$ with some mild assumptions on the gaps between the intervals. We also generalize Kadec’s stability theorem in some special and significant cases.

23. Tino Ullrich: Constructive sparsification of finite frames with application in optimal function recovery.

We present a new constructive subsampling technique for finite frames to extract (almost) minimal plain (non-weighted) subsystems which preserve a good lower frame bound. The technique is based on a greedy type selection of frame elements to positively influence the spectrum of rank one updates of a matrix. It is a modification of the 2009 algorithm by Batson, Spielman, Srivastava and produces an optimal size subsystem (up to a prescribed oversampling factor) without additional weights. It moreover achieves this in polynomial time and avoids the Weaver subsampling (based on the Kadison-Singer theorem) which has been applied in earlier work, yielding rather bad oversampling constants. In the second part of the talk we give applications for multivariate function recovery. Here we consider the particular problem of L_2 and L_∞ recovery from sample values. In this context, the presented subsampling technique allows to determine optimal (in cardinality) node sets even suitable for plain least squares recovery. It can be applied, for instance, to reconstruct functions in dominating mixed-smoothness Sobolev spaces, where we are able to discretize trigonometric polynomials with frequencies from a hyperbolic cross with nodes coming from an implementable subsampling procedure. In addition we may apply this to subspaces coming from hyperbolic cross wavelet subspaces. Numerical experiments illustrate the theoretical findings. Joint work with: Felix Bartel (Chemnitz), Martin Schaefer (Chemnitz)

24. Laura De Carli: Weaving Riesz bases, and piecewise weighted frames.

This talk consists of two parts loosely connected to one another. In the first part we discuss the properties of a family of Riesz bases on a separable Hilbert space obtained in the following way: For every Ω we let $\{e_n\}$ where $\{e_n\}$ is a Riesz basis of $L^2(\Omega)$ and $\{f_n\}$ is a set of unit vectors. We find necessary and sufficient conditions that ensure that $\{e_n\}$ and $\{f_n\}$ are Riesz bases, and we apply our results to the construction of exponential bases on domains of Ω . In the second part of the talk we present results on weighted Riesz bases and frames in finite or infinite-dimensional Hilbert spaces, with piecewise constant weights. We use our results to construct tight frames in finite-dimensional Hilbert spaces.

25. Kristina Oganessian: Hardy-Littlewood theorem in two dimensions.

We prove the Hardy-Littlewood theorem in two dimensions for functions whose Fourier coefficients obey general monotonicity conditions and, importantly, are not necessarily positive. The sharpness of the result is given by a counterexample, which shows that if one slightly extends the considered class of coefficients, the Hardy-Littlewood relation fails.

26. Yeli Niu: Jackson inequality on the unit sphere \mathbb{S}^d with dimension-free constant.

This is joint work with Feng Dai. Let ρ_n denote the rate of approximation by spherical polynomials of degree at most n in the L_p -metric on the d -dimensional unit sphere \mathbb{S}^d . Let ω_n denote the n -th order modulus of smoothness on the sphere introduced by Zeev Ditzian using the group of rotations. We prove the following Jackson inequality on the sphere \mathbb{S}^d : for each positive integer n and every $p \geq 1$, with the constant depending only on p . The key point here is that the constant is independent of the dimension d . The Jackson inequality on the sphere was previously established by Zeev Ditzian with the constant depending on d and n but going to exponentially fast as $n \rightarrow \infty$. For $p=1$ and the first order moduli of smoothness (i.e., ω_n), the Jackson inequality with dimension-free constant on the sphere was established by D. J. Newman and H. S. Shapiro in 1964, who also pointed out that their proof didn't work for higher-order moduli of smoothness.

27. Gustavo Garrigos: Recent results in Weak Chebyshev Greedy Algorithms.

The Weak Chebyshev Greedy Algorithm (WCGA) is a generalization to Banach spaces of the popular Orthogonal Matching Pursuit. For the latter, a fundamental theorem by T. Zhang establishes the optimal recovery of N -sparse signals after $O(N)$ iterations, under suitable RIP conditions in the dictionary. For the WCGA, however, several questions remain to be investigated. In 2014, Temlyakov proved a deep theorem establishing Lebesgue-type inequalities for the WCGA, which guarantee stable recovery after $O(N)$ iterations, with the exponent depending on the geometry of the Banach space, via the power-type of its modulus of smoothness, as well as on properties of the dictionary (which generalize RIP). In this talk we present recent work on Lebesgue-type inequalities for the WCGA, which extend the above theorem of Temlyakov. We obtain a new bound for the number of iterations in terms of the modulus of convexity of the dual space and similar properties of the dictionary, where this time the parameters are no longer necessarily of power type. In particular, when applied to the spaces ℓ_p , with $1 < p < \infty$, we show that it suffices with $O(N)$ iterations.

28. Han Feng: Generalization Analysis of deep neural networks for Classification.

Deep learning based on a neural networks is extremely efficient in solving classification problems in speech recognition, computer vision, and many other fields. But there is no enough theoretical understanding about this topic, especially the generalization ability of induced optimization algorithms. In this talk, we shall present the mathematical framework for binary classification problems. For target functions associated with a convex loss function, we provide rates of L_p -approximation and then present generalization bounds and learning rates for the excess misclassification error of the deep neural networks classification algorithm. Our analysis is based on efficient integral discretization and other tools from approximation theory.