Mathematisches Forschungsinstitut Oberwolfach

Report No. 17/2017

DOI: 10.4171/OWR/2017/17

Multiscale and High-Dimensional Problems

Organised by
Albert Cohen, Paris
Wolfgang Dahmen, Aachen
Ronald A. DeVore, College Station
Angela Kunoth, Köln

26 March – 1 April 2017

Abstract. High-dimensional problems appear naturally in various scientific areas. Two primary examples are PDEs describing complex processes in computational chemistry and physics, and stochastic/ parameter-dependent PDEs arising in uncertainty quantification and optimal control. Other highly visible examples are big data analysis including regression and classification which typically encounters high-dimensional data as input and/or output.

High dimensional problems cannot be solved by traditional numerical techniques, because of the so-called *curse of dimensionality*. Rather, they require the development of novel theoretical and computational approaches to make them tractable and to capture fine resolutions and relevant features. Paradoxically, increasing computational power may even serve to heighten this demand, since the wealth of new computational data itself becomes a major obstruction. Extracting essential information from complex structures and developing rigorous models to quantify the quality of information in a high dimensional setting constitute challenging tasks from both theoretical and numerical perspective.

The last decade has seen the emergence of several new computational methodologies which address the obstacles to solving high dimensional problems. These include adaptive methods based on mesh refinement or sparsity, random forests, model reduction, compressed sensing, sparse grid and hyperbolic wavelet approximations, and various new tensor structures. Their common features are the nonlinearity of the solution method that prioritize variables and separate solution characteristics living on different scales. These methods have already drastically advanced the frontiers of computability for certain problem classes.

This workshop proposed to deepen the understanding of the underlying mathematical concepts that drive this new evolution of computational methods and to promote the exchange of ideas emerging in various disciplines about how to treat multiscale and high-dimensional problems.

Mathematics Subject Classification (2010): 65 (Numerical Analysis), 41 (Approximations and Expansions).

Introduction by the Organisers

Complex scientific models like climate models, turbulence, fluid structure interaction, nanosciences and reliability control, demand finer and finer resolution in order to increase their reliability. This demand is not simply solved by increasing computational power. Indeed, higher computability even contributes to the problem by generating wealthy data sets for which efficient organization principles are not available. Extracting essential information from complex structures and developing rigorous models for quantifying the quality of information is an increasingly important issue. These tasks become even more demanding when the problem is high dimensional, in the sense that it involves a large number of variables.

Inherently high-dimensional problems appear naturally in various scientific disciplines. Prominent examples of such problems are: (i) PDEs that describe complex processes in computational chemistry and physics, such as the Fokker-Planck and the Schrödinger equations, (ii) stochastic or parameter-dependent PDEs used in simulation and optimal control and design, (iii) classification and regression problems arising in big-data analysis with large number of input/output variables. While significant advances have been made in "forward problems" trying to exploit sparsity in effectively recovering high-dimensional functions, corresponding inverse problems like state- or parameter estimation pose even greater challenges. One reason is that one usually has to cope with a strong undersampling - a small-data problem -, due to prohibitive cost or severe obstructions to acquiring observation data. An important issue is to properly formulate corresponding *data-assimilation* frameworks and to understand the role of model reduction and sparse recovery in this context.

The mathematical methods emerging to address these problems try to exploit in a subtle way the structure of the problem in order to extract the necessary information. They have several common features including the determination of whether the underlying objects have a sufficiently small information content to be computationally tractable, and how this content might be accessible through certain sparse representations. The numerical methods themselves are typically highly nonlinear with the ability of separating solution characteristics living on different length scales. Having to deal with the appearance and interaction of local features at different levels of resolution has, for instance, brought about spatially adaptive methods as a key methodology that has advanced the frontiers of computability for certain problem classes in numerical analysis. The current state of signal processing, learning theory, and numerical computation can be viewed as an evolution from the introduction of multiscale and adaptive methods to the current high dimensional methods based on concepts such as sparsity, anisotropy, model reduction, low-rank tensor methods, or random projections.

Multiscale techniques, such as wavelet decompositions, were introduced to manage the interaction of different length scales. In the very spirit of harmonic analysis they allow one to decompose complex objects into simple building blocks that again support analyzing multiscale features. Our first Oberwolfach Workshops "Wavelet and Multiscale Methods" held in July 2004, August 2007 served to bring together the main developers of multiscale decompositions for signal processing with those using these techniques for numerical methods for PDEs and thus contributed to the growth of both of these disciplines. While multiscale techniques were first exploited primarily for treating *explicitly* given objects, like digital signals and images or data sets, the use of such concepts proved important for recovering also *implicitly* given objects, like solutions of partial differential or boundary integral equations, as well. The close marriage of discretization, analysis and the solution process based on *adaptive* wavelet methods has led to significant theoretical advances as well as new algorithmic paradigms for linear and nonlinear stationary variational problems. Through thresholding, best N-term approximation, and adaptivity, multiscale techniques from nonlinear approximation theory and harmonic analysis become practically manageable. They now are a major component of modern signal processing and modern numerical computation.

Our last two workshops in August 2010 and "Multiscale and High-Dimensional Problems held in July/August 2013 recognized the increasing demand on finding numerical techniques which apply to high dimensional problems. They brought together various disciplines where such problems are encountered. Those workshops not only accelerated the advancement of nonlinear and multiscale methodologies but also provided beneficial cross-fertilizations between the various areas represented in the workshop, see the Oberwolfach Reports 34/2004, 36/2007, 33/2010, 39/2013. Among the several recognizable outcomes of the workshops were: (i) the emergence of compressed sensing as an exciting alternative to the traditional sensing-compression paradigm, (ii) fast online computational algorithms based on adaptive partition for mathematical learning, (iii) clarification of the role of coarsening in adaptive numerical methods for PDEs, (iv) injection of the notion of sparsity into stochastic models to identify computational paradigms that are more efficient than Monte Carlo techniques, (v) a coherent theory to explain why techniques like sparse representation and reduced modeling work and how they can be improved.

The current workshop *Multiscale and High-Dimensional Problems* organized by Albert Cohen (Université Paris VI), Wolfgang Dahmen (RWTH Aachen), Ronald A. DeVore (Texas A&M University) and Angela Kunoth (Universität zu Köln) was once again directed at multi-scale and high dimensional problems incorporating the new emerging aspects mentioned above. It focussed on the interaction of scientists from different disciplines and thereby result in more rapid developments of new methodologies in these various domains. It was also a bridge from theoretical foundations to applications, such as mechanical engineering, mathematical biology, quantum chemistry, signal and image processing, complex fluid flows. Examples of conceptual issues that were addressed in our workshop were:

- adaptive and nonlinear multilevel methods for high-dimensional PDEs, for parametric PDEs and PDEs with stochastic data;
- multilevel and high-dimensional meshless methods;
- incorporating anisotropy in analysis, estimation, compression and encoding;
- interaction of different scales and variables under relevant linear and nonlinear mappings;
- convergence theory and analysis for model reduction and low-rank methods;
- numerical aspects of compressed sensing;
- design and analysis of estimators in high dimensional machine learning;
- solution concepts for problems of high spatial dimension utilizing anisotropy;
- data assimilation and inversion concepts in high dimensional settings;
- tensor structures and tensor sparsity for high dimensional approximation problems.

In summary, the conceptual similarities that occurred in a variety of application domains suggested that a wealth of synergies and cross–fertilization could be exploited. These concepts are in our opinion not only relevant for the development of efficient solution methods for large scale and inherently high-dimensional problems but also for the formulation of rigorous mathematical models for quantifying the extraction of essential information from complex objects in many dimensions.

As in the previous workshops, the participants were experts in areas like nonlinear approximation theory, statistical learning theory, compressed sensing, tensor approximations hyperbolic cross approximation, finite elements, multigrid and spectral methods, harmonic analysis and wavelets, numerical fluid mechanics conservation laws, inverse problems stochastic PDEs, PDE-constrained control problems, and model reduction.

Acknowledgement:

The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, "US Junior Oberwolfach Fellows". Moreover, the MFO and the workshop organizers would like to thank the Simons Foundation for supporting Peter Binev in the "Simons Visiting Professors" program at the MFO.

Workshop: Multiscale and High-Dimensional Problems

Table of Contents

Abstracts

Polynomial approximation of smooth, multivariate functions on irregular domains

Ben Adcock

(joint work with Juan Manuel Cárdenas, Daan Huybrechs and Sebastián Scheuermann)

Many problems in scientific computing call for the approximation of a smooth function $f(\boldsymbol{y})$ defined on a d-dimensional tensor-product domain. A popular approach for doing this involves expanding f in an orthonormal basis of polynomials. Recent results have shown that this approximation can be computed stably and efficiently via discrete least-squares fitting, provided the samples are chosen randomly according to an appropriate measure. Moreover, it is often possible to achieve sample complexity estimates that are independent of the dimension d, and depending only polynomially on the dimension of the polynomial space.

However, in many practical problems $f(\boldsymbol{y})$ is either not defined over a tensor-product domain or can only be sampled over a subset of such a domain. In both cases, one effectively faces the challenge of approximating a smooth function over an irregular domain Ω. Such a domain may be nonconvex, non-simply connected, or not even explicitly defined.

In this work, we consider a method for approximating of smooth functions on irregular domains, known as *polynomial frame approximations*. This method proceeds as follows. First, the irregular domain Ω is embedded in a tensor-product domain D – for example, the unit cube $(-1,1)^d$ if Ω is compact. Second, an orthogonal polynomial basis is chosen on D. Third, M samples $f(\boldsymbol{y}_1),\dots,f(\boldsymbol{y}_M)$ are taken in the domain Ω, and finally, an approximation is computed in the polynomial system from these samples. This work primarily considers discrete least-squares approximation in a polynomial space of dimension $N \leq M$. However, other approaches may be possible, including interpolation ($N = M$) or compressed sensing ($N > M$).

The approach outlined above is extremely simple – in essence, all it requires is the knowledge of a bounding box D and an algorithm for determining whether or not a point $\boldsymbol{y} \in D$ also belongs to Ω – and can be applied to virtually any domain. However, its numerical implementation presents an immediate challenge. In transpires that the matrix of the algebraic least-squares problem is typically highly ill-conditioned. This is due to the fact that the orthonormal polynomial system is no longer a basis when restricted to Ω, but rather a frame. Frames are redundant systems of functions that span a Hilbert space, and this redundancy translates into finite ill-conditioning of the least-squares matrix. To illustrate how bad such ill-conditioning can be, we note in passing that if a hyperbolic cross polynomial space $\mathcal{P}_n^{\mathrm{HC}}$ of degree n is used, then the condition numbers grow faster than any power of n as $n \to \infty$.

Fortunately, the structure endowed on the approximation by the frame property means that effective regularization of the least-squares problem is possible. It is straightforward to show that polynomials corresponding to small singular values of the least-squares matrix are necessarily concentrated on $D\backslash\Omega$, meaning that their overall contribution to the polynomial approximation in Ω is minimal. This suggests a valid approach for regularization of the algebraic least-squares problem: namely, singular value thresholding.

Let $\Upsilon = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M\}$ be the set of $M \geq N$ points at which f is sampled, $\mathcal{P}$ be the N-dimensional polynomial space in which the approximation is computed, $\epsilon > 0$ be the SVD truncation parameter below which all singular values are discarded and $\tilde{f}$ be the approximation to f from $\mathcal{P}$ computed via the SVD-regularized least-squares fit. A result proved in [1] gives that

$$\|f - \tilde{f}\|_{L^2(\Omega)} \leq C(\Upsilon, \mathcal{P}, \epsilon) E_{\mathcal{P},\epsilon}(f),$$

where $C(\Upsilon, \mathcal{P}, \epsilon)$ is a constant depending on Υ,$\mathcal{P}$ and ϵ and $E_{\mathcal{P},\epsilon}(f)$ is an approximation error term given by

$$E_{\mathcal{P},\epsilon}(f) = \inf\left\{\|p - f\|_{L^\infty(\Omega)} + \epsilon\|p\|_{L^2(D)} : p \in \mathcal{P}\right\}.$$

Moreover, if $\boldsymbol{x}$ is the vector of coefficients of $\tilde{f}$ in the polynomial system then one has the stability result

$$\|\boldsymbol{x}\|_{\ell^2} \leq \frac{E_{\mathcal{P},\epsilon}(f)}{\epsilon}.$$

Up to the constant $C(\Upsilon, \mathcal{P}, \epsilon)$ (discussed below), these estimates imply that the approximation error and stability of the coefficients depend on how well f can be approximated on D by polynomials which do not grow too large on D. A typical result is as follows. Suppose that f is the restriction of a function belonging to the Sobolev space of dominating mixed smoothness $H^k_{\text{mix}}(D)$, and let $\mathcal{P} = \mathcal{P}^{\text{HC}}_n$ be the hyperbolic cross polynomial space. Then

$$E_{\mathcal{P},\epsilon}(f) \leq n^{3/2-k}\|f\|_{H^k_{\text{mix}}(D)} + \epsilon\|f\|_{L^2(D)}. \tag{1}$$

Thus, for smooth f, superalgebraic convergence can be expected down to the tolerance ϵ. Furthermore, the coefficients $\boldsymbol{x}$, although at most $1/\epsilon$ in magnitude, are necessarily order one in the limit $n \to \infty$.

For an accurate approximation, the constant $C(\Upsilon, \mathcal{P}, \epsilon)$ should be small. This depends on the number of sample points M and their distribution. To obtain concrete estimates, one approach involves drawing samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M$ independently according to the uniform measure on the (assumed measurable) set Ω. Using techniques from random matrix theory, one can then show that $C(\Upsilon, \mathcal{P}, \epsilon) \lesssim 1$ with high probability, provided

$$M \gtrsim (\mathrm{N}(\mathcal{P}, \Omega))^2 \log(N).$$

Here $\mathrm{N}(\mathcal{P}, \Omega)$ is the constant in the Nikolskii-type inequality for the polynomial space $\mathcal{P}$ over the domain Ω:

$$\mathrm{N}(\mathcal{P}, \Omega) = \sup\left\{\sqrt{\mathrm{Vol}(\Omega)}\|p\|_{L^\infty(\Omega)} : \|p\|_{L^2(\Omega)} = 1\right\}.$$

The behaviour of $\mathrm{N}(\mathcal{P}, \Omega)$ for general domains Ω and polynomial spaces $\mathcal{P}$ is unknown. However, in some case explicit estimates are available. For example, if Ω is a union of rectangles of volume at least $\lambda \mathrm{Vol}(\Omega)$ (for some $0 < \lambda < 1$) then

$$\mathrm{N}(\mathcal{P}_n^{\mathrm{HC}}, \Omega) \leq N^2/\lambda.$$

Hence for such domains the sample complexity for a hyperbolic cross polynomial approximation is at most quadratic in $N = |\mathcal{P}_n^{\mathrm{HC}}|$. See [1] for further details.

Conclusions and challenges. This work shows that accurate and numerically stable polynomial approximations of smooth functions can be computed over general irregular domains using a simple algorithm. Accuracy and stability are determined by how well f is approximated by polynomials that do not grow too large on the extended domain D, and sample complexity (for random samples drawn from the uniform measure) is determined by the Nikolskii-type inequality for the polynomial space. A key ingredient of these results is the fact that the polynomial system forms a frame over Ω, which implies stability and accuracy of the regularized least-squares approximation.

There are several challenges for future research. First, the bound (1) for the approximation error $E_{\mathcal{P},\epsilon}(f)$ is only valid for functions f which have mixed regularity over the extended domain D. This does not appear to be necessary. In the case of classical Sobolev regularity, it is enough to assume that $f \in H^k(\Omega)$. Standard Sobolev extension theorems can then be used to estimate $E_{\mathcal{P},\epsilon}(f)$. However, such estimates lead to approximation rates depending on $n^{-k/d}$, where d is the dimension (the curse of dimensionality). We suspect a similar extension argument could be used in the case of mixed regularity, but this relies on extension theorems for mixed Sobolev spaces.

Second, the sample complexity estimate is provably quadratic in N for suitable domains. Yet there is numerical evidence to suggest that it may in fact be linear [4]. This remains a conjecture. A related open problem is that of designing better measures from which to sample, with the aim of lowering the sample complexity. Another open problem is the question of proving sample complexity estimates for domains that cannot be written as a union of rectangles with nonvanishing volume; a class which includes triangles, simplices, spheres and many other instances relevant in practice.

Third and finally, this work has considered discrete least-squares fitting only. However, algorithms based on compressed sensing techniques have begun to gain prominence in high-dimensional polynomial approximation on tensor-product domains. The extension of this work to irregular domains is an interesting topic for future research.

References

[1] B. Adcock and D. Huybrechs. Approximating smooth, high-dimensional functions on irregular domains In preparation, 2017.
[2] B. Adcock and D. Huybrechs. Frames and numerical approximation. arXiv:1612.04464, 2016.
[3] B. Adcock and D. Huybrechs. On the resolution power of Fourier extensions for oscillatory functions. J. Comput. Appl. Math. 260:312-336, 2014.
[4] B. Adcock, D. Huybrechs & J. Martín–Vaquero. On the numerical stability of Fourier extensions. Found. Comput. Math. 14(4):635-687, 2014.
[5] B. Adcock & J. Ruan. Parameter selection and numerical approximation properties of Fourier extensions from fixed data. J. Comput. Phys. 273:453-471, 2014.

Submodular optimization: from discrete to continuous domains

Francis Bach

Submodular set-functions have many applications in combinatorial optimization, as they can be minimized and approximately maximized in polynomial time (in the ambient dimension and the inverse of the required precision). A key element in many of the algorithms and analyses is the possibility of extending the submodular set-function to a convex function, which opens up tools from convex optimization. Submodularity goes beyond set-functions and has naturally been considered for problems with multiple labels or for functions defined on continuous domains, where it corresponds essentially to cross second-derivatives being nonpositive, that is, if $H : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then it is submodular if and only if

$$\forall x \in \mathbb{R}^n,\ \forall (i,j) \in \{1,\dots,n\},\ i \neq j \Rightarrow \frac{\partial^2 H}{\partial x_i \partial x_j}(x) \leqslant 0.$$

In this work, we show that most results relating submodularity and convexity for set-functions [1, 2] can be extended to all submodular functions. In particular,

- we naturally define a continuous extension in a set of probability measures, based on a joint thresholding of the inverse cumulative distribution functions,
- we show that the extension is convex if and only if the original function is submodular,
- we prove that the problem of minimizing a submodular function is equivalent to a typically non-smooth convex optimization problem, and (d) propose another convex optimization problem with better computational properties (e.g., a smooth dual problem).

Most of these extensions from the set-function situation are obtained by drawing links with the theory of multi-marginal optimal transport, which provides also a new interpretation of existing results for set-functions. We then provide practical algorithms to minimize generic submodular functions on discrete domains, with associated convergence rates, with an application to computing proximal operators for non-convex regularizers.

Further information can be found in [3].

References

[1] S. Fujishige. *Submodular Functions and Optimization*. Elsevier, 2005.

[2] F. Bach. *Learning with Submodular Functions: A Convex Optimization Perspective*, volume 6 of *Foundations and Trends in Machine Learning*. NOW, 2013.

[3] F. Bach. Submodular functions: from discrete to continuous domains. Technical Report 1511.00394, arXiv, 2015.

Space-parameter-adaptive approximation of affine-parametric elliptic PDEs

Markus Bachmayr

(joint work with Albert Cohen, Wolfgang Dahmen, Dinh Dũng, Giovanni Migliorati, and Christoph Schwab)

We consider the diffusion problem $-\operatorname{div}(a\nabla u) = f$ on $D \subset \mathbb{R}^3$, $d \in \{1,2,3\}$, with $u|_{\partial D} = 0$, where a is an affine function of the parameters $y = (y_j)_{j\geq 1} \in U := [-1,1]^{\mathbb{N}}$, that is,

$$a(y) = \bar{a} + \sum_{j\geq 1} y_j \psi_j. \tag{1}$$

Our aim is to approximate the solution map $y \mapsto u(y)$, where $u(y) \in V := H^1_0(D)$, as an element of the space $\mathcal{V} := L^2(U, V, \mu)$ with μ the uniform measure on U. We assume $\bar{a}, \psi_j \in L^\infty(D)$, $j \geq 1$, to satisfy the *uniform ellipticity condition*

$$\operatorname{ess\,inf} \bar{a} > 0, \quad \Big\| \frac{1}{\bar{a}} \sum_{j\geq 1} |\psi_j| \Big\|_{L^\infty} < 1. \tag{UEA}$$

Then u is characterized by the $\mathcal{V}$-elliptic variational formulation

$$\int_U \int_D a\nabla u \cdot \nabla v \, dx \, d\mu(y) = \int_U \langle f, v\rangle_{V',V} \, d\mu(y), \quad v \in \mathcal{V}. \tag{2}$$

Let $L_\nu(y) := \prod_{j\geq 1} L_{\nu_j}(y_j)$ denote the $L^2(U,\mu)$-orthonormal product Legendre polynomials. With $\mathcal{F}$ denoting the set of compactly supported elements of $\mathbb{N}_0^{\mathbb{N}}$, $(L_\nu)_{\nu\in\mathcal{F}}$ is an orthonormal basis of $L^2(U,\mu)$. The approximations we consider are based on the Legendre expansion $u = \sum_{\nu\in\mathcal{F}} u_\nu L_\nu$, and are obtained by appropriately truncating the summation over ν to finitely many terms and using spatial approximations for the coefficients $u_\nu \in V$.

The main focus in this presentation are parametric expansions (1) corresponding to random fields of low smoothness. In treating such cases, functions ψ_j with multilevel structure turn out to be advantageous. In what follows, we consider the following example of such an expansion: let $\alpha > 0$ and let $\{\psi_\mu\}_{\mu\in\mathcal{P}}$ be wavelet-type basis functions on D scaled such that $\|\psi_\mu\|_{L^\infty} = c2^{-\alpha|\mu|}$ with a $c > 0$, where μ are scale-space indices with levels $|\mu|$. We assume that there are $\mathcal{O}(2^{d\ell})$ functions on level ℓ with finite overlap, that is, with $M > 0$, $\#\{\mu : |\mu| = \ell \text{ and } \psi_\mu(x) \neq 0\} \leq M$ for all x and ℓ. Let $\psi_j := \psi_{\mu_j}$ in (1), ordered by increasing level, so that $\|\psi_j\|_{L^\infty} \lesssim j^{-\alpha/d}$. In addition, let $\psi_\mu \in C^\kappa$ for a $\kappa > \lceil\alpha\rceil$, then $\sup_{y\in U} \|a(y)\|_{C^\alpha} < \infty$ for $\alpha \notin$

$\mathbb{N}$. If f, $\bar{a}$ and D have sufficient regularity, this also implies $\sup_{y\in U}\|u(y)\|_{H^{1+\beta}} < \infty$ for all $\beta < \alpha$, leading to a spatial approximation rate up to $\frac{\alpha}{d}$ in V for each single instance $u(y)$.

For $\alpha \le d$, the result from [5] that $(\|\psi_j\|_{L^\infty})_{j\ge1} \in \ell^p(\mathbb{N})$ with $0 < p < 1$ implies $(\|u_\nu\|_V)_{j\ge1} \in \ell^p(\mathcal{F})$ is not applicable. The following result, which is instead based on *pointwise* summability of $|\psi_j|$, is shown in [1].

Theorem 1 ([1]) *Assume that with $\rho_j > 0$, we have the weighted UEA*

$$(\mathrm{UEA}_\rho) \qquad \operatorname{ess\,inf} \bar{a} > 0, \quad \Big\|\frac{1}{\bar{a}}\sum_{j\ge1}\rho_j|\psi_j|\Big\|_{L^\infty} < 1\,.$$

Then with $\rho^\nu := \prod_{j\ge1}\rho_j^{\nu_j}$,

$$\sum_{\nu\in\mathcal{F}}\big(a_\nu^{-1}\rho^\nu\|u_\nu\|_V\big)^2 < \infty, \quad a_\nu := \prod_{j\ge1}\sqrt{2\nu_j+1}.$$

By Hölder's inequality, from this weighted ℓ^2-summability one can deduce ℓ^p-summability.

Corollary 2 ([1]) *Let $0 < p < 2$ and $0 < q < \infty$ be such that $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$. Moreover, let $\rho_j > 1$ and $(\rho_j^{-1})_{j\ge1} \in \ell^q(\mathbb{N})$. Then $(\|u_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$.*

In the above multiscale example, we can take $\rho_j \sim 2^{\beta|\mu_j|}$ for any $0 < \beta < \alpha$, that is, $(\rho_j^{-1})_{j\ge1} \in \ell^q(\mathbb{N})$ for any $q > d/\alpha$. One thus arrives at

$$\Big\|u - \sum_{\nu\in\Lambda_n} u_\nu\, L_\nu\Big\|_{\mathcal{V}} \le Cn^{-s}, \quad \text{for any } s = \frac{1}{p} - \frac{1}{2} < \frac{1}{q} < \frac{\alpha}{d}.$$

A first option for combining this with a spatial discretization is to fix a hierarchy of finite-dimensional subspaces $\{0\} = V_0 \subset V_1 \subset \cdots \subset V$, and to define for $\mathbf{n} = (n_\nu)_{\nu\in\mathcal{F}} \in \mathcal{F}$ the approximation $u_{\mathbf{n}} := \sum_{\nu\in\mathcal{F}} u_{\nu,n_\nu}\, L_\nu$ with $u_{\nu,n_\nu} \in V_{n_\nu}$. The convergence of $u_{\mathbf{n}}$ with respect to the total number of degrees of freedom $N(\mathbf{n}) := \sum_{\nu\in\mathcal{F}} n_\nu$ can be estimated based on summability of the higher-order norms $\|u_\nu\|_W := \|\Delta u_\nu\|_{L^2}$. With appropriately refined nested spaces V_n with $\dim(V_n) \sim n$, one has the approximation property

$$\min_{v_n\in V_n}\|v - v_n\|_V \lesssim n^{-\frac{1}{d}}\|v\|_W. \tag{3}$$

If $W = H^2(D)$ by elliptic regularity (e.g. when D is convex), uniformly refined finite element meshes suffice. The following summability result for W-norms is shown in [3].

Theorem 3 ([3]) *Assume that (UEA) holds, that $f \in L^2(D)$ and that $\bar{a}, \psi_j \in W^{1,\infty}(D)$, $j \ge 1$. If there exists a positive sequence $\rho = (\rho_j)_{j\ge1}$ such that*

$$\Big\|\frac{1}{\bar{a}}\sum_{j\ge1}\rho_j|\psi_j|\Big\|_{L^\infty} < 1, \quad \Big\|\sum_{j\ge1}\rho_j|\nabla\psi_j|\Big\|_{L^\infty} < \infty,$$

then $\sum_{\nu\in\mathcal{F}}\big(a_\nu^{-1}\rho^\nu\|u_\nu\|_W\big)^2 < \infty$.

Provided that one can choose $\rho_j > 1$ with $(\rho_j^{-1})_{j\geq 1}$ having ℓ^q-summability for some $q > 0$, as in Corollary 2 one can show ℓ^p-summability of $(\|u_\nu\|_W)_{\nu\in\mathcal{F}}$ with a $p < 2$. However, this is not possible in our example when $0 < \alpha \leq 1$, since we only have $\|\nabla\psi_j\|_{L^\infty} \lesssim 2^{-(\alpha-1)|\mu_j|}$. This issue can be circumvented by the following interpolation argument given in [3]: Let $0 < \theta < 1$, $Z = [V, W]_\theta$, and

$$\sum_{\nu\in\mathcal{F}} (a_\nu^{-1}\hat{\rho}^\nu\|u_\nu\|_V)^2 < \infty \quad \text{and} \quad \sum_{\nu\in\mathcal{F}} (a_\nu^{-1}\rho^\nu\|u_\nu\|_W)^2 < \infty.$$

Then with $\tilde{\rho}_j = \hat{\rho}_j^{1-\theta}\rho_j^\theta$, one also has $\sum_{\nu\in\mathcal{F}}(a_\nu^{-1}\tilde{\rho}^\nu\|u_\nu\|_Z)^2 < \infty$. In our example with $\alpha \leq 1$ and $0 < \theta < \beta < \alpha$, this can be applied with $\hat{\rho}_j \sim 2^{\beta|\mu_j|} \uparrow \infty$, $\rho_j \sim 2^{(\beta-1)|\mu_j|} \downarrow 0$ to obtain weighted summability with $\tilde{\rho}_j \sim 2^{(\beta-\theta)|\mu_j|} \uparrow \infty$. Optimizing $\mathbf{n} = (n_\nu)_{\nu\in\mathcal{F}}$ as detailed in [3], one subsequently arrives at the following result in our example: there exist $\mathbf{n}$ with $N = N(\mathbf{n}) \to \infty$ such that

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \lesssim N^{-s}, \quad \text{for any } s < \frac{\alpha}{2d}.$$

As we outline next, adaptive approximation of each u_ν can yield better convergence. We focus on spatial wavelet approximations: With $(\Psi_\lambda)_{\lambda\in\mathcal{S}}$ a wavelet Riesz basis of V, $(\Psi_\lambda \otimes L_\nu)_{(\lambda,\nu)\in\mathcal{S}\times\mathcal{F}}$ is a Riesz basis of $\mathcal{V}$. Based on the expansion $u = \sum_{(\lambda,\nu)\in\mathcal{S}\times\mathcal{F}} \mathbf{u}_{\lambda,\nu}\Psi_\lambda \otimes L_\nu$, we consider the best N-term approximations $u_{[N]}$ defined by retaining the indices (λ, ν) of the N largest $|\mathbf{u}_{\lambda,\nu}|$.

Corresponding approximation results are shown in [3] via summability of the norms $\|u_\nu\|_{B^\tau} := \|u_\nu\|_V + \|\Delta u_\nu\|_{L^\tau}$, $1 < \tau < 2$. We now assume that D is convex, so that elliptic regularity ensures $B^\tau = W^{2,\tau}(D)$. Under essentially the same conditions, we obtain stronger summability of these weaker norms.

Theorem 4 ([3]) *Let $1 \leq \tau < 2$. Assume $f \in H^{-1}(D) \cap L^\tau(D)$, that (UEA) holds, and that $\bar{a}, \psi_j \in W^{1,\infty}(D)$, $j \geq 1$. If there exist positive sequences $(\rho_j)_{j\geq 1}$, $(\hat{\rho}_j)_{j\geq 1}$ such that $\hat{\rho}_j/\rho_j > 1$ and $(\rho_j/\hat{\rho}_j)_{j\geq 1} \in \ell^q(\mathbb{N})$ with $\frac{1}{2} + \frac{1}{q} = \frac{1}{\tau}$ and*

$$\Big\|\frac{1}{\bar{a}}\sum_{j\geq 1}\hat{\rho}_j|\psi_j|\Big\|_{L^\infty} < 1, \quad \Big\|\sum_{j\geq 1}\rho_j|\nabla\psi_j|\Big\|_{L^\infty} < \infty,$$

then $\sum_{\nu\in\mathcal{F}}\big(a_\nu^{-1}\rho^\nu\|u_\nu\|_{B^\tau}\big)^\tau < \infty$.

For $\alpha \leq 1$, this can again be combined with an interpolation argument: Let $0 < \theta < 1$, $Z := [V, B^\tau]_\theta$. Then under the assumptions of Theorem 4, with $\tilde{\rho}_j := \hat{\rho}_j^{1-\theta}\rho_j^\theta$,

$$\sum_{\nu\in\mathcal{F}}\big(a_\nu^{-1}\tilde{\rho}^\nu\|u_\nu\|_Z\big)^\zeta < \infty, \qquad \frac{1}{\zeta} = \frac{1}{2} + \Big(\frac{1}{\tau} - \frac{1}{2}\Big)\theta.$$

For elements of Z, analogously to (3), best n-term approximations with respect to $(\Psi_\lambda)_{\lambda\in\mathcal{S}}$ converge with rate θ/d. Choosing weights as above, as shown in [3], for

$0 < \alpha \leq 1$ in our example one arrives at

$$\|u - u_{[N]}\|_{\mathcal{V}} \lesssim N^{-s}, \quad \text{for any } s < \begin{cases} \frac{2}{3}\alpha, & d = 1, \\ \frac{\alpha}{d} & d = 2, 3. \end{cases} \tag{4}$$

These results are numerically observed to be sharp.

Moreover, it is shown in [2] that the representation of the operator in (2) with respect to the basis $(\Psi_\lambda \otimes L_\nu)_{(\lambda,\nu)\in\mathcal{S}\times\mathcal{F}}$ is s^*-compressible (cf. [4]) with $s^* = \frac{1}{1+d/(2\gamma)}\frac{\alpha}{d}$, provided that $\psi_j \nabla\Psi_\lambda \in H^\gamma$ and Ψ_λ has $\lceil \gamma - 1 \rceil$ vanishing moments. Thus, with γ sufficiently large, approximations with convergence as in (4) can be computed by standard adaptive algorithms [4] using a number of operations that scales almost linearly with respect to N.

References

[1] M. Bachmayr, A. Cohen, G. Migliorati, Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients, ESAIM: M2AN 51(1), pp 321–339, 2017.

[2] M. Bachmayr, A. Cohen, W. Dahmen, Parametric PDEs: sparse or low-rank approximations?, arXiv:1607.04444

[3] M. Bachmayr, A. Cohen, Dinh D., and Ch. Schwab, Fully discrete approximation of parametric and stochastic elliptic PDEs, arXiv:1702.03671

[4] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods for elliptic operator equations: Convergence rates, Mathematics of Computation 70(233), pp 27–75, 2001.

[5] A. Cohen, R. DeVore and Ch. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic PDEs, Anal. and Appl. 9, 11-47, 2011.

Sampling and geometric optimization in high dimensions

Chandrajit Bajaj

Geometric optimization is the computational reduction technique of choice for a wide variety of model selection, ranking and assembly prediction problems. Moreover, optimization occurs naturally for solutions to rigid and flexible geometric shape similarity, complementarity matching problems (e.g. predicting multi-component assemblies, disaster reconstructions etc) . The optimization functional is often a multi-dimensional correlation integral while the search space is the product of transformations groups with dimension growth exponential in the number of movable components (e.g. $O(3^n)$ for an n-residue torsionally flexible molecule)[1]. In this talk, I present a solution framework for geometric optimization methods that combat the curse of high dimensionality, and also achieve adequate trade-offs between speed and accuracy. Fast approximate estimations to the geometric similarity or complementarity matching optimization problems take advantageof a new scheme of generating low-discrepancy samplings of the n-product configuration spaces[2], as well as the speedup utilization of approximate non-uniform fast Fourier transforms[1]. As applications of this to geometric optimization based prediction of multi-component assemblies, I present a general approximation algorithm for non-convex objective functions, and then describe provably polynomial

time approximation scheme (PTAS). The latter is by reduction to Semi-Definite Programming (SDP) for the special cases of rigid multi-piece 3D assemblies , and symmetric 3D spherical shell assemblies, given a constant number of primitive component molecules that make up the asymmetric unit. This prediction optimization is based on a method for generation of congruently tiled spherical arrangements using a new generative class of polyhedra[3].

Further information can be found in these publications [1, 2, 3].

References

[1] C. Bajaj, B. Bauer, R.K. Bettadapura, A. Vollrath Nonuniform Fourier Transforms for Rigid-Body and Multi-Dimensional Rotational Correlations SIAM Journal of Scientific Computing, vol 35 (4), B821-B845, 2013

[2] C. Bajaj, A. Bhowmick, E. Chattopadhyay and D. Zuckerman On Low Discrepancy Samplings in Product Spaces of Motion Groups arXiv:1411.7753v1, 2014

[3] M. Rasheed, C. Bajaj Highly Symmetric and Congruently Tiled Meshes for Shells and Domes Procedia Engineering, vol 124, p213-225, 2015

Diffusion coefficients estimation revisited: Stability and open problems

Andrea Bonito

(joint work with Albert Cohen, Ronald A. DeVore, Guergana Petrova, Gerrit Welper)

We consider the Dirichlet problem

$$-\mathrm{div}(a\nabla u_a) = f \quad \text{on} \quad D, \qquad u_a = 0 \quad \text{on } \partial D,$$

for a Lipschitz domain $D \subset \mathbb{R}^d$, $d \geq 1$. Here f belongs to L_∞ and $a \in \mathcal{A}$, where for $0 < \lambda \leq \Lambda$

$$\mathcal{A} := \{b \in L_\infty(D) \ : \ \lambda \leq b \leq \Lambda\}.$$

We propose to revisit the question of parameter estimation or identifiability in this context. This is to establish conditions under which, for a given f, the diffusion coefficient a is uniquely determined from the solution u_a to the above elliptic equation. We are also interested in the question of stable recovery, i.e. whether when u_b is close to u_a, say in the $H^1_0(D)$ norm, then b is close to a in some appropriate $L_p(D)$ norm.

Before describing our main result, a few comments are in order. We first note that we have at our disposal a single right hand side f (and therefore a single solution u_a) for the recovery of a. This is in contrast with the celebrated Calderón problem [3], where only boundary data are considered but the full knowledge of the Dirichlet to Neumann map is assumed. Along these lines, if instead we assume the full knowledge of the solution operator $S_a : H^{-1}(D) \to H^1_0(D)$, where $S_a f := u_a(f)$, then we have

$$\|S_a - S_b\|_{H^{-1}(D)\to H^1_0(D)} \sim \|a - b\|_{L_\infty(D)}.$$

Another important point, is that the diffusion coefficient a cannot be recovered if ∇u_a vanishes on a set of non-zero measure. However, in that case f must vanishes

as well. Therefore, in order to avoid this degeneracy, we assume that f is strictly positive.

Our main result reverses the forward estimate in [2] and guarantees the existence of a constant C only depending on D, d, θ, λ and Λ such that

$$\|a-b\|_{L_2(D)} \leq C\sqrt{1+(\|a\|_{H^s(D)}+\|b\|_{H^s(D)})^{\frac{1}{3s}}}\,\|u_a-u_b\|_{H_0^1}^{\frac{1}{6}-\frac{1-s}{6s\theta}},$$

provided $a, b \in \mathcal{A} \cap H^s(D)$ with $\frac{p}{2(p-1)} < s \leq 1$. We refer to [1] for more details and for improved stability estimates when additional smoothness on a, b and the domain D is assumed.

We conclude the talk by discussing open problems including:

- Are the values found for α sharp? When $d = 1$, the explicit expression of the solution u_a shows that $\alpha = \frac{1}{3}$ and cannot be improved. When a, b are piecewise constant in any dimension, we have $\alpha = 1$ but at the expense of having a stability constant C depending linearly on the number of pieces. The sharpness of α in generic cases is unknown.
- Is it possible to remove the positivity assumption on f by allowing for multiple right hand sides?
- Is it possible to recover (again using multiple right hand sides) tensor coefficients?

References

[1] A. Bonito, A. Cohen, R. DeVore, and G. Welper. Diffusion coefficients estimation for elliptic equations. SIAM J. Math. Anal., to appear.
[2] A. Bonito, R. DeVore, and R. Nochetto. Adaptive finite element methods for elliptic problems with discontinuous coefficients. SIAM J. Numer. Anal., 51, 2013, 3106–3134.
[3] A.P. Calderón. On an inverse boundary value problem. Seminar on Numerical Analysis and its Applications to Continuum Physics, Soc. Brasileira de Matematica, Rio de Janeiro, 1980, 65–73.

A sparse grid collocation method based on LaVallée Poussin kernel

Moulay Abdellah Chkifa

We present a new approach to polynomial approximation of functions in high dimension. The new scheme is a non-intrusive collocation method based on Smolyack formula applied with a polynomial scheme in one dimension using Fejér and LaVallée Poussin type sums. Smolyack formula allows us to transform a hierarchical collocation strategy in one dimension to a hierarchical collocation strategy in high dimension with possibly a straightforward computational formula and similar stability properties. More precisely, we let $(I_k)_{k\geq 0}$ be a sequence of operators defined each from $\mathcal{C}([-1,1],\mathbb{R})$ into $\mathbb{P}_k := \text{span}\{t^j : j = 0,\dots,k\}$ the space of polynomials of degree at most k and reproducing $\mathbb{P}_k$, i.e

$$I_k[f] \in \mathbb{P}_k, \qquad I_k[Q] = Q, \quad Q \in \mathbb{P}_k, \tag{1}$$

and we introduce the convention $I_{-1} = 0$. For a given set $\Lambda \subset \mathbb{N}^d$, the Smolyack formula is given by

$$I_\Lambda := \sum_{\nu \in \Lambda} \otimes_{j=1}^{d} (I_{\nu_j} - I_{\nu_j}) \tag{2}$$

The operator I_Λ defines a polynomial operator over $\mathcal{C}([-1,1]^d, \mathbb{R})$. In the case where $\Lambda \subset \mathbb{N}^d$ a lower set, i.e. $\nu \in \Lambda$ and $\mu \leq \nu$ implies $\mu \in \Lambda$, where the order relation $\nu := (\nu_1, \ldots, \nu_d) \leq \mu := (\mu_1, \ldots, \mu_d)$ means $\nu_1 \leq \mu_1, \ldots, \nu_d \leq \mu_d$, the operator I_Λ defines an operator from $\mathcal{C}([-1,1]^d, \mathbb{R})$ into

$$\mathbb{P}_\Lambda := \operatorname{span}\{y^\nu := y_1^{\nu_1} \ldots y_d^{\nu_d} : \nu \in \Lambda\}, \tag{3}$$

that reproduce $\mathbb{P}_\Lambda$, i.e.

$$I_\Lambda[f] \in \mathbb{P}_\Lambda, \qquad I_\Lambda[Q] = Q, \quad Q \in \mathbb{P}_\Lambda. \tag{4}$$

In the case where the operators I_k are hierarchical interpolation operators associated with the nested sections $(r_0, \ldots, r_{k+1})$ of a fixed sequence $R := (r_0, r_1, \ldots)$ of mutually distinct abscissas in [-1,1], the operators I_Λ are also interpolation operators. This case is studied in [1].

In this work, we consider operators I_k that are based on Lavallée Poussin sums and discuss the accuracy and stability of the obtained high dimensional scheme with respect to the numerical cost, which is mainly dominated by the number of collocations for problems in uncertainty quantification. The accuracy versus the number of collocations is used as a benchmark for comparing the new scheme to other sheme such as sparse hierarchical interpolation studied in [1] and least squares studied in [2].

References

[1] M.A. Chkifa, A. Cohen and C. Schwab. High-Dimensional Adaptive Sparse Polynomial Interpolation and Applications to Parametric PDEs. FoCM, 14, 2014, 601–633.

[2] M.A. Chkifa, A. Cohen, G. Migliorati, F. Nobile and R. Tempone. Discrete least squares polynomial approximation with random evaluations - application to parametric and stochastic elliptic PDEs ESAIM: M2AN, 49 3, 2015, 815–837.

Phase retrieval in infinite dimensions

Ingrid Daubechies

(joint work with Rima Al-Aifari, Jameson Cahill, Pete Casazza, Gaurav Thakur and Rachel (Rujie) Yin)

In the last 10 to 15 years, mathematical interest in phase retrieval has intensified again. In particular, progress has been made on the question whether and under what conditions a vector $f \in V$ can be recovered from the magnitudes $|\langle f, \varphi_\alpha \rangle|$ of its scalar products with all the elements in a frame $(\varphi_\alpha)_{\alpha \in A}$ for the vector space V. If the space V has finite dimension N, then it is known that the frame should have (approximately, up to an additive constant) at least $2N$ vectors if V is real, or $4N$ vectors of V is a complex vector space. Moreover, it is then also known

that if recovery of f is possible (i.e. if the map is injective), then it is is *stable*, i.e. that there is a positive constant C such that

$$\inf_{|c|=1} \| f - c\, g \|_{\ell_2}^2 \leq C \sum_{\alpha \in A} \left[\, |\langle f, \varphi_\alpha \rangle| - |\langle g, \varphi_\alpha \rangle| \, \right]^2 .$$

In the infinite-dimensional case, the situation is radically different. Although there are situations in which it can be shown that for a frame of functions $(\varphi_\alpha)_{\alpha \in A}$ with sufficient high redundancy, a function f is completely determined by the magnitudes $(|\langle f, \varphi_\alpha \rangle|)_{\alpha \in A}$, stability (in the sense of the inequality above) can *never* be achieved. Constructive counterexamples typically consist of functions that "naturally" decompose into several terms that are quite distinct and that are well-separated. It is then natural to consider a limited phase retrieval, in which each of the well-separated components can be retrieved, even if the relative phase between them cannot be. The talk presented several examples, together with theorems proving that this new, more limited phase retrieval, can be possible in infinite dimensions.

This presentation was based on joint work with Rima Al-Aifari, Jameson Cahill, Pete Casazza, Gaurav Thakur and Rachel (Rujie) Yin.

Computing a quantity of interest from observational data

Simon Foucart

(joint work with Ronald DeVore, Guergana Petrova, Przemyslaw Wojtaszczyk)

Scientific problems often feature observational data received in the form $w_1 = l_1(f), \ldots, w_m = l_m(f)$ of known linear functionals applied to an unknown function f from some Banach space $\mathcal{X}$, and it is required to either approximate f (the full approximation problem) or to estimate a quantity of interest $Q(f)$. In typical examples, the quantities of interest can be the maximum/minimum of f or some averaged quantity such as the integral of f, while the observational data consists of point evaluations. To obtain meaningful results about such problems, it is necessary to possess additional information about f, usually as an assumption that f belongs to a certain model class $\mathcal{K}$ contained in $\mathcal{X}$. This is precisely the framework of optimal recovery, which produced substantial investigations when the model class is a ball of a smoothness space, e.g. when it is a Lipschitz, Sobolev, or Besov class. This presentation is concerned with other model classes described by approximation processes, as studied in [1]. Its main contributions are: (i) for the estimation of quantities of interest, the production of numerically implementable algorithms which are optimal over these model classes, (ii) for the full approximation problem, the construction of linear algorithms which are optimal or near optimal over these model classes in case of data consisting of point evaluations. Regarding (i), when Q is a linear functional, the existence of linear optimal algorithms was established by Smolyak, but the proof was not numerically constructive. In classical recovery settings, it is shown here that such linear optimal algorithms can be produced by constrained minimization methods, and examples involving

the computations of integrals from the given data are examined in greater details. Regarding (ii), it is shown that linearization of optimal algorithms can be achieved for the full approximation problem, too, in the important situation where the l_j are point evaluations and $\mathcal{X}$ is a space of continuous functions equipped with the uniform norm. It is also revealed how the quasi-interpolation theory allows for the construction of linear algorithms which are near optimal.

References

[1] R. DeVore, G. Petrova, and P. Wojtaszcsyk. Data Assimilation in Banach spaces. P. Calcolo, To appear.

Stable Alternating Least Squares Approximation (SALSA)

Lars Grasedyck
(joint work with Sebastian Krämer)

Low rank tensor or matrix completion is a highly ill-posed inverse problem, particularly when the data model is not accurate or has to be chosen adaptively, and some sort of regularization is required in order to solve it. Here, we focus on the calibration of the data model in the sense that the model is vaguely fixed (the TT tensor train low rank format) but some critical parameters have to be adapted — in this case the many different ranks. For alternating optimization, we observe that existing rank adaption methods do not enable a continuous transition between manifolds of different ranks. We denote this flaw as *instability (under truncation)*, as results change dramatically when the (discrete) rank parameter is minimally changed. As a consequence of this flaw, arbitrarily small changes in the singular values of an iterate can have arbitrarily large influence on the further reconstruction. We therefore introduce a singular value based regularization to the standard alternating least squares (ALS), which is motivated by averaging in micro-steps. We prove its *stability* and derive a natural semi-implicit rank adaption strategy. We further prove that the standard ALS micro-steps are only stable on manifolds of fixed ranks, and only around points that have what we define as *internal tensor restricted isometry property iTRIP*. Finally, we provide numerical examples that show improvements of the reconstruction quality up to orders of magnitude in the new Stable ALS Approximation (SALSA) compared to standard ALS.

References

[1] L. Grasedyck, S. Krämer. Stable ALS Approximation in the TT-Format for Rank-Adaptive Tensor Completion. arXiv:1701.08045, 2017.

On foundational computational barriers in high-dimensional regularisation problems

Anders C. Hansen
(joint work with Alex Bastounis and Verner Vlacic)

The use of regularisation techniques in various areas of mathematics has become highly popular over the last decades and has in many cases changed the state of the art from linear techniques to non-linear approaches, typically via convex problems such as Basis Pursuit and Lasso [3, 4, 5, 6, 7, 8, 9, 15, 16]. There are six dominant techniques that involve l^1 and Total Variation (TV) minimisation, namely, one wants to compute

(1)

$$z \in \operatorname*{argmin}_x \mathcal{J}(x) \text{ such that } \|Ax - y\| \leq \delta, \qquad \delta \geq 0, \quad \text{Basis Pursuit (BP)}$$

$$z \in \operatorname*{argmin}_x \|Ax - y\|_2 \text{ such that } \mathcal{J}(x) \leq \tau, \qquad \tau > 0, \quad \text{Constrained Lasso (CL)}$$

$$z \in \operatorname*{argmin}_x \|Ax - y\|_2^2 + \lambda \mathcal{J}(x), \qquad \lambda > 0, \quad \text{Unconstrained Lasso (UL)}$$

where $A \in \mathbb{C}^{m \times N}$, $y \in \mathbb{C}^m$ and

$$\mathcal{J}(x) = \|x\|_1 \text{ or } \mathcal{J}(x) = \|x\|_{\mathrm{TV}}.$$

However, one must ask: can we always find algorithms that can compute accurate approximations to solutions of these problem? Determining the boundaries of what computers can achieve through existence and non-existence statements of algorithms has a rich history in mathematics. Famous cases include Hilbert's question [12] on the existence of algorithms for decision problems that led to Turing's seminal work [17]. Another slightly different question, more related to the issues addressed in this talk, is Smale's question on the existence of purely iterative generally convergent algorithms for polynomial root finding. The latter case is rather delicate as it was solved in the negative by McMullen [13], however, Doyle and McMullen showed a positive result that the quintic could be solved by purely iterative generally convergent algorithms by using several limits [10].

In this talk we establish the boundaries of what computers can achieve in regularisation, as described above, by asking and answering questions on existence of algorithms, similar to Smale's question above. This is done by providing new techniques and developments in the Solvability Complexity Index (SCI) hierarchy [11, 1] which yield results that are universal regardless of the computational model. The crucial problem is to handle inexact input.

Inexact input: The input in A and y may not be exact, but computed approximations. This can be caused in the simplest cases by simply having $\sqrt{2}$, $e^{2\pi i 5}$ or $\cos(3)$ as an input. These numbers can be obtained with arbitrary accuracy, however, they are never exact. Moreover, the frequent use of the discrete Fourier, wavelet or cosine transforms suggest that this is a daily encounter in practice. Moreover, an overwhelming amount of modern software used is based on floating point arithmetic. Thus, even if the input is based on rational numbers, there will

be inexactness due to the floating point representation. This leads to the following basic question.

Question 1 (**Existence of algorithms).** *Given any of the problems in (1), where the input may be given with some inaccuracy controlled by $\hat{\epsilon} > 0$, does there exists an algorithm that can compute an approximate solution, such that, for an arbitrary $\epsilon > 0$, the output will be no further than ϵ away from a true solution? The algorithm can choose $\hat{\epsilon}$ to be as small as desired (as a function of ϵ and the input) to produce the output.*

We are interested in computing a minimiser itself and not the real number value of the minimum the objective function. The latter problem of computing the minimal value is much easier to compute than the former, and there is a very well established literature addressing this question [2, 14].

We are of course assuming that the matrix $A \in \mathbb{C}^{m\times n}$, where $m \leq n$, is nice so that the condition number $\text{cond}(AA^*) \leq M < \infty$. In our case we take $M = 2$. Moreover, the inexactness is always assumed to be very much within the boundaries of feasibility, and can be made as small as one desires.

Arbitrary slowness and unknown error: A negative answer to Question 1 implies that for fixed dimensions and any small accuracy parameter $\epsilon > 0$, one can choose an arbitrary large time T, say $T = 50$ billion years, and find an input such that the algorithm will still after 50 billion years not have reached ϵ accuracy. However, it is impossible to determine when the algorithm should halt to achieve an ϵ accurate solution, and hence the algorithm will never be able to produce an output where one knows that the output is at least ϵ accurate. The largest ϵ for which this failure happens is called the *Breakdown-epsilon*, ϵ_{B}.

1. The main results

Given the great success of the many algorithms applied in different scenarios suggests that the answer to Question 1 must be yes. Yet, we have the following paradox.

Answer 1 (**Existence of algorithms).** *Given any of the problems in* (1)*, the answer to Question 1 is no, even when the input is bounded and well conditioned. This is universal regardless of the model of computation. Moreover, the answer to Question 1 is no for Unconstrained Lasso and Basis Pursuit with noise even when restricting to matrices that satisfy the Robust Nullspace Propert (RNP) (a key assumption in sparse recovery). In fact, a positive answer to Question 1 would imply decidability of well-known undecidable problems.*

Answer 1 seems like a paradox given the success of many algorithms used in practice, and thus raises the following basic question.

Question 2 (**Why do many algorithms work well).** *Given the paradox above, why do many algorithms perform very well in many real-world scenarios?*

To answer Question 2 one first has to examine the consequences of the above paradox and the subsequent questions that follow.

(i) The first consequence is that classes of problems become too big for one algorithm. However, by splitting into subclasses and by using the particular structure of a subclass, one may be able to construct an algorithm for the subclass in the sense of Question 1, however, different subclasses will require different algorithms.
(ii) If one has constructed an algorithm, it will only work (in the sense of Question 1) for certain subclasses of problem, the question is: which subclasses?
(iii) If a class of problems is such that it is impossible to design accurate algorithms for the class (in the sense of Question 1), what is the Breakdown-epsilon? And what is the algorithm realising the basic barrier? It could still be that the Breakdown-epsilon is good enough for the specific purpose of the computation.

Summing up the discussion above we end up with the following list of basic questions.

Question 3 (**Existence of algorithms for subclasses).** (i) *Given any of the problems in (1), which subclasses* Ω *of inputs* A *and* y *will provide a positive answer to Question 1?*
(ii) *Given an algorithm for solving any of the problems in* (1)*, for which subclasses* Ω *of input will the algorithm be accurate in the sense of Question 1?*
(iii) *Which subclasses* Ω *of inputs give negative answers to Question 1, and what is the Breakdown-epsilon?*

This talk addresses these questions, however, we can only scratch the surface, as this is a highly comprehensive classification program.

References

[1] J. Ben-Artzi, A. C. Hansen, O. Nevanlinna, and M. Seidel. Can everything be computed? - On the Solvability Complexity Index and Towers of Algorithms. *arXiv:1508.03280v1*, 2015.
[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
[3] J.-F. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: Total variation, wavelet frames, and beyond. *Journal of the American Mathematical Society*, 25(4):1033–1089, 2012.
[4] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
[5] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
[6] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.
[7] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, Jan. 2001.
[8] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.
[9] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
[10] P. Doyle and C. McMullen. Solving the quintic by iteration. *Acta Math.*, 163(3-4):151–180, 1989.

[11] A. C. Hansen. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. *J. Amer. Math. Soc.*, 24(1):81–124, 2011.
[12] D. Hilbert. *Principles of Mathematical Logic*. Ams Chelsea, 1950.
[13] C. McMullen. Families of rational maps and iterative root-finding algorithms. *Ann. of Math. (2)*, 125(3):467–493, 1987.
[14] Y. Nesterov. *Introductory lectures on convex optimization : a basic course*. Applied optimization. Kluwer Academic Publ., Boston, Dordrecht, London, 2004.
[15] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259 – 268, 1992.
[16] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1996.
[17] A. M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc.*, S2-42(1):230.

Novel results for the anisotropic sparse grid quadrature

Helmut Harbrecht
(joint work with Abdul-Lateef Haji-Ali, Michael Peters, Markus Siebenmorgen)

1. Introduction

The anisotropic sparse grid quadrature can be applied for high-dimensional integrands which are analytically extendable into an anisotropic tensor product domain. Taking into account this anisotropy, we end up with a dimension independent error versus cost estimate of the proposed quadrature. In addition, we provide a novel and sharp estimate for the cardinality of the underlying anisotropic index set. To validate the theoretical findings, we present numerical results which demonstrate the remarkable convergence behaviour of the anisotropic sparse grid quadrature in applications.

2. Anisotropic sparse grid quadrature

We introduce anisotropic sparse grid quadrature formulas which extend the original idea of Smolyak's construction from [4]. Consider an increasing sequence of univariate quadratures

$$Q_j : C([-1,1]) \to \mathbb{R}, \quad f \mapsto Q_j f = \sum_{i=1}^{N_j} w_{i,j} f(\xi_{i,j}),$$

where the number N_j of quadrature points satisfies $N_1 \leq N_2 \leq \cdots$ and $N_j \to \infty$ for $j \to \infty$. For given $j \in \mathbb{N}$, we further introduce the difference quadrature operator

$$\Delta_j := Q_j - Q_{j-1}, \quad \text{where} \quad Q_{-1} := 0.$$

Let $\mathbf{w} = (w_1, w_2, \ldots, w_m) \in \mathbb{R}^m_+$ denote a weight vector for the different parameter dimensions. We assume in the following that the weight vector is sorted in ascending order, i.e. $w_1 \leq w_2 \leq \cdots \leq w_m$. Otherwise, we would rearrange the particular dimensions accordingly.

We define the index set

$$X_{\mathbf{w}}(q,m) := \left\{ \mathbf{0} \le \boldsymbol{\alpha} \in \mathbb{N}^m : \sum_{n=1}^{m} \alpha_n w_n \le q \right\}. \tag{1}$$

The *anisotropic sparse grid quadrature operator* of level $q \in \mathbb{N}$ is thus defined by

$$\mathcal{A}_{\mathbf{w}}(q,m) := \sum_{\boldsymbol{\alpha} \in X_{\mathbf{w}}(q,m)} \Delta_{\alpha_1}^{(1)} \otimes \cdots \otimes \Delta_{\alpha_m}^{(m)}. \tag{2}$$

3. Main results

3.1. **Cardinality of the index set.** For computing the number of quadrature points which the quadrature operator $\mathcal{A}_{\mathbf{w}}(q,m)$ applies, we need a sharp estimate on the index set $X_{\mathbf{w}}(q,m)$ from (1). Existing estimates are the well-known Beged-Dov formula (cf. [1])

$$\#X_{\mathbf{w}}(q,m) \le \prod_{n=1}^{m} \frac{q + \|\mathbf{w}\|_1}{n w_n}$$

or the product estimate

$$\#X_{\mathbf{w}}(q,m) \le \prod_{n=1}^{m} \left(\left\lfloor \frac{q}{w_n} \right\rfloor + 1 \right),$$

which is related to the full tensor product quadrature operator. Nonetheless, both estimates are not very sharp and the following, new estimate is much better.

Lemma 1. *The cardinality of the set $X_{\mathbf{w}}(q,m)$ is bounded by*

$$\#X_{\mathbf{w}}(q,m) \le \prod_{n=1}^{m} \left(\frac{q}{n w_n} + 1 \right).$$

3.2. **Error estimate.** We should next provide an error estimate for the anisotropic sparse grid quadrature operator (2). To that end, we should specify the univariate quadratures Q_j. They are supposed to be the *Gauss-Legendre quadrature rule* on $\Gamma := [-1,1]$ with $N_j = \left\lceil \frac{1}{2}(j+2) \right\rceil$ quadrature points. The class of integrands $f_m : [-1,1]^m \to \mathbb{R}$ we consider are functions which admit an analytic extension into the region

$$\boldsymbol{\Sigma}_m = \bigtimes_{n=1}^{m} \{ z \in \mathbb{C} : \operatorname{dist}(z,\Gamma) \le \tau_n \}$$

with $\tau_n \ge c n^r$ $(r > 1)$.

Lemma 2. *Let the weight vector* $\mathbf{w}$ *in* (1) *be given by* $w_n = \log(\kappa_n)$*, where* $\kappa_n := \tau_n + \sqrt{1+\tau_n^2}$*. Then, for each* $\delta > 0$*, there exists a constant* $c(\delta)$*, independent of* m*, such that the error of the anisotropic sparse quadrature is bounded by*

$$\left| \big(\mathbf{I} - \mathcal{A}_{\mathbf{w}}(q,m) \big) f_m \right| \le c(\delta, \boldsymbol{\tau}) \exp\big(-q(1-\delta) \big) \|f_m\|_{C(\boldsymbol{\Sigma}_m)}$$

with $c(\delta,\boldsymbol{\tau}) = 4c(\delta) \big\| \{\tau_n^{-1}\}_n \big\|_{\ell^1}$*. Note that* $c(\delta) \to \infty$ *as* $\delta \to 0$*.*

3.3. **Cost complexity.** The number of quadrature points can be bounded by

$$N(q) := \operatorname{cost}\big(\mathcal{A}_{\mathbf{w}}(q,m)\big) \leq \#X_{\mathbf{w}}(q,m)^2.$$

Further, one has

$$\#X_{\mathbf{w}}(q,m) \leq c(r)\exp\left(\frac{q}{r}\log\big(\log(m)\big)\right) = c(r)\log(m)^{q/r}$$

with a constant $c(r)$ which is independent of m. By combining these estimates with Lemma 2, we derive the convergence rate

$$\big|\big(\mathbf{I}-\mathcal{A}_{\mathbf{w}}(q,m)\big)f_m\big| \leq c(\delta,\boldsymbol{\tau})N(q)^{-r(1-\delta)/(2\log(\log m))}. \tag{3}$$

However, this estimate is not dimension robust, i.e., it is not independent of the dimension m. By using a result of [2], we arrive at following cost complexity.

Theorem 3. *The error of the anisotropic sparse grid quadrature with $w_n = \log(\kappa_n)$ can be bounded in terms of the number $N(q)$ of quadrature points according to*

$$\big|\big(\mathbf{I}-\mathcal{A}_{\mathbf{w}}(q,m)\big)f_m\big| \leq c(\boldsymbol{\tau},\beta)N(q)^{-(\beta-1)/2} \tag{4}$$

for all $\beta < r$.

Notice that this rate of convergence is smaller than that in (3) whenever m is fixed.

4. Numerical results

For our numerical tests, we consider a simple quadrature problem, namely, we like to integrate

$$f_m\colon \Gamma^m \to \mathbb{R}, \quad f_m(\mathbf{y}) := 5\left(3+\sum_{n=1}^{m} n^{-s}y_n\right)^{-1}. \tag{5}$$

The parameter s is chosen according to $s = 2,3,4$. Respective reference solutions are computed for $m = 1000$ dimensions and verified by a quasi-Monte Carlo method.

To validate the rate of convergence (4) with respect to the number of quadrature points N, we approximate the integral (5) for the limit choice $\tau_n = n^{s-1}$ by the m-dimensional anisotropic sparse grid quadrature. In particular, to catch the inherent dimensionality for each choice of the parameter s, we consider $m = 10, 100, 1000$ dimensions. As found in Figure 1 on the next page, we obtain different rates of convergence, dependent on the choice of the parameter s. Especially, the inherent dimension is larger than $m = 100$ only in the case $s = 2$. We observe the rates of convergence $N^{-1.44}$ for $s = 2$, $N^{-2.21}$ for $s = 3$, and N^{-3} for $s = 4$. These are much better than $N^{-1/2}$, N^{-1}, and $N^{-3/2}$, respectively, which are predicted by Theorem 3. Indeed, the rates are at least twice as much as predicted, which issues from the fact that the factor $1/2$ in the exponent on the right hand side of (4) issues from the crude estimate $N(q) \leq \#X_{\mathbf{w}}(q,m)^2$, see [3] for details.

More advanced examples from the uncertainty quantification of boundary value problems with random input parameters can be found in [3].

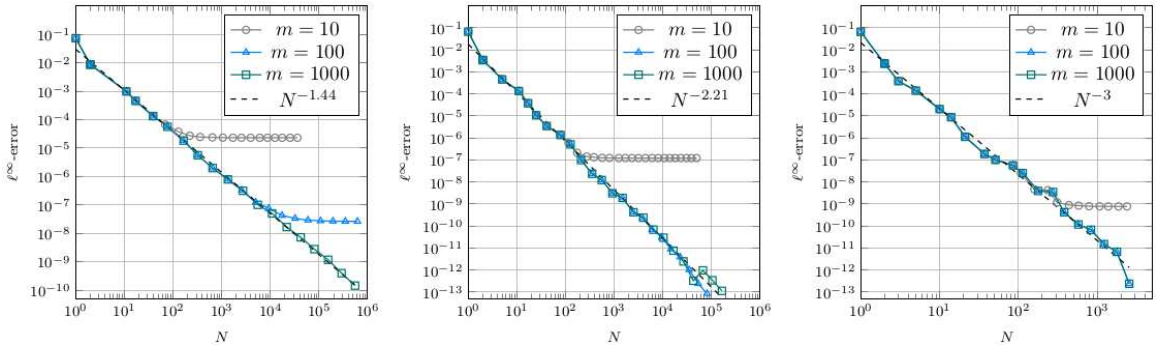

FIGURE 1. Rates of convergences for the $m = 10, 100, 1000$ dimensions and $s = 2$ (left), $s = 3$ (middle), and $s = 4$ (right).

REFERENCES

[1] A.G. Beged-Dov. Lower and upper bounds for the number of lattice points in a simplex. *SIAM Journal on Applied Mathematics*, 22(1):106–108, 1972.

[2] M. Griebel and J. Oettershagen. On tensor product approximation of analytic functions. *Journal of Approximation Theory*, 207(C):348–379, 2016.

[3] A.-L. Haji-Ali, H. Harbrecht, M.D. Peters and M. Siebenmorgen. Novel results for the anisotropic sparse grid quadrature. *arXiv:1509.05521*, 2015.

[4] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Doklady Akademii Nauk SSSR*, 4:240–243, 1963.

Optimal feedback control of semilinear parabolic equations: A "high"-dimensional HJB approach

KARL KUNISCH

(joint work with Dante Kalise)

A procedure for the numerical approximation of high-dimensional Hamilton-Jacobi-Bellman (HJB) equations associated to optimal feedback control problems for semilinear parabolic equations is proposed. Its main ingredients are a pseudospectral collocation approximation of the PDE dynamics, and an iterative method for the nonlinear HJB equation associated to the feedback synthesis. The latter is known as the Successive Galerkin Approximation. It can also be interpreted as Newton iteration for the HJB equation. At every step, the associated linear Generalized HJB equation is approximated via a separable polynomial approximation ansatz. The method requires a stabilizing control as initialisation. Its availability depends on the specific control system. If such a control can not be obtained, then the use of a discount factor and a continuation procedure as the discount factor tends to zero are proposed. Stabilizing feedback controls are obtained from solutions to the HJB equations for systems of dimension up to fourteen, i.e. the dimension of the pseudospectral approximation of the infinite dimensional dynamics has dimension fourteen. Further information can be found in [1].

References

[1] D. Kalise and K. Kunisch. Polynomial approximation of high-dimensional Hamilton-Jacobi-Bellman equations and applications to feedback control of semilinear parabolic PDEs. https://arxiv.org/abs/1702.04400

Optimal weighted least-squares methods for high-dimensional approximation

Giovanni Migliorati
(joint work with Albert Cohen)

We present some of the main results from [2] on the stability and accuracy of weighted discrete least squares for approximation in high-dimensional spaces. We consider the problem of reconstructing an unknown bounded function u defined on a domain $X \subset \mathbb{R}^d$ from noiseless or noisy samples of u at n points $(x^i)_{i=1,\dots,n}$. We measure the reconstruction error in a norm $L^2(X, d\rho)$ for some given probability measure $d\rho$. Given a linear space V_m with $\dim(V_m) = m \leq n$, we study in general terms the weighted least-squares approximations from the spaces V_m based on independent random samples. It is well known that least-squares approximations can be inaccurate and unstable when m is too close to n, even in the noiseless case.

The contribution of the analysis in [2] is twofold. From the theoretical perspective, we establish results in expectation and in probability for weighted least squares in general approximation spaces V_m. These results show that for an optimal choice of sampling measure $d\mu$ and weight w, which depends on the space V_m and on the measure $d\rho$, stability and optimal accuracy of the estimators are achieved under the mild condition that n scales linearly with m up to an additional logarithmic factor. The previous analyses in [1] and [3] have studied the stability and accuracy of standard (unweighted) discrete least squares in expectation and in probability, respectively. In contrast to [1] and [3], the analysis in [2] covers cases where the function u and its approximants from V_m are unbounded, which might occur for instance in the relevant case where $X = \mathbb{R}^d$ and $d\rho$ is the Gaussian measure.

From the numerical perspective, we propose a sampling method which allows one to generate independent and identically distributed samples from the optimal measure $d\mu$. This method becomes of interest in the multivariate setting where $d\mu$ is generally not of tensor product type. We illustrate this for particular examples of approximation spaces V_m of polynomial type, where the domain X is allowed to be unbounded and high or even infinite dimensional, motivated by certain applications to parametric and stochastic PDEs.

References

[1] A. Cohen, M. Davenport and D. Leviatan. On the stability and accuracy of least squares approximations. Found. Comput. Math. 13:819–834, 2013.

[2] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. submitted. Also available at arXiv:1608.00512
[3] G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone. Analysis of discrete L^2 projection on polynomial spaces with random evaluations. Found. Comput. Math. 14:419–456, 2014.

Effective boundary conditions: a general strategy and application to compressible flows over rough boundaries

Siegfried Müller
(joint work with Wolfgang Dahmen, Giulia Deolmi)

From several scenarios in nature it is well-known that microstructures on surfaces can significantly reduce drag. For instance, the skin of a shark exhibits small-scale structures that makes the shark a very fast maritime hunter.

This has been confirmed by experiments conducted in oil channels to study biological surfaces, e.g., shark-skin replicas, hairy surfaces such as seal fur experiencing significant drag reduction. Such observations lead engineers to mimic this effect for economical and ecological reasons in practical applications such as aviation.

Objectives. To gain a deeper insight in the underlying physical mechanisms of drag reduction and eventually permit predictions, simulations are performed that complement experimental investigations. Since resolving the microstructures requires a high resolution, numerical simulations are very expensive and, depending on the flow regime, are only feasible for small configurations. For a real application such as an airfoil the computational cost will be prohibitively high and a simulation will not be feasible in spite of an ever increasing computer power.

To deal with this type of problems a natural strategy is to resort to model reduction concepts. Some well-known strategies are *homogenization techniques*, *(heterogeneous) multiscale modeling* and multiscale finite element methods, all aiming to quantify the influence of small scale effects on the resolved macroscopic scale *without* directly resolving small scale structures. Typically, these concepts need to be adapted to the problem at hand. In fact, it should be noted that, strictly speaking, for the problems under consideration there is no clear (physical) scale separation so that a straightforward application of the heterogeneous multiscale method is delicate. Rather the range of relevant scales is too large to be resolved.

The central objective is to develop a new *computational* model reduction strategy that differs from the aforementioned methods. Our starting point is the formulation of an upscaling strategy where the micro-scale effect of a structured rough surface is modeled by means of effective boundary conditions given on a virtually smooth wall. For the derivation of these conditions the exact solution of the original problem on the rough domain is expanded in a *zeroth order solution* depending only on the macro-scale, i.e., the flow equations are solved in the artificial smooth

domain, and an upscaling term that depends on macro-scale *and* micro-scale variables in order to capture the micro-scale effects suppressed in the zeroth order solution. A natural idea is to plug this ansatz into the original equation and try to see under which conditions low order terms cancel to eventually arrive at the so-called *cell problem* which is typically much simpler than the original problem. From the asymptotic expansion at an artificial smooth wall located on top of the roughness we can then deduce the *effective boundary conditions* by means of a Taylor expansion in wall normal direction at the rough wall, where the mean of the solution of the cell problem enters as effective constant. Finally, the effective problem is solved on the smooth domain with effective boundary conditions.

Conceptual Background. We point out next how the approach relates to previous related work. In case of the steady incompressible Navier-Stokes equations a similar upscaling strategy has been developed and investigated by Achdou et al. [1] as well as Jäger and Mikelic [7] for small Reynolds numbers. In [1] a Navier wall law is derived from a Taylor expansion of velocity and pressure where the zeroth order solution is first solved on an *extension* of the rough domain. Applying an idea similar to [1] a Navier wall law is derived also in [2] for the steady Poisson problem.

Typically, only the influence in wall-normal direction is accounted for in the effective boundary conditions. This is justified as long as the flow is laminar. For instance, for turbulent flow the flow field is inherently three-dimensional and, thus, the roughness will most likely also affect the flow in streamwise and spanwise direction as well. In [6], it is suggested to solve two cell problems, corresponding to the effects in streamwise and spanwise direction, respectively. For this purpose the flow at an oblique angle is considered on the macro-scale, introducing in this way the spanwise effects. However, no systematic strategy is given for the choice of the angle.

What is Different here? The aims pursued in our work are yet different in three major respects. First, we are particularly interested in *compressible* flows over a rough surface for high Reynolds numbers, which corresponds to considering a regime that significantly differs from the one analyzed in the aforementioned literature. In [4] we already derived a similar upscaling strategy combining ideas from Achdou et al. as well as Jäger and Mikelic. However, and this is the second delineating issue, we target a roughness scale relative to the viscous sublayer thickness which is *larger* than in those works because of significantly larger velocities.

To account for this, the identification of the appropriate cell problem itself deserves some special attention. In order to make the underlying mechanisms transparent we outline in [3] for a deliberately general scalar problem a systematic way - in some sense a “recipe” - for determining a suitable cell problem for the desired roughness range which is then used to identify effective boundary conditions. We emphasize that these considerations do not intend to provide a rigorous analytical foundation - which in our opinion is out of reach for the regime of interest - but are to serve the following purposes:

First, in a given application it helps identifying some scaling effects which, for instance, determine whether a resulting cell problem is linear or nonlinear and which explain why macro-scale parameters enter the cell problem as parameters. In fact, targeting a roughness scale that is relatively large in comparison with the boundary layer thickness in the compressible regime comes at a price, namely, in contrast to the aforementioned work, we need to account for the influence of the *zeroth order solution in the cell problem* and these terms can no longer be discarded. This entails several impediments. First, the cell problem becomes *parameter-dependent*. This has been verified already in [4] where computations for different cell problems have been carried out. For a detailed discussion on the differences between our model and the models of Achdou et al. as well as Jäger and Mikelic we refer to [4].

Second, the discussion on the general scalar problem leads us in a natural way to formulate two types of effective boundary conditions - implicit and explicit ones - a point to be taken up below again.

As indicated earlier, a completely rigorous foundation of our approach, as given in [7] for a specific flow regime, is out of reach in the present framework. This brings us to the third point concerning the main objective, namely to develop *computationally viable* techniques for dealing with structured roughness for compressible flows, primarily focussing at this point on a numerical validation. The fact that the cell problem depends on the zeroth order solution at a first glance seems to preclude its original purpose. However, the availability of certified model reduction techniques for exactly the type of problems arising as cell problems allows us to efficiently query the solution of the cell problem for many parameter values using reduced models. A second issue of practical relevance is how to acquire the information needed for the cell problem, namely the zeroth order solution. For simple geometries one can resort e.g. to van Driest's solution which avoids the computation of the zeroth order solution. Moreover, if the cell problem would be the only place where this information enters, such a qualitative approximation is expected to suffice. Therefore, as one option, we derive effective boundary conditions of *implicit* type confining knowledge about the zeroth order solution only in the cell problem which, as indicated above, can then be handled by reduced basis concepts.

In more complex situations a good guess about the zeroth order solution will generally be missing and this strategy is no longer viable. Our point of view then is that computing the zeroth order solution in a smooth domain on a reasonably coarse mesh is, in principle, affordable. We therefore discuss this as a second option leading to *explicit* effective boundary conditions. In the context of laminar compressible flows both variants of our approach were tested highlighting, in particular, also the effect of different boundary conditions, cf. [3].

Application. For a first step towards analyzing turbulent flow scenarios our multiscale concept has been applied in [5] to a *turbulent, compressible* flow over a riblet surface. The zeroth-order solution and its derivatives can no longer be computed analytically using the Van Driest solution, as done in the laminar case.

These quantities are estimated numerically performing first a flat plate RANS computation. Moreover in the turbulent regime we encounter significantly larger Reynolds numbers than in the laminar case treated in [4, 3] and a correspondingly stronger convection.

References

[1] Y. Achdou, O. Pironneau, F. Valentin, Effective boundary conditions for laminar flows over periodic rough boundaries. J. Comp. Phys., 147, 1998, 187–218.

[2] D. Bresch, V. Milisic, High order multi-scale wall-laws, Part I: the periodic case Quarterly of Applied Mathematics, 68, 2010, 229–253.

[3] G. Deolmi, W. Dahmen, S. Müller. Effective boundary conditions: a general strategy and application to compressible flows over rough boundaries. Communications in Computational Physics, 21(2), 2017, 358–400.

[4] G. Deolmi, W. Dahmen, S. Müller. Effective boundary conditions for compressible flows over rough boundaries. Mathematical Models and Methods in Applied Sciences, 25(7), 2015, 25–52.

[5] G. Deolmi, W. Dahmen, S. Müller, M. Albers, P.S. Meysonnat, W. Schröder. Effective Boundary Conditions for Turbulent Compressible Flows over a Riblet Surface. Proceedings of HYP2016, Aachen (Germany), August 1-5, 2016, submitted.

[6] E. Friedmann, T. Richter, Optimal microstructures drag reducing mechanism of riblets. J. Math. Fluid Mech., 13, 2011, 429–447.

[7] W. Jäger, A. Mikelic, On the roughness-induced effective boundary conditions for an incompressible viscous flow. Journal of Differential Equations, 170, 2001, 96–122.

Highly localized summability kernels on the sphere induced by Newtonian kernels

Pencho Petrushev

(joint work with Kamen Ivanov)

We are interested in the problem for nonlinear n-term approximation of harmonic functions on the unit ball B^d in $\mathbb{R}^d$ from linear combinations of shifts of the fundamental harmonic function (Newtonian kernel) $\frac{1}{|x|^{d-2}}$ in dimension $d > 2$ or $\ln|x - y_j|$ when $d = 2$. More explicitly the *problem* is: For a given harmonic function f on B^d and $n \geq 1$ find n locations $\{y_j\}$ with $|y_j| > 1$ and $n+1$ coefficients $\{c_j\}$ so that

$$c_0 + \sum_{j=1}^{n} \frac{c_j}{|x - y_j|^{d-2}} \quad (d > 2) \qquad \text{or} \quad c_0 + \sum_{j=1}^{n} c_j \ln|x - y_j| \quad (d = 2)$$

approximates f with an optimal rate in the norm of the harmonic Hardy space $\mathcal{H}^p(B^d)$, $0 < p < \infty$. The poor localization of the fundamental harmonic function $|x|^{-d+2}$ $(d > 2)$ or $\ln|x|$ $(d = 2)$ creates a problem.

The solution of this problem naturally involves Besov spaces consisting of harmonic functions on B^d. The key to its solution, however, is the solution of the following fundamental problem:

Problem 1. *For arbitrary constants $M > 0$ (large) and $\varepsilon > 0$ (small) construct a function $\Phi_\varepsilon(\eta, x)$ on $\mathbb{S}^{d-1} \times B^d$ of the form*

$$\Phi_\varepsilon(\eta, x) = \sum_{j=1}^{K} \frac{b_j}{|x - y_j|^{d-2}} \quad (d > 2) \quad \text{or} \quad \Phi_\varepsilon(\eta, x) = \sum_{j=1}^{K} b_j \ln |x - y_j| \quad (d = 2)$$

with $|y_j| > 1$, $b_j \in \mathbb{R}$ depending on η, M, and d, and K depending only on M and d, such that

$$|\Phi_\varepsilon(\eta, x)| \le \frac{c\varepsilon^{-d+1}}{(1 + \varepsilon^{-1}|x - \eta|)^M}, \quad \forall x, \eta \in \mathbb{S}^{d-1},$$

and

$$\int_{\mathbb{S}^{d-1}} \Phi_\varepsilon(\eta, x) d\sigma(x) = 1, \ \ \forall \eta \in \mathbb{S}^{d-1}.$$

Here $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$ and $c > 0$ is a constant depending only on M and d.

As is well known the ℓth directional derivative operator $(\eta \cdot \nabla)^\ell$ is approximated well by the respective finite difference operator. Consequently, to solve Problem 1 it suffices to solve the following

Problem 2. *Let $M > d - 1$. For given $\varepsilon \in (0, 1]$ find $m + 1$ constants $b_\ell \in \mathbb{R}$ and $a > 1$ so that the restriction $F_\varepsilon(x \cdot \eta)$ of the function*

$$(1) \quad f_{\varepsilon,\eta}(x) = \sum_{\ell=0}^{m} b_\ell (\eta \cdot \nabla)^\ell \Big(\frac{1}{|x - a\eta|^{d-2}} \Big), \quad \eta \in \mathbb{S}^{d-1}, \ x \in \mathbb{R}^d \setminus \{a\eta\}, \quad \text{if } d > 2;$$

or

$$(2) \quad f_{\varepsilon,\eta}(x) = b_0 + \sum_{\ell=1}^{m} b_\ell (\eta \cdot \nabla)^\ell \ln \frac{1}{|x - a\eta|}, \quad \eta \in \mathbb{S}^1, \ x \in \mathbb{R}^2 \setminus \{a\eta\}, \quad \text{if } d = 2,$$

to $\mathbb{S}^{d-1}$ satisfies the conditions

$$(3) \qquad |F_\varepsilon(x \cdot \eta)| \le \frac{c\varepsilon^{-d+1}}{(1 + \varepsilon^{-1}\rho(x, \eta))^M}, \quad \forall x, \eta \in \mathbb{S}^{d-1},$$

and

$$(4) \qquad \int_{\mathbb{S}^{d-1}} F_\varepsilon(x \cdot \eta) d\sigma(\eta) = 1, \quad \forall x \in \mathbb{S}^{d-1}.$$

where $c > 0$ is a constant depending only on M and d. Above $x \cdot \eta$ stands for the inner product of $x, \eta \in \mathbb{S}^{d-1}$.

In this talk we focus on the solution of Problem 2. Note that using the well known Maxwell's identity

$$(\eta \cdot \nabla)^m |x|^{-d+2} = (-1)^m m! C_m^{(d-2)/2}(x \cdot \eta/|x|)|x|^{-m-d+2} \qquad (d > 2)$$

one can easily construct a function $F_\varepsilon(x \cdot \eta)$ that is the restriction to $\mathbb{S}^{d-1}$ of a function $f_{\varepsilon,\eta}(x)$ of the form (1) which obeys (3). However, this $F_\varepsilon(x \cdot \eta)$ will not satisfy condition (4). The situation is the same in dimension $d = 2$.

Our main result asserts that Problem 2 is completely solved by the function

$$F_\varepsilon(x\cdot\eta)=\frac{c^\star\varepsilon^{2m-1}}{|x-a\eta|^{2m+d-2}},\ \ x\in\mathbb{S}^{d-1},\ \ \text{with}\ \int_{\mathbb{S}^{d-1}}F_\varepsilon(x\cdot\eta)d\sigma(x)=1,$$

where $a=1+\varepsilon$, $m:=\lceil(M-d+2)/2\rceil$, and $c^\star>0$ is a normalization constant.

We also report on the solution of the same problem for the half-space, i.e. in the case when $\mathbb{S}^{d-1}$ is replaced by $\mathbb{R}^{d-1}:=\{x\in\mathbb{R}^d : x=(x_1,\dots,x_{d-1},0)\}$.

We have established this result: Let $m\geq 1$, $d\geq 2$, $\varepsilon>0$, and $\eta=(0,\dots,0,-1)$. The function

$$F_\varepsilon(x):=\frac{\varepsilon^{2m-1}}{|x-\varepsilon\eta|^{2m+d-2}},\quad x\in\mathbb{R}^{d-1},$$

obeys

$$0<F_\varepsilon(x)\leq\frac{c_1\varepsilon^{-d+1}}{(1+\varepsilon^{-1}|x|)^{2m+d-2}},\quad \forall x\in\mathbb{R}^{d-1},$$

and

$$\int_{\mathbb{R}^{d-1}}F_\varepsilon(x)dx\geq c_2>0,$$

where $c_1,c_2>0$ are constants depending only on m and d. Moreover, $F_\varepsilon(x)$ is the restriction to $\mathbb{R}^{d-1}$ of a harmonic function of the form

$$\sum_{\ell=1}^m b_\ell\Big(\frac{\partial}{\partial x_d}\Big)^\ell\frac{1}{|x-\varepsilon\eta|^{d-2}}\ \text{ if } d\geq 3\quad\text{or}\quad\sum_{\ell=1}^m b_\ell\Big(\frac{\partial}{\partial x_d}\Big)^\ell\ln\frac{1}{|x-\varepsilon\eta|}\ \text{ if } d=2.$$

Clustering high dimensional data or how to smoothly put complex people in small boxes : The change point problem

Dominique Picard

We consider the problem of clustering high dimensional data. We observe a matrix Y of size $n\times d$. Typically d is much larger than n (but not necessarily) and each column vector represents an individual denoted by Y_i, $i\leq n$, of dimension d.

For sake of simplicity, we assume that there is only two classes, i.e. that there exists $A\subset\{1,\dots n\}$ and two vectors of $\mathbb{R}^d$, θ_- and θ_+, such that $Y_i\sim N(\theta_-,\sigma^2 I)$ for $i\in A$, $Y_i\sim N(\theta_+,\sigma^2 I)$ for $i\in A^c$.

We assume that θ_- and θ_+ are unknown and sparse in the sense that they belong to the regularity set :

$$\Theta(s,L):=\{\theta\in\mathbb{R}^d,\ \sup_{K\in\mathbb{N}^*}K^{2s}\sum_{k\geq K}(\theta^k)^2\leq L^2\}.$$

Again, for sake of simplicity we suppose that A is of the form $\{1,\dots,n\tau\}$, for $\tau\in(0,1)$ unknown. For technical reason we also assume that in fact that $\tau\in]\varepsilon,1-\varepsilon[$ (for some known and fixed parameter $0<\varepsilon<1/2$) and we put :

$$\Delta^2=\sum_{l=1}^d(\theta_+-\theta_-)^2.$$

Our problem is to determine whether or not it is efficient to smooth the data i.e. to replace the vectors $Y_i := Y_i(d)$, $i \leq n$ by, for $T < d$, $Y_i(T)$, $i \leq n$, the vectors of $\mathbb{R}^T$, of the T first coordinates of Y_i. Then, if smoothing reveals to be useful, how to choose T ideally in an adaptive way (without knowing the regularity s).

We also propose an adaptive algorithm to estimate θ_+ and θ_- with minimax rates.

1. Smoothing rates

We consider the following family of algorithms (corresponding to the MLE, or Kmeans estimates in the general situation).

$$\hat{\tau}(T) = ArgMin_{t\in]\varepsilon,1-\varepsilon[}$$
$$\left\{ \sum_{j\leq nt}\sum_{\ell\leq T}(Y_j^\ell - \frac{1}{nt}\sum_{j\leq nt} Y_j^\ell)^2 + \sum_{j\geq nt+1}\sum_{\ell\leq T}(Y_j^\ell - \frac{1}{n(1-t)}\sum_{j\geq nt+1} Y_j^\ell)^2 \right\}$$

We prove the following Proposition.

Proposition 1. *We assume that θ_- and θ_+ belong to $\Theta(s, L)$, we stop the observation at $T \leq d$: $Y_j(T) = (Y_j^1, \ldots, Y_j^T)$ and assume that*

$$\Delta^2 \geq R[T^{-2s} \vee \frac{\sigma^2 T}{n}],$$

then for any κ, there exists $\gamma(\kappa, R)$ such that

$$P(|\hat{\tau} - \tau| \geq \kappa \frac{\sigma^2 T}{n\Delta^2}) \leq e^{-\gamma T}.$$

Remarks 1.

- *It is natural that the rate of convergence for τ is decreasing in Δ.*
- *Condition $\Delta^2 \gtrsim R[T^{-2s}]$ is necessary for identifiability: otherwise, because of the sparsity of the vector θ_+ and θ_-, $\sum_{l\leq T}(\theta_-^l - \theta_+^1)^2$ may be arbitrarily close to zero, leading to a model on the $Y_j(T)$'s observations in which τ has no proper meaning and cannot be estimated.*
- *We can prove that condition : $\Delta^2 \gtrsim [\frac{\sigma^2 T}{n}]$ is necessary for the MLE to converge.*
- *There is an obvious advantage to smoothing leading to better rates.*
- *If we now look at the conditions on Δ^2, we see that they are less restrictive, with better rates, as soon as T decreases subject to the condition $T^{-2s} \lesssim \Delta^2$.*
- *Hence, we need to minimize $\frac{\sigma^2 T}{n}$ subject to $T^{-2s} \lesssim \Delta^2$, leading to the 'usual' in nonparametric situations optimal choice $\rightarrow T_s \sim [\frac{n}{\sigma^2}]^{\frac{1}{1+2s}}$*
  $\rightarrow$ Rates and conditions then become :

$$\Delta^2 \gtrsim [\frac{n}{\sigma^2}]^{\frac{-2s}{1+2s}}, \quad Rate\ [\frac{n}{\sigma^2}]^{\frac{-2s}{1+2s}}\Delta^{-2}.$$

- *Rate and conditions could seem quite poor, but observe that very often* σ^2 *is of the form* $\frac{\sigma_0^2}{d}$*, where* σ_0 *is a fixed constant.*
- *In this case, the rate becomes much better :*

$$[\frac{nd}{\sigma_0^2}]^{\frac{-2s}{1+2s}}\Delta^{-2}$$

- *In this particular case,* $\sigma^2 = \frac{\sigma_0^2}{d}$*, the form of the optimal smoothing*

$$T_s := [\frac{nd}{\sigma_0^2}]^{\frac{1}{1+2s}}$$

proves that any adaptive smoothing on each individual signal Y_j *(thresholding or whatever) would lead instead to an optimal smoother of the form :*

$$T_{opt} = [\frac{d}{\sigma_0^2}]^{\frac{1}{1+2s}}$$

inevitably creating a loss of a factor polynomial in n *in the rates. This means that it is better to perform the smoothing globally (off-line).*

2. Adaptative choice for T

Form the two following pseudo-data in $\mathbb{R}^d$: $Z(1), Z(2)$

$$Z^\ell(1) = \frac{1}{n}\sum_{j=1}^{n} Y_j^\ell - \frac{2}{n}\sum_{j=1}^{n/2} Y_j^\ell, \ell = 1, \ldots, d$$

$$Z^\ell(2) = \frac{1}{n}\sum_{j=1}^{n} Y_j^\ell - \frac{2}{n}\sum_{j=n/2+1}^{n} Y_j^\ell, \ell = 1, \ldots, d$$

Consider the Lepski smoothers (c is a tuning constant)

$$\hat{T}(1) := \min\{k,\ \sum_{m=k'}^{l} [Z^m(1)]^2 \leq cl\frac{\sigma^2}{n}\log[d \vee n],\ \forall d \geq l \geq k' \geq k\},$$

$$\hat{T}(2) := \min\{k,\ \sum_{m=k'}^{l} [Z^m(2)]^2 \leq cl\frac{\sigma^2}{n}\log[d \vee n],\ \forall d \geq l \geq k' \geq k\}.$$

put

$$\hat{T} := \hat{T}(1) \vee \hat{T}(2)$$

Theorem 1. *We assume that* θ_+ *and* θ_- *are in* $\Theta(s, L)$*. We suppose that there exists a contant* $a > 0$ *such that*

$$\frac{n}{\sigma^2} \geq a \log d.$$

We put

$$\hat{T} := \hat{T}(1) \vee \hat{T}(2).$$

Then, if there exists a constant $R = R(L, \varepsilon)$ such that

$$\Delta^2 \geq R[\frac{\sigma^2 \log[d \vee n]}{n}]^{\frac{2s}{1+2s}}, \tag{1}$$

then for any γ, and for c large enough,

$$P(|\hat{\tau}(\hat{T}) - \tau| \geq \kappa[\frac{\sigma^2 \log[d \vee n]}{n}]^{\frac{2s}{1+2s}} \Delta^{-2}) \leq [d \vee n]^{-\gamma}. \tag{2}$$

3. Adaptation rates for θ_- and θ_+, case $\sigma^2 = \frac{\sigma_0^2}{d}$

We first detect the change using the procedure above, using $\hat{T}$, and then $\hat{\tau} = \hat{\tau}(\hat{T})$.

$$\hat{\tau}(\hat{T}) = ArgMin_{t \in]\varepsilon, 1-\varepsilon[}$$
$$\left\{ \sum_{j \leq nt} \sum_{\ell \leq \hat{T}} (Y_j^\ell - \frac{1}{nt} \sum_{j \leq nt} Y_j^\ell)^2 + \sum_{j \geq nt+1} \sum_{\ell \leq \hat{T}} (Y_j^\ell - \frac{1}{n(1-t)} \sum_{j \geq nt+1} Y_j^\ell)^2 \right\}$$

Then we estimate θ_- and θ_+, with the following procedure. (We denote $\theta_\pm$ for respectively either θ_- or θ_+:

$$\hat{\theta}_\pm = (\hat{\theta}_\pm^1, \ldots, \hat{\theta}_\pm^d),\ \hat{\theta}_\pm^k := \hat{\theta}_{\pm,k} I\{k \leq \hat{T}^*(\pm)\}$$

$$\hat{\theta}_{-,k} := \frac{1}{n\hat{\tau}(\hat{T})} \sum_{j=1}^{n\hat{\tau}(\hat{T})} Y_j^k \quad \hat{\theta}_{+,k} := \frac{1}{n(1-\hat{\tau}(\hat{T}))} \sum_{j=n\hat{\tau}(\hat{T})+1}^{n} Y_j^k$$

$$\hat{T}^*(\pm) := \min\{k, \sum_{m=k+1}^{l} [\hat{\theta}_{\pm,m}]^2 \leq cl \frac{\sigma^2}{n} \log[d \vee n],\ \forall l \geq k+2\}.$$

And we prove the following result meaning that (without condition on Δ this time) we are able to adaptively estimate θ_- and θ_+ with minimax rates up to logarithmic factors.

Theorem 2. *With the estimates defined above, then, for $s > 0$, $c > c_0$, we have that there exists a constant C such that :*

$$\sup_{\theta_\pm \in \Theta(s,L)} E\|\hat{\theta}_\pm - \theta_\pm\|_2^2 \leq C\{\frac{nd}{\log[n \vee d]}\}^{\frac{-2s}{1+2s}} \tag{3}$$

References

[1] D. Picard. Sparse estimation in the change point framework. Technical Report.

Invariant domain preserving approximations of hyperbolic systems

Bojan Popov
(joint work with Jean-Luc Guermond)

We present a numerical method to solve general hyperbolic systems in any space dimension using forward Euler time stepping and continuous finite elements on non-uniform grids. The properties of the method are based on the introduction of an artificial dissipation that is defined so that any convex invariant sets containing the initial data is an invariant domain for the method. The invariant domain property is proved for any hyperbolic system provided a CFL condition holds. The solution is also shown to satisfy a discrete entropy inequality for every admissible entropy of the system. The method is formally first-order accurate in space and can be made high-order in time by using Strong Stability Preserving algorithms. Further information for the new method can be found in [1] and details on implementation for the Euler system of gas dynamics can be found in [2]. A second-order accurate in space and local maximum principle preserving extension of the above method in the scalar case is given, for more details see [3]. Finally, we describe a a second-order extension of the method for the shallow water equation with topography. The method is well-balanced and positivity preserving and works well in the presence of dry states, see [4] for details. All of the above techniques are parameter free and can be used in any space dimension with continuous finite elements on unstructured meshes.

References

[1] J.-L. Guermond and B. Popov. Invariant domains and first-order continuous finite element approximation for hyperbolic systems. SIAM J. Numer. Anal. 54 (2016), no. 4, 2466–2489.
[2] J.-L. Guermond and B. Popov. Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations J. Comput. Phys. 321 (2016), 908–926.
[3] J.-L. Guermond and B. Popov. Invariant domains and second-order continuous finite element approximation for scalar conservation equations. SIAM J. Numer. Anal., submitted.
[4] J.-L. Guermond and B. Popov. Well-balanced second-order approximation of the shallow water equation with continuous finite elements. SIAM J. Numer. Anal., submitted.

Low rank regularity results for highdimensional PDEs

Reinhold Schneider

Langevin dynamics with a gradient field and corresponding Fokker Planck equation (here Smoluchovski equation) is described by a stochastic differential equation

$$d\mathbf{x} = \mathbf{b}(\mathbf{b})dt + dW_t \ .$$

Here W_t is a d-dimensional Brownian motion. We assume that the vector field $\mathbf{b}$ is a gradient field, i.e. $\mathbf{b}(\mathbf{x}) = \nabla V(\mathbf{x})$. The probability density at time $t \geq 0$ for a

system to be in state $\mathbf{x} = (x_1, \ldots, x_d)$ is described by the Fokker Planck equation

$$\partial_t p = \nabla \cdot (\nabla V p) + \Delta \tag{1}$$

The ground state corresponds to $\lambda_0 = 0$ is given explicitly by

$$\phi_0(\mathbf{x}) = \frac{1}{Z} e^{-V(\mathbf{x})} \quad , \quad Z := \int e^{-V} \ . \tag{2}$$

We consider only nearest neighbor interaction i.e. a potential of the form

$$V = \sum_{i=1}^{d-1} V_i \quad , \quad V_i(\mathbf{x}) = V_i(x_i, x_{i+1}) \ .$$

We consider a low rank approximation of the functions

$$e^{V_i(x_i, x_{i+1}} = \sum_{k=1}^{r} U_i(x_i) V_i(x_{i+1} \ .$$

One can see that all TT ranks r_i, $i = 1, \ldots, d-1$ of ϕ_0 are bounded by r. However in the generic case the canonical rank r_c may grow exponentially w.r.t. d, i.e. $r_c \sim r^d$.

We consider a low rank approximation of the first eigenfunctions ϕ_k, $k = 1, \ldots, K$, and claim that TT approximability can be proved by applying the arguments from [2] considering the matrix product state approximation for a 1D quantum lattice system. The importance of local (nearest neighbor) interaction has been known in quantum physics community and often referred to the *area law*.

First eigenfunctions of the backward Kolmogorov operator, which is the adjoint to Fokker Planck operator has been computed via the Transfer operator and Monte Carlo sampling in the framework of *variational conformation dynamics* (similar to Markov State Models) in [3]. The largest bio-molecular system has $d = 250$, where satisfying results has been obtained with ranks < 10. These computations requires long time trajectories of the SODE to sample from the equilibrium distributions. We have used (M)-ALS (or DMRG) scheme for the solution of the generalized eigenvalue problems, and highlight that the FP operator was not used, nor any low rank representation of the underlying transfer operator.

References

[1] R- Schneider and A. Uschmajew, *Approximation rates for the hierarchical tensor format in periodic Sobolev spaces*, Journal of Complexity **30** (2015), 56–71.

[2] I. Arad, A. Kitaev, Z. Landau and U. Vazirani, *An area law and sub-exponential algorithm for 1D systems* , arXiv:1301.1162 [quant-ph] (2013).

[3] F. Nüske, R. Schneider, F. Vitalini and F. Noé, *Variational Tensor Approach for Approximating the Rare-Event Kinetics of Macromolecular Systems*, Journal of Chemical Physics, (144), 1-14, 2016.

Hyperbolic cross approximation and beyond

Tino Ullrich

We give a brief historical overview of challenges and open problems of approximation theory with emphasis put on multivariate approximation. 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 anisotropic Sobolev and Besov classes, classes of functions with bounded mixed derivative and others. The simplest case of such a function class is the unit ball of the mixed Sobolev space of bivariate functions given by

$$\mathbf{W}_p^r := \left\{ f \in L_p \ : \ \|f\|_{\mathbf{W}_p^r} := \|f\|_p + \left\| \frac{\partial^r f}{\partial x_1^r} \right\|_p + \left\| \frac{\partial^r f}{\partial x_2^r} \right\|_p + \left\| \frac{\partial^{2r} f}{\partial x_1^r \partial x_2^r} \right\|_p \le 1 \right\}.$$

These classes are sometimes denoted as classes of functions with dominating mixed derivative since the condition on the mixed derivative is the dominating one. Babenko [2] was the first who introduced such classes and began to study approximation of these classes by the hyperbolic cross polynomials.

We note that the function classes with bounded mixed derivative are not only an interesting and challenging object for approximation theory. They also represent a suitable model in scientific computations. Bungartz and Griebel [3, 11] and their groups use approximation methods designed for these classes in elliptic variational problems. In addition, similar function classes are suitable models for the treatment of the electronic Schrödinger equation [14]. This makes approximation techniques developed for classes of functions with bounded mixed derivative a proper choice for the numerical treatment of those problems.

Another fundamental question is: How to approximate functions from these classes? In addition to the Kolmogorov n-width we report on the asymptotic behavior of linear widths. Interesting effects occur when studying the approximation of the class $\mathbf{W}_p^r$ in L_q if $p < q$. In contrast to the case $p = q$ the influence of the parameters p and q is always visible in the rate of the order of the linear widths. In fact, if either $1 < p \le q \le 2$ or $2 \le p \le q < \infty$ then we have

$$\lambda_m(\mathbf{W}_p^r, L_q) \asymp \left(\frac{(\log m)^{d-1}}{m} \right)^{r-1/p+1/q}, \quad m \in \mathbb{N}. \tag{1}$$

Then the optimal approximant is realized by a projection on appropriate linear subspace of the hyperbolic cross polynomials, which is not always the case, like for instance in the case $p < 2 < q$. Restricting the set of admissible rank m operators to such that are based on function evaluations (instead of general linear functionals), we observe a behavior which is clearly bounded below by λ_m. We call the corresponding asymptotic quantities sampling widths ϱ_m. However, this is not the end of the story. Already in the situation $\mathbf{W}_p^r$ in L_q we are able to determine sets of parameters where ϱ_m is equal to λ_m in the sense of order, and

others where ϱ_m behaves strictly worse (already in the main rate) [4]. However, the complete picture is still unknown. In that sense the situation $p = q$ (including $p = 2$) is of particular interest. The result

$$\varrho_m(\mathbf{W}_p^r, L_p) \lesssim \left(\frac{(\log m)^{(d-1)}}{m}\right)^r (\log m)^{(d-1)/2} \quad , \quad m \in \mathbb{N}\,,$$

has been a breakthrough since it improved on a standard upper bound by using a non-trivial technique. However, the exact order is still unknown even in case $p = 2$. The so far best-known upper bounds for sampling recovery are all based on sparse grid constructions.

High-dimensional approximation problems appear in several areas of science like for instance in quantum chemistry and meteorology. We will comment on some recent results on how the underlying dimension d affects the multivariate approximation error. The order of the approximation error is not longer sufficient for determining the *information based complexity* of the problem. We present some recent results to see the d-dependence of the constants in the approximation error estimates and the convergence rate of widths complemented by sharp *preasymptotical* estimates in the Hilbert space case. In case of high dimensions, the traditional estimate (1) becomes problematic. Fixing d and r the function $f_d(t) := t^{-r}\,(\log t)^{r(d-1)}$ is increasing on $[1, e^{d-1}]$ and decreasing on $[e^{d-1}, \infty)$. Hence, its maximum on $[1, \infty)$ is $\max_{t\geq 1} f_d(t) = f_d(e^{d-1}) = \left(\frac{d-1}{e}\right)^{r(d-1)}$, which increases super-exponentially in d. That means, for large d we have to wait "exponentially long" until the sequence $n^{-r}(\ln n)^{(d-1)r}$ decays, and even longer until it becomes less than one. In any case it is important to control the behavior of the constants $c(r, d)$ and $C(r, d)$ appearing in the lower and upper order estimate (1) in d and, in a second step, to establish *preasymptotic estimates* for small m. Results have been obtained recently by Kühn, Sickel, Ullrich [12]. There the authors considered the space $\mathbf{W}_2^r$ to be specifically normed as follows

$$\|f\|^2_{\mathbf{W}_2^r} := \sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{f}(\mathbf{k})|^2 \prod_{j=1}^{d} (1 + |k_j|^2)^r\,. \tag{2}$$

It has been shown in [12] that in this situation the "asymptotic constant" behaves exactly as follows

$$\lim_{m\to\infty} \frac{m^r}{(\log m)^{(d-1)r}} \cdot \lambda_m(\mathbf{W}_2^r, L_2) \;= \Big[\frac{2^d}{(d-1)!}\Big]^r\,. \tag{3}$$

This result is surprising from several points of view. First, the limit exists, second one can compute it explicitly and third, the number on the right-hand side decays exponentially in d. However, if d is large we may have to "wait" very long until that happens. Hence, the next question is what happens in the preasymptotical range, say for m less than 2^d. Here, we get the bound below in Theorem 1 (see [12, Thm. 4.17]). For similar results in more general classes as well as a non-periodic counterpart, see the longer arXiv version of [6].

Theorem 1 ([12]). *Let $d \geq 2$, and $r > 0$. Then for any $1 \leq m \leq 4^d$ we have the upper and lower estimate*

$$2^{-r}\Big(\frac{1}{2m}\Big)^{\frac{r}{2+\log_2(d/\log_2 n+1/2)}} \leq \lambda_m(\mathbf{W}_2^r, L_2) \leq \Big(\frac{e^2}{m}\Big)^{\frac{r}{2+\log_2 d}} .$$

Note, that there is no hidden constant in both bounds. This type of error decay reflects "quasi-polynomial" tractability, a notion recently introduced by Gnewuch, Woźniakowski [10]. The result in Theorem 1 is based on a refined estimate for the cardinality of the "smooth" hyperbolic cross

$$\Gamma(N,d) := \Big\{k \in \mathbb{Z}^d \ : \ \prod_{j=1}^{d}(1+|k_j|) \leq N\Big\} \quad , \quad N \in \mathbb{N}\,.$$

In [12, Thm. 4.9] it is shown

$$|\Gamma(N,d)| \leq e^2 N^{2+\log_2 d}\,. \tag{4}$$

In [7] the authors state cardinality bounds (without proof) for slightly modified hyperbolic crosses in d dimensions.

What concerns the approximation in the uniform norm L_∞ in case $r > 1/2$ we refer to the recent paper Cobos, Kühn, Sickel [8]. The authors obtained the asymptotic constant

$$\lim_{m\to\infty} \frac{m^{r-1/2}\lambda_m(\mathbf{W}_2^r, L_\infty)}{(\log m)^{(d-1)r}} = \frac{1}{\sqrt{2r-1}}\Big[\frac{2^d}{(d-1)!}\Big]^r .$$

References

[1] K.I. Babenko, On the approximation of periodic functions of several variables by trigonometric polynomials, Dokl. Akad. Nauk USSR, **132** (1960), 247–250; English transl. in Soviet Math. Dokl., **1** (1960).

[2] K.I. Babenko, On the approximation of a certain class of periodic functions of several variables by trigonometric polynomials, Dokl. Akad. Nauk USSR, **132** (1960), 982–985; English transl. in Soviet Math. Dokl., **1** (1960).

[3] H.-J. Bungartz and M. Griebel, Sparse grids, *Acta Numerica* **13** (2004), 147–269.

[4] G. Byrenheid and T. Ullrich, Optimal sampling recovery of mixed order Sobolev embeddings via discrete Littlewood-Paley type characterizations, Analysis Mathematica, to appear.

[5] G. Byrenheid and T. Ullrich, Sparse approximation with respect to the Faber-Schauder dictionary in spaces of functions with bounded mixed derivative, Preprint Uni Bonn.

[6] A. Chernov and D. Dũng, New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness, J. Complexity **1**(2016), 92–121, arXiv:1309.5170v3 [math.NA]. .

[7] A. Cohen, R. DeVore, S. Foucart, H. Rauhut, Recovery of functions of many variables via compressive sensing, in: Proc. SampTA 2011, Singapore, 2011.

[8] F. Cobos, T. Kühn, W. Sickel, Optimal approximation of Sobolev functions in the sup-norm, *J. Funct. Anal.* **270**(11), pp. 4196–4212, 2016.

[9] D. Dũng, V.N. Temlyakov, T. Ullrich, Hyperbolic cross approximation, Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, to appear, arxiv:1601.03978v2.

[10] M. Gnewuch and H. Woźniakowski, Quasi-polynomial tractability, *J. Complexity* **27** (3-4) (2011), 312–330.

[11] M. Griebel, Sparse grids and related approximation schemes for higher dimensional problems. In: Proceedings of Foundations of Computational Mathematics, Santander 2005, pp. 106–161. London Mathematical Society Lecture Notes Series, 331, Cambridge University Press, Cambridge (2006)
[12] T. Kühn, W. Sickel and T. Ullrich, Approximation of mixed order Sobolev functions on the d-torus–Asymptotics, preasymptotics and d-dependence, *Constr. Approx.*, 42:353–398, 2015.
[13] B.S. Mityagin, Approximation of functions in the space L^p and C on the torus, Mat. Sb., **58** (1962), 397–414.
[14] H. Yserentant, *Regularity and Approximability of Electronic Wave Functions*. Lecture Notes in Mathematics, Springer (2010).

The reduced basis method, adaptivity and high-dimensionality

Karsten Urban

The Reduced Basis Method (RBM) has become a widely known and used numerical method for solving parameter-dependent partial differential equations (PPDEs). In this talk, we focus on two challenges for the RBM:

(1) Usually, the RBM relies on a detailed discretization (sometimes called *truth*) which is the same for all parameters. In case of a strong parameter dependence, this might be unrealistic or by far too costly.
(2) The curse of dimensionality might hit the RBM in two places, namely for the parameters and for the domain underlying the PPDE (time and space).

With regards to (1), we report on new results from [1], where we have shown how to use adaptive discretizations instead of a common truth. This has severe implications in particular for the online-efficient error estimation. We propose to use a wavelet representation of the residual and show that this is online-efficient. The corresponding Greedy method for selecting the reduced basis functions offline is shown to be convergent. Numerical results underline the efficiency of the arising method.

With respect to (2), we first consider high-dimensional space-time domains. We review the approach introduced in [7, 8] for instationary problems, where we propose to use time as an additional variable. This yields much sharper error/residual-relations then using standard time-marching schemes. We extend this to the wave equation and time-depend transport equations based upon the results for the stationary case from [5].

The next issue is a high-dimensional physical domain. There are quite well-developed tensor-based methods. Within the framework of the RBM, we need a quantitatively efficient adaptive high-dimensional method. Hence, we extend the methods from [2, 3] to an Adaptive Wavelet-Galerkin Method (AWGM) known from [4] for the non-tensor case. We construct a Hierarchical Tensor (HT)-variant of AWGM and show convergence and quasi-optimality of the resulting HT-AWGM.

Finally, we consider infinite-dimensional parameter spaces within the RBM, i.e. parameter functions. Based upon [6], we suggest to use an adaptive wavelet

method to identify the significant ingredients of a parameter function online (similar to signal processing). This is combined with an adaptive offline Greedy method introduced above. Numerical results for option pricing within the Heston model are presented.

This talk is based upon joint work with M. Ali, S. Glas, K. Steih, A. Mayerhofer (all Ulm), A. Patera (MIT) as well as J. Brunken and K. Smetana (Münster).

References

[1] M. Ali, K. Steih and K. Urban. Reduced basis methods with adaptive snapshot computations. Adv. Comput. Math. 43(2), 2017 257–294.

[2] M. Bachmayr and W. Dahmen. Adaptive low-rank methods: problems on Sobolev spaces. SIAM J. Numer. Anal., 54(2), 2016, 744–796.

[3] M. Bachmayr and W. Dahmen. Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. Found. Comput. Math., 15(4), 2015, 839–898.

[4] A. Cohen, W. Dahmen and R.A. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. Math. Comp., 70(233), 2001, 27–75.

[5] W. Dahmen, C. Huang, C. Schwab and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. SIAM J. Numer. Anal., 50(5), 2012, 2420–2445.

[6] A. Mayerhofer and K. Urban. A Reduced Basis Method for Parameter Functions using Wavelet Approximations. in: Model Reduction of Parametrized Systems, P. Benner, A. Patera, M. Ohlberger, G. Rozza and K. Urban (eds.), Springer, to appear.

[7] K. Urban and A.T. Patera. A new error bound for reduced basis approximation of parabolic partial differential equations. C.R. Math Acad Sci Paris 350, no. 3-4 (2012), 203–207.

[8] K. Urban and A.T. Patera. An improved error bound for reduced basis approximation of linear parabolic problems. Math. Comp. 83, no. 288 (2014), 1599–1615.

Polynomial approximation via convex and nonconvex optimization of high-dimensional functions

Clayton Webster

In this talk, we present a compressed sensing approach to polynomial approximation of complex-valued functions in high dimensions. Of particular interest is the parameterized PDE setting, where the target function is smooth, characterized by a rapidly decaying orthonormal expansion, whose most important terms are captured by a lower (or downward closed) set. By exploiting this fact, we develop a novel weighted minimization procedure with a precise choice of weights, and a modification of the iterative hard thresholding method, for imposing the downward closed preference. We will also present theoretical results that reveal our new computational approaches possess a provably reduced sample complexity compared to existing compressed sensing, least squares, and interpolation techniques. In addition, the recovery of the corresponding best approximation using our methods is established through an improved bound for the restricted isometry property. Finally, we will also present a new theory for compressed sensing that reveals that nonconvex minimizations are at least as good as ℓ_1 minimization in exact recovery of sparse signals. Our theoretical recovery guarantees are developed

through a unified null space property based-condition that encompasses all currently proposed nonconvex functionals in literature. Several nonconvex functionals will be explored and the specific conditions in order to guarantee improved recovery will be given. Numerical examples, related to polynomial approximation of complex-valued functions in high dimensions, will be provided to support the new theory and demonstrate the computational efficiency of both the the new weighted convex minimization method as well as several nonconvex regularizations.

Interpolation of functions with moving discontinuities by transformed snapshots

Gerrit Welper

In comparison to elliptic and parabolic problems, our abilities to simulate parametric or stochastic hyperbolic PDEs is still fairly limited. Two major obstructions that render these problems challenging and have been addressed only rather recently in the literature are the

(1) robustness of PDE solvers for parametric/stochastic problems
(2) efficiency of approximations of the PDEs' solutions.

The first problem arises in the context of intrusive methods such as reduced basis, POD or stochastic Galerkin. One obtains discretizations that are tailored to the parametric nature of these problems for which one then has to find robust solution methods. However, the talk is mainly concerned with the second problem. In all of the above methods and also in non-intrusive methods such as stochastic collocation or empirical interpolation, one needs efficient ways to approximate the parametric solution $u(x,t,\xi)$ that depends on physical variables x and t as well as random or non-random parameters ξ. Three staple methods are reduced basis approximations, POD and polynomial chaos expansions. All three have proven very efficient in the elliptic and parabolic regime but are problematic for hyperbolic problems mainly due to shock discontinuities (and kinks as well).

Only very recently some work on the efficient approximation of parameter dependent jump discontinuities has emerged [1, 2, 3, 4]. A central problem is that these jumps in physical space induce a jump in the parameter variables which is very detrimental for the above mentioned methods. For example, polynomial chaos uses polynomial approximation in the parameter variables which is not well suited for efficient approximation of jumps.

To this end, we introduce transformations of the physical domain that "undo" the movement of the jumps or in other words align the jumps locations. Say we wish to approximate an unknown solution $u(\cdot,\xi)$ by several snapshots $u(\cdot,\eta_i)$. We first construct transforms $\phi(\xi,\eta_i)(x,t)$ that align the jump locations of each snapshot to the target's jump set so that the transformed snapshots $u(\phi(\xi,\eta_i)(x,t),\eta_i)$ have jumps independent of the parameter η_i. Then, in order to obtain an approximation of $u(\cdot,\xi)$, we interpolate these transformed snapshots with respect to the interpolation points η_i. By construction, the jumps are no longer "visible"

in parameter direction so that the interpolation converges with up to exponential rates, in fact, as we shall demonstrate, to the solution $u(\cdot, \xi)$.

In order to calculate ϕ, we minimize the error $\sup_{\xi \in \mathcal{P}_T} \|u(\cdot, \xi) - u_n(\cdot, \xi; \phi)\|_{L_1}$ with respect to ϕ over a training sample $\mathcal{P}_T$ of parameters, where $u_n(\cdot, \xi; \phi)$ is the TSI described above. On first sight this seems to be a complicated optimization problem and indeed, one can construct simple examples which have unacceptable local minima. Nonetheless, we can split the transform into a series of local contributions that allow us to avoid bad local minima, altogether: Given transforms $\phi(\xi_i, \xi_{i+1})$ for nearby parameters, we can compose them

$$\phi(\xi_i, \xi_j) := \phi(\xi_i, \xi_{i+1}) \circ \cdots \circ \phi(\xi_{j-1}, \xi_j)$$

to obtain transformations for parameters that are far apart. We therefore only need to find the local contributions $\phi(\xi_i, \xi_{i+1})$ which are not prone to unacceptable local minima. This is proven rigorously for some $1d$ cases and demonstrated numerically for some $2d$ problems.

The above outlined method works well if we can align all discontinuities which, however, is not always the case. For example, if the number of jumps changes in parameter a proper alignment is not possible. In contrast to the parameter dependent movement of the jumps which induce singularities for every parameter, changes of the number of jumps (and similar non-align-able cases) are local in parameter space. Therefore, in principle, additional localization methods such as *hp*-adaption in parameter space can resolve these difficulties. The drawback of this approach is that is does not scale well to high parameter dimensions. To this end, we argue that the TSI allows us to construct a "tensorized" *hp*-adaptive method that is more promising for higher parameter dimensions.

References

[1] P.G. Constantine and G. Iaccarino. Reduced order models for parameterized hyperbolic conservations laws with shock reconstruction. Stanford Center for Turbulence Research Annual Research Briefs 2012, 2012.

[2] J.-F. Gerbeau and D. Lombardi. Approximated Lax pairs for the reduced order integration of nonlinear evolution equations. Journal of Computational Physics, 265, 2014, 246–269.

[3] T. Taddei and S. Perotto and A. Quarteroni. Reduced basis techniques for nonlinear conservation laws. ESAIM: M2AN, 49(3), 2015, 787–814.

[4] G. Welper. Transformed snapshot interpolation. 2015, Preprint: `http://arxiv.org/abs/1505.01227`.

Multivariate approximation for analytic functions with Gaussian kernels

Henryk Woźniakowski

We study d-variate analytic functions defined on the whole space which belong to a reproducing kernel Hilbert space whose kernel is Gaussian with shape parameters. We find necessary and sufficient conditions on various notions of tractability in terms of these shape parameters.

A class of variational multiscale methods based on subspace decomposition

Harry Yserentant

(joint work with Ralf Kornhuber and Daniel Peterseim)

Numerical homogenization tries to approximate the solutions of elliptic partial differential equations with strongly oscillating coefficients by functions from modified finite element spaces. I presented in this talk a class of such methods that are very closely related to the method of Målqvist and Peterseim [2]. Like the method of Målqvist and Peterseim, the new methods do not make explicit or implicit use of a scale separation and rest upon the localization and smoothing properties of elliptic equations. Their comparatively simple analysis is based on the theory of additive Schwarz or subspace decomposition methods.

References

[1] R. Kornhuber, D. Peterseim, H. Yserentant: An analysis of a class of variational multiscale methods based on subspace decomposition. `arXiv:1608.04081v2 [math.NA]`.

[2] A. Målqvist, D. Peterseim: Localization of elliptic multiscale problems. *Math. Comp.* 83:2583–2603, 2014.

Reporter: Angela Kunoth