Mathematisches Forschungsinstitut Oberwolfach

Report No. 36/2007

Wavelet and Multiscale Methods

Organised by
Albert Cohen (Paris)
Wolfgang Dahmen (Aachen)
Ronald A. DeVore (Columbia)
Angela Kunoth (Bonn)

July 29th – August 4th, 2007

Abstract. Various scientific models demand finer and finer resolutions of relevant features. Paradoxically, increasing computational power serves to even heighten this demand. Namely, the wealth of available data itself becomes a major obstruction. Extracting essential information from complex structures and developing rigorous models to quantify the quality of information leads to tasks that are not tractable by standard numerical techniques. The last decade has seen the emergence of several new computational methodologies to address this situation. Their common features are the nonlinearity of the solution methods as well as the ability of separating solution characteristics living on different length scales. Perhaps the most prominent examples lie in multigrid methods and adaptive grid solvers for partial differential equations. These have substantially advanced the frontiers of computability for certain problem classes in numerical analysis. Other highly visible examples are: regression techniques in nonparametric statistical estimation, the design of universal estimators in the context of mathematical learning theory and machine learning; the investigation of greedy algorithms in complexity theory, compression techniques and encoding in signal and image processing; the solution of global operator equations through the compression of fully populated matrices arising from boundary integral equations with the aid of multipole expansions and hierarchical matrices; attacking problems in high spatial dimensions by sparse grid or hyperbolic wavelet concepts.

This workshop proposed to deepen the understanding of the underlying mathematical concepts that drive this new evolution of computation and to promote the exchange of ideas emerging in various disciplines.

Mathematics Subject Classification (2000): 16xx (Numerical Analysis and Scientific Computing).

Introduction by the Organisers

Complex scientific models like climate models, turbulence, fluid structure interaction, and nanosciences, demand finer and finer resolution in order to increase 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. This manifests itself through recent developments in various areas. Examples include regression techniques such as projection pursuit in stochastic modeling, the investigation of greedy algorithms in complexity theory, or compression techniques and encoding in signal and image processing. Further representative examples are the compression of fully populated matrices arising from boundary integral equations through concepts like multipole expansions, panel clustering or, more generally, hierarchical matrices, and adaptive solution techniques in numerical simulation based on continuous models such as partial differential or integral equations.

The mathematical methods emerging to address these problems have several common features including the nonlinearity of the solution methods as well as 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 multigrid methods as a key methodology that has advanced the frontiers of computability for certain problem classes in numerical analysis. In fact, the separation of frequencies plays an important role in preconditioning linear systems arising from elliptic partial differential equations so that the corresponding large scale systems could be solved with discretization error accuracy optimally in linear time.

A related but different concept for managing the interaction of different length scales centers on wavelet bases and multilevel decompositions. In the very spirit of harmonic analysis they allow one to decompose complex objects into versatile and simple building blocks that again support analyzing multiscale features.

While this ability was exploited first primarily for treating *explicitly* given objects, like digital signals and images or data sets, the use of such concepts for recovering also *implicitly* given objects, like solutions of partial differential or boundary integral equations, has become a major recent focus of attention. 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 and best N-term approximation based on wavelet expansions, concepts from nonlinear approximation theory and harmonic analysis become practically manageable. In our opinion, these ideas open promising perspectives not only for signal and image processing but also for the numerical analysis of differential and integral equations covering, in particular, such operator equations with stochastic data.

These two research areas have developed relatively independently of one another. Our first Oberwolfach Workshop 'Wavelet and Multiscale Methods' held in July 2004 sought to bring various disciplines utilizing multiscale techniques together by inviting leading experts and young emerging scientists in areas that rarely interact. That workshop not only accelerated the advancement of nonlinear and multiscale methodologies but also provided beneficial cross fertilizations to an array of diverse disciplines which participated in the workshop, see the Oberwolfach Report 34/2004. Among the several recognizable outcomes of the workshop 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.

The workshop *Wavelet and Multiscale Methods* organised by Albert Cohen (Paris), Wolfgang Dahmen (Aachen), Ronald A. DeVore (Columbia) and Angela Kunoth (Bonn) was held July 29th – August 4th, 2007. This meeting was well attended with over 50 participants with broad geographic representation from all continents. It was a nice blend of researchers with various backgrounds described in the following.

Compressed sensing, as being developed by Candes, Donoho, Vershynin, Gilbert, Strauss, and others advocates a fascinating alternative to the usual sensing and compression methodology. The classical model of limited bandwidth is replaced by sparsity models and the role of traditional sampling is played by sensing functionals that are typically based on random vectors. One can then prove that under certain circumstances by far fewer observations are needed to record all the information required to encode sparse signals. Adaptive methods for numerically solving a wide range of equations with proven optimality (in terms of the number of computations needed to achieve a prescribed error tolerance) originally involved coarsening procedures. The necessity of such coarsening was brought into question at the previous workshop and subsequent work of Stevenson has shown that it is possible to avoid coarsening for scalar elliptic problems through cautious bulk chasing.

As in the previous workshop, the participants are experts in areas like nonlinear approximation theory (e.g., DeVore, Temlyakov), statistics (e.g., Picard, Kerkyacharian), finite elements (e.g., Braess, Oswald, Xu), multigrid methods (e.g., Braess, Hackbusch), spectral methods (e.g., Canuto), harmonic analysis and wavelets (e.g., Cohen, Daubechies, Petrushev, Schneider, Stevenson), numerical fluid mechanics (e.g., Süli), conservation laws (e.g., Tadmor) or systems of stationary operator equations (e.g., Dahmen, Kunoth, Schwab). One of the main objectives of this workshop was to foster synergies by the interaction of scientists from different disciplines resulting in more rapid developments of new methodologies in these various domains. It also served to bridge theoretical foundations with applications.

Examples of conceptual issues that were advanced by the workshop were: convergence theory for adaptive multilevel methods for high-dimensional PDEs; extension of fast solution methods such as multigrid and multiscale methods to more complex models such as control problems involving partial differential equations, and partial differential equations with stochastic data; adaptive multiscale methods for coupled systems involving partial differential and integral equations; incorporating anisotropy in analysis, estimation, compression and encoding; adaptive treatment of nonlinear and time–dependent variational problems; interaction of different scales under nonlinear mappings, e.g., for flow problems and for problems with stochastic data.

We feel that our workshop propelled further advancement of several emerging areas: the numerical aspects of complete sensing including stability and optimality; deterministic methods for complete sensing based on coding theory; the design and analysis of universal estimators in nonparametric statistical estimation and machine learning — nonlinear multiscale techniques may offer much more efficient alternatives to schemes based on complexity regularization; solution concepts for problems of high spatial dimension by utilizing anisotropy, for instance, in mathematical finance, in quantum chemistry and electronic structure calculations.

In summary, we find that the conceptual similarities that occur in these diverse application areas suggested a wealth of synergies and cross-fertilization. These concepts are in our opinion not only relevant for the development of efficient solution methods for large scale problems but also for the formulation of rigorous mathematical models for quantifying the extraction of essential information from complex objects.

Workshop: Wavelet and Multiscale Methods

Table of Contents

Dietrich Braess (joint with Joachim Schöberl)
Reliable A Posteriori Error Estimates by the Hypercircle Method 2085

Martin Campos Pinto (joint with Kolja Brix and Wolfgang Dahmen)
Multilevel Preconditioners for one Discontinuous Galerkin Method 2088

Albert Cohen (joint with Wolfgang Dahmen and Ron DeVore)
Instance Optimality in Compressed Sensing 2091

Stephan Dahlke (joint with Massimo Fornasier, Thorsten Raasch, Rob Stevenson and Manuel Werner)
Adaptive Frame Schemes for Elliptic Operator Equations 2094

Ingrid Daubechies (joint with Eugene Brevdo and Shannon Hughes)
Multiscale Analysis of Vincent Van Gogh's Paintings 2097

Hans G. Feichtinger
Banach Gelfand Triples in Classical Fourier Analysis 2100

Anna C. Gilbert (joint with Martin Strauss, Joel Tropp and Roman Vershynin)
One Sketch for all: Fast Algorithms for Compressed Sensing 2101

Karlheinz Gröchenig (joint with Andreas Klotz)
Matrices with Off-Diagonal Decay and their Inverses 2103

C. Sinan Güntürk
Approximation and Interpolation by Power Series with ± 1 Coefficients . 2105

Wolfgang Hackbusch
Convolution of hp-Functions on Locally Refined Grids 2107

Helmut Harbrecht (joint with Reinhold Schneider and Christoph Schwab)
Sparse Second Moment Analysis for Elliptic Problems in Stochastic Domains 2111

Gitta Kutyniok (joint with S. Dahlke, T. Sauer, G. Steidl and G. Teschke)
Shearlets: A Wavelet-Based Approach to the Detection of Directional Features 2113

Peter Oswald (joint with S. Harizanov)
Nonlinear Multiresolution Analysis 2117

Roland Pabel
Wavelet Methods for PDE-Constrained Elliptic Control Problems with Dirichlet Boundary Control 2120

Pencho Petrushev (joint with George Kyriazis and Yuan Xu)
Decomposition of Weighted Besov and Triebel-Lizorkin Spaces and Nonlinear Approximation on the Ball 2124

Dominique Picard (joint with P. Baldi, G. Kerkyacharian and D. Marinucci)
Needlets on the Sphere and Applications to the Cosmological Microwave Background 2127

Bojan Popov (joint with Jean-Luc Guermond)
L_1 Approximation of Hamilton-Jacobi Equations 2129

Holger Rauhut (joint with Götz Pfander and Jared Tanner)
Identification of Sparse Operators 2130

Reinhold Schneider
Adaptive Coupled Cluster Method and CI Method for the Solution of the Electronic Schroedinger Equation 2133

Christoph Schwab (joint with Rob Stevenson)
Space Time Adaptive Wavelet Methods for Parabolic Problems 2137

Rob Stevenson (joint with Christoph Schwab)
An Adaptive Wavelet Method for Solving High-Dimensional Elliptic PDE's 2139

Endre Süli (joint with Christoph Schwab and Radu-Alexandru Todor)
Sparse Finite Element Approximation of High-Dimensional Transport-Dominated Diffusion Problems 2143

Anita Tabacco (joint with E. Cordero, F. De Mari and K. Nowak)
Analytic and Geometric Features of Reproducing Groups 2146

Eitan Tadmor
Edges and Hierarchy of Scales 2150

Vladimir Temlyakov
Relaxation in Greedy Approximation 2155

Karsten Urban (joint with Kai Bittner)
Can We Use Semiorthogonal Spline Wavelets for Adaptively Solving Nonlinear Problems? 2158

Przemysław Wojtaszczyk
Tresholding and Compressed Sensing 2162

Jinchao Xu (joint with Ming Wang)
Consistent Piecewise Polynomial Approximation of $H^m(\Omega)$ Functions in R^n for any $0 \leq m \leq n$ 2165

Abstracts

Reliable A Posteriori Error Estimates by the Hypercircle Method

Dietrich Braess
(joint work with Joachim Schöberl)

The talk is concerned with a posteriori error estimates for finite element solutions of elliptic differential equations. Specifically we want upper estimates that have no generic constant in the main term. For convenience, we restrict ourselves here to the Poisson equation in a two-dimensional domain Ω,

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{1}$$

and to linear elements on a partition $\mathcal{T}_h$ of Ω into triangles. Here the mixed method for the Poisson equation will also be important

$$\begin{aligned} \sigma &= \nabla u \qquad \text{in } \Omega, \\ \operatorname{div}\sigma &= -f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned} \tag{2}$$

A flux σ which satisfies the second equation in (2) is called *equilibrated*. The point of departure is the following theorem. Γ_D and Γ_N are the parts of the boundary with Dirichlet and Neumann boundary conditions, respectively. All norms are L_2 norms.

Theorem of Prager and Synge (Two-Energies-Principle).
Let $\sigma \in H(\mathrm{div})$, $\sigma \cdot n = 0$ on Γ_N
while $v \in H^1(\Omega)$, $v = 0$ on Γ_D and assume that

$$\operatorname{div}\sigma + f = 0. \tag{3}$$

Furthermore, let u be the solution of the Poisson equation. Then,

$$\|\nabla u - \nabla v\|^2 + \|\nabla u - \sigma\|^2 = \|\nabla v - \sigma\|^2. \tag{4}$$

There is much freedom in choosing v and σ. We also find the name *hypercircle method* in connection with the theorem. We emphasize that it is not restricted to the Poisson equation. We will refer to some other elliptic problems for which there are also theorems of Prager–Synge type, at the end of this abstract.

Let $v = u_h$ be a finite element solution for which an a posteriori estimate is wanted. The crucial step is the construction of an equilibrated flux σ. In contrast to Neittaanmäki and Repin [8] we perform the construction by computing a correction $\sigma^\Delta := \sigma - \nabla u_h$ to the given gradient of u_h, i.e., we use the information that we have a finite element solution.

Following [5] the computation will be performed for the *broken Raviart–Thomas space* of lowest order

$$\mathcal{RT}_{-1} := \{\tau \in L_2(\Omega);\ \tau|_T = a_T + b_T x,\ a_T \in \mathbb{R}^d,\ b_T \in \mathbb{R}\ \forall T\},$$

and the triangulation is the same as that for which the finite element solution was computed. The subspace of functions with continuous normal components is the usual space

$$\mathcal{RT}_0 := \mathcal{RT}_{-1} \cap H(\mathrm{div}).$$

Furthermore we denote the space of piecewise linear functions by $\mathcal{M}^0$.

The first step of the construction brings a separation of the data oscillation. Given f on the right-hand side of (1), let $\bar{f}$ be the piecewise constant function which results from the L_2 projection of f. The quantity

$$ch\|f - \bar{f}\|$$

is called the *data oscillation*. It reflects the error if f is replaced in (1) by $\bar{f}$, and it is found in most a posteriori error estimates. Since it is a term of higher order, we admit here a generic constant. Moreover, c depends only on the shape parameter of the triangulation.

Now we make an excursion to the mixed method by Raviart–Thomas

$$\begin{aligned} (\sigma_h, \tau) + (\mathrm{div}\,\tau, w_h) &= 0 && \forall \tau \in \mathcal{RT}_0 \\ (\mathrm{div}\,\sigma_h, v) &= -(\bar{f}, v) && \forall v \in \mathcal{M}^0. \end{aligned} \tag{5}$$

Note that σ_h is a piecewise linear function. Therefore, $\mathrm{div}\,\sigma_h$ is piecewise constant as well as $\bar{f}$ is by definition. Since we also test with functions in $\mathcal{M}^0$, it follows that

$$\mathrm{div}\,\sigma_h = -f \tag{6}$$

holds in the classical sense. In particular σ_h is equilibrated. It is easy to show that σ_h is even the equilibrated function in $\mathcal{RT}_0$ for which $\|\sigma - \nabla u_h\|$ is minimal.

The computation of the solution of (5), however, is considered as too expensive for an a posteriori error estimation. Fortunately there is a cheap local procedure that provides a suitable approximation.

Consider a vertex V of the triangulation $\mathcal{T}_h$ and let ω_V denote the patch of triangles around V:

$$\omega_V := \bigcup_{V \in \bar{T}} T\,.$$

The nodal basis function φ_V with $\varphi_V(V) = 1$ and support ω_V is inserted into the finite element equation

$$\int_{\omega_V} \nabla u_h \cdot \nabla \varphi_V = \int_{\omega_V} f \varphi_V\,.$$

By partial integration we see that the left-hand side equals $\sum_{E \subset \omega_V} \int_E [\nabla u_h \cdot n] \varphi_V$. We recall that $\mathrm{div}\,\sigma = -\bar{f}$ is the aim. Since all factors in the integrals are now piecewise linear or piecewise constant, we conclude that in the 2-dimensional case

$$\frac{1}{2} \sum_{E \subset \omega_V} [\nabla u_h \cdot n]_E\, |E| = \frac{1}{3} \sum_{T \subset \omega_V} \mathrm{div}\,\sigma_T\, |T|\,.$$

For this reason, we can shift one half of the jumps of $\nabla u_h \cdot n$ to obtain a Raviart–Thomas function σ_{ω_V} with one third of the required divergence in all triangles of

the patch. The algorithmic implementation for problems in 2-space has a simple geometric interpretation and is described in [4]. Otherwise merely small systems of algebraic equations have to be solved.

By repeating the procedure on all patches ω_V for all vertices V, we encounter each edge twice and each triangle three times. Hence, the sum $\sigma^\Delta := \sum_V \sigma_{\omega_V}$ yields a function $\sigma := \nabla u_h + \sigma^\Delta$ with $\operatorname{div} \sigma = -\bar{f}$. Now the theorem of Prager and Synge provides the guaranteed estimate

$$\|\nabla u - \nabla u_h\| \le \|\sigma^\Delta\| + ch\|f - \bar{f}\| \,. \tag{7}$$

This is the required upper estimate. The equivalence of the expression on the right-hand side of (7) with the residual estimator implies that the new estimator is not only reliable but also efficient.

We will comment on the differences to similar procedures in the literature and note that small changes may have much impact on the computing effort. Moreover, applications to quite different elliptic problems will be listed (without saying how our variant has to be adapted). We emphasize that often the connection between papers on this topic is not recognized at first glance, if the theorem of Prager and Synge, the hypercircle method, and the two-energies-method are not mentioned. Because of the lack of space, we will cite explicitly only one representer of closely related research.

Remarks.

1. The inequality (7) is written as an estimator for u_h, although it follows from (4) that the error of the mixed method by Raviart–Thomas on the same grid is also included. A comparison of the error of linear elements and of the Raviart–Thomas elements can be derived from this fact [4].

2. There is a strong similarity with the error estimators by local Neumann problems [2]. The construction in [2], however, is performed in infinite dimensional spaces. Moreover, the normal components of the equilibrated fluxes on edges are linear functions.

3. The construction of equilibrated fluxes in [6] refer to finer grids. The treatment of estimators for the Lamé equation in [9] is closer to our concept.

4. The approach of Repin (see, e.g., [8]) is directed to arbitrary approximate functions with respect to the elliptic problem, and their origin is not used for the construction. The latter is therefore more expensive.

5. Edge elements and the equations of magnetostatics are easily treated in the framework of the two-energies-principle [5].

6. A theorem of Prager–Synge type can be formulated for variational inequalities as found in obstacle problems or contact problems. An implementation without extra terms, however, is restricted to active sets with some regularity; see, e.g., [11].

7. When the theorem of Prager and Synge and (4) are applied for estimating finite element solutions of the mixed method, the construction proceeds in the opposite direction. Now the other term on the left-hand side of (4) must not be

too large. For this reason quadratic elements and not merely linear elements are used in [1].

8. The two-energies-principle has also been applied on the continuous level in order to justify or discard plate models [3, 7]. Admissible functions for the mixed method in 3-space are constructed from the solutions in the lower dimensional spaces.

References

[1] M. Ainsworth, *A posteriori error estimation for lowest order Raviart Thomas mixed finite elements.* Preprint 2007

[2] M. Ainsworth and T.J. Oden, A Posteriori Error Estimation in Finite Element Analysis. Wiley, Chichester 2000.

[3] A.L. Alessandrini, D.N. Arnold, R.S. Falk, and A.L. Madureira, *Derivation and justification of plate models by variational methods.* In "Plates and Shells, Quebec 1996", (M. Fortin, ed.), pp. 1-20. CRM Proceeding and Lecture Notes, vol. 21, American Mathematical Society, Providence, RI 1999.

[4] D. Braess, Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics. 3rd edition. Cambridge University Press 2007.

[5] D. Braess and J. Schöberl, *Equilibrated residual error estimator for Maxwell's equations.* Math. Comp. (to appear).

[6] R. Luce and B. Wohlmuth, *A local a posteriori error estimator based on equilibrated fluxes.* SIAM J. Numer. Anal. **42** (2004), 1394–1414.

[7] D. Morgenstern, *Herleitung der Plattentheorie aus der dreidimensionalen Elastizitätstheorie.* Arch. Ration. Mech. Anal. **4** (1959), 145-152.

[8] P. Neittaanmäki and S. Repin, Reliable Methods for Computer Simulation. Error control and a posteriori estimates. Elsevier. Amsterdam 2004.

[9] S. Nicaise, K. Witowski, and B. Wohlmuth, *An a posteriori error estimator for the Lamé equation based on $H(\mathrm{div})$-conforming stress approximation.* IANS report 2006/005, University of Stuttgart.

[10] W. Prager and J.L. Synge, *Approximations in elasticity based on the concept of function spaces.* Quart. Appl. Math. **5** (1947), 241–269.

[11] B. Wohlmuth, *A posteriori error estimator and error control for contact problems.* IANS report 12/2007, University of Stuttgart.

Multilevel Preconditioners for one Discontinuous Galerkin Method

Martin Campos Pinto

(joint work with Kolja Brix and Wolfgang Dahmen)

One major advantage of DG Finite Element schemes is the great freedom allowed in the design of the trial spaces, due to the complete lack of constraint between two neighboring elements. In this work, we consider the situation where we want to solve an elliptic model problem, say: find $u \in H_0^1(\Omega)$ such that

$$a(u,v) := \langle A\nabla u, \nabla v\rangle + \langle bu, v\rangle = \langle f, v\rangle,\ v \in H_0^1(\Omega), \tag{1}$$

A being a s.p.d. 2×2 matrix and b some nonnegative bounded function, using as trial functions *discontinuous* piecewise polynomials of variable order. Thus we set

$$V_h := \mathbb{P}_k(\mathcal{T}_h) = \{v \in L^2 : v_T \in \mathbb{P}_k(T), \forall T \in \mathcal{T}_h\},$$

where $\mathcal{T}_h$ denotes a (possibly) non-conforming and highly non-uniform mesh obtained by recursive refinements of some coarse initial triangulation of a bounded polygonal $\Omega \subset \mathbb{R}^2$. Thereby, $k = k(T)$ resp. $h = h(T)$, $T \in \mathcal{T}_h$, should be seen as a piecewise constant function on Ω representing the maximal degree of the elements on the triangle T, resp. the diameter of T. On such trial spaces, the Symmetric Interior Penalty Galerkin method introduced in the early 1970s reads as follows, see e.g. [1]: find $u_h \in V_h$ such that

$$a_h(u_h, v) = \langle f, v\rangle, \ \forall v \in V_h \tag{2}$$

where the mesh dependent, symmetric bilinear form a_h is given by

$$a_h(v,w) := \sum_{K\in\mathcal{T}_h} a(v,w)_K - \sum_{e\in\mathcal{E}_h}\int_e \big(\{\nabla w\}\cdot[v] + \{\nabla v\}\cdot[w]\big) + \sum_{e\in\mathcal{E}_h}\frac{\gamma}{|e|}\int_e [w]\cdot[v],$$

denoting as usual by $[\cdot]$, resp. $\{\cdot\}$, the jumps, resp. averages of piecewise smooth (polynomial) functions across the edges. For sufficiently large γ, this method is known to be well posed in V_h when equipped with the mesh dependent norm

$$|||v|||_h^2 := \sum_{K\in\mathcal{T}_h} a(v,v)_K + \sum_{e\in\mathcal{E}_h}\frac{1}{|e|}\|[v]\|^2_{L^2(e)} \tag{3}$$

i.e. for any v and w in V_h we have

$$c|||v|||_h^2 \le a_h(v,v) \quad \text{and} \quad a_h(v,w) \le C|||v|||_h|||w|||_h$$

(here and below, c and C denote generic constants independent of the mesh sizes h). Less has been shown, however, about the efficiency of the method. In particular, it is well known that -as it happens with conforming finite elements- the condition number $\kappa(A^\Phi) := \|A^\Phi\|\|(A^\Phi)^{-1}\|$ of the stiffness matrix $A^\Phi := (a_h(\phi_i,\phi_k))_{i,k\in\mathcal{I}_h}$ grows like h^{-2}, $h = \inf\{h(T) : T \in \mathcal{T}_h\}$ when $\Phi_h = \{\phi_i : i \in \mathcal{I}_h\}$ is a standard nodal basis of V_h. Preconditioning of elliptic operators, of course, has been thoroughly studied and in the conforming case, that is when the finite element space is embedded in $H^1_0(\Omega)$, several multilevel methods have been developed which are asymptotically optimal, see e.g. [2], [6], [9], [7]. Optimal results have also been obtained for certain classes of *nonconforming* elements (excluding DG), see e.g. [3], [10], [11]. We could summarize these findings as follows: building an optimal preconditioner for (2) is closely related with finding a stable splitting of the space V_h, which can either be a stable (e.g. wavelet-type) *basis* $\Psi = \{\psi_\lambda : \lambda \in \Lambda_h\}$ satisfying

$$c|||v|||_h^2 \le \sum_{\lambda\in\Lambda_h} |||d_\lambda\psi_\lambda|||_h^2 \le C|||v|||_h^2 \quad \text{for any} \quad v = \sum_{\lambda\in\Lambda_h} d_\lambda\psi_\lambda \in V_h,$$

in which case the preconditioner essentially consists in a change of basis, or more generally a *redundant* stable splitting, i.e. a collection $\mathcal{S}_h = \{V_\gamma : \gamma \in \Gamma_h\}$ of subspaces spanning V_h in such a way that

$$c|||v|||_h^2 \le \inf_{\substack{v_\gamma\in V_\gamma \\ v=\sum_{\gamma\in\Gamma_h} v_\gamma}} \left\{\sum_{\gamma\in\Gamma_h} |||v_\gamma|||_h^2\right\} \le C|||v|||_h^2$$

holds for any $v \in V_h$. In this case an optimal preconditioner is obtained through an additive Schwarz scheme based on the splitting, [9], [7].

Let us remind that in the case of conforming spaces, such (H^1-)stable splittings are very simple to build when the underlying triangulation $\mathcal{T}_h$ is obtained by recursive, shape regular refinements of some coarse $\mathcal{T}^0$, and for sake of simplicity we only consider piecewise affine elements, i.e. $V_h^c := \mathbb{P}_1(\mathcal{T}_h) \cap H_0^1$. In such cases indeed, it is possible to define $\mathcal{T}_{h,j}$, $j = 1, 2, \dots j_h$, as the successive unions of all level j triangles that appear in the refinement history leading to $\mathcal{T}_h$, and $\mathcal{N}_{h,j}^c$ as the vertices of $\mathcal{T}_{h,j}$ that lie in the interior of $\Omega_{h,j} := \cup\{T : T \in \mathcal{T}_{h,j}\}$. Next, for every $n \in \mathcal{N}_{h,j}^c$ we let $\phi_{j,n}^c$ denote the standard nodal (piecewise affine) continuous hat function at vertex n supported on the star of triangles in $\mathcal{T}_{h,j}$ sharing n. The multilevel collection of nodal bases

$$\mathcal{S}_h^c := \{\mathrm{Span}(\phi_{j,n}^c) : j = 1, 2, \dots j_h,\ n \in \mathcal{N}_{h,j}^c\} \tag{4}$$

is then a H_0^1-stable splitting of $\mathbb{P}_1(\mathcal{T}_h) \cap H_0^1(\Omega)$, see [7].

Clearly, the above described results apply in the DG setting, but up to now no bases -or splittings- were known to be stable in the DG norm (3). Inspired by the wavelet examples, we first tried to build stable bases possessing a multilevel structure, i.e. for which any local refinement of the mesh entails in adding a few functions to the basis, and found the following *continuity* constraint.

Theorem 1. *If Ψ_h is a multilevel and DG-stable basis of $V_h := \mathbb{P}_k(\mathcal{T}_h)$, then it contains one subset that is a stable basis of the conforming part $V_h \cap H_0^1(\Omega)$.*

In particular, this means that a multilevel, stable Ψ_h must contain continuous basis functions at any level, which is a rather bad news considering the full discontinuous structure of V_h. Moreover, such a constraint complicates the construction of a multilevel basis. Therefore we turned to redundant splittings, and found the following result which holds as long as $\mathcal{T}_h$ satisfies some grading (implying, in particular, that every edge has at most one hanging point), see [4] for further details.

Theorem 2. *Let Φ_h be one standard nodal basis of $V_h := \mathbb{P}_k(\mathcal{T}_h)$. Then*

$$\mathcal{S}_h := \mathcal{S}_h^c \cup \{\mathrm{Span}(\phi_i) : \phi_i \in \Phi_h\}$$

is a DG-stable splitting of the full space V_h, $\mathcal{S}_h^c$ denoting the conforming splitting of $\mathbb{P}_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ introduced in (4).

It should be emphasized that these results match very well the theoretical framework of the auxiliary space method and two-level techniques developed in [3], [10] and [11] for nonconforming elements. In particular, this implies that the corresponding additive Schwarz scheme solves the original problem (1) with a condition number that is bounded independently of the mesh sizes h. We also found that it is possible to build stable splittings with high order conforming parts, yielding probably better robustness of the condition numbers with respect to the diffusion coefficients.

References

[1] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., **39** (2002), pp. 1749–1779.

[2] J. H. Bramble, J. E. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comp., **55** (1990), pp. 1–22.

[3] S. Brenner, *Two-level additive Schwarz preconditioners for nonconforming finite element methods*, Math. Comp., **65** (1996), pp. 897–921.

[4] K. Brix, M. Campos Pinto and W. Dahmen, *A multilevel preconditioner for the interior penalty discontinuous Galerkin method*, preprint of IGPM, RWTH-Aachen, Germany.

[5] W. Dahmen, *Wavelet and Multiscale Methods for Operator Equations*, Acta Numerica, Cambridge University Press, **6** (1997), pp. 55–228.

[6] W. Dahmen and A. Kunoth, *Multilevel Preconditioning*, Numer. Math., **63** (1992), pp. 315–344.

[7] M. Griebel and P. Oswald, *On the abstract theory of additive and multiplicative Schwarz algorithms*, Numer. Math., **70** (1995), pp. 163–180.

[8] J. Gopalakrishnan and G. Kanschat, *A multilevel discontinuous Galerkin method*, Numer. Math., **95** (2003), pp. 527–550.

[9] P. Oswald, *Multilevel Finite Element Approximation - Theory and Applications*, Teubner Skripten zur Numerik, B.G. Teubner, Stuttgart, 1994.

[10] P. Oswald, *Preconditioners for nonconforming discretizations*, Math. Comp., **65** (1996), pp. 923–941.

[11] J. Xu, *The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids*, Computing, **56** (1996), pp. 215–235.

Instance Optimality in Compressed Sensing

Albert Cohen

(joint work with Wolfgang Dahmen and Ron DeVore)

The typical paradigm for obtaining a compressed version of a discrete signal represented by a vector $x \in \mathbb{R}^N$ is to choose an appropriate basis, compute the coefficients of x in this basis, and then retain only the k largest of these with $k < N$. If we are interested in a bit stream representation, we also need in addition to quantize these k coefficients.

Assuming, without loss of generality, that x already represents the coefficients of the signal in the appropriate basis, this means that we pick an approximation to x in the set Σ_k of k-sparse vectors

$$\Sigma_k := \{x \in \mathbb{R}^N \ : \ \#\,\mathrm{supp}(x) \le k\}, \tag{1}$$

where $\mathrm{supp}(x)$ is the support of x, i.e., the set of i for which $x_i \neq 0$, and $\#A$ is the number of elements in the set A. The best performance that we can achieve by such an approximation process in some given norm $\|\cdot\|_X$ of interest is described by the *best k-term approximation* error

$$\sigma_k(x)_X := \inf_{z \in \Sigma_k} \|x - z\|_X. \tag{2}$$

This approximation process should be considered as *adaptive* since the indices of those coefficients which are retained vary from one signal to another. On the other

hand, this procedure is stressed on the front end by the need to first compute all of the basis coefficients. The view expressed by Candès, Romberg, and Tao [2, 3] and Donoho [5] is that since we retain only a few of these coefficients in the end, perhaps it is possible to actually compute only a few *non-adaptive* linear measurements in the first place and still retain the necessary information about x in order to build a compressed representation. Similar ideas have appeared in data streaming see e.g. [4, 6].

These ideas have given rise to a very lively area of research called *compressed sensing* which poses many intriguing questions, of both a theoretical and practical flavor. Here, we focus our interest on the question of just how well compressed sensing can perform in comparison to best k-term approximation.

To formulate the problem, we are given a budget of n questions we can ask about x. These questions are required to take the form of asking for the values $\lambda_1(x), \ldots, \lambda_n(x)$ where the λ_j are fixed linear functionals. The information we gather about x can therefore by described by

$$y = \Phi x, \tag{3}$$

where Φ is an $n \times N$ matrix called the *encoder* and $y \in I\!R^n$ is the *information vector*. The rows of Φ are representations of the linear functionals λ_j, $j = 1, \ldots, n$.

To extract the information that y holds about x, we use a *decoder* Δ which is a mapping from $I\!R^n \to I\!R^N$. We emphasize that Δ is not required to be linear. Thus, $\Delta(y) = \Delta(\Phi x)$ is our approximation to x from the information we have retained. The main question that we want to address here is:

For a given norm $\|\cdot\|_X$ and $k < N$, what is the minimal value of n, for which there exists an encoding-decoding pair (Φ, Δ) such that

$$\|x - \Delta(\Phi x)\|_X \le C_0 \sigma_k(x)_X, \tag{4}$$

for all $x \in I\!R^N$, with C_0 a constant independent of k and N?

We shall say a pair (Φ, Δ) satisfying (4) is *instance optimal* of order k with constant C_0 for the space X. In particular, we want to understand under what circumstances the minimal value of n is roughly of the same order as k. We shall see below that the answer to this question strongly depends on the norm X under consideration.

The approximation accuracy of a compressed sensing matrix is determined by the null space

$$\mathbf{N} = \mathbf{N}(\Phi) := \{x \in I\!R^N : \Phi x = 0\}. \tag{5}$$

The importance of $\mathbf{N}$ is that if we observe $y = \Phi x$ without any a-priori information on x, the set of z such that $\Phi z = y$ is given by the affine space

$$\mathbf{F}(y) := x + \mathbf{N}. \tag{6}$$

The importance of the null space for studying instance-optimality is expressed by the following theorem of [1].

Theorem 1. *Given an $n \times N$ matrix Φ, a norm $\|\cdot\|_X$ and a value of k, then a sufficient condition that there exists a decoder Δ such that* (4) *holds with constant C_0 is that*

$$\|\eta\|_X \leq \frac{C_0}{2}\sigma_{2k}(\eta)_X, \quad \eta \in \mathbf{N}. \tag{7}$$

A necessary condition is that

$$\|\eta\|_X \leq C_0\sigma_{2k}(\eta)_X, \quad \eta \in \mathbf{N}. \tag{8}$$

The conditions (7) and (8) in this theorem essentially mean that no element in the null space can have most of its X norm concentrated on $2k$ of its coordinates. Theorem 1 can be used to characterize the amount of measurement n needed for instance-optimality in specific norms. We give below the corresponding results in the two cases $X = \ell^1$ and $X = \ell^2$ (see [1] for more detail).

Theorem 2. *There exists pairs (Φ, Δ) which are instance optimal of order k for the space ℓ^1, provided that $k \leq cn\log(N/n)$ where c is a fixed constant.*

Concrete example of such pairs are produced by matrices Φ which satisfies the so-called *restricted isometry property* of order $3k$

$$(1-\delta)\|x\|_{\ell^2}^2 \leq \|\Phi x\|_{\ell^2}^2 \leq (1+\delta)\|x\|_{\ell^2}^2, \quad x \in \Sigma_{3k}, \tag{9}$$

with $\delta < (\sqrt{2}-1)^2/3$, and by the decoder Δ defined by ℓ^1 minimization

$$\Delta(x) := \mathrm{Argmin}_{z \in \mathbf{F}(y)}\|z\|_{\ell^1}. \tag{10}$$

It is known that matrices which satisfy (9) exists provided that $k \leq cn\log(N/n)$ with $c = c(\delta)$, however all known constructions are so far based on probabilities, i.e. by proving that (9) holds with high probabiliy for certain classes random matrices. Typical examples are when the entries of Φ are i.i.d. Bernoulli variables $\frac{\pm 1}{\sqrt{n}}$ or Gaussians of variance $1/n$.

Theorem 3. *If a pair (Φ, Δ) is instance optimal of order $k = 1$ with constant C_0 for the space ℓ^2, then necessarily $n \geq c_0 N$ with $c_0 = \frac{1}{C_0^2}$.*

Theorem 3 shows that instance-optimality is not a viable concept in ℓ^2 since even with $k = 1$ a very large number of measurements might be necessary to achieve the desired accuracy. It is yet possible to recover instance-optimality in ℓ^2 with a small number of measurements closer to k, if we accept a slightly weaker statement involving probability. In such a statement, the matrix Φ is drawned from a probability law such as i.i.d. Bernoulli variables $\frac{\pm 1}{\sqrt{n}}$ or Gaussians of variance $1/n$ as entries.

Theorem 4. *Given any $b > 0$, there exists a $c > 0$ such that if $k \leq cn \log(N/n)$, there exists a decoder Δ such that for all $x \in \mathbb{R}^N$,*

$$\|x - \Delta(\Phi x)\|_{\ell^2} \leq C_0 \sigma_k(x)_{\ell^2} \tag{11}$$

holds with probability larger than $1 - e^{-bn}$.

It should be noted that similar results of instance optimality involving probability have been obtained in the theoretical computer science approach to compressed sensing [4, 6]. In these approaches the matrix Φ is of a different nature than Bernoulli or Gaussian. One specific interest of considering Gaussian matrices is that they are invariant with respect to an orthonormal change of coordinates. This allows us to perform the measurements on arbitrary signals which are known to be well compressed in a common orthogonal basis. However, the decoder involved in the proof of Theorem 4 does not have low algorithmic complexity. At the present stage it is an open problem to find a fast decoder enjoying the same instance optimality property.

References

[1] A. Cohen, W. Dahmen and R. DeVore, *Compressed sensing and best k-term approximation*, submitted, 2007.

[2] E. Candès, J. Romberg and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inf. Theory **52** (2006), 489–509.

[3] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure and Appl. Math. **59** (2006), 1207–1223.

[4] G. Cormode and S. Muthukrishnan, *Towards an algorithmic theory of compressed sensing*, Technical Report 2005-25, DIMACS, 2005, also, Proc. SIROCCO, 2006.

[5] D. Donoho, *Compressed Sensing*, EEE Trans. Information Theory **52** (2006), 1289–1306.

[6] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan and M. Strauss, *Near-optimal sparse fourier estimation via sampling*, ACM STOC (2002), 152–161.

Adaptive Frame Schemes for Elliptic Operator Equations

Stephan Dahlke

(joint work with Massimo Fornasier, Thorsten Raasch, Rob Stevenson and Manuel Werner)

The analysis of adaptive numerical schemes for operator equations is a field of enormous current interest. Especially, it has also turned out that adaptive schemes based on *wavelets* have several important advantages. The wavelet methodology differs from other schemes in so far as one uses a Riesz basis $\Psi = \{\psi_\lambda\}_{\lambda \in \mathcal{J}}$ for the entire solution Hilbert space H of the given operator equation

$$\mathcal{L}u = f, \tag{1}$$

where we assume that $\mathcal{L} : H \to H'$ is a linear isomorphism of H onto its normed dual H'. Prominent examples which fit into this setting are linear elliptic boundary value problems as well as boundary integral equations. Using the Riesz basis

property of Ψ, it can be shown that (1) is equivalent to the infinite–dimensional system

$$\mathbf{Lu} = \mathbf{f}, \tag{2}$$

where $\mathbf{u}$ and $\mathbf{f}$ are the wavelet coefficient arrays of the unknown solution u and the right–hand side f, respectively, and $\mathbf{L}$ is the standard representation of $\mathcal{L}$ in wavelet coordinates, see [2, 3] for details.

Since the operator $\mathcal{L}$ under consideration is usually defined on a bounded domain or a closed manifold $\Omega \subset \mathbb{R}^d$, a numerically stable construction of a wavelet basis on Ω is needed. Although there are by now several construction methods such as, e.g., [1, 6, 7], certain drawbacks such as a lack of numerical stability and/or sufficiently high smoothness are often unavoidable. A possible way out is to use a slightly weaker concept than Riesz bases, namely *frames*.

Definition 1. A *frame* for the Hilbert space X is a system $\Psi = \{\psi_\lambda\}_{\lambda\in\mathcal{J}} \subset X$ which satisfies the norm equivalence $\|g\|_X^2 \eqsim \sum_{\lambda\in\mathcal{J}} \left|\langle g, \psi_\lambda\rangle\right|^2$.

The first attempt to use frames for the adaptive discretization of (1) was made in [8]. For several theoretical and practical reasons, we use the following class of frames in X:

Definition 2. Let $H \subset X \subset H'$ be a Gelfand triple. A frame $\Psi = \{\psi_\lambda\}_{\lambda\in\mathcal{J}} \subset X$ is called a *Gelfand frame* for the Gelfand triple $H \subset X \subset H'$, if there exists a Gelfand triple of sequence spaces $\ell \subset \ell_2(\mathcal{J}) \subset \ell'$, such that the operators $F^* : \ell \to H$, $\mathbf{c} \mapsto \mathbf{c}^\top\Psi$ and $\widetilde{F} : H \to \ell$, $g \mapsto \langle g, \widetilde{\Psi}\rangle$ are bounded, where $\widetilde{\Psi}$ is the canonical dual frame of Ψ.

It has been shown in [4] how Gelfand frames can serve as an appropriate generalization of Riesz bases in the numerical discretization of operator equations like (1). More precisely, also in the Gelfand frame case it is possible to transform (1) into an equivalent infinite–dimensional system of the form (2). Due to the redundancy of the frame, the operator $\mathbf{L}$ will in general have a nontrivial kernel. However, $\mathbf{L}$ is still boundedly invertible on its range $\mathrm{ran}(\mathbf{L})$, so that the frame expansion of any coefficient array $\mathbf{u}$ solving (2) yields the unique solution u of (1).

In the spirit of [2, 3], the infinite–dimensional system (2) may be solved by a simple iterative scheme

$$\mathbf{u}^{(n+1)} = \mathbf{u}^{(n)} + \alpha_n\mathbf{r}^{(n)}, \tag{3}$$

with the residuals $\mathbf{r}^{(n)} = \mathbf{f} - \mathbf{Lu}^{(n)}$ and descent parameters $\alpha_n = \alpha$ (Richardson method) or $\alpha_n = \frac{\langle \mathbf{r}^{(n)}, \mathbf{r}^{(n)}\rangle}{\langle \mathbf{Lr}^{(n)}, \mathbf{r}^{(n)}\rangle}$ (steepest descent method), in both cases leading to a uniform per–step error reduction in the energy norm. These ideal iterations can be realized numerically by *approximately* applying $\mathbf{L}$ to the current finitely supported iterands [2, 3]. Convergence and work/accuracy balance for the approximate Richardson and steepest descent iteration have been investigated [5]. In general, convergence and work/accuracy balance results in the frame case have the following form, see [4, 8] for details:

Theorem 1. Given a *target accuracy* ε, the algorithm under consideration stops after a finite number of iterations and outputs a finitely supported coefficient array $\mathbf{u}_\varepsilon$ such that $\left\|\mathbf{P}(\mathbf{u}-\mathbf{u}_\varepsilon)\right\| \leq \varepsilon$, where $\mathbf{P} : \ell_2(\mathcal{J}) \to \mathrm{ran}(\mathbf{L})$ is the orthogonal projector. Whenever u can be approximated by N terms with accuracy $\mathcal{O}(N^{-s})$, then $\#\mathrm{supp}\mathbf{u}_\varepsilon$ and #flops to compute $\mathbf{u}_\varepsilon$ stay proportional to $\varepsilon^{-1/s}$.

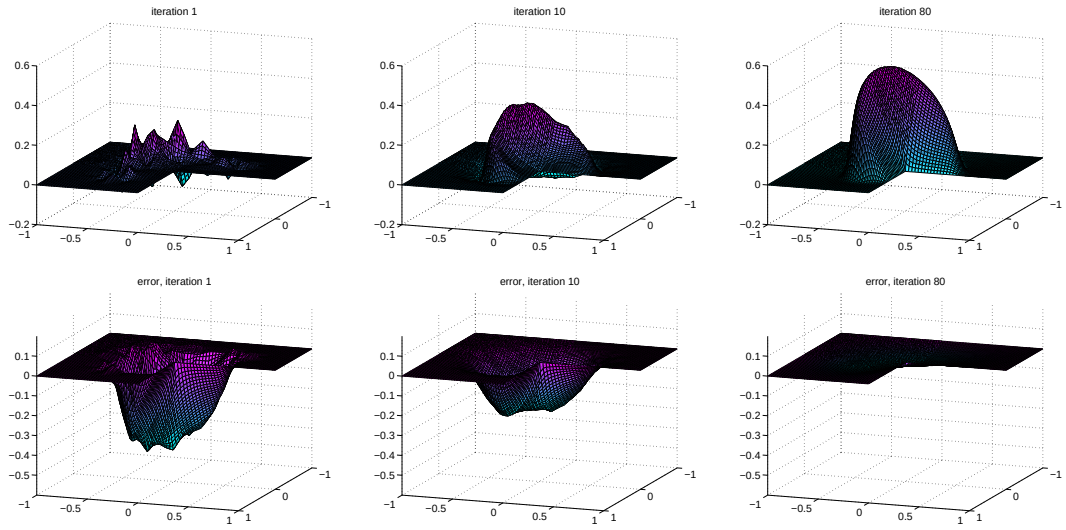


FIGURE 1. Some steepest descent iterands and corresponding errors for the Poisson equation example

In [4], an easy construction of Gelfand frames for $H = H_0^t(\Omega)$ was given, using an overlapping decomposition $\Omega = \bigcup_{i=1}^n \Omega_i$ with smoothly parametrized patches $\Omega_i = \kappa_i(\square)$, $\square := (0,1)^d$ and a reference Riesz basis $\Psi^\square \subset H_0^t(\square)$. These Gelfand frames allow the adaptive numerical treatment of linear elliptic boundary value problems of order $2t$ with homogeneous Dirichlet conditions. As an example, the Poisson equation $-\Delta u = f$ on the L–shaped domain $\Omega := (-1,0)\times(-1,1)\cup(-1,1)\times(0,1)$ was solved numerically [5], taking the reentrant corner singularity as exact solution u.

Moreover, since the Gelfand frames are obtained by aggregating local Riesz bases $\Psi^{(i)} \subset H_0^t(\Omega_i)$, domain decomposition preconditioners such as multiplicative and additive Schwarz methods can be realized in a straightforward fashion. For example, the additive Schwarz method can be considered as a Richardson iteration for the preconditioned system (2), i.e.

$$\mathbf{M}^{-1}\mathbf{L}\mathbf{u} = \mathbf{M}^{-1}\mathbf{f}, \tag{4}$$

where $\mathbf{M} = \mathrm{diag}(\mathbf{L}_1, \ldots, \mathbf{L}_m)$, and $\mathbf{L}_i$ denotes the standard representation of $\mathcal{L}$ with respect to $\Psi^{(i)}$. Thus, in order to perform a single iteration of such a method, the operators $\mathbf{L}_i$ have to be (approximately) inverted, thus on Ω_i an elliptic problem has to be solved. One of the main advantages of such an approach is that, depending on the structure of the algorithm at hand, the latter operations can be performed in parallel. Furthermore, for the local solves on Ω_i, efficient (adaptive)

wavelet Galerkin methods can be applied, which exclusively work in the case of discretizations with respect to wavelet *bases*.

Acknowledgement

The work of the first, third, and last author was supported through DFG, Grants Da 360/4–2, Da 360/4–3, and Da 360/7–1. The second author acknowledges the support provided through the Individual Marie Curie Fellowship (contract MEIF–CT–2004–501018). The work of the fourth author has been supported through the Netherlands Organization for Scientific Research.

References

[1] C. Canuto, A. Tabacco, and K. Urban, The wavelet element method, Part I: Construction and Analysis, Appl. Comput. Harmon. Anal. **6**, 1–52 (1999).

[2] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations — Convergence rates,* Math. Comput. **70**, 27–75 (2001).

[3] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods II: Beyond the elliptic case,* Journal of FoCM **2**, 203–245 (2002).

[4] S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations,* Adv. Comput. Math. doi:10.1007/s10444-005-7501-6 (2007).

[5] S. Dahlke, M. Fornasier, T. Raasch, R. Stevenson and M. Werner, *Adaptive frame methods for elliptic operator equations: the steepest descent approach,* IMA J. Numer. Anal, doi:10.1093/imamum/drl035 (2007).

[6] W. Dahmen and R. Schneider, *Composite wavelet bases for operator equations,* Math. Comput. **68**, 1533–1567 (1999).

[7] W. Dahmen and R. Schneider, *Wavelets on manifolds I. Construction and domain decomposition,* SIAM J. Math. Anal. **31**, 184–230 (1999).

[8] R. Stevenson, *Adaptive solution of operator equations using wavelet frames,* SIAM J. Numer. Anal. **41**, 1074–1100 (2003).

Multiscale Analysis of Vincent Van Gogh's Paintings

Ingrid Daubechies

(joint work with Eugene Brevdo and Shannon Hughes)

This was a report on the analysis by the "Princeton Team" of 101 high definition gray value scans of paintings by Vincent Van Gogh and other artists, in the framework of a workshop for Art Historians and Image Processors, held in May 2007 at the Van Gogh Museum in Amsterdam. This workshop was organized by Professor Rick Johnson, Electrical Engineering Department at Cornell University. Two other image analysis teams also participated: one led by Professor Eric Postma, Computer Science at the University of Maastricht, The Netherlands, and another led by Professors James Wang, College of Informations Sciences and Jia Li, Department of Statistics, Penn State University.

The Princeton researchers based their analysis on wavelet transforms of the high resolution gray-level images. More precisely, they divided every painting in rectangular patches of similar dimensions, 512 x 512 pixels wide (corresponding to roughly 7.4 cm x 7.4 cm), and then computed the wavelet transform for each

patch. They chose to work with a pair of complex wavelet filter banks, allowing for 6 different orientations [1, 2].

Before computing the wavelet transform of each patch, they equalized the collection of patches, so different patches had similar means and dynamic range in gray level distribution.

To analyze the wavelet transforms of the patches, they modeled the distribution of wavelet coefficients in every orientation and at every scale as a mixture of two zero-mean gaussian distributions (one wide, one narrow), associated with a hidden Markov tree, with two hidden states (one for each of the distributions). This model is based upon the intuition that locations in the picture where sharp edges are present correspond to wavelet coefficients that are of type W (for wide), i.e. distributed according to the wide distribution at every scale (and thus admitting quite large values); locations where the content depicted in the picture varies smoothly correspond to wavelet coefficients of type N, i.e. distributed according to the narrow distribution (so that all values are small). Less sharp edges can correspond to a hidden state of type N for fine scale coefficients, switching to W for coarser scales. Similar hidden Markov tree models have been successful in distinguishing different textures in images [3]. The parameters of the hidden Markov tree model included, for each scale and each orientation of the collection of wavelet coefficients, the variances of the W and N distributions (for that scale and orientation), the probability of switching from a coarser scale state W to state N at that scale (and in that orientation), and the probability for the other switch, from a coarser scale state N to state W. Once estimated by the EM algorithm, these parameters were combined into a feature vector that characterized the wavelet transform of each patch.

Machine learning algorithms showed that the features that dominated the classification between paintings by Van Gogh and other artists were mostly transition probabilities from type N to type W (going from coarser to finer scales), linked to orientation-dependent scale values. In other words, these features mostly identified the scales at which detail information "emerges", as one gradually zooms in, in Van Gogh paintings more so than in non-Van Gogh paintings. These characteristic scales turn out to be different for features in different directions; the relative strength of details in each scale and orientation seems characteristic for Van Gogh's style. One can then define an "essential m-feature vector", by restricting to only the m features dominant for classification. A "similarity distance" between paintings was defined by adding, for all pairings of a patch of one painting with a patch of the other, the (possibly weighted) distance between their essential m-feature vectors. Using a multidimensional scaling algorithm to arrange the paintings in space in accordance with these pairwise distances, we found that a good separation was obtained between paintings by Van Gogh and others in the dataset, even when using as few as 2 features. Additionally, stylistically similar Van Gogh paintings were found to tend to cluster in this analysis, with Van Gogh paintings that were stylistically less typical tending toward non-Van Gogh regions; the results of this analysis were therefore interpreted as a characterization of a painting's style.

However, it is also desirable to pinpoint paintings, such as copies or forgeries of true Van Goghs, that are stylistically similar to Van Goghs but are by another artistś hand. In order to do this, the Princeton team made a second analysis, now restricted to much finer scales, which was designed to measure the fluency of the brushstrokes. This analysis was based on patches of 128x128 pixels (roughly 1.85 cm x 1.85 cm); it was inspired by Eric Postma's earlier observation [4] that the infamous Walker forgeries of Van Gogh paintings typically had many more large-valued wavelet coefficients than true Van Gogh paintings. (In this earlier work, Postma used a type of wavelet different from the Princeton team's choice, but this is immaterial for this issue.) Since, in a two-dimensional wavelet transform, 15/16 of the wavelet coefficients pertain to the two finest scales, this suggested that wavelet transforms of non-authentic paintings would have many more large coefficients at the finest scales, i.e. that the painting would have many more prominent very fine scale details. Such abundance of superfine detail can be attributed to more hesitant brushstrokes, caused by a reduction in motion fluidity when copying another painting or another painter's manner. The second analysis technique used by the Princeton team thus checked the relative abundance of extremely fine detail. This feature did indeed separate copies and forgeries from most of the authentic, original Van Goghs; the wavelet transforms of the non-authentic paintings had a much larger population in the finest scale wavelet layers, corresponding to a wealth of "details" of the order of .25-.5 mm wide (2-4 pixels only, at the very limit of the spatial resolution in the dataset.) Surprisingly, a very small number of true Van Goghs were also marked out as "less fluent" by this analysis. Consultation with museum officials revealed that these were either copies that Van Gogh made after another painting, or paintings where, experimenting with technique, he had traced over his own brushstrokes again after the paint had dried. In both cases, the lack of fluency had therefore a natural explanation.

References

[1] N. G. Kingsbury, *Complex wavelets for shift invariant analysis and filtering of signals*, Applied and Computational Harmonic Analysis **10** (2001), 234–253.

[2] I. W. Selesnick, *A new complex-directional wavelet transform and its application to image denoising*, ICIP (2002) III:573–576

[3] H. Choi, J. Romberg, R. Baraniuk and N. G. Kingsbury, *Hidden Markov tree modelling of complex wavelet transforms*, Proc. ICASSP 2000, Instanbul, June 6-9, 2000.

[4] H. J. van den Herik, E. O. Postma(2000), *Discovering the Visual Signature of Painters: Future Directions for Intelligent Systems and Information Sciences*, in *The Future in Speech and Image Technologies, Brain Computers, WWW, and Bioinformatics* (editor: N. Kasabov), pp. 129-147, Physica Verlag (Springer-Verlag, Heidelberg-Tokyo-New York) 2000; ISBN 3-7908-1276-5.

Banach Gelfand Triples in Classical Fourier Analysis

Hans G. Feichtinger

It was the purpose of this talk to propagate certain tools that arose in the context of time-frequency analysis for the use in the context of classical Fourier analysis. In fact, it was pointed out that the relatively simple concept of Banach Gelfand triples allows to give a rather natural interpretation to approximation ideas that have been around already for a century, in the form of summability methods. Taking a new look into these topics may/should have also an influence on how we are teaching Fourier analysis to our students.

The central theme of the talk, which is in full available on the internet at under `www.univie.ac.at/nuhag-php/program/talks_show.php?name=Feichtinger` has been the concept of Banach Gelfand triples. They are defined as follows:

Definition: A triple $(\boldsymbol{B}, \boldsymbol{H}, \boldsymbol{B}')$, consisting of a Banach space $\boldsymbol{B}$, which is dense in some Hilbert space $\mathcal{H}$, which in turn is contained in $\boldsymbol{B}'$ is called a *Banach Gelfand triple (BGT)*. This triple of spaces is endowed with three norm topologies, and in addition the w^*-topology on the (big) dual space $\boldsymbol{B}'$.

Definition: A bounded linear operator T between two Banach Gelfand triples is called a BGT-homomorphism if it maps these spaces continuously into each other, at all three levels and with respect to all four topologies, i.e. also $w^* - w^*$-continuous on the dual spaces. Accordingly isomorphisms and automorphism are defined. They are called *unitary* if they are in addition unitary at the Hilbert space level.

In contrast to the concept of "rigged Hilbert spaces" occurring in the discussion of elliptic partial differential operators and quantum mechanics we propose to use the modulation space $\boldsymbol{M}_0^{1,1}(\mathbb{R}^d)$, also denoted by $\boldsymbol{S}_0(\mathbb{R}^d)$, is defined as the subspace of $\boldsymbol{L}^2(\mathbb{R}^d)$ with a short-time Fourier transform $V_g f$ with Gaussian window g being integrable, i.e. in $\boldsymbol{L}^1(\mathbb{R}^d \times \widehat{\mathbb{R}}^d)$. The $\boldsymbol{S}_0$-norm of f is by definition $\|V_g f\|_1$. This space is the minimal Banach space in $\boldsymbol{L}^2(\mathbb{R}^d)$ among all the Banach spaces on which time-frequency shifts act isometrically, i.e. with $\|M_\omega T_x f\|_{\boldsymbol{B}} = \|f\|_{\boldsymbol{B}}$, for all $t \in \mathbb{R}^d$ and $\omega \in \mathbb{R}^d$. Here T_t and M_ω are the time- and frequency shift operators respectively. We use the (original) symbol $\boldsymbol{S}_0(\mathbb{R}^d)$ for this space (because it is a so-called Segal algebra on $\mathbb{R}^d$, but it can easily be defined on general LCA groups as well). The fact that $\boldsymbol{S}_0(\mathbb{R}^d)$ is isometrically invariant under the Fourier transform (and hence by the usual transposition process the same is true for $\boldsymbol{S}_0'(\mathbb{R}^d)$) allows to formulate the following statement involving Banach Gelfand triples.

Theorem. The Fourier transform is a unitary Banach Gelfand triple automorphism on $(\boldsymbol{S}_0, \boldsymbol{L}^2, \boldsymbol{S}_0')(\mathbb{R}^d)$, and is uniquely characterized as such by the fact that it maps the "pure frequencies" χ_s (resp. characters of the LCA group $\mathbb{R}^d$, viewed as elements of $\boldsymbol{C}_b(\mathbb{R}^d) \subseteq \boldsymbol{S}_0'(\mathbb{R}^d)$, into the corresponding Dirac measures δ_s).

This example can be used to show typical facts about Banach Gelfand triple mappings in such a setting. While the DFT (discrete/finite Fourier transform) resp. FFT can be easily characterized as the unitary (up to the normalization factor $\sqrt{n}$) matrix which maps the discrete pure frequencies into unit vectors, or

in other words, the FFT is an orthogonal change of bases, with the pure frequencies (the system of eigenvectors to the shift operator) being the orthonormal basis.

In the setting of $\mathbb{R}^d$ one has a number of new problems. Where simple sums could be used to express the FFT and its inverse one has now integrals. Although this makes $\boldsymbol{L}^1(\mathbb{R}^d)$ look like a natural domain for $\mathcal{F}$ already $\mathcal{F}^{-1}$ makes problems. Classically summability methods are invoked, i.e. one multiplies $\hat{f}$ by some nice, classical kernel $\hat{h}$ (most of them are in fact members of $\boldsymbol{S}_0(\mathbb{R}^d)$). Since $\boldsymbol{L}^1 * \boldsymbol{S}_0 \subseteq \boldsymbol{S}_0$, one has $\hat{h} \cdot \hat{f} = \widehat{f * h} \in \mathcal{F}\boldsymbol{S}_0 = \boldsymbol{S}_0 \subseteq \boldsymbol{L}^1(\mathbb{R}^d)$, hence integral inversion is no problem. The Hilbert space level is the standard Plancherel theorem, but neither pure frequencies χ_s nor Dirac measures are in $\boldsymbol{L}_2(\mathbb{R}^d)$.

There are many other situations, see for example, in [3], in order to describe e.g. the Kohn-Nirenberg mapping as a unitary Banach Gelfand triple isomorphism. This result is based on a *kernel theorem* which identifies the linear operators from $\boldsymbol{S}_0(\mathbb{R}^d)$ into $\boldsymbol{S}_0'(\mathbb{R}^d)$ with distributional kernels in $\boldsymbol{S}_0'(\mathbb{R}^d)$. The Hilbert space result is the characterization of Hilbert-Schmidt operators via kernels in $\boldsymbol{L}^2(\mathbb{R}^{2d})$, while regularizing kernels in $\boldsymbol{S}_0(\mathbb{R}^d)$ characterize operators from $\boldsymbol{S}_0'(\mathbb{R}^d)$ into $\boldsymbol{S}_0(\mathbb{R}^d)$. In this context the KN-mapping is characterized as a unitary BGT-isomorphism which is characterized by the property that the TF-shift operators $M_\omega T_t$ is mapped onto the Dirac measure $\delta_{t,\omega}$ over $\mathbb{R}^d \times \widehat{\mathbb{R}}^d$.

References

[1] H. G. Feichtinger. On a new Segal algebra. *Monatsh. Math.*, 92:269–289, 1981.

[2] E. Cordero, H. G. Feichtinger, and F. Luef. Banach Gelfand triples for Gabor analysis, CIME report, Confernce 2006, to appear 2007.

[3] H. G. Feichtinger and W. Kozek. Quantization of TF lattice-invariant operators on elementary LCA groups. In H. Feichtinger and T. Strohmer, editors, *Gabor Analysis and Algorithms. Theory and Applications.*, Applied and Numerical Harmonic Analysis, pages 233–266, 452–488, Boston, MA, 1998. Birkhäuser Boston.

[4] H. G. Feichtinger and F. Luef. Wiener Amalgam Spaces for the Fundamental Identity of Gabor Analysis. *Collect. Math.*, 57 (Extra Volume (2006)):233–253, 2006.

[5] H. G. Feichtinger, F. Luef, and T. Werther. A Guided Tour from Linear Algebra to the Foundations of Gabor Analysis. In *Gabor and Wavelet Frames*, volume 10 of *IMS Lecture Notes Series*. 2007.

[6] H. G. Feichtinger and F. Weisz. The Segal algebra $S_0(\mathbb{R}^d)$ and norm summability of Fourier series and Fourier transforms. *Monatsh. Math.*, 148:333–349, 2006.

[7] H. G. Feichtinger and F. Weisz. Wiener amalgams and pointwise summability of Fourier transforms and Fourier series. *Math. Proc. Cambridge Philos. Soc.*, 140:509–536, 2006.

One Sketch for all: Fast Algorithms for Compressed Sensing

Anna C. Gilbert

(joint work with Martin Strauss, Joel Tropp and Roman Vershynin)

Compressed Sensing is a new paradigm for acquiring the compressible signals that arise in many applications. These signals can be approximated using an amount of information much smaller than the nominal dimension of the signal. Traditional approaches acquire the entire signal and process it to extract the information. The

new approach acquires a small number of nonadaptive linear measurements of the signal and uses sophisticated algorithms to determine its information content. Emerging technologies can compute these general linear measurements of a signal at unit cost per measurement.

This paper exhibits a randomized measurement ensemble and a signal reconstruction algorithm that satisfy four requirements:

(1) The measurement ensemble succeeds for all signals, with high probability over the random choices in its construction.
(2) The number of measurements of the signal is optimal, except for a factor polylogarithmic in the signal length.
(3) The running time of the algorithm is polynomial in the amount of information in the signal and polylogarithmic in the signal length.
(4) The recovery algorithm offers the strongest possible type of error guarantee. Moreover, it is a fully polynomial approximation scheme with respect to this type of error bound.

Emerging applications demand this level of performance. Yet no other algorithm in the literature simultaneously achieves all four of these desiderata.

References

[1] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1208–1223, 2006.

[2] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? Submitted for publication, Nov. 2004.

[3] A. Cohen, W. Dahmen, and R. DeVore. Compressed Sensing and best k-term approximation. Submitted for publication, July 2006.

[4] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for Compressed Sensing. In *Proc. of SIROCCO*, pages 280–294, 2006.

[5] D. L. Donoho. Compressed Sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, Apr. 2006.

[6] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk. Fast reconstruction of piecewise smooth signals from random projections. In *Proc. SPARS05*, Rennes, France, Nov. 2005.

[7] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss. Improved time bounds for near-optimal sparse Fourier representation via sampling. In *Proc. SPIE Wavelets XI*, San Diego, 2005.

[8] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin. Algorithmic linear dimension reduction in the ℓ_1 norm for sparse vectors. Submitted for publication, 2006.

[9] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk. Analog-to-information conversion via random demodulation. In *IEEE Dallas Circuits and Systems Workshop (DCAS)*, Dallas, Texas, Oct. 2006.

[10] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss. Random sampling for analog-to-information conversion of wideband signals. In *IEEE Dallas Circuits and Systems Workshop (DCAS)*, Dallas, Texas, Oct. 2006.

[11] M. Rudelson and R. Veshynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Proc. 40th Ann. Conf. Information Sciences and Systems*, Princeton, Mar. 2006.

[12] D. Takhar, J. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Proc. IS&T/SPIE Symposium on Electronic Imaging*, 2006.

[13] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via Orthogonal Matching Pursuit. Submitted for publication, Apr. 2005. Revised, Nov. 2006, 2006.

[14] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. Compressive imaging for video representation and coding. In *Proc. Picture Coding Symposium 2006*, Beijing, China, Apr. 2006.

Matrices with Off-Diagonal Decay and their Inverses

Karlheinz Gröchenig
(joint work with Andreas Klotz)

We study the off-diagonal decay of infinite matrices. This property is important in many applications, ranging from the properties of dual frames and regularity questions of pseudodifferential operators to equalization algorithms in wireless communications.

The best known result deals with the inverse of banded matrices: *If A is invertible on $\ell^2(\mathbb{Z}^d)$ and $a_{kl} = 0$ for $|k-l| > N$, then A^{-1} has exponential decay* [2]

$$|(A^{-1})_{kl}| \leq Ce^{-\epsilon|k-l|}\,.$$

In this case, the properties of the matrix are not preserved exactly by the inverse. The abstract reason is that both exponential decay and bandedness are preserved by matrix multiplication, but not by limits.

In order to obtain symmetry between A and A^{-1} one has to use *Banach algebras* of matrices. The fundamental results are due to Jaffard and Journée [4] for polynomial decay and to Gohberg, Kurbatov, Baskakov, and others for ℓ^1-decay, see for instance [1].

If A is invertible on $\ell^2(\mathbb{Z}^d)$ and $|a_{kl}| \leq C(1 + |k - l|)^{-s}$ for $s > d$, then also $(A^{-1})_{kl}| \leq C'(1 + |k - l|)^{-s}$ (same exponent s)

Likewise, if A is invertible on $\ell^2(\mathbb{Z}^d)$ and $|a_{kl}| \leq h(k - l)$ for some $h \in \ell^1$, then also $|(A^{-1})_{kl}| \leq H(k - l)$ for some $H \in \ell^1$.

In the last few years, these results have been investigated intensively, and many variations (weights, additional parameters, different types of decay conditions) have been studied, see e.g. [3] or Q. Sun's work.

The abstract concept behind the scenes is that of inverse-closedness, which is defined as follows: Let $\mathcal{A} \subseteq \mathcal{B}$ be two (involutive) Banach algebras with common identity. Then $\mathcal{A}$ is called *inverse-closed* in $\mathcal{B}$, if

$$a \in \mathcal{A} \text{ and } a^{-1} \in \mathcal{B} \quad \Longrightarrow \quad a^{-1} \in \mathcal{A}\,.$$

The mentioned results of Jaffard, Baskakov, Gohberg state that a certain Banach algebra of matrices is inverse-closed in $\mathcal{B}(\ell^2)$, the bounded operators on ℓ^2.

Inverse-closedness is a strong property with many implications, e.g., invariance under holomorphic calculus, but most constructions of inverse-closed subalgebras are somewhat adhoc.

The purpose of this talk was to explain two methods to the systematic construction of subalgebras $\mathcal{A}$ of infinite matrices in $\mathcal{B}(\ell^2)$ that are inverse-closed. If $\mathcal{A}$ is defined by some type of decay condition, . then the inverse of a matrix in $\mathcal{A}$, if it exists, satisfies automatically the same decay conditions.

The systematic construction of inverse-closed subalgebras exhibits a nice interplay between *approximation theory* and the theory of *operator algebras and Banach algebras*.

Smooth Subalgebras. Consider the Banach algebra C^n of all n-times differentiable functions on $[0,1]$. If a function $f \in C^n$ does not vanish anywhere, then its inverse $1/f$ is again n-times differentiable, $f \in C^n$, by the quotient rule.

This observation can by transferred to arbitrary Banach algebras by using derivations instead of derivatives. Let $\mathcal{A}$ be a Banach algebra and $\delta : \mathcal{A} \to \mathcal{A}$ a (closed) derivation, i.e., δ satisfies the product rule $\delta(AB) = \delta(A)B + A\delta(B)$.

Now define the subalgebra of n-smooth elements by

$$C^n(\mathcal{A}) = \bigcap_{k=1}^{n} \operatorname{dom} \delta^k$$

Theorem 1 (Bratteli)**.** If $A \in C^n(\mathcal{A})$ and A is invertible in $\mathcal{A}$, then $A^{-1} \in C^n(\mathcal{A})$, in other words, $C^n(\mathcal{A})$ is inverse-closed in $\mathcal{A}$.

This is an abstract statement about any Banach algebra. The connection to matrices is established by looking at a particular derivation, namely the commutator with the (unbounded) diagonal matrix X with entries $X_{kl} = k\delta_{kl}, k, l \in \mathbb{Z}$. Set $\delta(A) = [X, A] = XA - AX$, then the entries of $\delta(A)$ are $\delta(A)_{kl} = (k-l)a_{kl}, k, l \in \mathbb{Z}$. By Bratteli's Theorem we have that $\delta^n(A) \in \mathcal{A}$, if and only if $\left((k-l)^n a_{kl}\right) \in \mathcal{A}$. This means rougly, that the entries of the original matrix A satisfy polynomial off-diagonal decay $|a_{kl}| \leq C\,|k-l|^{-n}$. This observation can be used to make some shortcuts in Jaffard's Theorem.

The analogy between spaces of smooth functions and spaces of smooth elements in Banach algebras can be turned into a program and leads to many more inverse-closed Banach algebras of "smooth" elements.

For instance, by imitating the fractional smoothness of Hölder-Lipschitz spaces, one may consider fractional smoothness in a Banach algebra $\mathcal{A}$ with a derivation δ as follows. Let $\alpha_h : \mathcal{A} \to \mathcal{A}, h \in \mathbb{R}$ be the automorphism group generated by δ. We say that $A \in C^s(\mathcal{A})$ for $0 < s < 1$, if $\|\alpha_h(A) - A\|_{\mathcal{A}} \leq C|h|^s$. One then obtains the following result:

Theorem 2. If $A \in C^s(\mathcal{A})$ and A is invertible in $\mathcal{A}$, then $A^{-1} \in C^s(\mathcal{A})$. In other words, $C^s(\mathcal{A})$ is inverse-closed in $\mathcal{A}$.

Algebras by Approximation Properties. A second idea is to imitate the definition of approximation spaces. For this recall the characterization of the Hölder spaces C^s on the torus by approximation with trigonometric polynomials. Let $\sigma_n(f)$ denotes the minimal error made when approximating f by a trigonometric polynomial of degree n. Then $f \in C^s$ if and only if $\sigma_n(f) = \mathcal{O}(n^{-s})$.

To generalize the construction of approximation spaces to Banach algebras, assume that $\mathcal{T}_n$ is a nested sequence of subspaces of $\mathcal{A}$ satisfying the following properties: (1) Each $\mathcal{T}_n$ contains that identity element e, (2) $\mathcal{T}_n \subseteq \mathcal{T}_{n+1}$, and (3) $\mathcal{T}_m \cdot \mathcal{T}_n \subseteq \mathcal{T}_{m+n}$

Let $\sigma_n(A)$ be the (linear) approximation error $\sigma_n(A) = \inf_{B\in\mathcal{T}_n} \|A - B\|_{\mathcal{A}}$. Then we can define the *approximation space* $\mathcal{E}_s^p(\mathcal{A})$ as usual by

$$\|A\|_{\mathcal{E}_s^p} := \Big(\sum_{n=0}^{\infty}(n^s\sigma_n(A))^p\frac{1}{n}\Big)^{1/p}$$

It was shown by Almira and Luther that $\mathcal{E}_s^p(\mathcal{A})$ is a subalgebra of $\mathcal{A}$.

The connection to inverse-closedess is established in the following new result.

Theorem 3. $\mathcal{E}_s^p(\mathcal{A})$ is inverse closed in $\mathcal{A}$.

To apply this abstract result on Banach algebras to matrices, we choose the natural sequence of of approximation subspaces, namely the banded matrices $\mathcal{T}_n = \{A : a_{kl} = 0 \text{ for } |k-l| > n\}$. Thus the main theorem says that if a matrix is approximated well by banded matrices, then its inverse is also approximated well by banded matrices. Clearly good approximation by banded matrices is also a measure for the off–diagonal decay of a matrix.

This program may and will be pursued much further in the work of A. Klotz.

The talk gave an overview how ideas from classical approximation may be modified to yield systematic construction procedures for inverse-closed algebras of matrices with off-diagonal decay.

References

[1] A. G. Baskakov. Wiener's theorem and asymptotic estimates for elements of inverse matrices. *Funktsional. Anal. i Prilozhen.*, 24(3):64–65, 1990.

[2] S. Demko, W. F. Moss, and P. W. Smith. Decay rates for inverses of band matrices. *Math. Comp.*, 43(168):491–499, 1984.

[3] K. Gröchenig and M. Leinert. Symmetry and inverse-closedness of matrix algebras and symbolic calculus for infinite matrices. *Trans. Amer. Math. Soc.*, 358:2695–2711, 2006.

[4] S. Jaffard. Propriétés des matrices "bien localisées" près de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7(5):461–476, 1990.

Approximation and Interpolation by Power Series with ± 1 Coefficients

C. Sinan Güntürk

This talk consists of two parts. The first part concerns the results published by the author in [1]. The motivation of this paper comes from the following "fair duel problem" of S. Konyagin [2]: There are two duellists A and B who will shoot at each other (only one at a time) using a given ± 1 sequence $q = (q_n)_{n\geq 0}$ which specifies whose turn it is to shoot at time n. The shots are independent and identically distributed random variables with outcomes hit or miss. Each shot hits (and therefore kills) its target with a small unknown probability ϵ, which is arbitrary but fixed throughout the duel. The "fair duel" problem is to find an ordering q, which is independent of ϵ, and is as fair as possible in the sense that the probability of survival for each duellist is as close to $1/2$ as possible. (The problem makes sense only when we ask q to be universal, i.e., independent of ϵ. Otherwise, for any fixed known $\epsilon \leq 1/2$, an ordering can be found such that the

probability of survival is exactly equal to 1/2 for both duellists.) We measure the fairness of an ordering q by its *bias* function $B_q(\epsilon)$, defined to be

$$B_q(\epsilon) := \mathbb{P}\{A \text{ survives}\} - \mathbb{P}\{B \text{ survives}\},$$

and ask that $B_q(\epsilon) \to 0$ as fast as possible as $\epsilon \to 0$. It is an elementary calculation that the bias is given by

$$B_q(\epsilon) = \epsilon \sum_{n=0}^{\infty} q_n (1-\epsilon)^n.$$

At first, it may appear as the best ordering should be to simply alternate between the two duellists, i.e., to set $q_n = (-1)^n$, for which $B_q(\epsilon) = \epsilon/(2-\epsilon) \asymp \epsilon$. However, this naive option is quickly ruled out as, for instance, the 4-periodic sequence given by $q_0 = 1$, $q_1 = -1$, $q_2 = -1$, $q_3 = 1$ yields $B_q(\epsilon) \asymp \epsilon 2$. Continuing in this fashion, it is tempting to think that the Thue-Morse sequence on the alphabet $\{-1, +1\}$ might perhaps be the optimal sequence. For the Thue-Morse sequence, one has

$$B_{\mathrm{TM}}(\epsilon) = \epsilon \prod_{n=0}^{\infty} \left(1 - (1-\epsilon)^{2^n}\right),$$

where the infinite product $\prod \left(1 - z^{2^n}\right) = \sum q_n z^n$ can in fact be taken as the definition of this sequence. It is not difficult to show that there is a positive constant $c > 0$ such that $B_{\mathrm{TM}}(\epsilon) \sim e^{-c(\log \epsilon)2}$.

It turns out that one can do much better. The following result is proven in [1]: Let $0 \le \mu < 1 \le M < \infty$ be arbitrary and $\mathcal{R}_M := \{z \in \mathbb{C} : |1-z| < M(1-|z|)\}$. There exist constants $C_1 := C_1(\mu, M) > 0$ and $C_2 := C_2(\mu, M) > 0$ such that for any power series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \ a_n \in [-\mu, \mu], \ \forall n,$$

there exists a power series with ± 1 coefficients, i.e.,

$$Q(z) = \sum_{n=0}^{\infty} q_n z^n, \ q_n \in \{-1, +1\}, \ \forall n,$$

which satisfies

$$|f(z) - Q(z)| \le C_1 e^{-C_2/|1-z|}$$

for all $z \in \mathcal{R}_M$. A result by Borwein-Erdélyi-Kós [3] shows that this type of decay rate is best possible in the sense that modulo the constant in the exponent, the lower bound will be achieved infinitely often as $z \to 1$.

The special case $f \equiv 0$ corresponds to the fair duel problem and one obtains a universal ordering q for which

$$|B_q(\epsilon)| \le C\sqrt{\epsilon}\, e^{-\frac{\pi 2}{24\epsilon}}.$$

This sequence is recursively defined and its computation involves special numerical methods. The first 50 values of this sequence in 0/1 format is

10010101101010100101101001010101101001011010100101...

The above result concerned how well power series with ± 1 coefficients could approximate more general power series around the point $z = 1$. In the second part of the talk, we consider the interpolation properties of ± 1 power series and ask if the graph of such a series can be passed through any number of generic points whose abscissa are sufficiently close to 1 and ordinate close to 0. The following result is reported:

Given any positive integer M, there exists $c = c_M > 0$ such that for all M distinct numbers $x_1, x_2, \ldots, x_M \in (1-c, 1)$, there exists $\delta = \delta(x_1, ..., x_M) > 0$ such that for all $y_1, y_2, \ldots, y_M \in [-\delta, \delta]$, there exists a ± 1 power series $Q(x) = \sum q_n x^n$ of the real variable x whose graph goes through all the points (x_i, y_i), i.e.,

$$y_i = \sum_{n=0}^{\infty} q_n x_i^n; \ i = 1, ..., M.$$

It is not surprising that the result holds on a left neighborhood of the point 1 whose length depends on M, as a simple volume covering argument imposes the constraint that $x_1 x_2 \cdots x_M \geq 1/2$. We also note that this result has consequences in terms of non-separable Bernoulli convolutions.

References

[1] C.S. Güntürk, *Approximation by Power Series with* ± 1 *Coefficients*, International Mathematics Research Notices **26** (2005), 1601–1610.

[2] S. Konyagin, personal communication, 2003.

[3] P. Borwein, T. Erdélyi and G. Kós, *Littlewood-type problems on* $[0, 1]$, Proc. London. Math. Soc. **79** (1999), 22–46.

Convolution of hp-Functions on Locally Refined Grids

Wolfgang Hackbusch

Usually, the fast evaluation of a convolution integral requires that the involved functions have a simple structure based on an equidistant grid in order to apply the fast Fourier transform. Here we discuss the efficient performance of the convolution of *hp*-functions in certain locally refined grids. More precisely, the convolution result is projected into some given *hp*-space (Galerkin approximation). The overall cost is $\mathcal{O}(p^2 N \log N)$, where N is the sum of the dimensions of the subspaces containing f, g and the resulting function, while p is the maximal polynomial degree.

We consider the convolution integral

$$\omega_{\text{exact}}(x) := (f * g)(x) := \int_{\mathbb{R}} f(y) g(x - y) \mathrm{d}y \tag{1}$$

for *hp*-functions f, g of bounded support. We do not compute the exact result ω_{exact}, but its L^2-orthogonal projection $\omega := P\omega_{\text{exact}}$ into a certain subspace of *hp*-functions.

Convolutions involving a kernel function $f = k$ occur for instance when integral operators $Kg(x) = \int_{\mathbb{R}} k(x-y)g(y)\mathrm{d}y$ are to be evaluated. In [1] and [2] one finds applications where the convolutions are *not* derived from integral operators. In the case of integral operators, the kernel function k is often assumed to satisfy special (smoothness) conditions. This allows various approximations and various methods for its efficient numerical treatment (e.g., [7], [8]). Here we make no assumptions about f, g except that they belong to certain *hp*-finite element subspaces.

Convolution of locally refined piecewise constant functions has been considered in [3]. A variant of the method with "mass conservation" in the piecewise constant case can be found in [5]. The mass conservation holds in general for approaches with polynomial degrees $p \geq 1$. The particular case of $p = 1$ is discussed in [4].

The present article concentrates on the algorithmic aspects when large polynomial degrees appear as it is generally assumed in the *hp*-case, where coarser grid sizes are compensated by higher polynomials degrees.

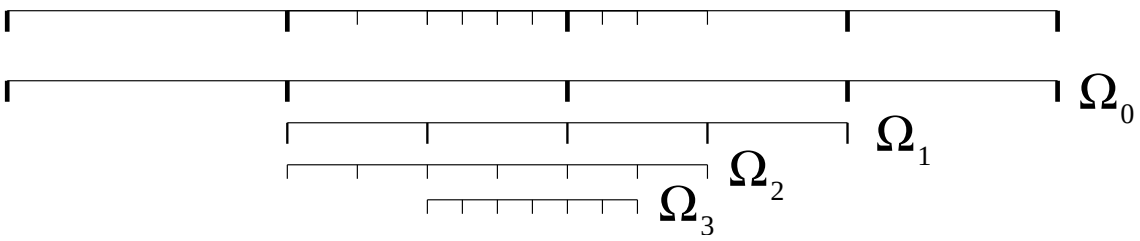


FIGURE 1. Refined grid (first line) composed by local refinements at the levels 0-3 in the zones Ω_ℓ

The *hp*-structure is based on nested refinement zones

$$\mathbb{R} \supset \Omega_0 \supset \ldots \supset \Omega_{\ell-1} \supset \Omega_\ell \supset \ldots \supset \Omega_L \tag{2}$$

(cf. Figure 1; $\Omega_{L+1} := \emptyset$) corresponding to step sizes $h_\ell = 2^{-\ell}h$ with a fixed (coarsest) step size $h = h_0$. More precisely, Ω_ℓ must be a nonempty interval consistent with the h_ℓ-mesh, i.e., $\Omega_\ell = [i_{a,\ell}h_\ell, i_{b,\ell}h_\ell]$ for some $i_{a,\ell}, i_{b,\ell} \in \mathbb{Z}$. The nestedness (2) can be rewritten as $i_{a,0}h_0 \leq i_{a,1}h_1 \leq \ldots \leq i_{a,L}h_L < i_{b,L}h_L \leq \ldots \leq i_{b,1}h_1 \leq i_{b,0}h_0$.

The infinite grid $\mathcal{M}_\ell := \left\{I_\nu^\ell : \nu \in \mathbb{Z}\right\}$ of level $\ell \in \mathbb{N}_0$ containing the intervals

$$I_\nu^\ell := [\nu h_\ell, (\nu+1)\, h_\ell) \qquad \text{for } \nu \in \mathbb{Z},\ \ell \in \mathbb{N}_0.$$

The geometric mesh for the *hp*-functions consists of the intervals in the set

$$\mathcal{M} := \left\{ I_\nu^\ell \in \mathcal{M}_\ell : I_\nu^\ell \subset \overline{\Omega_\ell \backslash \Omega_{\ell+1}},\ 0 \leq \ell \leq L \right\},$$

i.e., inside of $\Omega_0 \backslash \Omega_1$ the h_0-mesh is used, inside of $\Omega_1 \backslash \Omega_2$ the h_1-mesh, ... and finally $\Omega_L \backslash \Omega_{L+1} = \Omega_L$ is filled with the h_L-mesh.

Furthermore, we associate to each interval $I_\nu^\ell \in \mathcal{M}$ with a polynomial degree[1] $p \in \mathbb{N}_0$ and define the space $\mathcal{S} = \mathcal{S}(\mathcal{M})$ of *hp*-functions by all functions ψ with

$$\psi|_{I_\nu^\ell} \quad \text{is a polynomial of degree } \le p \qquad \text{for all } I_\nu^\ell \in \mathcal{M} \tag{3}$$

and $\psi = 0$ outside of Ω_0. Note that no continuity of ψ is required.

We allow that the two factors f and g of the convolution belong to *different* *hp*-spaces characterised by different refinement zones. For the resulting projection $\omega := P\omega_{\text{exact}}$ (ω_{exact} from (1)) a third *hp*-space may be defined. We denote these three spaces by the superscripts "f, g, ω". Therefore, we have to replace the sets Ω_ℓ, $\mathcal{M}$, $\mathcal{S}$ by

$$\Omega_\ell^f,\ \Omega_\ell^g,\ \Omega_\ell^\omega, \quad \mathcal{M}^f,\ \mathcal{M}^g,\ \mathcal{M}^\omega,\ \mathcal{S}^f,\ \mathcal{S}^g,\ \mathcal{S}^\omega.$$

s similar to the fast wavelet transformation, and a discrete convolution which is solved by the Fast Fourier Transform.

Now, the problem can be formulated.

PROBLEM. Given $f \in \mathcal{S}^f$ and $g \in \mathcal{S}^g$, we want to compute the (exact) projection $\omega = P(f * g)$, where P is the L^2-orthogonal projection onto the subspace $\mathcal{S}^\omega \subset L^2(\mathbb{R})$.

The computational details are as follows.

The (orthonormal) system of basis functions $\Phi_{i,\alpha}^\ell$ is given by the Legendre polynomials of degree α, mapped affinely from $(-1,1)$ onto $I_i^\ell := [ih_\ell, (i+1)h_\ell)$ and normalised such that $\int_{-1}^1 \left(\Phi_{i,\alpha}^\ell(x)\right)^2 \mathrm{d}x = 1$. Here I_i^ℓ varies in $\mathcal{M}$, and $0 \le \alpha \le p$. Hence, the discrete subspace is $\mathcal{S}(\mathcal{M}) = \operatorname{span}\left\{\Phi_{i,\alpha}^\ell : I_i^\ell \in \mathcal{M},\ 0 \le \alpha \le p\right\}$.

The functions f and g have representations

$$f = \sum_\ell f_\ell, \quad f_\ell = \sum_{i,\kappa} f_{i,\kappa}^\ell \Phi_{i,\kappa}^\ell \in \mathcal{S}_\ell \quad g = \sum_\ell g_\ell, \quad g_\ell = \sum_{i,\kappa} g_{i,\kappa}^\ell \Phi_{i,\kappa}^\ell \in \mathcal{S}_\ell,$$

where $\mathcal{S}_\ell := \operatorname{span}\{\Phi_{i,\alpha}^\ell : i \in \mathbb{Z},\ 0 \le \alpha \le p\}$ $(\ell \in \mathbb{N}_0)$.

Their convolution leads to $f*g = \sum_{\ell'=0}^{L^f} \sum_{\ell=0}^{L^g} f_{\ell'} * g_\ell = \sum_{\ell' \le \ell} f_{\ell'} * g_\ell + \sum_{\ell < \ell'} g_\ell * f_{\ell'}$. In the following we concentrate on the first sum with $\ell' \le \ell$.

The projection $P(f*g)$ onto the *hp*-space $\mathcal{S}^\omega$ has the representation

$$P(f*g) = \sum_{\ell'',i,\alpha} \omega_{i,\alpha}^{\ell''} \Phi_{i,\alpha}^{\ell''} \qquad \text{with } \omega_{i,\alpha}^{\ell''} = \left\langle f*g, \Phi_{i,\alpha}^{\ell''} \right\rangle.$$

The dependence of the desired coefficients $\omega_{i,\alpha}^{\ell''}$ on the coefficients of f and g is given by

$$\omega_{i,\alpha}^{\ell''} = \sum_{j,k\in\mathbb{Z}} \sum_{\beta,\kappa=0}^{p} f_{j,\beta}^{\ell'} g_{k,\kappa}^{\ell} \gamma_{(i,\alpha),(j,\beta),(k,\kappa)}^{\ell'',\ell',\ell},$$

$$\gamma_{(i,\alpha),(j,\beta),(k,\kappa)}^{\ell'',\ell',\ell} := \iint \Phi_{i,\alpha}^{\ell''}(x)\, \Phi_{j,\beta}^{\ell'}(y)\, \Phi_{k,\kappa}^{\ell}(x-y)\mathrm{d}x\mathrm{d}y = \left\langle \Phi_{i,\alpha}^{\ell''}, \Phi_{j,\beta}^{\ell'} * \Phi_{k,\kappa}^{\ell} \right\rangle.$$

[1] We may use different degrees $p_\nu^\ell \in \mathbb{N}_0$,

Let $\mathbf{P}_\ell$ the mapping from $\psi \in L^2(\mathbb{R})$ into the corresponding coefficients $\psi^\ell_{i,\kappa} = \left\langle \psi, \Phi^\ell_{i,\kappa} \right\rangle$ of the projection onto $\mathcal{S}_\ell$, which is denoted by $P_\ell(\psi) = \sum_{i,\kappa} \left\langle \psi, \Phi^\ell_{i,\kappa} \right\rangle \Phi^\ell_{i,\kappa}$.

The critical problem is the computation of the projected convolution $P_{\ell'}(f_{\ell'} * g_\ell)$, where $\ell' \le \ell$ indicates that ℓ' corresponds to a coarser grid than ℓ. The naive approach is the representation of $f_{\ell'}$ at the fine level ℓ: $f_{\ell'} \longmapsto f_\ell$ (same function, other representation). Then $f_\ell * g_\ell$ leads to a standard discrete convolution and the result can be coarsened to level ℓ'. However, the disadvantage is that by $f_{\ell'} \longmapsto f_\ell$ the data size is increased by the factor $2^{\ell-\ell'}$. Instead, the g-data must be transported from level ℓ to ℓ' as explained in

LEMMA. Let $\psi_{\ell'} = \sum_{i\in\mathbb{Z}} \sum_{\beta=0}^{p} \psi^{\ell'}_{i,\beta} \Phi^{\ell'}_{i,\beta} \in \mathcal{S}_{\ell'}$. Then

$$\left(\psi^{\ell'}_{i,\beta}\right)_{i\in\mathbb{Z},0\le\beta\le p} * \left(\Gamma^{\ell',\ell}_{i,(\alpha,\beta)}\right)_{i\in\mathbb{Z},0\le\alpha,\beta\le p} = \mathbf{P}_{\ell'}\left(\psi_{\ell'} * g_\ell\right) \qquad \text{for } \ell \ge \ell'$$

holds, where $\Gamma^{\ell',\ell}_{i,(\alpha,\beta)} := \sum_{k,\kappa} g^\ell_{k,\kappa}\, \gamma^{\ell',\ell',\ell}_{k-i2^{\ell-\ell'},(\alpha,\beta,\kappa)}$. The discrete convolution on the left-hand side is defined by $\left(\sum_{j\in\mathbb{Z}} \sum_{\beta=0}^{p} \psi^{\ell'}_{j,\beta} * \Gamma^{\ell',\ell}_{j-i,(\alpha,\beta)}\right)_{i\in\mathbb{Z},0\le\alpha\le p}$.

The Γ-coefficients can be computed by a simple recursion starting with $\ell' = \ell$. The final algorithm for computing projections $P_{\ell''}(\sum_{\ell'\le\ell} f_{\ell'} * g_\ell)$ distinguishes the cases (A) $\ell'' \le \ell' \le \ell$, (B) $\ell' < \ell'' \le \ell$, (C) $\ell' \le \ell < \ell''$. In each case, the calculation involves a loop which is similar to the fast wavelet transform using some mask coefficients, and a discrete convolution which can be performed by the Fast Fourier Transform.

Details are in [6].

References

[1] M.V. Fedorov, H.-J. Flad, G.N. Chuev, L. Grasedyck, and B.N. Khoromskij, *A structured low-rank wavelet solver for the Ornstein-Zernike integral equation*. Computing **80** (2007), 47–73.

[2] W. Hackbusch: *On the efficient evaluation of coalescence integrals in population balance models*, Computing **78** (2006) 145–172.

[3] W. Hackbusch: *Fast and exact projected convolution for non-equidistant grids*, Computing **80** (2007) 137–168.

[4] W. Hackbusch: *Fast and exact projected convolution of piecewise linear functions on non-equidistant grids*, Preprint 110, Max-Planck-Institut für Mathematik, Leipzig, 2006.

[5] W. Hackbusch: *Approximation of coalescence integrals in population balance models with local mass conservation*, Numer. Math. (electronically published)

[6] W. Hackbusch: *Convolution of hp-functions on locally refined grids - extended version*, Preprint 38, Max-Planck-Institut für Mathematik, Leipzig, 2007.

[7] W. Hackbusch and S. Börm: *$\mathcal{H}^2$-matrix approximation of integral operators by interpolation*, Appl. Numer. Math. **43** (2002) 129–143.

[8] D. Potts, G. Steidl, and A. Nieslony: *Fast convolution with radial kernels at nonequispaced knots*, Numer. Math. **98** (2004) 329–351.

Sparse Second Moment Analysis for Elliptic Problems in Stochastic Domains

Helmut Harbrecht
(joint work with Reinhold Schneider and Christoph Schwab)

Introduction. The rapid development of scientific computing and numerical analysis in recent years allows the efficient numerical solution of large classes of partial differential equation models with high accuracy, *provided* that the problem's input data are known exactly. Often, however, exact input data for numerical simulation in engineering is not known. The practical significance of highly accurate numerical solution of differential equation models in engineering must thus address how to account for uncertain input data.

If a statistical description of the input data is available, one can mathematically describe data and solutions as random fields and aim at the computation of corresponding deterministic statistics of the unknown random solution u. Here, we consider elliptic boundary value problems on uncertain domains D. In particular, we like to compute the expectation

$$\mathrm{E}_u(\mathbf{x}) = \int_\Omega u(\mathbf{x},\omega)dP(\omega), \quad \mathbf{x} \in D,$$

the two-point correlation

$$\mathrm{Cor}_u(\mathbf{x},\mathbf{y}) = \int_\Omega u(\mathbf{x},\omega)u(\mathbf{y},\omega)dP(\omega), \quad \mathbf{x},\mathbf{y} \in D,$$

and from it the variance $\mathrm{Var}_u(\mathbf{x}) = \mathrm{Cor}_u(\mathbf{x},\mathbf{x}) - \mathrm{E}_u^2(\mathbf{x})$.

The goal of computation is as follows: *given mean and two-point correlation of the boundary perturbation field, compute, to leading order, the mean and the two-point correlation of the random solution of the boundary value problem.*

Partial differential equations on stochastic domains. Starting point of our modelling is the boundary value problem

$$-\Delta u(\mathbf{x},\omega) = f(\mathbf{x}), \quad \mathbf{x} \in D_\omega, \quad \omega \in \Omega$$

with Dirichlet boundary conditions

$$u(\mathbf{x},\omega) = g(\mathbf{x}), \quad \mathbf{x} \in \partial D_\omega, \quad \omega \in \Omega.$$

Herein, the domain D_ω is bounded by the boundary ∂D_ω that is defined as the perturbation of a fixed nominal boundary ∂D:

$$\partial D_\omega = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}(\mathbf{x},\omega) = \mathbf{x} + \varepsilon\kappa(\mathbf{x},\omega)\mathbf{n}(\mathbf{x}),\ \mathbf{x} \in \partial D\}, \quad \omega \in \Omega$$

where $\mathbf{n}$ denotes the ouward normal vector to D. For a fixed small $\varepsilon > 0$ one can linearize the problem under additional assumptions on the smoothness of $\kappa(\mathbf{x},\omega)$ with respect to ω, necessary to ensure that the perturbed boundary is still admissible (e.g. Lipschitz). We fix $\mathbf{x} \in \mathbb{R}^n$ with $P(\{\mathbf{x} \in D_\omega\}) = 1$ and expand $u(\mathbf{x},\omega)$

in terms of a second order *shape Taylor expansion*

$$u(\mathbf{x},\omega) = \overline{u}(\mathbf{x}) + \varepsilon du(\mathbf{x})[\kappa(\mathbf{x},\omega)] + \frac{\varepsilon^2}{2} d^2u(\mathbf{x})[\kappa(\mathbf{x},\omega),\kappa(\mathbf{x},\omega)] + \mathcal{O}(\varepsilon^3), \tag{1}$$

see [2]. Herein, du denotes the *local shape derivative*

$$\Delta du = 0 \text{ in } D, \qquad du = \kappa \frac{\partial(g-u)}{\partial \mathbf{n}} \text{ on } \partial D. \tag{2}$$

The second order local shape derivative d^2u needs not to be explicitly known to derive the following estimates.

Assuming that $E_\kappa = 0$ we arrive in view of the shape Taylor expansion (1) at (cf. [2])

$$E_u(\mathbf{x}) = \overline{u}(\mathbf{x}) + \mathcal{O}(\varepsilon^2), \qquad \mathrm{Cor}_u(\mathbf{x},\mathbf{y}) = \varepsilon^2 \mathrm{Cor}_{du}(\mathbf{x},\mathbf{y}) + \mathcal{O}(\varepsilon^3).$$

Herein, $\overline{u}$ and Cor_{du} satisfy the boundary value problems

$$-\Delta \overline{u} = f \text{ in } D, \qquad \overline{u} = g \text{ on } \partial D \tag{3}$$

and

$$\begin{aligned} (\Delta_{\mathbf{x}} \otimes \Delta_{\mathbf{y}})\mathrm{Cor}_{du} &= 0 \text{ in } D \times D, \\ \mathrm{Cor}_{du} &= \mathrm{Cor}_\kappa \left[\frac{\partial(g-\overline{u})}{\partial \mathbf{n}} \otimes \frac{\partial(g-\overline{u})}{\partial \mathbf{n}} \right] \text{ on } \partial D \times \partial D. \end{aligned} \tag{4}$$

Consequently, the expectation of the random solution is with leading order given by the Poisson equation (3). The two-point correlation, the second order statistical moment, is to leading order determined by a partial differential equation in tensor product form, based on the equation for the local shape derivative (2).

Solution by boundary element methods. To numerically solve the deterministic problems (3) and (4) we can use a variational boundary integral equation approach combined with a wavelet discretization. The sparse tensor product approximation is advantageous for the efficient computation of Kronecker products of operators. Thus, the boundary value problem (4) has been formulated as boundary integral equation and approximated by a sparse tensor product wavelet discretization, combined with the wavelet matrix compression strategy from [1]. That way, the algorithmical complexity to compute the first two statistical moments scales essentially (i.e., except for logarithmical terms) proportionally to N, the number of degrees of freedom needed to parametrize the unperturbed, nominal boundary ∂D. That is, the complexity of solving (4) essentially scales like that of solving the mean field equation (3).

Solution by finite element methods. We shall consider a traditional and widely used multilevel hierarchy

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset H^1(D).$$

The sparse tensor product space $\widehat{V}_J \subset H^{1,1}(D \times D)$ is defined via the complementary spaces

$$W_j := V_{j+1} \ominus V_j \subset H^1(D)$$

according to

$$\widehat{V}_J = \bigoplus_{j+j' \le J} W_j \otimes W_{j'} = \sum_{j+j' \le J} V_j \otimes V_{j'} \subset H^{1,1}(D \times D).$$

A basis is given by tensor products of the hierarchical bases of W_j, like e.g. wavelet bases.

Instead of a basis, one can consider the collection of appropriate normed tensor products of the basis functions in V_j. According to [3] this collection forms a frame for the sparse tensor product space. The discretization of boundary value problems by frames and the solution of operator equations in frame coordinates is well understood and quite similarly to the basis case. The algorithms developed in [4], especially the applications of tensor product operators, can be modified to the case of multilevel frames. It turns out that, in order to solve equations of the type (4) efficiently, it suffices to provide standard multigrid hierarchies and associated finite elements together with prolongations and restrictions, see [3]. The computational complexity is essentially the one required to solve the mean field equation (3) by the finite element method.

References

[1] W. Dahmen, H. Harbrecht, and R. Schneider. Compression techniques for boundary integral equations — optimal complexity estimates. *SIAM J. Numer. Anal.* **43** (2006) 2251–2271.

[2] H. Harbrecht, R. Schneider, and C. Schwab. Sparse Second Moment Analysis for Elliptic Problems in Stochastic Domains. *Preprint No. 304*, Berichtsreihe des SFB 611, Universität Bonn, 2006.

[3] H. Harbrecht, R. Schneider, and C. Schwab. Multilevel frames for sparse tensor product spaces. *Bericht 07-14*, Berichtsreihe des Mathematischen Seminars, Christian-Albrechts-Universität zu Kiel, 2007.

[4] C. Schwab and R. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.* **95** (2003), 707–734.

Shearlets: A Wavelet-Based Approach to the Detection of Directional Features

Gitta Kutyniok

(joint work with S. Dahlke, T. Sauer, G. Steidl and G. Teschke)

1. Continuous Shearlet Systems

In data analysis, one main focus of current research is on the development of directional representation systems which precisely detect orientations of singularities like edges in a 2-D image while providing optimally sparse representations. Several approaches have been suggested in the last years such as the ridgelets [1], the curvelets [2], the contourlets [5], and many others.

The shearlet systems are the first directional representation systems, which not only possess the above mentioned properties (cf. [7, 9]), but are moreover equipped with a rich mathematical structure similar to wavelets. The main idea for the construction is employing a two-parameter dilation group, where one parameter ensures the multiscale property, whereas the second parameter provides a means to detect directions. For each $a > 0$ and $s \in \mathbb{R}$, let A_a denote the *parabolic scaling matrix* and S_s denote the *shear matrix* of the form

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix} \qquad \text{and} \qquad S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix},$$

respectively. Then the *(continuous) shearlet system* generated by $\psi \in L^2(\mathbb{R}^2)$ is defined by

$$\{\psi_{ast} = T_t D_{S_s A_a} \psi = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(\cdot - t)) : a \in \mathbb{R}^+,\, s \in \mathbb{R},\, t \in \mathbb{R}^2\},$$

and the associated *Continuous Shearlet Transform* of some $f \in L^2(\mathbb{R}^2)$ is given by

$$\mathcal{SH}_\psi f : \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{C}, \quad \mathcal{SH}_\psi f(a, s, t) = \langle f, \psi_{ast} \rangle.$$

A function $\psi \in L^2(\mathbb{R}^2)$ is called a *continuous shearlet*, if it satisfies the *admissibility condition* $\int_{\mathbb{R}^2} |\hat{\psi}(\xi_1, \xi_2)|^2 / \xi_1^2 \, d\xi < \infty$. In this case, each function $f \in L^2(\mathbb{R}^2)$ can be reconstructed from its shearlet coefficients $\{\langle f, \psi_{ast} \rangle : (a, s, t) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2\}$.

The shearlet systems can also be viewed from a group theoretic point of view. The associated locally compact group – the so-called *Shearlet group* $\mathbb{S}$ – is defined to be the set $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$ endowed with the multiplication

$$(a, s, t) \cdot (a', s', t') = (aa', s + s'\sqrt{a}, t + S_s A_a t'),$$

which is isomorphic to a semi-direct product of the dilation group with $\mathbb{R}^2$. Letting $\sigma : \mathbb{S} \to \mathcal{U}(L^2(\mathbb{R}^2))$ be the unitary representation of this group given by

$$\sigma(a, s, t)\psi(x) = a^{-\frac{3}{4}} \psi(A_a^{-1} S_s^{-1}(x - t)),$$

the link with shearlet systems is established by the relation $\psi_{ast} = \sigma(a, s, t)\psi$.

This rich mathematical structure enables, for instance, the application of uncertainty principles to tune the accuracy of the transform [3] and of special Fourier methods to investigate the discretization process of the Continuous Shearlet Transform [10]. For more information on the theory of shearlets we refer to the webpage www.shearlet.org.

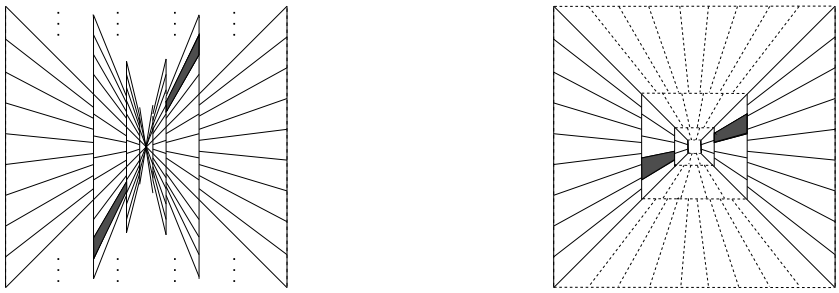
For a large class of continuous shearlets, the Continuous Shearlet Transform precisely detects the wavefront set of distribution. In fact, if $\psi_1 \in L^2(\mathbb{R})$ is a continuous wavelet, $\hat{\psi}_1 \in C^\infty(\mathbb{R})$, and supp $\hat{\psi}_1 \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$, if $\psi_2 \in L^2(\mathbb{R})$ is such that $\hat{\psi}_2 \in C^\infty(\mathbb{R})$ and supp $\hat{\psi}_2 \subseteq [-1, 1]$, and if we define $\psi \in L^2(\mathbb{R}^2)$ by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$

then the wavefront set of a distribution f is the closure of the set of points (s, t) where $\mathcal{SH}_\psi f(a, s, t) = \langle f, \psi_{ast} \rangle$ is of slow decay as $a \to 0$ [9].

2. Discrete Shearlet Systems

For the implementation of the shearlet transform, the shearlet systems need first to be discretized. Two special discrete shearlet systems have already been constructed in [8], which form Parseval frames for $L^2(\mathbb{R}^2)$. The tiling of the frequency plane induced by those systems is illustrated in Figure 1, where the shaded area indicates the support of one of the generating functions.



(a) (b)

Figure 1. The tiling of the frequency domain induced by the discrete shearlets (a) and the discrete shearlets on the cone (b).

For a general discretization strategy, we employ the fact that the Continuous Shearlet Transform is related with group theory as explained above. This property provides us with a link to another central problem in applied analysis on how to measure the smoothness of a given function. Classical approaches are, e.g., based on (strong or weak) derivatives (Hölder and Sobolev spaces), or moduli of smoothness (Besov spaces). However, by means of the concept of square-integrable group representations it is possible to derive a unified approach to many different smoothness measures: they can all be restated in terms of the decay of a particular transform associated with the representation. Moreover, by discretizing the representation in a judicious way, one obtains frames for these smoothness spaces which can therefore be interpreted as the natural building blocks for the underlying transformation. All these relationships have been clarified in the so-called *coorbit space theory* which has been derived by Feichtinger and Gröchenig in a series of papers (see, for instance, [6]).

In [4], we use this approach to derive a new class of smoothness spaces and associated (Banach) shearlet frames. First we remark that the Shearlet group needs to be enlarged to $\mathbb{R} \setminus \{0\} \times \mathbb{R} \times \mathbb{R}^2$ with group multiplication given by $(a, s, t)\,(a', s', t') = (aa', s + s'\sqrt{|a|}, t + S_s A_a t')$ and A_a appropriately adapted, so that the associated representation, which generates the shearlet system, is in fact square-integrable. Denoting this larger group also by $\mathbb{S}$, we let $\psi \in \mathcal{S}(\mathbb{R}^2)$ with some additional support conditions on $\hat{\psi}$, $w \in L_1^{loc}(\mathbb{R}^2)$, and $1 \leq p \leq \infty$. Then the *shearlet coorbit space* $\mathcal{SC}_{p,w}$ is defined by

$$\mathcal{SC}_{p,w} = \{f \in \mathcal{SC}_{1,w}^{\sim} : \mathcal{SH}_\psi f \in L_{p,w}(\mathbb{S})\},$$

where $\mathcal{SC}_{1,w} = \{f \in L_2(\mathbb{R}^2) : \mathcal{SH}_\psi f \in L_{1,w}(\mathbb{S})\}$ and $\mathcal{SC}_{1,w}^{\sim}$ denotes its anti-dual, i.e., the space of all continuous conjugate-linear functionals on $\mathcal{SC}_{1,w}$. The $\mathcal{SC}_{p,w}$-norm of an $f \in \mathcal{SC}_{p,w}$ is given by $\|f\|_{\mathcal{SC}_{p,w}} = \|\mathcal{SH}_\psi f\|_{L_{p,w}}$. This definition is in fact independent of the choice of ψ, and the Schwartz space is contained in $\mathcal{SC}_{p,w}$ for particular weights.

The relation to discrete shearlet systems is now established in the following way. Let $U \subset \mathbb{S}$ be a neighborhood of e which is "small enough", and let $w \in L_1^{loc}(\mathbb{R}^2)$. We prove that provided $\{(a_i, s_i, t_i)\}_{i\in I} \subset \mathbb{S}$ is U-dense and relatively separated, the sequence $\{\mathcal{SH}_\psi f(a_i, s_i, t_i)\}_{i\in I}$ is a Banach frame for $\mathcal{SC}_{p,w}$, hence

$$\|f\|_{\mathcal{SC}_{p,w}} \cong \|\{\mathcal{SH}_\psi f(a_i, s_i, t_i)\}_{i\in I}\|_{\ell_{p,w}}.$$

Moreover, there exists a bounded, linear reconstruction operator $\mathcal{R}$ from $\ell_{p,w}$ to $\mathcal{SC}_{p,w}$ such that

$$\mathcal{R}(\{\mathcal{SH}_\psi f(a_i, s_i, t_i)\}_{i\in I}) = f.$$

Interestingly, one particular example of a sequence satisfying the hypotheses of this result agrees with the discretization chosen in [8] (cf. Figure 1 (a)).

3. Shearlet Multiresolution Analysis

The question remains open whether it is possible to construct an associated multiresolution analysis with finitely supported filters leading to a fast decomposition. In [11], we show that this indeed can be achieved by constructing a (non-stationary) bivariate adaptive directional subdivision scheme in the following way.

Considering the shearlet dilation matrices $M_\varepsilon := S_{-\varepsilon} A_{\frac{1}{2}}$, $\varepsilon \in \{0, 1\}$, we observe that these indeed preserve the lattice structure in contrast to a rotation matrix. In fact, for all $j \in \mathbb{Z}$, we have $M_\varepsilon(4^{-j}\mathbb{Z} \times 2^{-j}\mathbb{Z}) = 4^{-(j+1)}\mathbb{Z} \times 2^{-(j+1)}\mathbb{Z}$. Now defining two expanding matrices by $W_\varepsilon := M_\varepsilon^{-1}$ and choosing two finitely supported masks $a_\varepsilon \in \ell(\mathbb{Z}^2)$, $\varepsilon \in \{0, 1\}$, we can define the associated subdivision schemes by

$$S_\varepsilon c = \sum_{\alpha\in\mathbb{Z}^2} a_\varepsilon \left(\cdot - W_\varepsilon\, \alpha\right)\, c\left(\alpha\right), \qquad \varepsilon \in \{0, 1\},\ c \in \ell(\mathbb{Z}^2).$$

The two subdivision schemes S_0 and S_1 are now applied to some data $c \in \ell(\mathbb{Z}^2)$ in an iterative way, thereby providing the opportunity to adaptively change the orientation of the data during the subdivision process. In this sense we might view the subdivision process as a binary tree, in which the direction of the refined data is dependent on the branch we choose. Convergence certainly needs to apply to each branch of the tree. Therefore, we call this subdivision scheme *convergent in* $C\left(\mathbb{R}^2\right)$, if for any sequence ε in $\{0, 1\}^{\mathbb{N}}$ the associated sequence of subdivision schemes applied to δ converges to some nonzero uniformly continuous function f_ε. If the masks are chosen for this to be satisfied, the limit functions f_ε, $\varepsilon \in \{0, 1\}^{\mathbb{N}}$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$, fulfill the *refinement equation*

$$f_\varepsilon = \sum_{\alpha\in\mathbb{Z}^2} a_{\varepsilon_1}(\alpha)\, f_{\widehat{\varepsilon}}\left(W_{\varepsilon_1} \cdot - \alpha\right), \qquad \widehat{\varepsilon} := (\varepsilon_2, \varepsilon_3, \dots).$$

Moreover, we derive a complete characterization of convergence of these subdivision schemes in terms of the restricted joint spectral radius and membership of the z-transform of the masks in a quotient ideal of the Laurent polynomials.

Returning to the initial question concerning a shearlet multiresolution analysis, we employ the fact that each multiresolution analysis is associated with a subdivision scheme and introduce scaling spaces based on the previously constructed directional subdivision schemes. Using the above mentioned refinement equations, we prove that these indeed provide a multiresolution analysis structure. This then leads to a fast decomposition in a very natural way, where we focus on the interpolatory case.

References

[1] E. J. Candès and D. L. Donoho, *Ridgelets: a key to higher-dimensional intermittency?,* Phil. Trans. R. Soc. Lond. A. **357** (1999), 2495–2509.

[2] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with C^2 singularities*, Comm. Pure Appl. Math. **56** (2004), 219–266.

[3] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke, *The uncertainty principle associated with the Continuous Shearlet Transform,* Int. J. Wavelets Multiresolut. Inf. Process., to appear.

[4] S. Dahlke, G. Kutyniok, G. Steidl and G. Teschke, *Shearlet Coorbit Spaces and associated Banach frames*, preprint, 2007.

[5] M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multiresolution image representation*, IEEE Trans. Image Process. **14** (2005), 2091–2106.

[6] H. G. Feichtinger and K. Gröchenig, *Banach spaces related to integrable group representations and their atomic decomposition I,* J. Funct. Anal. **86** (1989), 307–340.

[7] K. Guo and D. Labate, *Optimally sparse multidimensional representations using shearlets,* SIAM J Math. Anal. **39** (2007), 298–318.

[8] K. Guo, G. Kutyniok, and D. Labate, *Sparse multidimensional representations using anisotropic dilation und shear operators,* in Wavelets und Splines (Athens, GA, 2005), G. Chen and M. J. Lai, eds., Nashboro Press, Nashville, TN (2006), 189–201.

[9] G. Kutyniok and D. Labate, *Resolution of the wavefront set using continuous shearlets,* preprint, 2006.

[10] G. Kutyniok and D. Labate, *Construction of regular and irregular shearlet frames,* J. Wavelet Theory and Appl. **1** (2007), 1–10.

[11] G. Kutyniok and T. Sauer, *Adaptive directional subdivision schemes and shearlet multiresolution analysis*, preprint, 2007.

Nonlinear Multiresolution Analysis

Peter Oswald

(joint work with S. Harizanov)

From an algorithmical point of view, multiresolution analysis is governed by two sets of operators acting between consecutive spaces V_j, $j \geq 0$, representing different resolution levels: Restrictions (also called decimation operators) $R_j : V_j \to V_{j-1}$ and prolongations (depending on context, also called interpolation, prediction, or subdivision operators) $P_j : V_{j-1} \to V_j$. The associated pyramid algorithm

$$v_{j-1} := R_j v_j, \qquad d_j := v_j - P_j v_{j-1} \quad (\iff \quad v_j = P_j v_{j-1} + d_j \ )$$

leads to a multiscale representation

$$v_J \quad \leftrightarrow \quad \{v_0, d_1, \ldots, d_J\}$$

of fine-scale data v_J by coarse-scale data v_0 and details d_j from all intermediate resolution levels. While most of the existing multiscale analysis is carried out when the V_j are nested linear spaces of exponentially growing dimension, and the R_j, P_j are linear operators acting between them, during the last 10 years many applications have used nonlinear setups. E.g., normal multiresolution for geometry compression [7], edge-adapted image analysis schemes [1], morphological pyramid algorithms [6, 5], subdivision schemes for the parametrization of manifolds [9, 10], etc. all use nonlinear P_j and/or R_j.

The talk draws attention to this emerging area of research on nonlinear multiresolution analysis. Although some papers have already appeared (e.g., on the convergence and smoothness analysis of special cases of nonlinear schemes), much needs still to be done. In particular, the investigation of stability of nonlinear pyramid algorithms and subdivision schemes has only begun. Using the language of dynamical systems, a (nonlinear) subdivision scheme

$$u_0 \longmapsto u_1 := P_1 u_0 \longmapsto \ldots \longmapsto u_j := P_j u_{j-1} \longmapsto \ldots$$

is called *Lyapunov stable at* $\tilde{u}_0$ if for any $\epsilon > 0$ there is a $\delta > 0$ such that for all u_0 in a δ-neighborhood of $\tilde{u}_0$, the u_j stay in a ϵ-neighborhood of $\tilde{u}_j$ for all $j > 0$, i.e.,

$$\|u_0 - \tilde{u}_0\| \le \delta \qquad \Longrightarrow \qquad \|u_j - \tilde{u}_j\| \le \epsilon \qquad \forall\, j > 0.$$

For simplicity, we have assumed that all V_j are subsets of the same normed space. A more quantitative, stronger definition is *Lipschitz stability at* $\tilde{u}_0$: There is a $C < \infty$ such that for all u_0 in some neighborhood of $\tilde{u}_0$

$$\|u_j - \tilde{u}_j\| \le C\|u_0 - \tilde{u}_0\| \qquad \forall\, j > 0.$$

Similarly, the pyramid transform governed by the P_j is called Lipschitz stable if

$$\|v_j - \tilde{v}_j\| \le C(\|v_0 - \tilde{v}_0\| + \sum_{k=1}^{j} \|d_k - \tilde{d}_k\|) \qquad \forall\, j > 0. \tag{1}$$

This notion of stability is closely related to error control in lossy compression algorithms using the given nonlinear pyramid transform.

So far, only scattered results are known. In [3], a theory for stationary univariate nonlinear subdivision in $\ell_p(\mathbf{Z})$ spaces has been developed. The basic assumption is that the nonlinear subdivision operator S is, for each $v \in \ell_p(\mathbf{Z})$, written in the form $Sv = S(v)v$, where $S(v)$ is a linear subdivision operator (w.r.t. dyadic dilation on $\mathbf{Z}$). Then, sufficient conditions for the convergence of the subdivision sequence $v_j := S^j v$ as $j \to \infty$, the smoothness of the limits in Hölder-Sobolev spaces, and the Lipschitz stability of the subdivision scheme can be formulated in terms of properties of the associated family $\{S(v)\}$ of linear subdivision operators. The crucial condition for Lipschitz stability is Lipschitz continuity w.r.t. v:

$$\|S(v) - S(u)\| \le C\|v - u\|, \tag{2}$$

where the constant $C < \infty$ may depend on $\max(\|u\|, \|v\|)$ in a monotone way (the norms are induced by $\ell_p(\mathbf{Z})$). This condition holds for WENO-type subdivision schemes but is unfortunately not satisfied for many other schemes such as the PPH scheme [2] and median subdivision [5]. For the dyadic interpolatory PPH scheme given by the formulas $(Sv)_{2i} = v_i$ and

$$(Sv)_{2i+1} = \frac{v_i + v_{i+1}}{2} - \frac{(\Delta^2 v_i \cdot \Delta^2 v_{i+1})_+}{4(\Delta^2 v_i + \Delta^2 v_{i+1})}, \qquad \Delta^2 v_i := v_{i+1} - 2v_i + v_{i-1},$$

Lipschitz stability has been established in [2] by proving a contraction property of the form

$$\|\Delta^2(S^2 v - S^2 u)\| \le \frac{3}{4}\|\Delta^2(v-u)\| \qquad \forall\, u, v \in \ell_\infty(\mathbf{Z}). \tag{3}$$

For this particular scheme, even Lipschitz stability (1) for the full pyramid transform has been shown [2]. Although stability is also discussed in other papers [1, 4, 11], further rigorous results on Lipschitz stability as defined above are missing.

We are currently working towards systematic generalizations of these scattered results (joint work with S. Harizanov, PhD student at Jacobs University Bremen). A benchmark case under consideration is median interpolation subdivision, originally introduced in [5] for heavy tail noise removal from univariate time series. Convergence and Hölder smoothness results for this scheme have been established in [5, 8, 12]. However, the Lipschitz stability problem is still open. Besides a refinement of conditions such as (3), a systematic use of tools from dynamical systems theory seems necessary to achieve further progress.

References

[1] S. Amat, F. Arandiga, A. Cohen, R. Donat, G. Garcia, M. v. Oehsen, *Data compression with ENO schemes: A case study*, Appl. Comput. Harmon. Anal. **11** (2001), 273–288.

[2] S. Amat, J. Liandrat, *On the stability of the PPH nonlinear multiresolution*, Appl. Comput. Harmon. Anal. **18** (2005), 198–206.

[3] A. Cohen, N. Dyn, B. Matei, *Quasilinear subdivision schemes with applications to ENO interpolation*, Appl. Comput. Harmon. Anal. **15** (2003), 89–116.

[4] I. Daubechies, O. Runborg, W. Sweldens, *Normal multiresolution approximation of curves*, Constr. Approx. **20** (2004), 399–463.

[5] D. L. Donoho, T. P.-Y. Yu, *Nonlinear pyramid transform based on median interpolation*, SIAM J. Math. Anal. **31** (2000), 1030–61.

[6] J. Goutsias, H. Heijmans, *Nonlinear multiresolution signal decomposition schemes - Part I: Morphological pyramids, Part II: Morphological wavelets*, IEEE Trans. Image Proc. **9** (2000), 1862–76, 1897–1913.

[7] I. Guskov, K. Vidimce, W. Sweldens, P. Schröder, *Normal meshes*, In: Computer Graphics (SIGGRAPH'00 Proceedings), New York, ACM Press, 2000, pp 259–268.

[8] P. Oswald, *Smoothness of a nonlinear subdivision scheme*, Adv. Comput. Math. **20** (2004), 401–423.

[9] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, P. Schröder, *Multiscale representation of mani-fold valued data*, SIAM J. Multiscale Model. Sim. **4** (2005), 1201–32.

[10] J. Wallner, *Smoothness analysis of subdivision schemes by proximity*, Constr. Approx. **24** (2006), 289–318.

[11] J. Wallner, N. Dyn, *Convergence and C^1 analysis of subdivision schemes on manifolds by proximity*, Comput. Aided Geom. Design **22** (2005), 593–622.
[12] G. Xie, T. P.-Y. Yu, *Smoothness analysis of nonlinear subdivision schemes of homogeneous and affine invariant type*, Constr. Approx. **22** (2005), 219–254.

Wavelet Methods for PDE-Constrained Elliptic Control Problems with Dirichlet Boundary Control

Roland Pabel

Given a domain $\Omega = (0,1)^d$ with spatial dimension $d \geq 2$ and fixing the observation $\Gamma_Y \subset \partial\Omega$ and control boundaries $\Gamma \subset \partial\Omega$, the control problem is stated as follows:

For some given data y_{Γ_Y} and f, minimize

$$\mathcal{J}(y,u) = \frac{1}{2}\|y - y_{\Gamma_Y}\|^2_{H^s(\Gamma_Y)} + \frac{\omega}{2}\|u\|^2_{H^t(\Gamma)} \tag{1}$$

where the state y *and the* control u *are coupled through the linear elliptic boundary value problem*

$$\begin{aligned} -\nabla\cdot(\mathbf{a}\nabla y) + a_0\, y &= f && \text{in } \Omega, \\ y &= u && \text{on } \Gamma, \\ (\mathbf{a}\nabla y)\cdot\mathbf{n} &= 0 && \text{on } \partial\Omega\setminus\Gamma. \end{aligned} \tag{2}$$

Here we are not only faced with a problem formulation which is much more complex than calculating the solution of a single elliptic PDE alone, but also with the added complexity of the minimization formulation involving evaluation of Sobolev norms of fractional order in the cost functional (1). The intrinsic space of the trace of a function of the Sobolev space $H^1(\Omega)$ is the space $H^{1/2}(\partial\Omega)$ and as such we have to consider fractional Sobolev norms, i.e. the natural case is $s = t = 1/2$.

Wavelet methods allow us to tackle both problems simultaneously. Norm equivalences in the wavelet setting are used for optimal preconditioning of discretized systems [2, 4] and here also for evaluating the Sobolev norms as precisely as possible.

The construction of boundary adapted biorthogonal wavelets on the interval $[0,1]$ based on B-Splines (piecewise polynomials) was first set out in [3]. Additionally, we use specialized constructions including basis transformations for lowering the overall absolute condition numbers, for details see [6]. Wavelet bases for the higher dimensional domain $\Omega \subset \mathbb{R}^d$ are realized here by tensor product construction of these univariate bases.

We will make use of the *Riesz basis* property of a wavelet basis Ψ indexed by an infinite set $\mathcal{I}$, i.e. for a (sufficiently smooth) function $v = \mathbf{v}^T\Psi \in L_2$ holds $\|v\|_{H^s} \sim \|\mathbf{D^s}\,\mathbf{v}\|_{\ell_2(\mathcal{I})}$ for a range of values $s \in (s_*, s^*)$, not only as a means of optimal preconditioning (see e.g. [2]) but also for evaluating the Sobolev norms of the cost functional. To this end, we introduce *Riesz Operators* $R_{H^s} : H^s \to H^{-s}$ with $\langle Rv, w\rangle_{H^{-s}\times H^s} := (v,w)_{H^s}$. Since the exact H^s norm is inaccessible for $s \notin \mathbb{Z}$, we use different constructions based on scaling or interpolation instead,

see e.g. [6]. We then get norm equivalences $\|v\|_{H^s} \sim \|\mathbf{R}^{1/2}\,\mathbf{D}^{\mathbf{s}}\,\mathbf{v}\|_{\ell_2(I\!I)}$ with smaller absolute constants than before.

Our boundary value problem (2) is now expressed as a saddle point problem using the Lagrange multiplier method. To this end, we introduce the trace operator onto the control boundary, $B : H^1(\Omega) \to H^{1/2}(\Gamma)$, with which the Dirichlet boundary conditions are expressed. The norm on the observation boundary Γ_Y in (1) is evaluated with help of another trace operator, $T : H^1(\Omega) \to H^{1/2}(\Gamma_Y)$.

Fixing wavelets and index sets for the spaces $X = H^1(\Omega)$, $Y = H^s(\Gamma_Y)$ and $Q = H^t(\Gamma)$, our control problem in infinite wavelet coordinates then reads:
Given $(\mathbf{f}, \mathbf{y}_{\Gamma_Y}) \in \ell_2(I\!I_X \times I\!I_Y)$, *find* $(\mathbf{y}, \mathbf{p}, \mathbf{u}) \in \ell_2(I\!I_X \times I\!I_Q \times I\!I_Q)$, *such that*

$$\mathbf{J}(\mathbf{y}, \mathbf{u}) = \frac{1}{2}\|\mathbf{R}_{\Gamma_Y}^{1/2}(\mathbf{T}\mathbf{y} - \mathbf{y}_{\Gamma_Y})\|^2_{\ell_2(I\!I_Y)} + \frac{\omega}{2}\|\mathbf{R}_{\Gamma}^{1/2}\mathbf{u}\|^2_{\ell_2(I\!I_Q)}, \tag{3}$$

is minimized subject to

$$\mathbf{L}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} := \begin{pmatrix}\mathbf{A} & \mathbf{B}^T\\ \mathbf{B} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}\mathbf{f}\\ \mathbf{u}\end{pmatrix}. \tag{4}$$

The necessary and sufficient conditions for this optimal control problem can be derived by appending the constraint (4) to the functional (3) by means of Lagrangian multipliers $(\mathbf{z}, \boldsymbol{\mu})^T$ and calculating the first variation of this Lagrangian functional with respect to the now five unknowns $(\mathbf{y}, \mathbf{p}, \mathbf{z}, \boldsymbol{\mu}, \mathbf{u})^T$:

$$\begin{aligned}\mathbf{L}\begin{pmatrix}\mathbf{y}\\ \mathbf{p}\end{pmatrix} &= \begin{pmatrix}\mathbf{f}\\ \mathbf{u}\end{pmatrix},\\ \mathbf{u} &= \omega^{-1}\,\mathbf{R}_{\Gamma}^{-1}\,\boldsymbol{\mu},\\ \mathbf{L}^T\begin{pmatrix}\mathbf{z}\\ \boldsymbol{\mu}\end{pmatrix} &= \begin{pmatrix}-\mathbf{T}^T\mathbf{R}_{\Gamma_Y}\,(\mathbf{T}\,\mathbf{y} - \mathbf{y}_{\Gamma_Y})\\ \mathbf{0}\end{pmatrix}.\end{aligned} \tag{5}$$

There are several approaches possible for solving these coupled equations. Eliminating the control variable $\mathbf{u}$ from (5) gives rise to an *All-In-Solver* which is numerically unfavorable, see [6]. An *Inexact Gradient* method was proposed in [5] and numerical results for this setup were given in [6], but the speed of this method depends to a large extent on choosing the right step size parameter. We present an *Inexact Conjugate Gradient (ICG)* method which relinquishes the step size parameter and gives rise to optimal convergence results.

The ICG method is based on the fact that the optimal control u^* is the unique global minimizer of the reduced cost functional [5], i.e.

$$\delta\mathbf{J}(\mathbf{u}^*) \equiv \mathbf{Q}\mathbf{u}^* - \mathbf{g} = 0,$$

where $\mathbf{Q}$ is a symmetric positive definite matrix,

$$\mathbf{Q} = \mathbf{S}^{-1}\mathbf{B}\mathbf{A}^{-T}\mathbf{T}^T\mathbf{R}_{\Gamma_Y}\mathbf{T}\mathbf{A}^{-1}\mathbf{B}^{-T}\mathbf{S}^{-T} + \omega\,\mathbf{R}_{\Gamma}, \quad \mathbf{S} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T,$$

and the affine term $\mathbf{g}$ depends only on $\mathbf{f}$ and $\mathbf{y}_{\Gamma_Y}$,

$$\mathbf{g} = -\mathbf{S}^{-1}\mathbf{B}\mathbf{A}^{-T}\mathbf{T}^T\mathbf{R}_{\Gamma_Y}\left(\mathbf{T}\,\widetilde{\mathbf{f}} - \mathbf{y}_{\Gamma_Y}\right), \quad \widetilde{\mathbf{f}} := \mathbf{A}^{-1}\left(\mathbf{I} - \mathbf{B}^T\mathbf{S}^{-1}\mathbf{B}\mathbf{A}^{-1}\right)\mathbf{f}\,.$$

The algorithmic description to compute these analytic expressions up to a given tolerance η can be found in Figures 1 and 2.

Our properly scaled discretized systems $\mathbf{A}, \mathbf{L}$ have uniformly bounded condition numbers. The SOLVE algorithms can be selected as CG or UZAWA methods, which thus reduce the residual error by any factor in uniformly bounded iteration numbers. The linear equations are thus solved up to discretization error accuracy with complexity and in time proportional to the number of unknowns.

$\textsc{ApplyQ}[\eta, \mathbf{u}] \to (\mathbf{Q}\,\mathbf{u})_\eta$

$\quad \textsc{Solve}\ \left[\frac{1}{2}c_P\eta, \mathbf{L}, \begin{pmatrix}\mathbf{0}\\ \mathbf{u}\end{pmatrix}\right] \to \begin{pmatrix}\mathbf{y}_\eta\\ \mathbf{p}_\eta\end{pmatrix}$

$\quad \textsc{Solve}\ \left[\frac{1}{2}c_A\eta, \mathbf{L}^T, \begin{pmatrix}-\mathbf{T}^T\mathbf{R}_{\Gamma_Y}\mathbf{T}\,\mathbf{y}_\eta\\ \mathbf{0}\end{pmatrix}\right] \to \begin{pmatrix}\mathbf{z}_\eta\\ \boldsymbol{\mu}_\eta\end{pmatrix}$

$\quad \textsc{Return}\ (\mathbf{Q}\,\mathbf{u})_\eta := \omega\,\mathbf{R}_\Gamma\,\mathbf{u} - \boldsymbol{\mu}_\eta$

FIGURE 1. Algorithm APPLYQ calculates the matrix product $\mathbf{Qu}$ up to given error accuracy η: $\|\mathbf{Qu} - (\mathbf{Qu})_\eta\|_{\ell_2(I\!I_Q)} \le \eta$.

$\textsc{InexactRHS}\ [\eta, \mathbf{f}, \mathbf{y}_{\Gamma_Y}] \to \mathbf{g}_\eta$

$\quad \textsc{Solve}\ \left[\frac{1}{2}c_P\eta, \mathbf{L}, \begin{pmatrix}\mathbf{f}\\ \mathbf{0}\end{pmatrix}\right] \to \begin{pmatrix}\widetilde{\mathbf{f}}_\eta\\ \mathbf{p}_\eta\end{pmatrix}$

$\quad \textsc{Solve}\ \left[\frac{1}{2}c_A\eta, \mathbf{L}^T, \begin{pmatrix}-\mathbf{T}^T\mathbf{R}_{\Gamma_Y}(\mathbf{T}\,\widetilde{\mathbf{f}}_\eta - \mathbf{y}_{\Gamma_Y})\\ \mathbf{0}\end{pmatrix}\right] \to \begin{pmatrix}\mathbf{z}_\eta\\ \boldsymbol{\mu}_\eta\end{pmatrix}$

$\quad \textsc{Return}\ \mathbf{g}_\eta := \boldsymbol{\mu}_\eta$

FIGURE 2. Algorithm INEXACTRHS calculates the right hand side $\mathbf{g}$ up to given error accuracy η: $\|\mathbf{g} - \mathbf{g}_\eta\|_{\ell_2(I\!I_Q)} \le \eta$.

The ICG algorithm described in Figure 3 is based upon a nested-iteration strategy and solves the control problem with complexity proportional to the number of unknowns on the highest level J. The numerical results in Table 1 show that the internal CG steps k_J are uniformly bounded as well as the mean number of steps necessary to solve the primal (#P-It/k_J) and adjoint (#A-It/k_J) systems. There was no Inexact-CG step on level $J = 9$ necessary because the residual error was already below $0.01 * 2^{-9}$. The third column shows the error in the control declining from level to level en par with the residual error $\mathbf{r}_J^{(k_J)}$ in the first column.

The results show that wavelet discretizations are suitable for exact numerical representation of control problems with Dirichlet boundary control. The optimal

INEXACTCG $[j, J] \to \mathbf{u}^j$
(i) SET $j := j$, $\mathbf{u}^j := \mathbf{u_0}$
(ii) WHILE $j \leq J$
SET $h_j := 2^{-j}$
ASSEMBLE $\mathbf{f}^j, \mathbf{y}^j_{\Gamma_Y}$
INEXACTRHS$[h_j, \mathbf{f}^j, \mathbf{y}^j_{\Gamma_Y}] \to \mathbf{g}^j$
USE REGULAR CG WITH APPLYQ$[c_Q h_j, \cdot]$
TO CALCULATE $\|\mathbf{Q}\,\mathbf{u}^j - \mathbf{g}^j\|_{\ell_2(\mathbb{I}_Q)} \leq h_j$
SET $j := j + 1$ AND PROLONGATE $\mathbf{u}^j$
(iii) RETURN $\mathbf{u}^J$

FIGURE 3. The INEXACTCG algorithm computes $\mathbf{u}_J$ up to discretization error accuracy $h_J := 2^{-J}$.

J	$\|\mathbf{r}_J^{(k_J)}\|_{\ell_2}$	$\|\mathbf{u}^* - \mathbf{u}_J^{(k_J)}\|_{\ell_2}$	k_J	$\frac{\#\text{P-It}}{k_J}$	$\frac{\#\text{A-It}}{k_J}$
4	4.1123e−03	4.1231e−03	1	–	–
5	8.2395e−06	8.6744e−05	2	3	1
6	1.8710e−05	8.0957e−05	1	3	1
7	9.5404e−07	2.3294e−05	1	2	1
8	1.4663e−06	6.5335e−06	1	1	1
9	1.1902e−05	– "" –	–	–	–
10	2.6704e−07	1.8761e−06	1	2	1

TABLE 1. Inexact CG example results. The stopping criteria on each level is $\|\mathbf{r}_J\|_{\ell_2} \leq 0.01 * 2^{-j}$. We use a direct solver on level $J = 4$.

preconditioning ensures low absolute condition and iteration numbers and the norm equivalence property in combination with Riesz operators can improve the accuracy of the model substantially. Changes in the domain Ω and the boundaries Γ, Γ_Y can be implemented via a fictitious domain approach and by exchanging trace operators.

REFERENCES

[1] C. Burstedde and A. Kunoth, *Fast Iterative Solution of Elliptic Control Problems in Wavelet Discretizations*, J. Comp. Appl. Maths. **196(1)** (2006), 299–319
[2] W. Dahmen, *Wavelet and Multiscale Methods for Operator Equations*, Acta Numerica (1997), 55–228
[3] W. Dahmen, A. Kunoth and K. Urban, *Biorthogonal Spline-Wavelets on the Interval – Stability and Moment Conditions*, Appl. Comput. Harm. Anal. **6** (1999), 132–196
[4] A. Kunoth, *Wavelet Methods – Elliptic Boundary Value Problems and Control Problems*, Teubner Verlag, Advances in Numerical Mathematics (2001)

[5] A. Kunoth, *Fast Iterative Solution of Saddle Point Problems in Optimal Control Based on Wavelets*, Comput. Optim. Appl. **22** (2002), 225–259
[6] R. Pabel, *Wavelet Methods for PDE Constrained Elliptic Control Problems with Dirichlet Boundary Control*, doi:10.2370/236_232 (2007)

Decomposition of Weighted Besov and Triebel-Lizorkin Spaces and Nonlinear Approximation on the Ball

Pencho Petrushev
(joint work with George Kyriazis and Yuan Xu)

Localized bases and frames allow to decompose functions and distributions in terms of building blocks of simple nature and have numerous advantages over other means of representation. In particular, they enable one to encode smoothness and other norms in terms of the coefficients of the decompositions. Meyer's wavelets [6] and the φ-transform of Frazier and Jawerth [1, 2, 3] provide such building blocks for decomposition of Triebel-Lizorkin and Besov spaces in the classical case on $\mathbb{R}^d$.

Our aim is to develop similar tools for decomposition of weighted Triebel-Lizorkin and Besov spaces on the unit ball B^d in $\mathbb{R}^d$ $(d > 1)$ with weights

$$\omega_\mu(x) := (1 - |x|^2)^{\mu - 1/2}, \quad \mu \geq 0,$$

were $|x|$ is the Euclidean norm of $x \in B^d$. These include $L_p(B^d, \omega_\mu)$, the Hardy spaces $H_p(B^d, \omega_\mu)$, and weighted Sobolev spaces. For our purposes we develop localized frames which can be viewed as an analogue of the φ-transform of Frazier and Jawerth on B^d.

For the construction of our frame elements we use orthogonal polynomials in the weighted space $L_2(\omega_\mu) := L_2(B^d, \omega_\mu)$. Denote by Π_n the space of all algebraic polynomials of degree n in d variables and by V_n the subspace of all polynomials of degree n which are orthogonal to lower degree polynomials in $L_2(\omega_\mu)$. We have the orthogonal polynomial decomposition

$$L_2(\omega_\mu) = \bigoplus_{n=0}^{\infty} V_n, \qquad V_n \subset \Pi_n.$$

As is shown in [13] the orthogonal projector $\mathrm{Proj}_n : L_2(\omega_\mu) \mapsto V_n$ can be written as

$$(\mathrm{Proj}_n f)(x) = \int_{B^d} f(y) \mathsf{P}_n(x, y) \omega_\mu(y) dy,$$

where, for $\mu > 0$, the kernel $\mathsf{P}_n(x, y)$ has the representation

$$\mathsf{P}_n(x, y) = b_d^\mu b_1^{\mu - \frac{1}{2}} \frac{n + \lambda}{\lambda} \times \int_{-1}^{1} C_n^\lambda \left(\langle x, y \rangle + u \sqrt{1 - |x|^2} \sqrt{1 - |y|^2} \right) (1 - u^2)^{\mu - 1} du. \tag{1}$$

Here $\langle x, y \rangle$ is the Euclidean inner product in $\mathbb{R}^d$, C_n^λ is the n-th degree Gegenbauer polynomial, $\lambda = \mu + \frac{d-1}{2}$, and the constants b_d^μ, $b_1^{\mu - \frac{1}{2}}$ are defined by $(b_d^\gamma)^{-1} :=$

$\int_{B^d}(1-|x|^2)^{\gamma-1/2}dx$. For a representation of $\mathsf{P}_n(x,y)$ in the limiting case $\mu=0$, see (4.2) in [11].

Evidently,

$$K_n(x,y) := \sum_{j=0}^{n} \mathsf{P}_j(x,y) \tag{2}$$

is the kernel of the orthogonal projector of $L_2(\omega_\mu)$ onto the space $\bigoplus_{\nu=0}^{n} V_\nu$.

A key role in this study plays the fact (established in [11]) that if the coefficients on the right in (2) are "smoothed out" by sampling a compactly supported C^∞ function, then the resulting kernel has nearly exponential localization around the main diagonal $y=x$ in $B^d\times B^d$. More precisely, let

$$L_n(x,y) := \sum_{j=0}^{\infty} \widehat{a}\Big(\frac{j}{n}\Big)\mathsf{P}_j(x,y),$$

where the "smoothing" function $\widehat{a}$ is admissible in the sense of the following definition:

Definition: A function $\widehat{a}\in C^\infty[0,\infty)$ is called admissible of type
(a) if $\operatorname{supp}\widehat{a}\subset[0,2]$ and $\widehat{a}(t)=1$ on $[0,1]$, and of type
(b) if $\operatorname{supp}\widehat{a}\subset[1/2,2]$.
We introduce the distance

$$d(x,y) := \arccos\Big\{\langle x,y\rangle + \sqrt{1-|x|^2}\sqrt{1-|y|^2}\Big\} \quad \text{on } B^d$$

and set

$$W_\mu(n;x) := \Big(\sqrt{1-|x|^2}+n^{-1}\Big)^{2\mu}, \quad x\in B^d.$$

The main result in [11, Theorem 4.2] asserts that for any $k>0$ there exists a constant $c_k>0$ depending only on k, d, μ, and $\widehat{a}$ such that

$$|L_n(x,y)| \le c_k \frac{n^d}{\sqrt{W_\mu(n;x)}\sqrt{W_\mu(n;y)}(1+n\,d(x,y))^k}, \quad x,y\in B^d.$$

The kernels L_n are our main ingredient in constructing *analysis* and *synthesis* needlet systems $\{\varphi_\xi\}_{\xi\in\mathcal{X}}$ and $\{\psi_\xi\}_{\xi\in\mathcal{X}}$, indexed by a multilevel set $\mathcal{X}=\cup_{j=0}^{\infty}\mathcal{X}_j$. This is a pair of dual frames whose elements have nearly exponential localization on B^d and provide representation of every distribution f on B^d or $f\in L_p(\omega_\mu)$:

$$f = \sum_{\xi\in\mathcal{X}} \langle f,\varphi_\xi\rangle\psi_\xi. \tag{3}$$

The superb localization of the frame elements prompted us to term them *needlets*.

Our main interest lies with distributions in the weighted Triebel-Lizorkin (F-spaces) and Besov spaces (B-spaces) on B^d. These spaces are naturally defined via spectral decompositions (see [9, 12] for the general idea). To be specific, let

$$\Phi_0(x,y) := 1 \quad \text{and} \quad \Phi_j(x,y) := \sum_{\nu=0}^{\infty} \widehat{a}\Big(\frac{\nu}{2^{j-1}}\Big)\mathsf{P}_\nu(x,y), \quad j\ge 1,$$

where $\mathsf{P}(\cdot,\cdot)$ is from (1) and $\widehat{a}$ is admissible of type (b) such that $|\widehat{a}| > 0$ on $[3/5, 5/3]$.

The F-space $F^{s\rho}_{pq}$ with $s, \rho \in \mathbb{R}$, $0 < p < \infty$, $0 < q \le \infty$, is defined as the space of all distributions f on B^d such that

$$\|f\|_{F^{s\rho}_{pq}} := \Big\| \Big(\sum_{j=0}^{\infty} (2^{sj} W_\mu(2^j;\cdot)^{-\rho/d} |\Phi_j * f(\cdot)|)^q \Big)^{1/q} \Big\|_p < \infty, \tag{4}$$

where $\Phi_j * f(x) := \langle f, \overline{\Phi(x,\cdot)} \rangle$. Here and in what follows $\|\cdot\|_p := \|\cdot\|_{L_p(\omega_\mu)}$.

The corresponding scales of weighted Besov spaces $B^{s\rho}_{pq}$ with $s, \rho \in \mathbb{R}$, $0 < p, q \le \infty$, are defined via the (quasi-)norms

$$\|f\|_{B^{s\rho}_{pq}} := \Big(\sum_{j=0}^{\infty} \Big(2^{sj} \|W_\mu(2^j;\cdot)^{-\rho/d} \Phi_j * f(\cdot)\|_p \Big)^q \Big)^{1/q}. \tag{5}$$

Unlike in the classical case on $\mathbb{R}^d$, we have introduced an additional parameter ρ, which allows considering different scales of Triebel-Lizorkin and Besov spaces. To us most natural are the spaces $F^s_{pq} := F^{ss}_{pq}$ and $B^s_{pq} := B^{ss}_{pq}$, which embed correctly with respect to the smoothness parameter s. A "classical" choice would be to consider the spaces F^{s0}_{pq} and B^{s0}_{pq}, where the weight $W_\mu(2^j;\cdot)$ is excluded from (4)-(5). The introduction of the parameter ρ enables us to treat these spaces simultaneously.

One of our main results [5] is the characterization of the F-spaces in terms of the size of the needlet coefficients in the decomposition (3), namely,

$$\|f\|_{F^{s\rho}_{pq}} \sim \Big\| \Big(\sum_{j=0}^{\infty} 2^{sjq} \sum_{\xi \in \mathcal{X}_j} |\langle f, \varphi_\xi \rangle| W_\mu(2^j;\xi)^{-\rho/d} |\psi_\xi(\cdot)|^q \Big)^{1/q} \Big\|_p.$$

Similarly for the Besov spaces $B^{s\rho}_{pq}$ we have the characterization

$$\|f\|_{B^{s\rho}pq} \sim \Big(\sum_{j=0}^{\infty} 2^{sjq} \Big[\sum_{\xi \in \mathcal{X}_j} \Big(W_\mu(2^j;\xi)^{-\rho/d} \|\langle f, \varphi_\xi \rangle \psi_\xi\|_p \Big)^p \Big]^{q/p} \Big)^{1/q}.$$

Further, the weighted Besov spaces are applied to nonlinear n-term approximation from needlets on B^d. Consider nonlinear n-term approximation for a single needlet system $\{\psi_\eta\}_{\eta \in \mathcal{X}}$ (i.e. $\varphi_\eta = \psi_\eta$). Suppose Σ_n is the nonlinear set of all functions g of the form $g = \sum_{\xi \in \Lambda} a_\xi \psi_\xi$, where $\Lambda \subset \mathcal{X}$, $\#\Lambda \le n$, and Λ may vary with g. Let $\sigma_n(f)_p$ denote the error of best $L_p(\omega_\mu)$-approximation to $f \in L_p(\omega_\mu)$ from Σ_n, i.e.

$$\sigma_n(f)_p := \inf_{g \in \Sigma_n} \|f - g\|_p.$$

We consider approximation in $L_p(\omega_\mu)$, $0 < p < \infty$. Suppose $s > 0$ and let $1/\tau := s/d + 1/p$. Denote briefly $B^s_\tau := B^{ss}_{\tau\tau}$. Our main result asserts that if $f \in B^s_\tau$, then

$$\sigma_n(f)_p \le c n^{-s} \|f\|_{B^s_\tau} \quad \text{[Jackson estimate]}.$$

The results presented here are contained in [5, 11]. They are part of a broader undertaking for needlet characterization of Triebel-Lizorkin and Besov spaces on

nonclassical domains, including the multidimensional unit sphere [7, 8] and the interval [4, 10] with Jacobi weights. Our results generalize the results in the univariate case from [4] (with $\alpha = \beta$), where needlet characterizations of F- and B-spaces on the interval are obtained.

References

[1] M. Frazier, B. Jawerth, Decomposition of Besov Spaces, Indiana Univ. Math. J. 34 (1985), 777–799.

[2] M. Frazier, B. Jawerth, A discrete transform and decompositions of distribution, J. of Funct. Anal. 93 (1990), 34–170.

[3] M. Frazier, B. Jawerth, and G. Weiss, *Littlewood-Paley Theory and the Study of Function Spaces*, CBMS 79 (1991), AMS.

[4] G. Kyriazis, P. Petrushev, and Yuan Xu, Jacobi decomposition of weighted Triebel-Lizorkin and Besov spaces, math.CA/0610624.

[5] G. Kyriazis, P. Petrushev, and Yuan Xu, Decomposition of weighted Triebel-Lizorkin and Besov spaces on the ball, math.CA/0703403.

[6] Y. Meyers, *Ondelletes et OpérateursI: Ondelletes*, Hermann, Paris, 1990.

[7] F. Narcowich, P. Petrushev, and J. Ward, Localized tight frames on spheres, SIAM J. Math. Anal. 38 (2006), 347-692.

[8] F. Narcowich, P. Petrushev, and J. Ward, Decomposition of Besov and Triebel-Lizorkin spaces on the sphere, J. Funct. Anal. 238 (2006), 530-564.

[9] J. Peetre, *New thoughts on Besov spaces*, Duke Univ. Math. Series. Durham, N.C., 1993.

[10] P. Petrushev, Y. Xu Localized polynomial frames on the interval with Jacobi weights, J. Four. Anal. Appl. 11 (2005), 557–575.

[11] P. Petrushev, Y. Xu, Localized polynomial frames on the ball, Consrt. Approx, to appear.

[12] H. Triebel, *Theory of function spaces*, Monographs in Math., Vol. 78, Birkhäuser, Verlag, Basel, 1983.

[13] Yuan Xu, Summability of Fourier orthogonal series for Jacobi weight on a ball in $\mathbb{R}^d$, Trans. Amer. Math. Soc. 351 (1999), 2439-2458.

Needlets on the Sphere and Applications to the Cosmological Microwave Background

Dominique Picard

(joint work with P. Baldi, G. Kerkyacharian and D. Marinucci)

The Cosmological Microwave Background (CMB) is a relic radiation which provides a picture of the Universe at the time of the so-called recombination, estimated to have occurred approximately 3×10^5 years after the Big Bang, that is, approximately 1.3×10^{10} years ago (or in other words, when the Universe was less than 0.005% of its current age, much before the formation of the first stars or galaxies). Analysing the CMB will provide a lot of informations about the structure of the Universe. However, the CMB as many other cosmological maps are usually provided with large parts of missing observations (Point sources, stars, galaxies, the milky way, for instance, mask the CMB.)

Statistical Issues

(1) The observation of the CMB is generally modelised as a random field $(T_\xi)_{\xi\in S^2}$ where S^2 is the unit sphere of $\mathbb{R}^3$.
(2) Two important statistical issues are :
   (a) Can $\xi \mapsto T_\xi$ be considered as a Gaussian field ?
   (b) Is $\xi \mapsto T_\xi$ is a rotation- invariant field ?
   i.e. $\forall \rho \in O(3),\ \forall \xi, \eta \in S^2\ E(T_\xi, T_\eta) = E(T_{\rho(\xi)}, T_{\rho(\eta)}), \quad ET_\xi = c \quad ?$
(3) Unfortunately, from the statistical point of view, we have only *ONE* observation of the field (T_ξ will not change in one year or in one hundred years) and only on some points of the sphere (outside the masks).

Decomposition of the Field

Under the hypothesis of isotropy of the field, the spaces H_k of spherical harmonics of order k are eigenspaces of the covariance operator

$$f(x) \in L_2(S^2) \mapsto Kf(x) = \int_{S^2} K(\langle x, y\rangle) f(y) dy$$

where $K(x, y) := E(T_x, T_y)$ and

$$\forall f \in H_k,\ Kf = C_k f.$$

The sequence C_k is called the angular power spectrum of the field T_ξ.

So, if $P_i^k,\ 1 \le i \le 2k+1$ is an orthonormal basis of spherical harmonics of H_k, we get a decomposition of the field into uncorrelated atoms :

$$T_\xi = \sum_{k\in\mathbb{N}} \sum_{1\le i\le 2k+1} (\int_{S^2} T_u P_i^k(u) d\sigma(u)) P_i^k(\xi) = \sum_{k\in\mathbb{N}} \sum_{1\le i\le 2k+1} Z_i^k P_i^k(\xi)$$

and the variables $Z_i^k = \int_{S^2} T_u P_i^k(u) d\sigma(u)$ are uncorrelated and such that

$$Var(Z_i^k) = E(\int_{S^2} T_u P_i^k(u) d\sigma(u))^2 = C_k.$$

The Z_i^k's provide ideal atoms to build tests for isotropy or gaussiannity.

Unfortunately this is not really feasible since the evaluation of the Z_i^k's is corrupted by the mask.

In this paper we investigate the statistical properties of the so-called needlets (ψ_{jk}). These are a family of spherical wavelets which were introduced by Narcowich, Petrushev and Ward [6]. Needlets enjoy several properties which are not shared by other spherical wavelets. First they enjoy good localization properties in frequency: needlets are compactly supported in the frequency domain with a bounded support which depends explicitly on a user-chosen parameter. On the other hand, needlets enjoy excellent localization properties in real space, with an exponential decay of the tails.

As a major consequence of the localization property both in the frequency and in the space domain, the needlet coefficients are asymptotically uncorrelated as the frequency tends to ∞ for any fixed angular distance.

If we define the needlet (random) coefficients associated to the CMB field :

$$\beta_{j,k} = \int_{S^2} T_u \psi_{j,k}(u) d\sigma(u).$$

Under the assumptions of isotropy and gaussianity, the $\beta_{j,k}$'s are gaussian, centered and above all asymptotically uncorrelated. These coefficients are not corrupted provided that they are evaluated outside the mask, and yield much more robust test statistics for isotropy and gaussianity.

References

[1] P. Baldi, G. Kerkyacharian D. Marinucci , D. Picard *Asymptotics for Spherical Needlets* arXiv:math/0606599

[2] P. Baldi, G. Kerkyacharian D. Marinucci , D. Picard *High frequency asymptotics for wavelet-based tests for gaussianity and isotropy on the torus* arXiv:math/0606154

[3] P. Baldi, G. Kerkyacharian D. Marinucci , D. Picard *Subsampling Needlet coefficients on the Sphere* arXiv:math/0706.4169

[4] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, N. Vittorio *Spherical Needlets for CMB Data Analysis* arXiv:math/0707.0844

[5] G.Kerkyacharian, P. Petrushev, D. Picard and T. Willer, *Needlets algorithms for estimation in inverse problems* Electronic Journal of Statistics, (2007) Vol. 1, 30–76

[6] Narcowich, F. J. and Petrushev, P. and Ward, J. D. *Localized tight frames on spheres* SIAM J. Math. Anal. **38** (2006), 574–594

L_1 Approximation of Hamilton-Jacobi Equations

Bojan Popov

(joint work with Jean-Luc Guermond)

L_1-based minimization methods for stationary Hamilton-Jacobi equations

$$H(x, u, Du) = 0, \;\; x \in \Omega, \quad \text{with} \quad u|_{\partial\Omega} = 0$$

are developed. The first case considered is of a one dimensional linear equation. An algorithm is developed to solve the ill-posed boundary value problem and an optimal error estimate is proved. In the nonlinear case, the convergence theory and numerical algorithms are extended to Hamilton-Jacobi equations with convex and uniformly continuous hamiltonians. The main result is proven in the case of a 2D bounded domain with a Lipschitz boundary. The general assumption is that the viscosity solution u of the problem is unique, $u \in W^{1,\infty}(\Omega)$, and the gradient Du is of bounded variation. We approximate the solution to this problem using continuous finite elements and by minimizing the residual in L_1. In the case of a convex and uniformly continuous hamiltonian, it is shown that upon introducing an appropriate entropy the sequence of approximate solutions based on quasi-uniform shape regular finite element triangulations converges to the unique viscosity solution u. The main features of the methods are that they are of arbitrary polynomial order and do not have any artificial viscosity. The fact that residual in minimized in L_1 is a key.

Numerical examples, computational complexity and possible applications to other hyperbolic equations or systems, including time dependent problems, will be discussed.
All references are available at www.math.tamu.edu/∼popov/preprints.html

References

[1] J.-L. Guermond and B. Popov, *Linear advection with ill-posed boundary conditions via L1-minimization*, Numerical Analysis and Modeling **4** (2007), 39–47.
[2] J.-L. Guermond and B. Popov, *L^1-minimization methods for Hamilton-Jacobi equations: the one-dimensional case*, submitted.
[3] J.-L. Guermond and B. Popov, *L^1-approximation of stationary Hamilton-Jacobi equations*, submitted.
[4] J.-L. Guermond, F. Marpeau, and B. Popov, *A fast algorithm for solving first-order PDEs by L^1-minimization*, submitted.

Identification of Sparse Operators

Holger Rauhut
(joint work with Götz Pfander and Jared Tanner)

Motivated by the channel estimation problem in communication engineering (wireless communication and sonar) we consider the problem of identifying a matrix $\Gamma \in \mathbb{C}^{n\times m}$ from its action Γh on a single vector $h \in \mathbb{C}^m$. Clearly, without further knowledge Γ is completely determined only by its action on n basis vectors in $\mathbb{C}^m$, and our task seems impossible. However, physical considerations suggest that in certain practical situations (see also below) Γ can be well-represented by a short linear combination of a few basic matrices; in other words it has a sparse representation. In this situation one can exploit connections to sparse approximation and compressed sensing [5, 7] to efficiently reconstruct Γ.

Given a suitable set $\mathbf{\Psi}$ of N "elementary" matrices $\Psi_j \in \mathbb{C}^{n\times m}$, $j = 1, \dots, N$ (a matrix dictionary) we say that $\Gamma \in \mathbb{C}^{n\times m}$ has a k-sparse representation if

$$\Gamma = \sum_j x_j \Psi_j$$

for a vector $x \in \mathbb{C}^N$ whose support has at most cardinality k, formally $\|x\|_0 := |\{k, x_k \neq 0\}| \leq k$. The action of such a matrix Γ on a vector $h \in \mathbb{C}^m$ can be written as

$$\Gamma h = \sum_j x_j \Psi_j h \;=\; \Psi_h x$$

with the matrix $\mathbf{\Psi}_h = (\Psi_1 h | \dots | \Psi_N h) \in \mathbb{C}^{n\times N}$. Identification of Γ clearly amounts to reconstructing the sparse vector x from Γh. Unfortunately, the obvious approach of determining the vector x with shortest support (i.e. minimal $\|x\|_0$) that is consistent with the observation, $\Gamma h = \mathbf{\Psi}_h x$, yields an NP-hard combinatorial problem[2] and, thus, is not feasible in practice.

Several tractable alternative recovery algorithms have been proposed so far, most notably ℓ_1-minimization (Basis Pursuit), on which we will concentrate here.

Instead of solving a combinatorial optimization problem we consider the minimizer of the problem

$$\min_x \|x\|_1 = \sum_{j=1}^{N} |x_j| \quad \text{subject to} \quad \Gamma h = \mathbf{\Psi}_h x. \tag{1}$$

This minimization problem can be solved efficiently with convex optimization techniques [4].

Obvious questions concern the choice of h, and the maximal sparsity k that allows for recovery of x resp. identification of Γ by ℓ_1-minimization.

Our first result in this direction [11] deals with dictionaries of random matrices. Although in practice rather deterministic dictionaries will appear, it nevertheless provides some intuition of what can be expected, in particular, the maximal recoverable sparsity.

Theorem 1. Let h be a non-zero vector in $\mathbb{R}^m$.

(a) Let all entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$, $j = 1, \ldots, N$, be chosen independently according to a standard normal distribution (Gaussian ensemble); or

(b) let all entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$, $j = 1, \ldots, N$, be independent Bernoulli ± 1 variables (Bernoulli ensemble).

Then there exists a positive constant c such that

$$k \leq c \, \frac{n}{\log\left(\frac{N}{n\varepsilon}\right)} \tag{2}$$

implies that with probability at least $1 - \varepsilon$ all matrices Γ having a k-sparse representation with respect to $\mathbf{\Psi} = \{\Psi_j\}$ can be recovered from Γh by Basis Pursuit (1).

The proof of this theorem is based on estimating the so called restricted isometry constants [6] of the random matrix $\mathbf{\Psi}_h$.

We will now concentrate on the matrix dictionary of time–frequency shifts, which appears naturally in the channel identification problem in wireless communications [3] or sonar [13]. Due to physical considerations wireless channels may indeed be modeled by sparse linear combinations of time–frequency shifts $M_\ell T_p$, where the translation operators T_p and modulation operator M_ℓ on $\mathbb{C}^n$ are given by

$$(T_p h)_q = h_{p+q \mod n} \quad \text{and} \quad (M_\ell h)_q = e^{2\pi i \ell q/n} h_q.$$

The system of time–frequency shifts $\mathbf{G} = \{M_\ell T_p : \ell, p = 0, \ldots, n-1\}$ forms a basis of $\mathbb{C}^{n\times n}$ and for any non-zero h, the vector dictionary $\mathbf{G}_h = (M_\ell T_p h)_{\ell,p=0,\ldots,n-1}$ is a Gabor system [9]. Below, we focus on the so-called Alltop window h^A [1, 10] with entries

$$h_q^A := \frac{1}{\sqrt{n}} e^{2\pi i q^3/n}, \quad q = 0, \ldots, n-1, \tag{3}$$

and the randomly generated window h^R with entries

$$h_q^R := \frac{1}{\sqrt{n}}\epsilon_q, \quad q = 0, \dots, n-1, \tag{4}$$

where the ϵ_q are independent and uniformly distributed on the torus $\{z \in \mathbb{C}, |z| = 1\}$.

Invoking existing recovery results [8, 14] and results on the coherence of the Gabor systems $\mathbf{G}_{h^A}$ [10] and $\mathbf{G}_{h^R}$ [11], we obtain

Theorem 2. (a) Let $n \geq 5$ be prime and h^A be the Alltop window defined in (3). If $k < \frac{\sqrt{n}+1}{2}$ then BP recovers from Γh^A all matrices Γ having a k-sparse representation with respect to the time–frequency shift dictionary.

(b) Let n be even and choose h^R to be the random unimodular window in (4). Let $t > 0$ and suppose

$$k \leq \frac{1}{4}\sqrt{\frac{n}{C\log(n) + t}} + \frac{1}{2}$$

with $C = 2\log(4) \approx 2.77$. Then with probability of at least $1 - e^{-t}$ BP recovers from Γh^R all matrices $\Gamma \in \mathbb{C}^{n\times n}$ having a k-sparse representation.

Although this theorem provides a first recovery result, it is not yet satisfactory as the maximal sparsity which guarantees recovery is quite small – on the order of $\sqrt{n}$ – compared to (2), where it is of the order $n/\log(N/n)$ which in our case, $N = n^2$, is $n/\log(n)$. By passing from worst case analysis to a probability model on the sparse coefficient vector x one can apply recent work by Tropp based on the coherence [15] in order to achieve an improvement. Indeed, if the support set Λ is chosen at random as well as the signs of the non-zero coefficients x_j, $j \in \Lambda$, then for both h^A and h^R one has recovery with high probability of the true coefficient vector x provided

$$k \leq c\frac{n}{\log(n)^{1+u}}$$

for some $c, u > 0$ (governing the probability of recovery). We refer to [11] for a precise formulation.

In case of the randomly generated vector h^R we were able to improve further on the above recovery results by removing the randomness in the coefficient vector x [12].

Theorem 3. Let $\Gamma \in \mathbb{C}^{n\times n}$ be k-sparse with respect to the time-frequency shift dictionary $\mathbf{G}$. Choose h^R at random. There exists a constant $C > 0$ such that

$$k \leq C\frac{n}{\log(n/\epsilon)} \tag{5}$$

implies that with probability at least $1 - \epsilon$ Basis Pursuit (1) recovers Γ from Γh^R.

The above theorem is based on a careful analysis of the singular values of a submatrix consisting of k columns of $\mathbf{G}_{h^R}$ [12]. It would be interesting to investigate an analog for the deterministic Alltop window h^A.

Numerical experiments for both h^A and h^R in [11] suggest that recovery is possible with high probability for most signals provided $k \leq \frac{n}{2\log(n)}$.

References

[1] W. O. Alltop. Complex sequences with low periodic correlations. *IEEE Trans. Inf. Theory*, 26(3):350–354, 1980.

[2] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13(1):57–98, 1997.

[3] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. Commm.*, 11:360–393, 1963.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.

[5] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[6] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[7] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[8] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, 2006.

[9] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Appl. Numer. Harmon. Anal. Birkhäuser Boston, Boston, MA, 2001.

[10] T. Strohmer and R. W. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.*, 14(3):257–275, 2003.

[11] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. Preprint, 2007.

[12] G. Pfander, H. Rauhut. Sparsity in time-frequency representations. in preparation, 2007.

[13] M. Stojanovic. *Underwater Acoustic Communications*, volume 22 pages 688-689. John Wiley & Sonce, 1999.

[14] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.

[15] J. A. Tropp. Random subdictionairies of general dictionairies. *preprint*, 2006.

Adaptive Coupled Cluster Method and CI Method for the Solution of the Electronic Schroedinger Equation

Reinhold Schneider

The electronic Schrödinger equation plays a fundamental role in molecular physics. It describes the stationary non-relativistic quantum mechanical behavior of an N electron system in the electric field generated by the nuclei. The *Coupled Cluster Method* has been developed for the numerical computation of the ground state energy and wave function. It provides a powerful tool for high accuracy electronic structure calculations. The present paper aims to provide a convergence analysis of this method. Under additional assumptions quasi-optimal convergence of the projected coupled cluster solution to the full CI solution and also to the exact wave function can be shown in the Sobolev H^1 norm. The error of the ground state energy computation is obtained by an Aubin Nitsche type approach.

Recent developments in science and technology, in particular in chemistry, molecular biology, material sciences or semi-conductor devices, are requesting reliable computations of molecular behavior. On an atomic or molecular length scale, the physics is governed by the laws of quantum mechanics. Consequently, numerical modeling of such processes should be based on *first principles* of quantum mechanics. The basic equation for a quantitative description of atomistic and molecular phenomena of interest is the *electronic Schrödinger equation*. It describes the stationary and non-relativistic behavior of an ensemble of N electrons in an electric field resulting from fixed nucleons,

$$\mathcal{H}\Psi = \sum_{i=1}^{N}[-\frac{1}{2}\Delta_i - \sum_{j=1}^{M}\frac{Z_j}{|\mathbf{x}_i - \mathbf{R}_j|} + \frac{1}{2}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{x}_i - \mathbf{x}_j|}]\Psi = E\Psi \ .$$

The ground state energy of a given N-electron system, which is the eigenvalue of the lowest eigenstate of the Schrödinger-Hamilton operator is of major interest,

$$E_0 = \min_{\langle\Psi,\Psi\rangle=1}\langle\mathcal{H}\Psi,\Psi\rangle \ .$$

Since the corresponding wave function depends on $3N$ spatial variables and N two-valued spin variables

$$\Psi(\mathbf{x}_1, s_1; \ldots; \mathbf{x}_N, s_N) \quad , \quad \Psi : \mathbb{R}^{3N} \otimes S^N \to \mathbb{C} \text{ resp. } \mathbb{R} \quad , \quad \langle\Psi,\Psi\rangle = 1 \ ,$$

not to mention its lack of smoothness, its numerical approximation is a rather difficult and challenging task. Usually it is approximated by sums anti-symmetric tensor products $\Psi = \sum_{k=1}^{\infty} c_k\Psi_k$ with Slater determinants Ψ_k defined by

$$\Psi_k(\mathbf{x}_1, s_1, \ldots, \mathbf{x}_N, s_N) = \varphi_{1,k} \wedge \ldots \wedge \varphi_{N,k} = \frac{1}{\sqrt{N!}}\det(\varphi_{i,k}(\mathbf{x}_j, s_j))_{i,j=1}^{N} \quad ,$$

$\langle\varphi_{i,k}, \varphi_{j,k}\rangle = \delta_{i,j}$, with N orthonormal functions φ_i, $i = 1, \ldots, N$ called *spin orbitals* $\varphi_i : \mathbb{R}^3 \times \{\pm\frac{1}{2}\} \to \mathbb{C}(\mathbb{R})$, $i = 1, \ldots, N$. In contrast to the CI (Configuration Interaction) method, which consists in a linear parametrization and a corresponding Galerkin method, the *Coupled Cluster* (CC)-method is based on an exponential parametrization of the wave function,

$$\Psi = e^T\Psi_0 \quad , T = \sum_{i=1}^{N} T_i = \sum_{\mu\in\mathcal{J}} t_\mu X_\mu$$

for given reference determinant Ψ_0 and excitation operators X_μ and unknown amplitudes t_μ satisfying the *amplitude equations*

$$0 = \langle\Psi_\mu, e^{-T}H\Psi\rangle = \langle\Psi_\mu, e^{-T}He^T\Psi_0\rangle =: \mathbf{f}_\mu(\mathbf{t}) \ , \quad \mathbf{t} = (t_\nu)_{\nu\in\mathcal{J}} \ , \ \mu,\nu \in \mathcal{J}.$$

Restricting the ansatz and the above amplitude equations to a relevant subset of excitation $\mathcal{J}_h \subset \mathcal{J}$, usually consisting of single and double excitations, yields to the *projected CC method*, e.g CCSD. In contrast to truncated CI methods, like CISD, the projected is *size consistent*, due to exponential parametrization. Nowadays the projected CC approach constitutes the method of choice for high resolution wave function computation in electronic structure calculation, at least in cases where dynamical correlation is dominating. However, it performs poorly

in some cases, where a single reference Slater determinant is not sufficient. Due to the authors knowledge, the authors paper [4] was the first attempt to analyse the *(Projected Coupled) Cluster method* rigorously from a numerical analysis point of view. For the analysis of the CC method, as well as in practice, presented in [4], the (approximate) Hartree Fock wave function is considered as a reference Slater determinant. Based on the established convergence of the full CI solution towards the exact wave function by the full CI solution, in a second step, the convergence of the projected CC solution to the full CI solution is considered in detail. In particular, a weighted l_2-norm for the *coupled cluster amplitudes* $\|\mathbf{t}\|_V$ has been introduced, which are shown to be equivalent to the H^1 Sobolev norm of the approximate wave functions in a certain extent. With this notion at hand, the projected CC method can be interpreted as a Galerkin method for a non-linear function $\mathbf{f} : V \to V'$ seeking $\mathbf{t}_h \in V_h$ s.t.

$$\langle \mathbf{f}(\mathbf{t}_h), \mathbf{v}_h \rangle = 0 \ \forall \ \mathbf{v}_h \in V_h \subset V.$$

Under an additional assumption of strict monotonicity of the amplitude function, quasi-optimal convergence of the projected CC method is concluded.

Theorem ([4]) (a priori estimate) If $\mathbf{f}(\mathbf{t}) = \mathbf{0}$ and $\mathbf{f}$ is (locally) strictly monotone at $\mathbf{t}$ and the solution of the Galerkin scheme $\mathbf{t}_h$ satisfies $\|\mathbf{t} - \mathbf{t}_h \leq \delta$ then it satisfies

$$\|\mathbf{t} - \mathbf{t}_h\|_V \lesssim \inf_{\mathbf{v} \in \mathbb{R}^{\sharp J}} \|\mathbf{t} - \mathbf{v}_h\|_V.$$

and

$$\|\Psi - \Psi_h\|_{H^1} \lesssim \inf_{\mathbf{v} \in \mathbb{R}^L} \|\Psi - e^{\sum_{\mu \in \mathcal{J}_h} v_\mu X_\mu} \Psi_0\|_{H^1}.$$

Since convergence problems are reported for the CC method in presence of statical correlation, convergence might depend on the actual molecular configuration. Therefore setting such an assumption seems to be reasonable. Whether this might be necessary or not is not clear yet.

Since the ground state energy is a functional of the amplitudes

$$J(\mathbf{t}) := E(\mathbf{t}) := \langle \Psi_0, H(1 + T_2 + \frac{1}{2} T_1^2) \Psi_0 \rangle \ ,$$

in order to estimate the convergence of the computed eigenvalue to the exact ground state energy, the *dual weighted residual* approach, introduced by Rannacher et al., has been applied for the derivation of sharp a priori estimates. Defining the *Lagrange functional*

$$L(\mathbf{t}, \mathbf{a}) := J(\mathbf{t}) - \langle \mathbf{f}(\mathbf{t}), \mathbf{a} \rangle \ , \ \mathbf{t} \in V' \ , \ \mathbf{a} \in V \ ,$$

its stationary points are given by the solution of the *amplitude equation* $\mathbf{f}(\mathbf{t}) = \mathbf{0}$ and the *dual equation* for $\mathbf{a} \in V$

$$\mathbf{f}'[\mathbf{t}]^\top \mathbf{a} = -(J'[\mathbf{t}]) \in V'.$$

The error of the energy functional obtained by the Galerkin solutions $\mathbf{t}_h, \mathbf{a}_h$ can be expressed by the corresponding residuals

$$|J(\mathbf{t}) - J(\mathbf{t_h})| = \frac{1}{2}(\langle \mathbf{r}(\mathbf{t}_h), \mathbf{a} - \mathbf{b}_h) \rangle + \frac{1}{2} \langle \mathbf{r}^*(\mathbf{t}_h, \mathbf{a}_h), (\mathbf{t} - \mathbf{u}_h) \rangle + \mathcal{R}^3 \ , \ \forall \mathbf{u}_h, \mathbf{b}_h \in V_h \ ,$$

with a cubic remainder term $\mathcal{R}^3$, which implies the following result.

Theorem ([4]) The error in the energy $E = J(\mathbf{t})$ and the discrete energy $E_h = J(\mathbf{t}_h)$ can be estimated by

$$\begin{aligned} |E - E_h| &\lesssim \|\mathbf{t} - \mathbf{t}_h\|_V \|\mathbf{a} - \mathbf{a}_h\|_V + (\|\mathbf{t} - \mathbf{t}_h\|_V)^2 \\ &\lesssim \inf_{\mathbf{u}_h \in V_h} \|\mathbf{t} - \mathbf{u}_h\|_V \inf_{\mathbf{b}_h \in V} \|\mathbf{a} - \mathbf{b}_h\|_V + \\ &\quad + (\inf_{\mathbf{u}_h \in V_h} \|\mathbf{t} - \mathbf{u}_h\|_V)^2. \end{aligned}$$

Originally, the *dual weighted residual method* has been developed for the design of local a posteriori error estimators for adaptive finite element methods, which are used to optimize the finite element meshes for computation of certain functionals of the solution. However adaptive finite element method (FEM) techniques cannot be transferred immediately to the solution of the electronic Schrödinger equation. Perhaps, instead of truncating with respect to the excitation level one needs nonlinear approximation procedures for optimizing the basis sets, and even more significant, for an automatic selection of those amplitudes which contributes most to the ground energy. An alternative is provided by nonlinear best n-term approximation strategies introduced by Cohen, Dahmen, DeVore.

The amplitude equations

$$0 = \mathbf{f}_\mu(\mathbf{t}) = \langle \Psi_\mu, e^{-T} H e^T \Psi_0 \rangle = \langle \Psi_\mu, [\mathcal{F}, T] \Psi_0 \rangle + \langle \Psi_\mu, e^{-T} U e^T \Psi_0 \rangle$$

are of the form $\mathbf{f}(\mathbf{t}) = \mathbf{F}\mathbf{t} - \Phi(\mathbf{t}) = \mathbf{0}$ which can be solved iteratively by an augmented Newton type scheme $\mathbf{t}^{n+1} := \mathbf{F}^{-1}\Phi(\mathbf{t}^n)$, which in turn has to be performed only approximatively according to best n-term strategies.

The same strategy applies also to the CI method still in an almost size extensive way. In two recent papers on adaptive best n-term algorithm for the computation of the lowest eigenvalue and eigenstate has been developed [1, 2]. Regularity providing best n-term convergence rates has been obtained by Yserentant and also in [3].

References

[1] W. Dahmen, T. Rohwedder, R. Schneider, A. Zeiser. *Adaptive eigenvalue computation - complexity estimates*, Preprint IGPM RWTH Aachen 2007

[2] T. Rohwedder, R. Schneider, A. Zeiser. *Perturbed preconditioned inverse iteration for operator eigenvalue problems with applications to adaptive wavelet discretization* submitted to Adv. Comp. Math

[3] H.-J. Flad, W. Hackbusch, R. Schneider, *Best N term approximation for electronic wavefunctions I One electron reduced density matrix*, *M2AN Math. Model. Numer. Anal.* **40**, 2006

Best N term approximation for electronic wavefunctions II, Jastrow factors, *M2AN Math. Model. Numer. Anal.*, 2007

[4] R. Schneider. *Analysis of the projected Coupled Cluster Method in Electronic Structure Calculation*, submitted to *Numer. Math.*, CAU Preprint Kiel 06-6 2006

Space Time Adaptive Wavelet Methods for Parabolic Problems

Christoph Schwab
(joint work with Rob Stevenson)

In a bounded polyhedral domain $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary, and in a finite time interval $(0,T)$, let

$$V \hookrightarrow H \cong H^* \hookrightarrow V^* \tag{1}$$

be an evolution triple with dense injection. Assume given a one parameter family of (pseudo) differential operators $A(t) \in \mathcal{L}(V,V^*)$, of constant order $2\rho \in (0,2]$. The family $\{A(t)\}_{0\le t\le T}$ is assumed uniformly elliptic in the sense that the associated bilinear form

$$a(t;\eta,\zeta) := {}_{V^*}\big\langle A(t)\eta,\zeta\big\rangle_V : V\times V \to \mathbb{R}$$

satisfies for some $M_a, \alpha > 0$, $\lambda \in \mathbb{R}$ and for almost every $t \in (0,T)$, $\eta,\zeta \in V$:

$$|a(t;\eta,\zeta)| \le M_a\|\eta\|_V\|\zeta\|_V, \quad a(t;\eta,\eta) \ge \alpha\|\eta\|_V^2 - \lambda\|\eta\|_H^2\,. \tag{2}$$

Typically, $V = \widetilde{H}^\rho(\Omega)$, $H = L^2(\Omega)$, $V^* = H^{-\rho}(\Omega)$, $0 < \rho \le 1$. For given $g \in L^2(0,T;V^*)$, $h \in H$, we consider the **parabolic problem**:

$$\dot u(t) + A(t)\,u(t) = g(t) \text{ in } V^*, \quad u(0) = h \text{ in } H\,. \tag{P}$$

In **weak form**: given $(g,h) \in \mathcal{Y}^*$, find

$$u \in \mathcal{X}: \qquad b(u,v) = f(v) \qquad \forall v = (v_1,v_2) \in \mathcal{Y} \tag{W}$$

where

$$\mathcal{X} = L^2(0,T;V)\cap H^1(0,T;V^*) \subset C^0([0,T];H), \quad \mathcal{Y} = L^2(0,T;V)\times H\,, \tag{3}$$

$$b\big(u,(v_1,v_2)\big) := \int_0^T \big\{{}_{V^*}\big\langle \dot u, v_1\big\rangle_V + a(t;u,v_1)\big\}\,dt + {}_H\big\langle u(0),v_2\big\rangle_H\,, \tag{4}$$

$$f(v) := \int_0^T {}_{V^*}\big\langle g(t), v_1(t)\big\rangle_V\,dt + {}_H\big\langle h, v_2\big\rangle_H\,. \tag{5}$$

Examples:

(1) Diffusion problem in $\Omega = (0,1)^d$ with possibly large d with variable, but uniformly in Ω elliptic coefficients. Here $V = H_0^1(\Omega)$. See [3].
(2) $A(t)$ infinitesimal generator of anisotropic Feller-Lévy process X_t in $\Omega = (0,1)^d$: $V = \widetilde{H}^{s_1,\dots,s_d}(\Omega)$, $H = L^2(\Omega)$, $(A\eta)(\zeta) := \mathcal{E}(\eta,\zeta)$, $0 < s_i \le 1$, $i = 1,\dots,d$ where $\mathcal{E}(\cdot,\cdot)$ denotes the (possibly nonsymmetric) Dirichlet Form of the process $\{X_t\}_{t\ge 0}$, see [4].
(3) Stokes Problem in $\Omega\times(0,T)$ in velocity-formulation with

$$V = \big\{u \in H_0^1(\Omega)^d : \operatorname{div} u = 0 \text{ in } L^2(\Omega)\big\}\,, \qquad H = L^2(\Omega)^d.$$

Regarding the well-posedness of (W), we have

Proposition 1. There exist constants $\beta > 0$, $M_b < \infty$ s.t. the form $b(\cdot,\cdot)$ is continuous

$$\forall u \in \mathcal{X},\ v \in \mathcal{Y}: \quad |b\big(u,(v_1,v_2)\big)| \le M_b\,\|u\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}, \tag{6}$$

coercive

$$\inf_{0\neq u\in\mathcal{X}}\ \sup_{0\neq v\in\mathcal{Y}} \quad \frac{b\big(u,(v_1,v_2)\big)}{\|u\|_{\mathcal{X}}\ \|v\|_{\mathcal{Y}}} \ge \beta, \tag{7}$$

and injective

$$\forall 0 \neq v \in \mathcal{Y}: \quad \sup_{0\neq u\in\mathcal{X}} \frac{b\big(u,(v_1,v_2)\big)}{\|u\|_{\mathcal{X}}\ \|v\|_{\mathcal{Y}}} > 0\,. \tag{8}$$

In particular,

$$Bu := \big(\dot u + A(t)u,\ u(0)\big) : \mathcal{X} \to \mathcal{Y}^* \text{ is an isomorphism}. \tag{9}$$

To discretize the operator equation (P), we require therefore Riesz bases of the spaces $\mathcal{X}$ and $\mathcal{Y}$. These can be obtained by tensorizing wavelet bases *which are, properly scaled, Riesz bases in the evolution triple (1) and in* $H^1(0,T)$, $L^2(0,T)$ *and* $H^1(0,T)^*$, *respectively.*

Proposition 2. Assume given $\Sigma = \{\sigma_\mu : \mu \in \nabla_x\} \subset V$ and $\Theta = \{\theta_\lambda : \lambda \in \nabla_t\} \subset H^1(0,T)$ which are, properly scaled, Riesz bases of V, H, V^* and of $H^1(0,T)$, $L^2(0,T)$, $\widetilde{H}^{-1}(0,T) = (H^1(0,T))^*$, respectively. Then

$$\left\{(t,x) \longmapsto \frac{\theta_\lambda(t)\,\sigma_\mu(x)}{(\|\theta_\lambda\|^2_{L^2(0,T)}\|\sigma_\mu\|^2_V + \|\theta_\lambda\|^2_{H^1(0,T)}\|\sigma_\mu\|^2_{V^*})^{\frac12}} : (\lambda,\mu) \in \nabla_t \times \nabla_x\right\}$$

is a Riesz Basis in $\mathcal{X}$ and

$$(t,x) \longmapsto \Big\{\Big(\frac{\theta_\lambda(t)\,\sigma_\mu(x)}{\|\theta_\lambda\|_{L^2(0,T)}\|\sigma_\mu\|_V}, 0\Big) : (\lambda,\mu) \in \nabla_t \times \nabla_x\Big\} \times \Big\{\Big(0, \frac{\sigma_\mu(x)}{\|\sigma_\mu\|_H}\Big) : \mu \in \nabla_x\Big\}$$

is a Riesz Basis in $\mathcal{Y}$.

Proposition 3. In the tensor basis $\Theta \otimes \Sigma := \{\theta_\lambda\,\sigma_\mu : (\lambda,\mu) \in \nabla_t \times \nabla_x\}$ the bi-infinite matrix

$$\boldsymbol{B} := b\big(\Theta\otimes\Sigma,\ (\Theta\otimes\Sigma)\times\Sigma\big) = \begin{bmatrix} \|\Theta\otimes\Sigma\|^{-1}_{L^2\otimes V} & 0 \\ 0 & \|\Sigma\|^{-1}_H \end{bmatrix} \times$$

$$\times \begin{bmatrix} {}_{L^2}\big\langle \dot\Theta, \Theta\big\rangle_{L^2} \otimes {}_H\big\langle \Sigma,\Sigma\big\rangle_H + \displaystyle\int_0^\top a(t;\Theta\otimes\Sigma,\Theta\otimes\Sigma)dt \\ {}_H\big\langle \Theta(0)\otimes\Sigma,\Sigma\big\rangle_H \end{bmatrix} \|\Theta\otimes\Sigma\|^{-1}_{\mathcal{X}} \tag{10}$$

is a ***bounded isomorphism*** *in* $\mathcal{L}\big(\ell_2(\nabla_t\times\nabla_x),\ \ell_2(\nabla_t\times\nabla_x)\times\ell_2(\nabla_x)\big)$; *here,* $\|\Sigma\|_H$ *denotes the mass matrix of the basis* $[\Sigma]_H$ *in* H *and analogously for the other expressions in (10).*

Now assume that Θ, Σ are spline-wavelet bases in $(0,T)$ resp. in Ω with sufficient smoothness and sufficiently many vanishing moments and with isotropic support.

Proposition 4. i) If $A \in \mathcal{L}(V, V^*)$ is independent of t satisfying (2) and is s^*-computable with respect to the basis $[\Sigma]_V$, then the bi-infinite matrix $\boldsymbol{B}$ in (10) is s^*-computable with respect to the bases $[\Theta \otimes \Sigma]_{\mathcal{X}}$, $[\Theta \otimes \Sigma \times \Sigma]_{\mathcal{Y}}$.

ii) If $A(t)$ satisfying (2) and is smooth in $t \in [0,T]$, then there exists a quadrature scheme in (x,t) s.t. the matrix $\boldsymbol{B}$ in (10) is s^*-computable.

If the wavelets σ_μ in the basis $[\Sigma]_H$ are tensor products of univariate spline wavelets of order $p+1$, s^* compressibility of the matrix corresponding to the diffusion operator has been shown in [3].

Then there are variants of both adaptive wavelet algorithms of Cohen, Dahmen & DeVore in [1, 2] which converge with optimal rates in linear complexity to any solution $u \in \mathcal{A}_\infty^{s^*}$ of (W) with convergence rate $N^{-s/d}$ for wavelet bases Σ consisting of spline wavelets in $\mathbb{R}^d$ with isotropic support and with convergence rate N^{-s} (i.e. *without the curse of dimensionality*) for wavelet bases Σ consisting of tensor products of univariate spline wavelets in $(0,1)^d$ (see [3]) for $0 \le s < \min(s^*, p+1-\rho)$ in work and memory, up to some absolute constant equal to the support size N of the solution vector.

References

[1] A. Cohen, W. Dahmen and R. De Vore: Adaptive wavelet methods for elliptic operator equations – Convergence rates, Math. Comp. **70** (2001) 27–75.

[2] A. Cohen, W. Dahmen and R. De Vore: Adaptive Wavelet Methods II - Beyond the Elliptic case, Found. Comput. Math., **2** (2002) 203–245.

[3] Ch. Schwab and R. Stevenson: Adaptive Wavelet Algorithms for Elliptic PDEs on Product Domains, Report 2006-10 Seminar for Applied Mathematics, ETH Zürich (to appear in Mathematics of Computation 2007) http://www.sam.math.ethz.ch/reports/2006/10

[4] E.W. Farkas, N. Reich and Ch. Schwab: Anisotropic stable Lévy copula processes – analytical and numerical aspects, Mathematical Models and Methods in the Applied Sciences, Vol. 17, No. 9 (September 2007) 1405 - 1443.

An Adaptive Wavelet Method for Solving High-Dimensional Elliptic PDE's

Rob Stevenson

(joint work with Christoph Schwab)

In $\Omega = (0,1)^n$, for given $f \in H^{-1}(\Omega)$, we study the numerical solution of the problem: find $u \in H_0^1(\Omega)$ such that

$$a(u,v) := \int_\Omega c_0 uv + \sum_{m=1}^n c_m \partial_m u \partial_m v = f(v) \qquad \forall v \in H_0^1(\Omega) \tag{1}$$

where $c_0 \ge 0$ and $c_m > 0$ $(m = 1, \dots, n)$ are constants.

PDE's such as (1) in spatial domains of high dimension arise in numerous areas. We only mention mathematical finance (valuation of derivative contracts on large baskets), elliptic homogenization problems with multiple separated length scales, deterministic methods for stochastic PDE's, the multi-electron Schrödinger equation in molecular dynamics and problems from molecular biology.

With a standard, piecewise polynomial approximation procedure in $(0,1)^n$, the error in energy norm

$$|\!|\!| \cdot |\!|\!| := a(\cdot,\cdot)^{\frac{1}{2}} \tag{2}$$

on $H^1_0(\Omega)$ behaves at best as $\sim N^{-(d-1)/n}$, where d is the polynomial order and N is the number of degrees of freedom. The rate $(d-1)/n$ being inversely proportional to the space dimension n is known as *curse of dimensionality*.

Using that $\Omega = (0,1)^n$ is a tensor product domain, the curse of dimensionality can be circumvented by applying hyperbolic cross approximation or sparse grids ([BG04]). With this approach, for any fixed n the error behaves as $\sim (\log N)^{n-1} N^{-(d-1)}$, or with some small modification even without the log-factor, *assuming* that $u \in \bigcap_{k=1}^n \bigotimes_{m=1}^n H^{d+\delta_{km}}(0,1)$. The largest order of a (mixed) partial derivative that is involved in this regularity constraint increases linearly with nd.

The key to overcome regularity restrictions is to apply *nonlinear approximation*. Let $\{\psi_\lambda : \lambda \in \nabla\}$ be a Riesz basis for $L_2(0,1)$ consisting of *wavelets* of order d, such that $\{2^{-|\lambda|}\psi_\lambda : \lambda \in \nabla\}$ is a Riesz basis for $H^1_0(0,1)$. As usual, here $|\lambda| \in \mathbb{N}_0$ denotes the *level* of ψ_λ, and we assume that for $\ell \in \mathbb{N}_0$, $\#\{\lambda \in \nabla : |\lambda| = \ell\} = 2^\ell$. Now for any n, the properly scaled *tensor product basis*

$$\boldsymbol{\Psi} := \Big\{ \boldsymbol{\psi}_{\boldsymbol{\lambda}} := \bigotimes_{m=1}^n \psi_{\boldsymbol{\lambda}_m} / |\!|\!| \bigotimes_{m=1}^n \psi_{\boldsymbol{\lambda}_m} |\!|\!| : \boldsymbol{\lambda} \in \boldsymbol{\nabla} := \nabla^n \Big\} \tag{3}$$

is a Riesz basis for $H^1_0(0,1)^n$, even uniformly in $c_0 \geq 0$ and $c_m > 0$ $(m = 1,\dots,n)$ when this space is equipped with the energy norm $|\!|\!| \cdot |\!|\!|$. This means that the quotient

$$\sup_{0 \neq \mathbf{v} \in \ell_2(\boldsymbol{\nabla})} \frac{|\!|\!| \sum_{\boldsymbol{\lambda} \in \boldsymbol{\nabla}} \mathbf{v}_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}} |\!|\!|^2}{\|\mathbf{v}\|^2_{\ell_2(\boldsymbol{\nabla})}} \Big/ \inf_{0 \neq \mathbf{v} \in \ell_2(\boldsymbol{\nabla})} \frac{|\!|\!| \sum_{\boldsymbol{\lambda} \in \boldsymbol{\nabla}} \mathbf{v}_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}} |\!|\!|^2}{\|\mathbf{v}\|^2_{\ell_2(\boldsymbol{\nabla})}} \tag{4}$$

is bounded, uniformly in $c_0 \geq 0$ and $c_m > 0$ $(m = 1,\dots,n)$. With $|\boldsymbol{\lambda}| := (|\boldsymbol{\lambda}_1|,\dots,|\boldsymbol{\lambda}_n|)$, for any $\ell \in \mathbb{N}_0$, $\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \boldsymbol{\nabla}, \||\boldsymbol{\lambda}|\|_1 \leq \ell\}$ spans the corresponding sparse grid space of dimension $N = \sum_{k=0}^\ell 2^k \binom{n-1+k}{n-1}$.

Instead of restricting ourselves to sparse-grid approximation, we consider approximations to u from the span of $\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_N\}$, where $\boldsymbol{\Lambda}_N \subset \boldsymbol{\nabla}$ is *any* subset with $\#\boldsymbol{\Lambda}_N = N$. Because of the boundedness of the quotient from (4), approximating u by $\sum_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_N} \mathbf{v}_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}}$ in $|\!|\!| \cdot |\!|\!|$ is equivalent to approximating its representation $\mathbf{u}$ with respect to $\boldsymbol{\Psi}$ by $(\mathbf{v}_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_N}$ in $\|\cdot\|_{\ell_2(\boldsymbol{\nabla})}$. We have $\inf_{\{\mathbf{v} \in \ell_2(\boldsymbol{\nabla}) : \operatorname{supp} \mathbf{v} \subset \boldsymbol{\Lambda}_N\}} \|\mathbf{u} - \mathbf{v}\|_{\ell_2(\boldsymbol{\nabla})} = \|\mathbf{u} - \mathbf{u}|_{\boldsymbol{\Lambda}_N}\|_{\ell_2(\boldsymbol{\nabla})}$, where $\mathbf{u}|_{\boldsymbol{\Lambda}_N}$ is the vector in $\ell_2(\boldsymbol{\nabla})$ that coincides with $\mathbf{u}$ on its support being $\boldsymbol{\Lambda}_N$. For approximating $\mathbf{u}$ in $\|\cdot\|_{\ell_2(\boldsymbol{\nabla})}$, a best choice for

$\mathbf{\Lambda}_N$ is one such that $\mathbf{u}|_{\mathbf{\Lambda}_N}$ is a *best N-term approximation* to $\mathbf{u}$, denoted as $\mathbf{u}_N$, meaning that it contains the N largest coefficients in modulus of $\mathbf{u}$.

The class

$$\mathcal{A}^s_\infty := \{\mathbf{v} \in \ell_2(\mathbf{\nabla}) : \|\mathbf{v}\|_{\mathcal{A}^s_\infty} := \sup_{\varepsilon>0} \varepsilon[\min\{N \in \mathbb{N}_0 : \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\mathbf{\nabla})} \le \varepsilon\}]^s < \infty\}$$

gathers under one roof all $\mathbf{v}$ whose best N-term approximations converge to $\mathbf{v}$ with rate $s > 0$. Note that $\|\mathbf{v}\|_{\mathcal{A}^s_\infty} \ge \sup_{\varepsilon<\|\mathbf{v}\|_{\ell_2(\mathbf{\nabla})}} \varepsilon = \|\mathbf{v}\|_{\ell_2(\mathbf{\nabla})}$. Although $\mathcal{A}^s_\infty$ is non-empty for any s, as it contains any finitely supported vector, in view of the order d of polynomial reproduction being applied, the representation $\mathbf{v}$ of an arbitrarily smooth v cannot be expected to be in $\mathcal{A}^s_\infty$ for $s > d - 1$. On the other hand, in [Nit06], Nitsche showed that for sufficiently smooth wavelets, e.g., spline wavelets, for $0 < s < d - 1$ and with $\tau = (s + \frac{1}{2})^{-1}$

$$\mathbf{v} \in \mathcal{A}^s_\tau \Longleftrightarrow v \in \bigcap_{k=1}^n \bigotimes_{m=1}^n{}_\tau B^{s+\delta_{mk}}_\tau(L_\tau(0,1))$$

where $\mathcal{A}^s_\tau := \{\mathbf{v} \in \ell_2(\mathbf{\nabla}) : \sum_{N\in\mathbb{N}}(N^s\|\mathbf{v}-\mathbf{v}_N\|_{\ell_2(\mathbf{\nabla})})^\tau N^{-1} < \infty\}$, $B^{s+\delta_{mk}}_\tau(L_\tau(0,1))$ is the Besov space measuring "$s + \delta_{mk}$ orders of smoothness in L_τ", and $\bigotimes_\tau$ denotes the so-called "τ tensor product". Note that $\mathcal{A}^s_\tau$ is even (slightly) smaller than $\mathcal{A}^s_\infty$.

The upshot of this result is that $\bigcap_{k=1}^n \bigotimes_{m=1}^n{}_\tau B^{s+\delta_{mk}}_\tau(L_\tau(0,1))$ is much larger, with an increasing difference when s and n get larger, than $\bigcap_{k=1}^n \bigotimes_{m=1}^n H^{s+\delta_{km}}(0,1)$, membership of which is needed to guarantee the same rate with sparse grid approximation. In particular, for two and three space dimensions, in [Nit05] it was shown that for general, sufficiently smooth f, the solution u of (1) is in $\bigcap_{k=1}^n \bigotimes_{m=1}^n{}_\tau B^{s+\delta_{mk}}_\tau(L_\tau(0,1))$ for any s. This means that the rate of convergence of best N-term approximation is never restricted by the regularity of u for whatever order d. The proof in [Nit05] makes use of the splitting of u into known singular functions and a smooth remainder, which is also available for (1) in more than three dimensions in [Dau88b]. It indicates the huge potential of best N-term approximation in tensor product bases.

Above considerations concern best N-term approximations that, however, are not feasible in practice, already because $\mathbf{u}$ is not given explicitly. It can be found as the solution of the infinite matrix-vector problem

$$\mathbf{A}\mathbf{u} = \mathbf{f}, \tag{5}$$

with "stiffness matrix" $\mathbf{A} = [a(\boldsymbol{\psi}_{\boldsymbol{\lambda}'}, \boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda},\boldsymbol{\lambda}'\in\mathbf{\nabla}}$ and $\mathbf{f} = [f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda}\in\mathbf{\nabla}}$. This infinite matrix problem is equivalent to (1). In [CDD01, CDD02], optimal *adaptive* algorithms were introduced for solving (5). It was shown that whenever for some $s > 0$, $\mathbf{u}$ happens to belong to $\mathcal{A}^s_\infty$, then, under two assumptions discussed below, the sequence of approximations produced by these algorithms converge to $\mathbf{u}$ with *this rate* s, requiring a number of operations equivalent to their length. The first

assumption is that one has available a routine **RHS** that given an $\varepsilon > 0$ produces an approximation $\mathbf{f}_\varepsilon := \mathbf{RHS}[\varepsilon]$ with $\|\mathbf{f} - \mathbf{f}_\varepsilon\| \leq \varepsilon$ and

$$\mathbf{rhs}_s := \sup_{0<\varepsilon<\|\mathbf{f}\|_{\ell_2(\boldsymbol{\nabla})}} \varepsilon\big[\#\text{ operations required by the call } \mathbf{RHS}[\varepsilon]\big]^s < \infty. \tag{6}$$

The second assumption is that for some $s^* > s$, $\mathbf{A}$ is s^*-computable, which is a quantitative measure how well $\mathbf{A}$ can be approximated by computable sparse matrices. This assumption implies that $\mathbf{A} : \mathcal{A}^s_\infty \to \mathcal{A}^s_\infty$ is bounded, and thus that $\|\mathbf{f}\|_{\mathcal{A}^s_\infty} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_\infty}$. So in any case if we could realize (quasi-) best N-term approximations for $\mathbf{f}$ in $\mathcal{O}(N)$ operations, then the resulting routine **RHS** would satisfy $\mathbf{rhs}_s < \infty$, with in particular $\mathbf{rhs}_s \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_\infty}$.

In the tensor product basis setting, applying as a building block univariate biorthogonal spline wavelets, in [SS07] we proved that $\mathbf{A}$ is s^*-computable for some $s^* > d - 1$, which s^* is thus larger than any s for which $\mathbf{u} \in \mathcal{A}^s_\infty$ might be expected. We showed this result in a more general setting than we consider here. Most importantly, we allowed general, i.e., non-separable, smooth variable coefficients in the differential operator, meaning that we had to design and analyze a quadrature scheme. Our results imply that for any n, and $s \in [0, d-1]$, the adaptive wavelet schemes produce an approximation to $\mathbf{u} \in \mathcal{A}^s_\infty$ within any given *tolerance* $0 < \varepsilon \lesssim \|\mathbf{f}\|_{\ell_2(\boldsymbol{\nabla})}$ in $\|\cdot\|_{\ell_2(\boldsymbol{\nabla})}$ with a *support length* not exceeding

$$D_1 \varepsilon^{-1/s} |\mathbf{u}|^{1/s}_{\mathcal{A}^s_\infty},$$

taking a *number of operations* not exceeding

$$D_2 \varepsilon^{-1/s} |\mathbf{u}|^{1/s}_{\mathcal{A}^s_\infty} + D_3 \varepsilon^{-1/s} \mathbf{rhs}_s^{1/s},$$

where D_1, D_2 and D_3 are some constants, independent of $c_0 \geq 0$ and $c_m > 0$ $(m = 1, \ldots, n)$. Only knowing that $\mathbf{u} \in \mathcal{A}^s_\infty$, up to the factors D_1 or $D_2 + D_3[\mathbf{rhs}_s/|\mathbf{u}|_{\mathcal{A}^s_\infty}]^{1/s}$, this length or number of operations are indeed the best that generally can be expected.

What we did not analyze, however, is the dependence of D_1, D_2, D_3 and the quotient from (4) on the space dimension n. Concerning the latter, note that instead of approximating $\mathbf{u}$, our ultimate goal is to approximate u in $|||\cdot|||$ within some given tolerance with, up to some constant factor, the smallest linear combination of wavelets. The quotient from (4) is nothing else than the condition number $\kappa(\mathbf{A})$ of $\mathbf{A}$, which for biorthogonal wavelets can be expected to grow exponentially with n. Since in any case also D_2 is an increasing function of $\kappa(\mathbf{A})$, it may be that, although optimal for any fixed n, the method has only practical value for relatively small values of n.

In view of this, in the current work we apply univariate $L_2(0,1)$-*orthonormal, piecewise polynomial* wavelets as introduced in [DGH96]. In this case, $\kappa(\mathbf{A})$ is bounded *uniformly in* n, and $c_0 \geq 0$ and $c_m > 0$ $(m = 1, \ldots, n)$. Thanks to both the L_2-orthonormality, and the fact that the wavelets are piecewise polynomial, the stiffness matrix $\mathbf{A}$ appears to be very close to a sparse matrix. We give a detailed description of an adaptive wavelet algorithm for which aforementioned

statements are valid with

$$D_1, \frac{D_2}{n}, D_3 \text{ constants, } \textit{independent of } n \text{ and } c_0 \geq 0,\ c_m > 0\,(m = 1, \ldots, n).$$

So only the constant involved in the operation count may grow with the space dimension, but only linearly.

The algorithm we use is a modification developed in [GHS07] of the adaptive wavelet method from [CDD01]. With this modification, the recurrent coarsening of the iterands is avoided. An additional quantitative improvement will be obtained by the use of a modified approximate matrix vector routine **APPLY** that will use a posteriori information to optimize the accuracies with which the columns of the infinite stiffness matrix are approximated.

References

[BG04] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, 70:27–75, 2001.

[CDD02] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[Dau88b] M. Dauge. *Elliptic boundary value problems on corner domains*, volume 1341 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1988. Smoothness and asymptotics of solutions.

[DGH96] G.C. Donovan, J.S. Geronimo, and D.P. Hardin. Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets. *SIAM J. Math. Anal.*, 27(6):1791–1815, 1996.

[GHS07] T. Gantumur, H. Harbrecht, and R.P. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 77:615–629, 2007.

[Nit05] P.-A. Nitsche. Sparse approximation of singularity functions. *Constr. Approx.*, 21(1):63–81, 2005.

[Nit06] P.-A. Nitsche. Best N-term approximation spaces for tensor product wavelet bases. *Constr. Approx.*, 41(1):49–70, 2006.

[SS07] Ch. Schwab and R.P. Stevenson. Adaptive wavelet algorithms for elliptic PDEs on product domains. *Math. Comp.*, 2007. To appear.

Sparse Finite Element Approximation of High-Dimensional Transport-Dominated Diffusion Problems

Endre Süli

(joint work with Christoph Schwab and Radu-Alexandru Todor)

Partial differential equations with non-negative characteristic form arise in numerous mathematical models in science. In problems of this kind, the exponential growth of computational complexity as a function of the dimension d of the problem domain, the so-called "curse of dimension", is exacerbated by the fact that the problem may be transport-dominated. We develop the numerical analysis of stabilized sparse tensor-product finite element methods for such high-dimensional, non-self-adjoint and possibly degenerate second-order partial differential equations, using piecewise polynomials of degree $p \geq 1$. Our convergence analysis is based on new high-dimensional approximation results in sparse tensor-product spaces. By

tracking the dependence of the various constants on the dimension d and the polynomial degree p, we show that in the case of elliptic transport-dominated diffusion problems the error-constant exhibits exponential decay as $d \to \infty$ for $p \geq 1$. When the characteristic form of the partial differential equation is non-negative, under a mild condition relating p to d, the error constant is shown to grow no faster than $\mathcal{O}(d^2)$. In any case, the sparse stabilized finite element method exhibits an optimal rate of convergence with respect to the mesh size h_L, up to a factor that is polylogarithmic in h_L.

1. Model problem

Suppose that $\Omega := (0,1)^d$, $d \geq 2$, and that $a = (a_{ij})_{i,j=1}^d$ is a symmetric positive semidefinite matrix with entries $a_{ij} \in \mathbb{R}$, $i,j = 1,\ldots,d$. In other words, $a^\top = a$ and $\xi^\top a\, \xi \geq 0$ for all $\xi \in \mathbb{R}^d$. Suppose further that $b \in \mathbb{R}^d$ and $c \in \mathbb{R}$, and let $f \in \mathrm{L}^2(\Omega)$. Let $\Gamma \subset \partial\Omega$ denote the union of all $(d-1)$-dimensional open faces of Ω. We shall consider the partial differential equation

$$-a : \nabla\nabla u + b \cdot \nabla u + cu = f(x), \quad x \in \Omega, \tag{1}$$

subject to suitable boundary conditions on Γ which will be stated below.

Here $\nabla\nabla u$ is the $d \times d$ Hessian matrix of u whose (i,j) entry is $\partial^2 u/\partial x_i\, \partial x_j$, $i,j = 1,\ldots,d$. For two $d \times d$ matrices A and B, we define $A : B := \sum_{i,j=1}^d A_{ij} B_{ij}$. The real-valued polynomial $\alpha \in \mathcal{P}^2(\mathbb{R}^d;\mathbb{R})$ of degree ≤ 2 defined by

$$\xi \in \mathbb{R}^d \;\mapsto\; \alpha(\xi) := \xi^\top a\, \xi \in \mathbb{R}$$

is called the *characteristic polynomial* or *characteristic form* of the differential operator $u \;\mapsto\; \mathcal{L}u := -a : \nabla\nabla u + b \cdot \nabla u + cu$ featuring in (1) and, under our hypotheses on the matrix a, following Oleĭnik and Radkevič, the equation (1) is referred to as a *partial differential equation with non-negative characteristic form.*

2. Summary of the main results

Let $I = (0,1)$ and consider the sequence of partitions $\{\mathcal{T}^\ell\}_{\ell \geq 0}$, where $\mathcal{T}^0 = \{I\}$ and $\mathcal{T}^{\ell+1}$ is obtained from $\mathcal{T}^\ell := \{I_j^\ell : j = 0,\ldots,2^\ell - 1\}$ by halving each of the intervals I_j^ℓ; the mesh-size in the partition $\mathcal{T}^\ell$ is $h_\ell := |I_j^\ell| = 2^{-\ell}$. We consider the finite-dimensional linear subspace $\mathcal{V}^{\ell,p}$ of $\mathrm{H}^1(0,1)$ consisting of all continuous piecewise polynomials of degree $p \geq 1$ on $\mathcal{T}^\ell$, $\ell \geq 0$. For $\ell \geq 0$ let $\mathcal{V}_0^{\ell,p}$ be the subspace of $\mathcal{V}^{\ell,p}$ defined by $\mathcal{V}_0^{\ell,p} := \mathcal{V}^{\ell,p} \cap \mathrm{H}_0^1(0,1)$, consisting of all continuous piecewise polynomials on $\mathcal{T}^\ell$ of degree p that vanish at both endpoints of the interval $[0,1]$. We shall use $\mathcal{V}_{(0)}^{\ell,p}$ to denote $\mathcal{V}_0^{\ell,p}$ or $\mathcal{V}^{\ell,p}$, as the case may be. We shall write $\mathrm{H}_{(0)}^1(0,1)$ to denote $\mathrm{H}_0^1(0,1)$ or $\mathrm{H}^1(0,1)$, depending on whether or not homogeneous Dirichlet boundary conditions are imposed at the endpoints of I.

For $\ell \geq 0$ and $p \geq 1$, consider the projector $P^{\ell,p} : \mathrm{H}^1(0,1) \to \mathcal{V}^{\ell,p}$ defined by

$$(P^{\ell,p}u)(x) := u(0) + (P_0^{\ell,p}u)(x), \quad (P_0^{\ell,p}u)(x) := \int_0^x (\Pi^{\ell,p-1}u')(\xi)\,\mathrm{d}\xi, \quad x \in [0,1],$$

where $\Pi^{\ell,p-1} : \mathrm{L}^2(0,1) \to \mathcal{V}^{\ell,p-1}$ is the $\mathrm{L}^2(0,1)$-orthogonal projector onto $\mathcal{V}^{\ell,p-1}$.

It is easily seen that $(P^{\ell,p}u)(1) = u(1)$ for all $\ell \geq 0$ and all $p \geq 1$; furthermore, $P^{\ell,p}|_{\mathrm{H}^1_0(0,1)} = P_0^{\ell,p}$. The projector $P_{(0)}^{\ell,p}$ has the following approximation property: for any v in $\mathrm{H}^{k+1}(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, $k \geq 1$, we have that

$$|v - P_{(0)}^{\ell,p} v|_{\mathrm{H}^s(0,1)} \leq \left(\frac{h_\ell}{2}\right)^{t+1-s} \frac{1}{p^{1-s}} \sqrt{\frac{(p-t)!}{(p+t)!}}\, |v|_{\mathrm{H}^{t+1}(0,1)}, \quad 1 \leq t \leq \min(p,k),$$

where $h_\ell = 2^{-\ell}$, $\ell \geq 0$, $p \geq 1$, $s \in \{0,1\}$. Clearly, $\mathcal{V}_{(0)}^{\ell,p} = P_{(0)}^{\ell,p}\, \mathrm{H}^1_{(0)}(0,1)$ for all $\ell \geq 0$ and $p \geq 1$. Now, let $Q_{(0)}^{\ell,p} := P_{(0)}^{\ell,p} - P_{(0)}^{\ell-1,p}$, $\ell \geq 1$, and $Q_{(0)}^{\ell,p} := P_{(0)}^{0,p}$ for $\ell = 0$. We define the *hierarchical increment spaces* $\mathcal{W}_{(0)}^{\ell,p} := Q_{(0)}^{\ell,p}\, \mathrm{H}^1_{(0)}(0,1)$, $\ell = 0, 1, \ldots$. Hence, for any pair of integers $L \geq 0$ and $p \geq 1$, $\mathcal{V}_{(0)}^{L,p} = \bigoplus_{\ell=0}^{L} \mathcal{W}_{(0)}^{\ell,p}$.

Consider, on $\Omega = (0,1)^d$, the *sparse tensor-product finite element space*

$$\hat{V}_{(0)}^{L,p} := \bigoplus_{|\ell|_1 \leq L} \mathcal{W}_{(0)}^{\ell_1,p} \otimes \cdots \otimes \mathcal{W}_{(0)}^{\ell_d,p}, \qquad \ell = (\ell_1, \ldots, \ell_d).$$

Given $\mathfrak{I} = \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, d\}$, let $\mathfrak{I}^c = \{j_1, j_2, \ldots, j_{d-k}\}$ denote the (possibly empty) complement of $\mathfrak{I}$ with respect to $\{1, 2, \ldots, d\}$; for non-negative integers α and β we then denote by $|u|_{\mathrm{H}^{\alpha,\beta,\mathfrak{I}}(\Omega)}$ the seminorm

$$\sum_{(\alpha)_i \leq \alpha_i \leq \alpha} \cdots \sum_{(\alpha)_k \leq \alpha_k \leq \alpha} \sum_{(\beta)_j \leq \beta_j \leq \beta} \cdots \sum_{(\beta)_{d-k} \leq \beta_{d-k} \leq \beta} \left\| \left(\frac{\partial^{\alpha_1}}{\partial x_{i_1}^{\alpha_1}} \cdots \frac{\partial^{\alpha_k}}{\partial x_{i_k}^{\alpha_k}} \right) \left(\frac{\partial^{\beta_1}}{\partial x_{j_1}^{\beta_1}} \cdots \frac{\partial^{\beta_{d-k}}}{\partial x_{j_{d-k}}^{\beta_{d-k}}} \right) u \right\|$$

where $\|\cdot\|$ is the $\mathrm{L}^2(\Omega)$-norm and, for $i = 1, \ldots, k$,

$$(\alpha)_i := \begin{cases} \alpha & \text{if } Ox_i \text{ is an elliptic co-ordinate direction,} \\ 0 & \text{if } Ox_i \text{ is a hyperbolic co-ordinate direction,} \end{cases}$$

with analogous definition of $(\beta)_j$, $j = 1, \ldots, d-k$.

Theorem 1. Let $\Omega = (0,1)^d$, $s \in \{0,1\}$, $k \geq 1$, and $p \geq 1$. Let, further, $\mathrm{C}^\infty_{(0)}(\bar{\Omega})$ denote the set of all $\mathrm{C}^\infty(\bar{\Omega})$-functions that vanish on Γ_0, the elliptic part of Γ. Then, for $1 \leq t \leq p$, there exist constants $\underline{c}_{p,t}$, $\kappa(p,t,s,L) > 0$, independent of d, and κ monotonic decreasing in $L \geq 1$, such that, for any $u \in \mathrm{C}^\infty_{(0)}(\bar{\Omega})$, any $L \geq 1$ and any $d \geq 1$, we have

$$\inf_{v \in \hat{V}_{(0)}^{L,p}} |u - \hat{P}_{(0)}^{L,p} u|_{\mathrm{H}^s(\Omega)} \leq d^{1+\frac{s}{2}} \underline{c}_{p,t} (\kappa(p,t,s,L))^{d-1+s} L^{\nu(s)} 2^{-(t+1-s)L} |u|_{\mathcal{H}^{k+1}(\Omega)},$$

where, for $s = 0$, the seminorm $|\cdot|_{\mathrm{H}^s(\Omega)}$ coincides with the $\mathrm{L}^2(\Omega)$-norm and $\nu(0) = d - 1$, while for $s = 1$ the seminorm $|\cdot|_{\mathrm{H}^s(\Omega)}$ is the $\mathrm{H}^1(\Omega)$-seminorm and $\nu(1) = 0$, and where $|u|_{\mathcal{H}^{k+1}(\Omega)} := \max_{1 \leq k \leq d} \max_{\mathfrak{I} \subseteq \{1,2,\ldots,d\},\, |\mathfrak{I}| = k} |u|_{\mathrm{H}^{t+1,s,\mathfrak{I}}(\Omega)}$.

It is shown in [1] that when $\Gamma = \Gamma_0$ (i.e. when $a = a^\top > 0$, and so the problem is elliptic in Ω,) the constant $\kappa(p,t,s,L)$ is in the range $(0,1)$ for all $t = p \geq 1$, $s \in \{0,1\}$ and $L \geq 1$. In the case when $\Gamma_0 \subsetneq \Gamma$ (i.e. the hyperbolic part of the boundary is nonempty), for $s = 0$, $\kappa(p,p,0,L) \in (0,1)$ for all $p \geq 2$ and all $L \geq 2$,

while for $s = 1$, $\kappa(p, p, 1, L) \in (0, 1)$ when $p = 2$ and $d \leq 7$, when $p = 3$ and $d \leq 71$, and when $p = 4$ and $d \leq 755$. Finally it is shown that the, potentially harmful, polylogarithmic factor $L^{d-1} = |\log_2 h_L|^{d-1}$ can be absorbed into the factor $(\kappa(p, t, 0, L))^{d-1}$ for $p = t = 2$ and $L \leq 5$, for $p = t = 3$ and $L \leq 29$, and for $p = t \geq 4$ and $L \leq 397$. Thus, in most cases of practical interest, the error constant, for $p = t$ fixed and L fixed, exhibits exponential decay to 0 as $d \to \infty$.

Based on this approximation result, in [1] we also develop a streamline-diffusion type stabilized finite element method for (1) over the sparsified tensor-product finite element space $\hat{V}^{L,p}_{(0)}$, and we show that for each (fixed) $p \geq 1$ the method exhibits an optimal order of convergence in h_L (up to a polylogarithmic factor in h_L) in the streamline-diffusion norm $|||\cdot|||_{\mathrm{SD}}$ defined by

$$|||v|||^2_{\mathrm{SD}} := \|\sqrt{a}\,\nabla v\|^2_{\mathrm{L}^2(\Omega)} + c\|v\|^2_{\mathrm{L}^2(\Omega)} + \delta_L\|b \cdot \nabla v\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}(1 + c\delta_L)\textstyle\int_{\Gamma_+}|\beta||v|^2\,\mathrm{d}s,$$

where δ_L is the streamline-diffusion parameter, $\beta = b \cdot n$, n is the unit outward normal to Γ and $\Gamma_+ \subset \Gamma$ is the hyperbolic outflow boundary. As in Theorem 1, the error constant of the method exhibits exponential decay to 0 as $d \to \infty$.

Our results represent an extension and generalization of an observation made by Griebel in [2], where exponential decay of the error constant was observed for Poisson's equation on $\Omega = (0, 1)^d$, as $d \to \infty$, for $p = 1$.

References

[1] C. Schwab, E. Süli, and R.-A. Todor, *Sparse finite element approximation of high-dimensional transport-dominated diffusion problems.* M2AN, submitted for publication, (2007). Available from: `http://web.comlab.ox.ac.uk/oucl/publications/natr/`

[2] M. Griebel, *Sparse grids and related approximation schemes for higher dimensional problems.* Foundations of Computational Mathematics 2005. Luis-Miguel Pardo, Allan Pinkus, Endre Süli, Michael Todd, editors. Cambridge University Press (2006) 106–161.

Analytic and Geometric Features of Reproducing Groups

Anita Tabacco

(joint work with E. Cordero, F. De Mari and K. Nowak)

The analysis of oriented features in two-dimensional images requires objects more flexible than the wavelets arising from the tensor product of one-dimensional wavelets. To solve this problem, frame systems of directional functions, with excellent angular selectivity, have been constructed and used in signal processing. Among these, two-dimensional directional wavelets well-known and heavily used in the last few years are the so-called curvelets and shearlets. The former obtain directional selectivity by a construction that requires a rotation operator, the latter gets the directional property by means of a shearing operation. They both represent different solutions to the same problem of image recovering and both have advantages and disadvantages. A natural question is whether there are other constructions, involving other operations on the plane, that guarantee the reproducibility of a signal and enjoy more flexible properties than the ones above.

Moreover, a complete classification of such systems (at least on the plane) would provide a clear panorama of all the instruments available. We want to study such problems; the starting point is given by continuous representations in every dimensions. We would like to understand analytic, geometric and algebraic features which distinguish such reproducing formulae.

Precisely, in this report we study the reproducing formula

$$f = \int_H \langle f, \mu_e(h)\phi\rangle \mu_e(h)\phi \, dh, \qquad f \in L^2(\mathbb{R}^d), \tag{1}$$

where H is a Lie subgroup of the semidirect product G of the (double cover of the) symplectic group $Sp(d,\mathbb{R})$ and the Heisenberg group $\mathbb{H}^d$, or, to be precise, its quotient modulo the (irrelevant) center. The map $h \mapsto \mu_e(h)\phi$ arises from the restriction to H of the (extended) metaplectic representation μ_e of G as applied to a fixed and suitable window $\phi \in L^2(\mathbb{R}^d)$. Formula (1) is otherwise known as resolution of the identity, and must be interpreted in the weak sense. Such formulae appear pervasively in mathematics and physics, both in pure and in applied areas such as coherent states, harmonic analysis and group representations, wavelet and Gabor analysis, signal processing, and so on [1, 2, 4, 8].

Among the most widely used and studied versions of (1) are the Calderón-Grossman-Morlet formula, on which wavelet theory rests, and Gabor's formula, at the heart of signal processing. It is though *not* as widely known that both these formulae arise by restricting the integral in (1) to suitable subgroups of G [4, 7].

The main question is: for which subgroups H of G does there exist an analyzing window, that is, a function $\phi \in L^2(\mathbb{R}^d)$ such that the above reproducing formula holds *for all* $f \in L^2(\mathbb{R}^d)$? Clearly, one looks for invariants or other general properties that will decide whether a group H enjoys the property, and is thus called a *reproducing* group, or not. Further, one seeks conditions that single out the analyzing windows, those for which (1) holds and, consequently, are named *reproducing*. A complete classification of reproducing subgroups when $d = 1$ is given in [7] and various examples in higher dimension have been worked out in [3, 4, 5]. Let us fix some notation. The symplectic group is

$$Sp(d,\mathbb{R}) = \left\{ g \in GL(2d,\mathbb{R}) : \ {}^t g J g = J \right\}, \quad J = \begin{bmatrix} 0 & I_d \\ -I_d & 0 \end{bmatrix},$$

where J defines the standard symplectic form $\omega(x,y) = {}^t x J y$ on $\mathbb{R}^{2d}$. The Heisenberg group $\mathbb{H}^d$ is $\mathbb{R}^{2d+1} = \mathbb{R}^{2d} \times \mathbb{R}$ equipped with the product

$$(z,t)\cdot(z',t') = (z+z', t+t' - \frac{1}{2}\omega(z,z')). \tag{2}$$

We often refer to $\mathbb{R}^{2d}$ as the phase space and we write $z = (x,\xi)$ when we separate the space component x from the frequency component ξ. We denote the translation and modulation operators on $L^2(\mathbb{R}^d)$ by $T_x f(t) = f(t-x)$ and $M_\xi f(t) = e^{2\pi i\langle \xi, t\rangle} f(t)$. The Schrödinger representation of $\mathbb{H}^d$ on $L^2(\mathbb{R}^d)$ is

$$\rho(x,\xi,t)f(y) = e^{2\pi i t} e^{\pi i\langle x,\xi\rangle} e^{2\pi i\langle \xi, y\rangle} f(y-x) = e^{2\pi i t} e^{\pi i\langle x,\xi\rangle} T_x M_\xi f(y).$$

The metaplectic representation μ is a representation of the double cover of the symplectic group. Up to a sign, μ can be considered as a representation of the symplectic group, as we briefly indicate. For any given $A \in Sp(d, \mathbb{R})$, $\mu(A)$ is the intertwining unitary operator (whose existence is guaranteed by the Stone–von Neumann theorem) between ρ and ρ_A, where ρ_A is the unitary representation of $\mathbb{H}^d$ obtained by composing ρ with the automorphism of $\mathbb{H}^d$ defined by A, that is

$$\rho_A : \mathbb{H}^d \to \mathcal{U}(L^2(\mathbb{R}^d)), \qquad (z,t) \mapsto \rho(Az,t).$$

Observe that the reproducing formula (1) is insensitive to phase factors e^{is}, that is, it remains unchanged under the mapping $\phi \mapsto e^{is}\phi = \rho(0,s)\phi$. Hence the role of the center of the Heisenberg group is irrelevant and the "true" group under consideration is $\mathbb{R}^{2d} \rtimes Sp(d, \mathbb{R})$, which we denote again by G.

The affine action of G on $\mathbb{R}^{2d}$ is given by

$$g(x,\xi) = ((q,p), A)\,(x,\xi) = A\,{}^t(x,\xi) + {}^t(q,p). \tag{3}$$

The many connections of these themes with wavelet theory have been pointed out in [4], where we have introduced the notion of *admissible* group. This is very closely related to a parallel notion developed in [8] in a different setting, and it brings geometry to the forefront through a purely analytic construct, the Wigner distribution. The latter is the time-frequency representation of $\phi \in L^2(\mathbb{R}^d)$

$$W_\phi(x,\xi) = \int_{\mathbb{R}^d} e^{-2\pi i\langle \xi, y\rangle} \phi(x + \frac{y}{2})\overline{\phi(x - \frac{y}{2})}\, dy, \tag{4}$$

where $\langle x, \xi\rangle$ is the inner product of $x, \xi \in \mathbb{R}^d$. Its most relevant feature in this context is that it intertwines the representation μ_e with the affine action:

$$W_{\mu_e(g)\phi}(x,\xi) = W_\phi\left(g^{-1}(x,\xi)\right), \qquad g \in G, \tag{5}$$

where $g(x,\xi)$ is defined in (3). The relation between (1) relative to a subgroup $H \subset G$ and the properties of its orbits is shown in the following result [4, Thm. 1], which gives an analytic condition that a subgroup has to enjoy.

THEOREM 1. If there exists a function ϕ such that the mapping

$$h \to W_\phi(h^{-1}(x,\xi)) \tag{6}$$

is in $L^1(H)$ for a.e. $(x,\xi) \in \mathbb{R}^{2d}$, and such that

$$\int_H |W_\phi(h^{-1}(x,\xi))|\, dh \le M, \quad \text{for a.e. } (x,\xi) \in \mathbb{R}^{2d}, \tag{7}$$

then H is reproducing if and only if the following admissibility condition holds:

$$\int_H W_\phi(h^{-1}(x,\xi))\, dh = 1 \quad \text{for a.e. } (x,\xi) \in \mathbb{R}^{2d}. \tag{8}$$

A group for which there exists a ϕ such that conditions (6)–(8) are satisfied is called an *admissible* group, whereas ϕ is an *admissible analyzing function.*

Let us now consider some geometric features of an admissible subgroup (we refer to [5] for the proofs and related results). The group G and its affine action on phase space play a prominent role, and we get dimensional bounds as simple

consequences of the topology of its stabilizers. Indeed, as it follows by a direct adaptation of a result in [8], they must be (almost all) compact. Since the orbits are at most $2d$-dimensional, it is enough to look at the largest possible compact stabilizers. If H has no nontrivial compact subgroups, then clearly $\dim H \leq 2d$. Here, as it is customary, $\mathrm{Stab}_z(H) = \{h \in H : hz = z\}$ and hz is as in (3).

PROPOSITION 2. [8, Prop. 2.1 and 2.3] *Let $H \subset G$ be admissible. Then*

(i) *the stabilizer* $\mathrm{Stab}_z(H)$ *is compact for a.e. $z \in \mathbb{R}^{2d}$;*
(ii) *if $H \subset Sp(d, \mathbb{R})$, then H is not unimodular.*

The maximal compact stabilizers, among all possible subgroups of G, turn out to be isomorphic to the unitary group $U(d)$, whence the first upper bound $\dim H \leq d^2 + 2d$. Since the restriction of (1) to $\mathbb{R}^{2d} \subset \mathbb{H}^d$ yields Gabor reproducing formula, $\mathbb{R}^{2d}$ is a reproducing, and in fact admissible, subgroup of G. Thus, its semidirect product with $U(d)$ is again an admissible group, and an example of maximal dimension. Indeed, adding compact semidirect factors to a group preserves both reproducibility and admissibility, as observed formally in the following proposition.

PROPOSITION 3. Let $H = H_0 \rtimes K$ with $H_0, K \subset G$, K compact. Then

(i) if H_0 is reproducing, then H is reproducing;
(ii) if H_0 is admissible, then H is admissible.

Moreover, if one looks at subgroups of $Sp(d, \mathbb{R})$ alone, thereby considering only linear transformations of phase space (*versus* affine), then a stabilizer must fix a vector in phase space and is therefore isomorphic to a subgroup of $U(d-1)$. Hence the upper bound $\dim H \leq d^2 + 1$. This is also sharp, because the semidirect product $F = \mathbb{H}_e^{d-1} \rtimes U(d-1)$ of the extended Heisenberg group and the unitary group embeds into $Sp(d, \mathbb{R})$, has dimension $d^2 + 1$ and is admissible for all $d \geq 2$, as we show. We also prove directly that F is reproducing and we describe necessary and sufficient conditions for its reproducing windows.

Let us summarize these results.

THEOREM 4. Let $H \subset G$ be admissible. Then

(i) $\dim H \leq d^2 + 2d$;
(ii) if H contains no nontrivial compact subgroups, then $\dim H \leq 2d$;
(iii) if $H \subset Sp(d, \mathbb{R})$, then $\dim H \leq d^2 + 1$.

Moreover the upper bounds are sharp.

Finally, in a forecoming paper [6], we study a unified approach to Gabor and wavelet reproducing formulae.

REFERENCES

[1] S. T. Ali, J. P. Antoine and J. P. Gazeau, *Coherent states, wavelets and their generalizations*, Springer-Verlag, New York, 2000.
[2] D. Bernier and K. F. Taylor, *Wavelets from square-integrable representations*, SIAM J. Math. Anal. **27** (1996), 594–608.

[3] E. Cordero, F. De Mari, K. Nowak, and A. Tabacco, *Reproducing subgroups for the metaplectic representation*, in Operator Theory: Advances and Applications **164** (2006), 227–244, 2006.
[4] E. Cordero, F. De Mari, K. Nowak, and A. Tabacco, *Analytic features of reproducing groups for the metaplectic representation*, J. Fourier Anal. Appl. **12** (2006), 157–180.
[5] E. Cordero, F. De Mari, K. Nowak, and A. Tabacco, *Dimensional upper bounds for admissible subgroups for the metaplectic representation*, Preprint n. 11-2007, Dipartimento di Matematica, Politecnico di Torino.
[6] E. Cordero, F. De Mari, and A. Tabacco, *A unified approach to Gabor and wavelet reproducing formulae*, in preparation.
[7] F. De Mari and K. Nowak, *Analysis of the affine transformations of the time-frequency plane* Bull. Austral. Math. Soc., **63** (2001), 195–218.
[8] R. S. Laugesen, N. Weaver, G. L. Weiss, and E. N. Wilson, *A characterization of the higher dimensional groups associated with continuous wavelets*, J. Geom. Anal. **12** (2002),89–102.

Edges and Hierarchy of Scales

Eitan Tadmor

1. Introduction

I will describe two, not unrelated scenarios, where edges can be treated through different scales. The first part of the talk is devoted to detection of edges in piecewise smooth data. Spectral edge detection is achieved by *separation of scales*. The data is given in terms of its spectral representation. The objective is to recover the location and amplitudes of jump discontinuities, from possibly incomplete and noisy spectral information. We utilize a general family of edge detectors based on concentration kernels. Each kernel forms an approximate derivative of the delta function, which detects edges by separation of scales.

Edges are also the noticeable feature in images. The second part of the talk is devoted to a novel representation of general images as hierarchy of edges, by *decompoistion of scales*. The resulting decomposition is essentially nonlinear. The questions of convergence, localization and adaptivity will be discussed and numerical results will illustrate applications to synthetic and real images.

2. Separation of scales

This part of the talk is a summry of joint works with A. Gelb [3, 4], S. Engelberg [2] and J. Zou [10]. A detailed account can be found in the recent Acta Numerica review article [9].

We are interested in processing piecewise smooth functions from their spectral information. The prototype example will be one-dimensional functions that are smooth except for finitely many jump discontinuities. The location and amplitudes of these discontinuities are not correlated. Thus, a piecewise smooth f is in fact a collection of several intervals of smoothness which do *not* communicate among themselves. The jump discontinuities can be viewed as the edges of these intervals of smoothness. Similarly, two-dimensional piecewise smooth functions consist of finitely many edges which lie along simple curves, separating two-dimensional local

regions of smoothness. We are concerned here with *edge detection*: detecting the location and amplitudes of the edges. Often, these are the essential features sought in piecewise smooth data. Moreover, they define the regions of smoothness and are therefore essential for the *reconstruction* of the underlying function f inside its different regions of smoothness.

There are many classical algorithms to detect edges and reconstruct the data in between those edges, based on *local* information. For example, suppose that the values of a one-dimensional f are given at equidistant grid-points, $f_\nu = f(\nu\Delta x)$. Then, the first-order differences, $\Delta f_\nu := f_{\nu+1} - f_\nu$, can detect edges where $\Delta f_\nu = \mathcal{O}(1)$, by separating them from smooth regions where $\Delta f_\nu = \mathcal{O}(\Delta x)$. Similarly, piecewise linear interpolants can recover the point-values of $f(x)$ up to order $\mathcal{O}\big((\Delta x)^2\big)$. Of course, these are only asymptotic statements that may greatly vary with the dependence of the $\mathcal{O}$-terms on the *local* smoothness of f in the immediate neighborhood of x. We may do better, therefore, by taking higher-order differences, $\Delta^r f_\nu$, where $\mathcal{O}(1)$-edges are better separated from $\mathcal{O}\big((\Delta x)^r\big)$-regions of smoothness, Similarly, reconstruction of f using r-order approximations, with $r = 2, 3, \ldots$ and so on. In practice, higher accuracy is translated into higher resolution extracted from the information on a given grid. But, as the order of accuracy increases, the stencils involved become wider and one has to be careful not to extract smoothness information *across* edges, since different regions separated by edges are completely independent of each other. An effort to extract information from one region of smoothness into another one, will result in spurious oscillations, spreading from the edges into the surrounding smooth regions, preventing uniform convergence. This is, in general terms, the *Gibbs phenomenon*, which is the starting point of the present discussion.

The prototype for spectral information we are given on f, is the set of its N Fourier coefficients, $\{\widehat{f}(k)\}_{|k|\le N}$. These are *global* moments of f. It is well known that the Fourier projection, $S_N f = \sum_{|k|\le N} \widehat{f}(k) e^{ikx}$ forms a highly accurate approximation of f provided that f is *sufficiently smooth*; for example,

$$|S_N f - f| \lesssim e^{-\eta \sqrt[\alpha]{N}}.$$

Here, the root exponent α is tied to *global* smoothness of f of order $\alpha \ge 1$. But this high accuracy is lost with piecewise smooth f's, due to spurious oscillations which are formed around the edges of f. It is in this context of the Fourier projections, that the formation of spurious oscillations became known as *the* Gibbs phenomenon. This is precisely because of the global nature of $S_N f$, which extracts smoothness information *across* the internal edges of f. The Gibbs' phenomenon is also responsible for a *global* loss of accuracy: first-order oscillations spread *throughout* the regions of smoothness. The highly accurate content in the spectral data, $\{\widehat{f}(k)\}_{|k|\le N}$, is lost in the Fourier projections, $S_N f$.

Our aim is to detect edges and reconstruct piecewise smooth functions, while regaining the high accuracy encoded in their spectral data. Our fairly general framework for edge detection is based on *concentration kernels*: these are partial

sums of the form

$$K_N^\sigma f(x) := \frac{\pi i}{c_\sigma} \sum_{|k|\le N} sgn(k)\sigma\Big(\frac{|k|}{N}\Big)\widehat{f}(k)e^{ikx}, \qquad c_\sigma := \int_0^1 \frac{\sigma(\xi)}{\xi}d\xi.$$

The $K_N^\sigma f(x)$ approximates the *local* jump function, $K_N^\sigma f(x) \approx f(x+) - f(x-)$. Consequently, $K_N^\sigma f$ tends to concentrate near edges, where $f(x+) - f(x-) \neq 0$, which are separated from smooth regions where $K_N^\sigma f \approx 0$. We can express $K_N^\sigma f(x)$ as a convolution with the Fourier projection of f, that is,

$$K_N^\sigma f(x) = \mathcal{K}_N^\sigma * (S_N f)(x), \qquad \mathcal{K}_N^\sigma(x) := -\frac{1}{c_\sigma}\sum_{k=1}^{N} \sigma\Big(\frac{|k|}{N}\Big)\sin kx.$$

Here, $\mathcal{K}_N^\sigma(x)$ are the corresponding *concentration kernels* which enable us to convert the global moments of $S_N f$ into local information about its edges — both their locations and their amplitudes. The choice of concentration factor, σ, is at our disposal. We discuss a few prototype examples of concentration factors and assess the different behavior of the corresponding edge detectors, $K_N^\sigma f$. We also mention a series of *extensions* which show how concentration kernels apply in more general set-ups. We discuss the *discrete* framework, applying concentration kernels as edge detectors in the Fourier interpolants, $I_N f = \sum_{|k|\le N} \widehat{f}_k e^{ikx}$. We also show how concentration kernels can be used to detect edges in *non-periodic* projections, $S_N f = \sum \widehat{f}(k)C_k(x)$, based on general Gegenbauer expansions. Fianlly, we summarize [2], showing how the concentration factors could be adjusted to deal with *noisy data*, by taking into account the noise variance, $\eta \gg 1/N$, in order to detect the underlying $\mathcal{O}(1)$-edges; see figure 1. We also demonstrate how to deal with *incomplete data*: we decribe the results of [10], which show how concentration kernels based on partial information, $\{\widehat{f}(k)\}_{k\in K}$, can be complemented by a compressed sensing approach to form effective edge detectors.

Concentration kernels, $\mathcal{K}_N^\sigma(x)$, are approximate *derivatives* of the delta function. Convolution with such kernels, $\mathcal{K}_N^\sigma * (S_N f)$ yield edge detectors by *separation of scales*, separating between smooth and nonsmooth parts of f. We can improve the edge detectors by *enhancement* of this separation of scales. In particular, we use nonlinear *limiters* which assign low- and high-order concentration kernels in regions with different characteristics of smoothness. The result is *parameter-free*, high-resolution edge detectors for one-dimensional piecewise smooth functions.

Next we turn to the two-dimensional set-up. Concentration kernels can be used to separate scales in the x_1 and x_2 directions. Enhancements and limiters are shown to greatly reduce, though not completely eliminate, the Cartesian staircasing effect. We show how concentration kernels are used to detect edges from incomplete two-dimensional data. So far, we have emphasized the role of separation of scales in edge detectors based on concentration kernels, $\mathcal{K}_N^\sigma * (S_N f)(\mathbf{x})$. But how do we actually *locate* those $\mathcal{O}(1)$ edges? to this end one seeks the zero-level set $\mathbf{x} = (x_1, x_2)$ of $\nabla_{\mathbf{x}} \mathcal{K}_N^\sigma * (S_N f)(\mathbf{x})$. Depending on our choice of the concentration factors, $\sigma(\cdot)$, this leads to a large class of two-dimensional edge detectors which

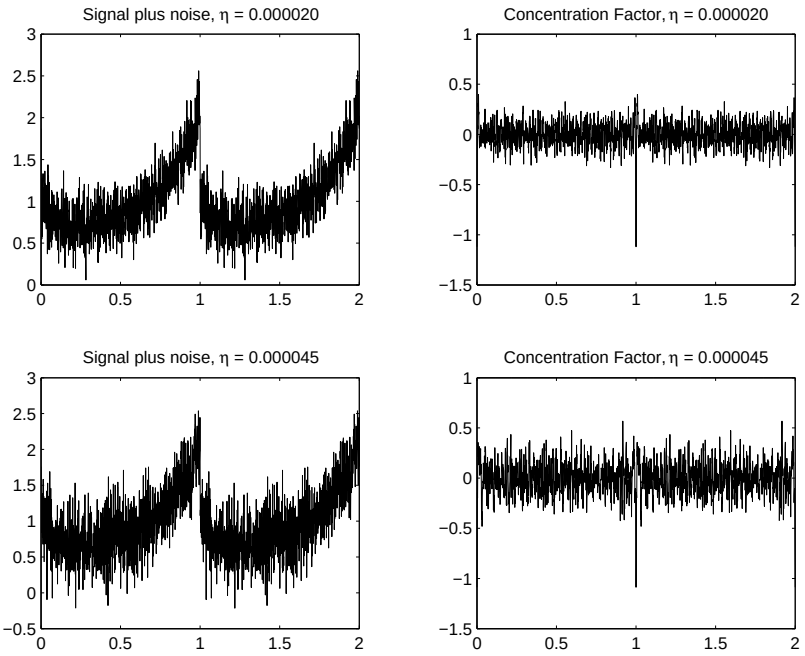


FIGURE 1. Noisy data on the left. The edge was detected from the spectra information using (BV, L^2)-based concentration factors outlined in [2].

generalize the popular two-dimensional zero-crossing method associated with discrete Laplacian stencils.

3. Decomposition of scales

The second part of our talk describes the development of a novel *multiscale* image decomposition introduced with S. Nezzar and L. Vese in [8].

The starting point is variational decomposition of an image, $f = u_0 + v_0$, where $[u_0, v_0]$ is the minimizer of a J-functional,

$$J(f, \lambda_0; X, Y) = \inf_{u+v=f} \Big\{ \|u\|_X + \lambda_0 \|v\|_Y^2 \Big\}.$$

Here, u_0 should capture 'essential features' of f which are to be separated from the spurious components absorbed by v_0, and λ_0 is a *fixed* threshold which dictates separation of scales. Such minimizers are standard tools for image manipulations — denoising, deblurring, compression, ...,e.g., [6, 7, 5].

We focus our attention on the particular case of [7] J-minimizer with $(X, Y) = (BV, L^2)$ decomposition where u_0 and v_0 are sought to separate texture from edges. But in what sense texture is different from an edge? our approach is that the distinction is scale dependent. Texture contains significant edges when "viewed" closer. Accordingly, we suggested to decompose v_0 with a refined scale of , say, $2\lambda_0$, leading to finer decomposition $v_0 = u_1 + v_1$ at scale $2\lambda_0$. To proceed, we

FIGURE 2. Successive decompositions of an image of a woman with $\lambda_0 = .0005$.

iterate the *refinement step*

$$[u_{j+1}, v_{j+1}] = \arg\inf J(v_j, \lambda_0 2^j),$$

leading to the *hierarchical decomposition*, $f = \sum_{j=0}^{k} u_j + v_k$. The resulting hierarchical decomposition, $f \sim \sum_j u_j$, is essentially nonlinear. Numerical results of synthetic and real images in [8], [1] illustrate the superiority of the hierarchical decomposition. An example is illustrated in figure 2 below.

References

[1] A. Baudes, B. Coll and J.-M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul. **4** (2005), no. 2, 490–530.

[2] S. Engelberg and E. Tadmor, *Recovery of edges from spectral data with noise—a new perspective*, Preprint.

[3] A. Gelb and E. Tadmor *Detection of Edges in Spectral Data II. Nonlinear Enhancement* SIAM Journal on Numerical Analysis **38** (2000) 1389–1408.

[4] A. Gelb and E. Tadmor *Adaptive edge detectors for piecewise smooth data based on the minmod limiter* Journal of Scientific Computing **23**(2-3) (2006) 279–306.

[5] Y. Meyer, Oscillating Patterms in Image Processing and Nonlinear Evolution Equations, University Lecture Series, AMS **22** (2002).

[6] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, IEEE, Los Alamitos, CA, 1985, pp. 22–26.

[7] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, **60** (1992) 259–268.

[8] E. Tadmor, S. Nezzar & L. Vese, *A multiscale image representation using hierarchical (BV,L2) decompositions* ., Multiscale Modeling & Simulation **2** (2004) 554–579.

[9] E. Tadmor, *Filters, mollifiers and the computation of the Gibbs phenomenon Computing certain invariants of topological spaces of dimension three*, Acta Numerica **16** (2007) 305–378.

[10] E. Tadmor and J. Zou, *Novel edge detection methods for incomplete and noisy spectral data*, Preprint.

Relaxation in Greedy Approximation

Vladimir Temlyakov

Abstract. We study greedy algorithms in a Banach space from the point of view of convergence and rate of convergence. There are two well studied approximation methods: the Weak Chebyshev Greedy Algorithm (WCGA) and the Weak Relaxed Greedy Algorithm (WRGA). The WRGA is simpler than the WCGA in the sense of computational complexity. However, the WRGA has limited applicability. It converges only for elements of the closure of the convex hull of a dictionary. In this paper we study algorithms that combine good features of both algorithms the WRGA and the WCGA. In construction of such algorithms we use different forms of relaxation. First results on such algorithms have been obtained in a Hilbert space by A. Barron, A. Cohen, W. Dahmen, and R. DeVore. Their paper was a motivation for the research reported here.

Main results. Let X be a Banach space with norm $\|\cdot\|$. We say that a set of elements (functions) $\mathcal{D}$ from X is a dictionary (symmetric dictionary) if each $g \in \mathcal{D}$ has norm bounded by one ($\|g\| \le 1$),

$$g \in \mathcal{D} \quad \text{implies} \quad -g \in \mathcal{D},$$

and the closure of the span of $\mathcal{D}$ coincides with X. We denote the closure (in X) of the convex hull of $\mathcal{D}$ by $A_1(\mathcal{D})$. We introduce a new norm, associated with a dictionary $\mathcal{D}$, in the dual space X' by the formula

$$\|F\|_{\mathcal{D}} := \sup_{g\in\mathcal{D}} F(g), \quad F \in X'.$$

We discuss greedy algorithms with regard to $\mathcal{D}$. For a nonzero element $f \in X$ we denote by F_f a norming (peak) functional for f:

$$\|F_f\| = 1, \qquad F_f(f) = \|f\|.$$

The existence of such a functional is guaranteed by Hahn-Banach theorem. Let $\tau := \{t_k\}_{k=1}^{\infty}$ be a given sequence of nonnegative numbers $t_k \le 1$, $k = 1, \dots$. We define first the Weak Chebyshev Greedy Algorithm (WCGA) (see [4]) that is a generalization for Banach spaces of Weak Orthogonal Greedy Algorithm defined and studied in [3] (see also [2] for Orthogonal Greedy Algorithm).

Weak Chebyshev Greedy Algorithm (WCGA). We define $f_0^c := f_0^{c,\tau} := f$. Then for each $m \ge 1$ we inductively define

1). $\varphi_m^c := \varphi_m^{c,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^c}(\varphi_m^c) \ge t_m \|F_{f_{m-1}^c}\|_{\mathcal{D}}.$$

2). Define

$$\Phi_m := \Phi_m^{\tau} := \operatorname{span}\{\varphi_j^c\}_{j=1}^m,$$

and define $G_m^c := G_m^{c,\tau}$ to be the best approximant to f from Φ_m.

3). Denote

$$f_m^c := f_m^{c,\tau} := f - G_m^c.$$

We define now the generalization for Banach spaces of the Weak Relaxed Greedy Algorithm studied in [4] (see [3] for the case of a Hilbert space).

Weak Relaxed Greedy Algorithm (WRGA). We define $f_0^r := f_0^{r,\tau} := f$ and $G_0^r := G_0^{r,\tau} := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m^r := \varphi_m^{r,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^r}(\varphi_m^r - G_{m-1}^r) \geq t_m \sup_{g\in\mathcal{D}} F_{f_{m-1}^r}(g - G_{m-1}^r).$$

2). Find $0 \leq \lambda_m \leq 1$ such that

$$\|f - ((1-\lambda_m)G_{m-1}^r + \lambda_m\varphi_m^r)\| = \inf_{0\leq\lambda\leq 1} \|f - ((1-\lambda)G_{m-1}^r + \lambda\varphi_m^r)\|$$

and define

$$G_m^r := G_m^{r,\tau} := (1-\lambda_m)G_{m-1}^r + \lambda_m\varphi_m^r.$$

3). Denote

$$f_m^r := f_m^{r,\tau} := f - G_m^r.$$

Remark 1. *It follows from the definition of WCGA and WRGA that the sequences* $\{\|f_m^c\|\}$ *and* $\{\|f_m^r\|\}$ *are nonincreasing sequences.*

Both of the above algorithms use the functional $F_{f_{m-1}}$ in a search for the mth element φ_m from the dictionary to be used in approximation. The construction of the approximant in the WRGA is different from the construction in the WCGA. In the WCGA we build the approximant G_m^c in a way to maximally use the approximation power of the elements $\varphi_1, \dots, \varphi_m$. The WRGA by its definition is designed for approximation of functions from $A_1(\mathcal{D})$. In building the approximant in the WRGA we keep the property $G_m^r \in A_1(\mathcal{D})$. We call the WRGA *relaxed* because at the mth step of the algorithm we use a linear combination (convex combination) of the previous approximant G_{m-1}^r and a new element φ_m^r. The relaxation parameter λ_m in the WRGA is chosen at the mth step depending on f. Recently, the following modification of the above idea of relaxation in greedy approximation has been studied in [1]. Let a sequence $r := \{r_k\}_{k=1}^\infty$, $r_k \in [0,1)$, of relaxation parameters be given. Then at each step of our new algorithm we build the mth approximant of the form $G_m = (1-r_m)G_{m-1} + \lambda\varphi_m$. With an approximant of this form we are not limited to approximation of functions from $A_1(\mathcal{D})$ as in the WRGA. Remarkable results on the approximation properties of such an algorithm in a Hilbert space have been obtained in [1]. We will discuss here a realization of the above new idea of relaxation in the case of Banach spaces. We study the Greedy Algorithm with Weakness parameter t and Relaxation r (GAWR(t,r)). We give a general definition of the algorithm in the case of a weakness sequence τ.

GAWR(τ, r). Let $\tau := \{t_m\}_{m=1}^\infty$, $t_m \in [0,1]$, be a weakness sequence. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \geq t_m \|F_{f_{m-1}}\|_{\mathcal{D}}.$$

2). Find $\lambda_m \geq 0$ such that

$$\|f - ((1-r_m)G_{m-1} + \lambda_m\varphi_m)\| = \inf_{\lambda\geq 0} \|f - ((1-r_m)G_{m-1} + \lambda\varphi_m)\|$$

and define

$$G_m := (1 - r_m)G_{m-1} + \lambda_m \varphi_m.$$

3). Denote

$$f_m := f - G_m.$$

In the case $\tau = \{t\}$, $t \in (0, 1]$, we write t instead of τ in the notation. We note that in the case $r_k = 0$, $k = 1, \dots$, when there is no relaxation, the GAWR$(\tau, 0)$ coincides with the Weak Dual Greedy Algorithm [5], p.66. We will also consider here a relaxation of the X-greedy algorithm (see [5], p.39) that corresponds to $r = 0$ in the definition that follows.

X-Greedy Algorithm with Relaxation r (XGAR(r)). We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ and $\lambda_m \geq 0$ are such that

$$\|f - ((1 - r_m)G_{m-1} + \lambda_m \varphi_m)\| = \inf_{g \in \mathcal{D}, \lambda \geq 0} \|f - ((1 - r_m)G_{m-1} + \lambda g)\|$$

and

$$G_m := (1 - r_m)G_{m-1} + \lambda_m \varphi_m.$$

2). Denote

$$f_m := f - G_m.$$

We note that, practically, nothing is known about convergence and rate of convergence of the X-greedy algorithm. It will be seen from the results that follow that relaxation helps to prove convergence results for the XGAR(r).

The following version of relaxed greedy algorithm has been studied in [6].

Weak Greedy Algorithm with Free Relaxation (WGAFR). Let $\tau := \{t_m\}_{m=1}^{\infty}$, $t_m \in [0, 1]$, be a weakness sequence. We define $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \geq t_m \|F_{f_{m-1}}\|_{\mathcal{D}}.$$

2). Find w_m and $\lambda_m \geq 0$ such that

$$\|f - ((1 - w_m)G_{m-1} + \lambda_m \varphi_m)\| = \inf_{\lambda \geq 0, w} \|f - ((1 - w)G_{m-1} + \lambda \varphi_m)\|$$

and define

$$G_m := (1 - w_m)G_{m-1} + \lambda_m \varphi_m.$$

3). Denote

$$f_m := f - G_m.$$

We consider here approximation in uniformly smooth Banach spaces. For a Banach space X we define the modulus of smoothness

$$\rho(u) := \sup_{\|x\| = \|y\| = 1} \left(\frac{1}{2}(\|x + uy\| + \|x - uy\|) - 1\right).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u \to 0} \rho(u)/u = 0.$$

We formulate two typical results of the paper [6].

Theorem 1. *Let a sequence $r := \{r_k\}_{k=1}^{\infty}$, $r_k \in [0,1)$, satisfy the conditions*

$$\sum_{k=1}^{\infty} r_k = \infty, \quad r_k \to 0 \quad \text{as} \quad k \to \infty.$$

Then the GAWR(t, r) and the XGAR(r) converge in any uniformly smooth Banach space for each $f \in X$ and for all dictionaries $\mathcal{D}$.

Theorem 2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Take a number $\epsilon \ge 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \le \epsilon, \quad f^\epsilon / A(\epsilon) \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$. Then we have for the WGAFR

$$\|f_m\| \le \max\left(2\epsilon, C(q,\gamma)(A(\epsilon)+\epsilon)(1+\sum_{k=1}^{m} t_k^p)^{-1/p}\right), \quad p := q/(q-1).$$

The setting in Theorem 2 with two functions f and f^ϵ covers the following noisy data setting. Let $A(\epsilon) = A$ and the target function f^ϵ is such that $f^\epsilon / A \in A_1(\mathcal{D})$. The task is to approximate f^ϵ from the noisy data f of it.

References

[1] A. Barron, A. Cohen, W. Dahmen, and R. DeVore, *Approximation and learning by greedy algorithms*, Manuscript (2005), 1–27.

[2] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in Computational Mathematics **5** (1996), 173–187.

[3] V.N. Temlyakov, *Weak Greedy Algorithms*, Advances in Computational Mathematics **12** (2000), 213-227.

[4] V.N. Temlyakov, *Greedy Algorithms in Banach spaces*, Advances in Computational Mathematics **14** (2001), 277-292.

[5] V.N. Temlyakov, *Nonlinear Methods of Approximation*, Foundations of Computational Mathematics **3** (2003), 33–107.

[6] V.N. Temlyakov, *Relaxation in Greedy Approximation*, Constructive Approximation (to appear).

Can We Use Semiorthogonal Spline Wavelets for Adaptively Solving Nonlinear Problems?

Karsten Urban

(joint work with Kai Bittner)

We consider the nonlinear boundary value problem

$$-\Delta u = \mathcal{F}u \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega,$$

with a given (nonlinear) operator $\mathcal{F} : H_0^1(\Omega) \to H^{-1}(\Omega)$. Its variational formulation amounts finding $u \in H_0^1(\Omega)$ such that

$$a(u,v) = \langle \mathcal{F}u, v\rangle \quad \text{for all } v \in H_0^1(\Omega),$$

where the bilinear form is given by $a(u,v) := (\nabla u, \nabla v)_{L_2}$ and $\langle \cdot, \cdot \rangle$ denotes the standard duality pairing. Using a Galerkin method based upon a wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\}$ of $H_0^1(\Omega)$ requires the computation (or at least, the approximation) of terms like $a(u^\varepsilon, \psi_\mu)$ and $\langle \mathcal{F}u^\varepsilon, \psi_\mu \rangle$ for a given (finite) approximation u^ε.

It is known, that a sufficiently good finite approximation of $\big(a(u^\varepsilon, \psi_\mu)\big)_{\mu \in \mathcal{J}}$ can be computed with linear cost, if the entries $a(\psi_\lambda, \psi_\mu)$ of the stiffness matrix can be computed at unit cost. This is the case if the wavelets ψ_λ are spline functions of compact support (cf. [8]).

This situation changes for the nonlinear terms $\langle \mathcal{F}u^\varepsilon, \psi_\mu \rangle$ as can easily be seen by the simple example $\mathcal{F}u := u^2$. For a finite input

$$u^\varepsilon = \sum_{\lambda \in \Lambda} u_\lambda^\varepsilon \ \psi_\lambda,$$

we obtain

$$\langle \mathcal{F}u^\varepsilon, \psi_\mu \rangle = \sum_{\lambda \in \Lambda} \sum_{\nu \in \Lambda} u_\lambda^\varepsilon u_\nu^\varepsilon \langle \psi_\lambda \psi_\nu, \psi_\mu \rangle,$$

which requires $\mathcal{O}\big((\#\Lambda)^2\big)$-operations (thus non-optimal) if one computes these values in a straightforward way. Moreover, the terms $\langle \psi_\lambda \psi_\nu, \psi_\mu \rangle$ are integrals of wavelets on fairly different levels. Such integrals can be computed with an amount of work that is linear in the level difference of the involved functions (see e.g. [3]), which again is non-optimal. Hence, one is interested in finding a numerical scheme that is able to approximate $\langle \mathcal{F}u^\varepsilon, \psi_\mu \rangle$ up to any desired accuracy with (asymptotically) optimal complexity.

In [12] (and later improvements in [1, 9]) an optimal adaptive wavelet evaluation scheme for nonlinear functions has been introduced and analyzed. All these schemes are based upon *biorthogonal* wavelet bases which offer the advantage that decomposition and reconstruction are based upon finite masks since primal and dual functions are compactly supported. On the other hand, it is known that the conditioning of these bases causes some quantitative problems for their numerical use. We display the condition numbers in the right part of Table 2. Note that these condition numbers enter the constants in basically all estimates for the corresponding numerical scheme. This is the reason why several improved constructions of biorthogonal wavelet bases on the interval have been introduced, see e.g. [2, 4, 13, 14]. Another alternative is to use semiorthogonal spline wavelets, sometimes also called prewavelets. They offer some advantages, namely:

- Scaling functions and wavelets are splines so that all fast spline algorithms can be used e.g. for point evaluation.
- The wavelets are levelwise orthogonal which gives rise to quantitative good stability properties, see the left part of Table 2.
- The support length of semiorthogonal wavelets is shorter then the corresponding biorthogonal one of the same order.

However, there are also two main drawbacks, namely:

- The dual functions are in the same space. For the variational formulations, one would like the duals to characterize the dual space $H^{-1}(\Omega)$. The duals

	semiorth.				$\rho^{\Omega}_{\Psi_j}$ biorth.		
m	$\rho^{\mathbb{R}}_{\Phi}$	$\rho^{\mathbb{R}}_{\Psi}$	$\rho^{[0,1]}_{\Phi_j}$	$\rho^{[0,1]}_{\Psi_j}$	$d=\tilde{d}$, $\mathbb{R}$	$\tilde{d}=d+10$, $\mathbb{R}$	$[0,1]$
2	3.0	2.3	3.0	2.3	10	4.2	4.1
3	7.5	3.5	7.6	3.5	80	16.0	16.0
4	18.5	5.9	19.3	6.0	∞	64.0	64.1
5	45.7	10.4	49.8	11.7	∞	256.2	264
6	112.8	18.7	130.4	21.2	∞	≥ 1024	≥ 1024
7	278.4	33.9	345.0	50.3	∞	≥ 4096	≥ 4096
8	686.9	61.6	920.7	158.4	∞	≥ 16384	≥ 16384

TABLE 2. Condition numbers for scaling functions and wavelets from [7] ($\rho^{\mathbb{R}}_{\Phi}$ and $\rho^{\mathbb{R}}_{\Psi}$) for $L_2(\mathbb{R})$, and from [6] ($\rho^{[0,1]}_{\Phi_j}$ and $\rho^{[0,1]}_{\Psi_j}$) for $L_2([0,1])$ and $j \leq 11$ in dependence of the spline order d. For comparison, we also display corresponding numbers for biorthogonal spline wavelets on $\mathbb{R}$ from [10] and on $[0,1]$ from [4].

being in the same space, this is not possible. However, this does not affect approximation and stability properties of the adaptive method, hence we do not view this as a serious problem.

- The dual functions are *globally supported* which in particular means that the decomposition is based upon an infinite mask. This is a crucial drawback.

In this talk, we answer the question if we can circumvent the problem of the global duals in order to construct an efficient adaptive semiorthogonal wavelet scheme for evaluating nonlinear functions. As a core algorithm we have used the method in [12] which may be summarized as follows:

Algorithm 1 (Dahmen, Schneider, Xu: 2000)**.** Input: u^ε, Λ, ε

1. *Prediction*: Predict the significant indices $\hat{\Lambda} \subset \mathcal{J}$;
2. *Reconstruction*: Determine a 'local scaling function representation' of u^ε (permits a fast computation of function values);
3. *Quasi-Interpolation*: Based on prediction, compute a quasi-interpolant $g = A(\mathcal{F}u^\varepsilon)$;
4. *Decomposition*: Compute the significant wavelet coefficients d_λ, of g.

Let us briefly describe the steps.

The *prediction* is typically based upon a subdivision of the domains Ω with respect to so called support cubes that are related to each wavelet index λ. The second ingredient is a computable error function. Then, a wavelet ψ_λ is rated significant if the error functional exceeds a certain value related to the desired accuracy. Since the error functional is linked to the quasi-interpolant, this will have to be adapted for semiorthogonal spline wavelets.

In the *reconstruction*, the input in terms of the wavelet expansion is transformed to a sparse single scale representation over all levels. This simply requires the refinement equation which in turns is based upon the mask coefficients of the

d	$\gamma_{d,\ell}$	C_Q
1	1	1
2	1	1
3	$-\frac{1}{8}$, $\frac{5}{4}$, $-\frac{1}{8}$	$\frac{3}{2}$
4	$-\frac{1}{6}$, $\frac{4}{3}$, $-\frac{1}{6}$	$\frac{5}{3}$
5	$\frac{47}{1152}$, $-\frac{107}{288}$, $\frac{319}{192}$, $-\frac{107}{288}$, $\frac{47}{1152}$	$\frac{179}{72}$
6	$\frac{13}{240}$, $-\frac{7}{15}$, $\frac{73}{40}$, $-\frac{7}{15}$, $\frac{13}{240}$	$\frac{43}{15}$

TABLE 3. Coefficients for the quasi-interpolant.

primal functions. These are compactly supported semiorthogonal spline wavelets here and thus no changes are needed.

The *quasi-interpolation* gives typically a linear combination of dual scaling functions hence, we need an alternative here. We have used an operator introduced by Zheludev [15] defined by

$$q(f) := \sum_{\ell=0}^{\lceil \frac{d-1}{2} \rceil} (-1)^\ell \beta_{d,\ell}\, \tilde{\Delta}_1^{2\ell} f\Big(\frac{d}{2}\Big)$$

with $\beta_{d,\ell}$ defined by

$$\left(\frac{2\arcsin\frac{t}{2}}{t}\right)^d = \sum_{\ell=0}^{\infty} \beta_{d,\ell} t^{2\ell},$$

and $\tilde{\Delta}_h^k$ is the symmetric difference defined in the usual way $\Delta_h^1 f(x) := f(x + \frac{h}{2}) - f(x - \frac{h}{2})$ and $\tilde{\Delta}_h^{k+1} := \Delta_h^1 \tilde{\Delta}_h^k$. The above representation is not well suited for numerical application, but there is an alternative representation as

$$q(f) = \sum_{\ell=1}^{d-1} \gamma_{d,\ell}\, f\Big(\ell - \mu(d)\frac{1}{2}\Big), \qquad \mu(d) := d \bmod 2,$$

where the coefficients $\gamma_{d,\ell}$ are displayed in Table 3 The constant C_Q is the constant for the corresponding error estimate. We extend this operator to an adaptive version and present its error analysis. This also gives rise to a new computable error functional which is then used for the prediction.

Finally, the *decomposition* is based upon the dual mask coefficients which are infinitely many. However, since primals and duals span the *same* space, we can perform an efficient change of bases from primals to duals by using the mass matrix of the primals which is sparse.

Hence, the answer to the question raised in the title of the talk is *yes*! This part of the talk is based upon the paper [5].

Finally, we show the use of the new adaptive quasi-interpolant in an industrial application, namely the analysis and correction of optical surfaces produced by the company Zeiss, Oberkochen (Germany). Since here particular importance is

in the real-time application, we also show the main parts of the used software tools FLENS and LAWA, http://flens.sourceforge.net and http://lawa.sourceforge.net.

References

[1] A. Barinka, W. Dahmen, and R. Schneider. Fast computation of adaptive wavelet expansions. *Numer. Math.*, 105(4):549–589, 2007.

[2] T. Barsch. *Adaptive Multiskalenverfahren für elliptische partielle Differentialgleichungen — Realisierung, Umsetzung und numerische Ergebnisse (in german)*. PhD thesis, RWTH Aachen, Shaker-Verlag, Aachen, 2001.

[3] S. Bertoluzza, C. Canuto, and K. Urban. On the adaptive computation of integrals of wavelets. *Appl. Numer. Math.*, 34(1):13–38, 2000.

[4] K. Bittner. Biorthogonal spline wavelets on the interval. In G. Chen and M.-J. Lai, editors, *Wavelets and Splines: Athens 2005*, pages 93–104. Nashboro Press, Brentwood, TN, 2006.

[5] K. Bittner and K. Urban Adaptive Wavelet Methods using Semiorthogonal Spline Wavelets: Sparse Evaluation of Nonlinear Functions Preprint, Univ. of Ulm 2005, to appear in Appl. Comp. Harm. Anal.

[6] C. K. Chui and E. Quak. Wavelets on a bounded interval. In D. Braess and L. L. Schumaker, editors, *Numerical Methods in Approximation Theory*, volume 9, pages 53–75. Birkhäuser, Basel, 1992.

[7] C. K. Chui and J. Wang. On compactly supported spline wavelets and a duality principle. *Trans. Amer. Math. Soc.*, 330:903–915, 1992.

[8] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet schemes for elliptic operator equations — convergence rates. *Math. Comput.*, 70(233):27–75, 2001.

[9] A. Cohen, W. Dahmen, and R. Devore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41(5):1785–1823 (electronic), 2003.

[10] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 45:485–560, 1992.

[11] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline-wavelets on the interval — stability and moment conditions. *Appl. Comp. Harm. Anal.*, 6:132–196, 1999.

[12] W. Dahmen, R. Schneider, and Y. Xu. Nonlinear functionals of wavelet expansions — adaptive reconstruction and fast evaluation. *Numer. Math.*, 86:49–101, 2000.

[13] S. Grivet-Talocia and A. Tabacco. Wavelets on the interval with optimal localization. *Math. Models Methods Appl. Sci.*, 10(3):441–462, 2000.

[14] A. Kunoth and J. Sahner. Wavelets on manifolds: an optimized construction. *Math. Comp.*, 75(255):1319–1349 (electronic), 2006.

[15] V. A. Zheludev. Local quasi-interpolatory splines and Fourier transforms. *Dokl. Akad. Nauk SSSR*, 282:1293–1298, 1985. English transl. in *Soviet Math. Dokl.*, 31:573–577, 1985.

Tresholding and Compressed Sensing

Przemysław Wojtaszczyk

The aim of this talk is to present some simple observations about compressed sensing. The main problem in compresssed sensing is to find k–sparse (i.e. with at most k non zero coefficients) vector $\mathbf{x} \in \mathbb{R}^N$ when we make d nonadaptive linear measurements. We think of N as very large. So formally

$$\mathbf{\Phi} = (\Phi_j)_{j=1}^N \text{ with } \Phi_j \in \mathbb{R}^d \tag{1}$$

and $\mathbf{\Phi}(\mathbf{x}) = \sum_{j=1}^{N} x_j \Phi_j \in \mathbb{R}^d$. We look for (non-linear) $\Delta : \mathbb{R}^d \to \mathbb{R}^N$ such that $\Delta\mathbf{\Phi}(\mathbf{x}) = \mathbf{x}$ for every k–sparse vector $\mathbf{x} \in \mathbb{R}^N$. Usually $\mathbf{\Phi} = \mathbf{\Phi}(\omega)$ is a random matrix.

Example I will concentrate on is a properly normalised Bernoulli measurement matrix i.e. $\mathbf{\Phi}(\omega) = \left(d^{-1/2}\epsilon_j^i(\omega)\right)_{j,i}$ with $j = 1, \ldots, N$ and $i = 1, \ldots, d$ and ϵ_j^i are independent Bernoulli random variables. I will concentrate on this example, but other examples also work.

A possible Δ is ℓ_1 minimization i.e.

$$\Delta_1(y) = \text{Argmin}\{\|z\|_1 \ : \ \mathbf{\Phi}(z) = y\} \tag{2}$$

where $\|\cdot\|_1$ is the ℓ_1 norm of the vector. The above scheme was put forward recently by D. Donoho, E. Candes, T. Tao in a series of papers (see e.g. [1] [2]) They isolated and studied conditions k-**RIP** (see also [5], [6]) : there is a $0 < \delta < 1$ such that for every set $A \subset \{1, \ldots, N\}$ of k elements and all possible scalars we have

$$(1-\delta)\Big(\sum_{j\in A} |a_j|^2\Big)^{1/2} \le \|\sum_{j\in A} a_j \Phi_j\| \le (1+\delta)\Big(\sum_{j\in A} |a_j|^2\Big)^{1/2} \tag{3}$$

They have shown the importance of this conditions to compressed sensing by showing the following two results:

THEOREM 1 *If deterministic* $\mathbf{\Phi}$ *satisfies (3) then for every* k*–sparse vector* $\mathbf{x} \in \mathbb{R}^N$ *we have* $\mathbf{x} = \Delta_1\mathbf{\Phi}(\mathbf{x})$.

THEOREM 2 *Except for a set of* very small *probability* $\mathbf{\Phi}(\omega)$ *satisfies (3)provided* $d \ge Ck \ln N$.

Unfortunatelly the map Δ_1 is computationally costly.

Recently, A. Gilbert, J. Tropp [4] suggested to use OMP. Given dictionary $\{\Phi_j\}_{j=1}^N \subset \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^d$ we define

(1) $r_0 = \mathbf{y}$
(2) $j_1 = \text{Argmax}|\langle r_0, \Phi_j\rangle|$ and $r_1 = \mathbf{y} - P_1(\mathbf{y})$ with P_1 is the orthogonal projection onto $\text{span}\{\Phi_{j_1}\}$
(3) next inductively $j_{s+1} = \text{Argmax}|\langle r_s, \Phi_j\rangle|$ and $r_{s+1} = \mathbf{y} - P_{s+1}(\mathbf{y})$ where P_{s+1} is orthogonal projection onto $\text{span}\{\Phi_{j_1}, \Phi_{j_2} \ldots \Phi_{j_{s+1}}\}$

When $\mathbf{y} = \mathbf{\Phi}(\mathbf{x})$ then we define

$$\Delta_{OMP}^s(\mathbf{y}) = \sum_{\mu=1}^{s} a_\mu e_{j_\mu} \in \mathbb{R}^N$$

where $P_s(\mathbf{y}) = \sum_{\mu=1}^{s} a_\mu \Phi_{j_\mu}$

THEOREM 3 (Gilbert–Tropp) *If* $\mathbf{\Phi}(\omega)$ *is random and* $d \ge ck \log N$ *then for each* k*–sparse vector* $\mathbf{x} \in \mathbb{R}^N$ *there exists a set of* ω*'s of probability* very close to 1 *such that for those* ω*'s when we run OMP for* $\mathbf{y}(\omega) = \mathbf{\Phi}(\mathbf{x}) = \sum_{j=1}^{N} x_j \Phi_j$ *we get* $r_k = 0$ *i.e.* $\Delta_{OMP}^k(\mathbf{y}) = \mathbf{x}$.

For $x \in \mathbb{R}^N$ and $\mathbf{y}(\omega) = \mathbf{\Phi}(\omega)(\mathbf{x}) \in \mathbb{R}^d$ we have

$$\begin{aligned}\mathbb{P}\{|\langle \mathbf{y}(\omega), \Phi_j(\omega)\rangle - x_j| &\leq \frac{a\|\mathbf{y}(\omega)\|}{\sqrt{k}} \text{ for } j = 1, \dots, N\} \\ &\geq 1 - N^{-b}\end{aligned}$$

provided $d \geq C(a,b)k \log N$. This shows that big coefficients can be recovered with big probability by simple tresholding. In particular we have

COROLLARY 1 *If $\mathbf{x} \in \mathbb{R}^N$ is k-sparse with $|x_j| = 0$ or $|x_j| = 1$ and $d \geq Ck \log N$ and $\mathbf{y} = \mathbf{\Phi}(\mathbf{x})$, then tresholding recovers $\mathbf{x}$ with big probability i.e.*

(1) *we find the set A of coordinates such that $|\langle \mathbf{y}, \Phi_j\rangle| \geq \frac{\|\mathbf{y}\|}{2\sqrt{k}}$*
(2) *we write $\mathbf{y} = \sum_{\mu \in A} a_\mu \Phi_\mu$*
(3) *we get $\mathbf{x} = \sum_{\mu \in A} a_\mu e_\mu \in \mathbb{R}^N$*

This suggests the following modification of OMP which we will call Tresholding OMP, TOMP for short.

Fix k. For a set of indices $A_s \subset \{1, 2, \dots, N\}$ we define by $P_{A_s} = P_s$ the orthogonal projection onto $\{\Phi_j\}_{j \in A_s}$. If $A_s = \emptyset$ then $P_s = 0$. We have $\mathbf{x} = \sum_{j=1}^N x_j e_j$. As usual $\mathbf{y} = \mathbf{\Phi}(\mathbf{x})$. We start with $A_0 = \emptyset$, $\mathbf{y}^0 = P_0(\mathbf{y}) = 0$ and $r^0 = \mathbf{y} - \mathbf{y}^0$. Given A_s, r^s and y^s we put

$$B_{s+1} = \{j \ : \ |\langle r^s, \Phi_j\rangle| \geq \frac{1}{2\sqrt{k}}\|r^s\|\}. \tag{4}$$

and define $A_{s+1} = A_s \cup B_{s+1}$, $\mathbf{y}^{s+1} = P_{s+1}(\mathbf{y})$ and $r^{s+1} = \mathbf{y} - \mathbf{y}^{s+1}$. For this algorithm we can easily modify the argument of Gilbert and Tropp [4] and prove

THEOREM 4 *If $\mathbf{\Phi}(\omega)$ is random and $d \geq ck \log N$ then for each k–sparse vector $\mathbf{x} \in \mathbb{R}^N$ there exists a set of ω's of probability* very close to 1 *such that for those ω's when we run TOMP for $\mathbf{y}(\omega) = \mathbf{\Phi}(\mathbf{x}) = \sum_{j=1}^N x_j \Phi_j$ we get $r^s = 0$ for some $s \leq k$ i.e. we recover $\mathbf{x}$.*

An algorithm very similar to TOMP was used by Donoho-Tsiag-Drori-Starck [3] for somewhat different purposes.

We consider the case when we measure $\mathbf{y}$ with some error, so instead of $\mathbf{y}$ we see $V = \mathbf{\Phi}(\mathbf{x}) + b$ with a *fixed* vector $b \in \mathbb{R}^d$. The argument from [4] easily extends to give

THEOREM 5 *If $\mathbf{\Phi}(\omega)$ is random, $\mathbf{x}$ is k–sparse and $d \geq ck \log N$ then there exists a set of ω's of probability* very close to 1 *such that for those ω's when we run TOMP for $V(\omega)$ we get k-sparse $\tilde{\mathbf{x}}$ with $\|\mathbf{x} - \tilde{\mathbf{x}}\| \leq 2\|b\|$.*

Remark Clearly the same if b is random, independent of $\mathbf{\Phi}$ with uniformly bounded norm. The same holds for OMP.

REFERENCES

[1] E. Candes and T. Tao, *Decoding by linear programing*, IEEE trans. Information Theory **51**, 4203-4215, 2005
[2] D.L. Donoho, *Compressed sensing*, IEEE Trans. Information Theory **52**, 1289-1306, 2006
[3] A.L Donoho, Y. Tsiag, I. Drori and J-L Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit* (to appear)

[4] A.C. Gilbert and J.A. Tropp, *Signal recovery from random measurements via orthogonal matching persuit* (to appear)
[5] S. Mendelson, A. Pajor and N. Tomczak-Jaegermann, *Uniform uncertainty principle for Bernoulli and subgaussian ensambles* (to appear)
[6] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*,Communications in Pure and Applied Math. (to appear)

Consistent Piecewise Polynomial Approximation of $H^m(\Omega)$ Functions in $\mathbf{R}^n$ for any $0 \le m \le n$

Jinchao Xu
(joint work with Ming Wang)

This talk addresses the question of constructing piecewise polynomial approximation spaces (that have locally supported bases) for the the Sobolev space $H = H^m(\Omega)$, $\Omega \subset R^n$ for any n and m. A universal construction will be given for the case that $0 \le m \le n$.

Internal approximations. Let W_h be a class of approximation space of H^m. W_h is said to be an internal approximation of H^m if $W_h \subset H^m$ and, otherwise, an external approximation.

Well-known internal approximation spaces include smooth splines, wavelets (e.g. of Daubechies) and conforming finite elements (FE). Construction of such spaces are well-known for the case of $n = 1, m \ge 0$. For $n > 1, m \ge 0$, such constructions are also easy for special domains $\Omega = \Pi_{i=1}^n (a_i, b_i)$ using tensor product of spaces for $n = 1$. For $n > 1$ and general polyhedral domain with simplicial partition, it is complicated and very high degree of polynomials are required when $m \ge 2$. When $n = 2$ and $m = 2$, 5th order polynomial in each triangle (Argyris element) is needed. In more general cases, one often needs polynomial of degree $(m-1)2^n + 1$ (see [4] and references cited therein) to construct spline and finite element subspaces of H^m. Such a high degree makes these spaces very complicated and difficult to use in practice. Lower order constructions such as super-splines (difficult to get locally supported basis) and composite elements (see [1, 5, 6]), are also possible but complicated.

External approximation and nonconforming finite element spaces. For the external approximation such as nonconforming finite elements, the approximation space W_h is not in general a subspace of $H^m(\Omega)$: $W_h \not\subset H^m(\Omega)$. Given $\mathcal{T}_h = \{T\}$, a partition of Ω, consider the inner product and norm:

$$(u, v)_{m,h} = \sum_{|\alpha| \le m} (\partial_h^\alpha u, \partial_h^\alpha v), \quad \|v\|_{m,h} = (v, v)_{m,h}^{1/2},$$

where ∂_h^α are the piecewise derivatives: $\partial_h^\alpha v_h(x) = \partial^\alpha v_h(x), x \in T$.

The completely discontinuous approximation spaces are trivial to construct, they are hardly useful because of lack of smoothness consistency. The so-called discontinuous Galerkin (DG) methods do use completely discontinuous polynomials, but they need to use many penalty terms to enforce smoothness approximately (see [3]) to achieve consistency. A more desirable external approximation space

W_h should impose some minimal smoothness requirement in the spaces so that the limiting space exactly recovers the original target space $H^m(\Omega)$, and no additional penalty parameters are needed to enforce the smoothness when applied to, e.g., numerical solution of partial differential equations.

An appropriate external approximation space may be motivated by a pure approximation theoretical point of view, namely, the limiting space of $\{W_h\}$, as $h \to 0$, should exactly recover $H^m(\Omega)$ in the following sense [2, 7]. First of all, every function in $H^m(\Omega)$ can be approximated by functions in W_h:

$$\lim_{h\to 0} \inf_{v_h \in W_h} \|v - v_h\|_{m,h} = 0, \ \forall v \in H^m(\Omega).$$

Secondly, only the function in $H^m(\Omega)$ can be approximated by functions in W_h: if for any $\{v_{h_k}\} \subset W_{h_k}$ and $h_k \to 0$ as $k \to \infty$, $\partial^\alpha v_{h_k} \rightharpoonup v^\alpha$ for all $|\alpha| \le m$ (weakly in $L^2(\Omega)$), then $v^0 \in H^m(\Omega)$ and $v^\alpha = \partial^\alpha v^0$, for all $|\alpha| \le m$. The spaces $\{W_h\}$ satisfying the above two properties will be called *consistent approximation* spaces of H^m.

Another motivation of appropriate external approximation is that $\{W_h\}$ are convergent nonconforming finite element spaces for $2m$-th order elliptic boundary value problems, which, according to [7], turn out to be equivalent that $\{W_h\}$ are consistent approximation spaces of H^m.

In the existing literature, nonconforming finite element spaces have been constructed mostly for $H^2(\Omega)$ with $\Omega \subset R^2$. Among these nonconforming elements, the so-called Morley element appears to be most peculiar because of the following unusual properties: (1) it is a consistent approximation for $H^2(\Omega)$, but it is not always a consistent approximation for $H^1(\Omega)$, and (2) it is not even continuous. One question that motivated our studies is as follows: "*Is the Morley element an isolated incident or does it have a more general implication?*" As it turns out, the Morley element belongs to a family of nonconforming finite elements that can be constructed in a universal and elegant fashion.

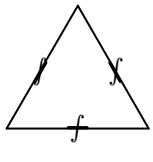
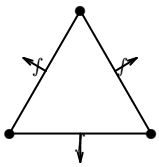
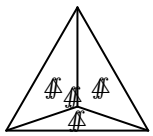
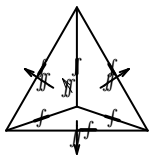
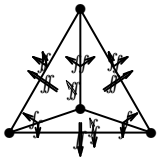
New nonconforming finite element of lowest degree for $H^m(\Omega)$ in R^n ($n \ge m \ge 1$). Define the FE triple (T, P_T, D_T^m) as follows: the geometric shape of the element T is an n-simplex, the shape function space P_T is $P_m(T)$ (the polynomial of total degree m) and the set of the degrees of freedom D_T^m consists of the integral averages of normal derivatives of order $m-k$ on all subsimplexes of dimension $n-k$ for $1 \le k \le m$. The global FE space M_h^m is then the piecewise P_m that is continuous on all d.o.f. (w.r.t. simplicial partition $\mathcal{T}_h$).

The dimension of P_T is equal to the number of degrees of freedom in D_T^m thanks to the well-known Vandermonde combinatorial identity that

$$\sum_{k=1}^{m} C_{n+1}^{n-k+1} C_{m-1}^{m-k} = C_{n+m}^{m}.$$

For $m = 1$ and $n = 1$, we obtain the well-known conforming P_1 elements, the only conforming element in this family of elements. For $m = 1$ and $n \ge 2$, we obtain the well-known nonconforming P_1 elements. For $m = 2$, we recover the

TABLE 1. Some examples (degrees of freedom)

	$m=1$	$m=2$	$m=3$
$n=1$			
$n=2$			
$n=3$			

Morley element for $n=2$ and its generalization to $n \geq 2$ (see [8]). For $m=3$ and $n=3$, we obtain a new cubic element on a simplex that has 20 degrees of freedom.

Lemma 1 (Weak continuity)**.** For any $1 \leq k \leq m$, $|\alpha| = m-k$ and $v_h \in M_h^n$, $\int_F [\partial^\alpha v_h] = 0$ on every internal $(n-k)$-dimensional subsimplex F.

To prove the lemma, we only need to prove that $\int_F \partial^\alpha v_h$ is determined by the d.o.f. on F. For $k=m, \alpha=0$, $\int_F v_h$ is a d.o.f. for $n-m$ subsimplex F. Now for $k=m-1$, we need to prove that $\int_F \nabla v_h$ is determined by the d.o.f for $n-m+1$ subsimplex F. We write $\int_F \nabla v_h = \sum_i \int_F \frac{\partial v_h}{\partial \nu_i}\nu_i + \sum_j \int_F \frac{\partial v_h}{\partial \tau_j}\tau_j$. Here each ν_i is normal to F, and $\int_F \frac{\partial v_h}{\partial \nu_i}$ is already a d.o.f. Each τ_j is tangent to F, and $\int_F \frac{\partial v_h}{\partial \tau_j}$ can be expressed in terms of the d.o.f. on ∂F ($n-m$ subsimplexes).

The proof of the above lemma can then be completed by induction.

Lemma 2 (Unisolvent Property)**.** D_T^m is P_T-unisolvent.

To prove this, let $p \in P_m$ with all d.o.f.'s being zero, we need to show that $p \equiv 0$. For any $|\alpha| = m$, $\partial^\alpha p =$const. Hence by Green's formula

$$\partial^\alpha p = \frac{1}{|T|}\int_T \partial^\alpha p = \frac{1}{|T|}\sum_{|\beta_j|=m-1} c_j \int_{\partial T} \partial^{\beta_j} p = 0.$$

By induction, we can show $\partial^\alpha p = 0$ for all $|\alpha| = k$ with $k = m-1, m-2, \cdots, 0$.

Lemma 3. For any $|\alpha| \leq m$,

$$(\partial_h^\alpha v_h, \phi) = (-1)^{|\alpha|}(v_h, \partial^\alpha \phi) + \mathcal{O}(h)\|v_h\|_{m,h}\|\phi\|_m, \quad \phi \in C_0^\infty(\Omega), v_h \in M_h^m.$$

Again, the above lemma can be proved by induction. $\alpha = 0$ is trivial. Assume true for $|\alpha| = \ell < m$. Then

$$\begin{aligned}(\partial_h^{\alpha+e_i} v_h, \phi) &= -(\partial_h^\alpha v_h, \partial_i \phi) + \sum_{T\in\mathcal{T}_h}\int_{\partial T} \phi \partial_h^\alpha v_h \nu_i \\ &= (-1)^{|\alpha|+1}(v_h, \partial^{\alpha_h+e_i}\phi) + \sum_F \int_F [\partial_h^\alpha v_h]\phi\nu_i + \mathcal{O}(h)\|v_h\|_{m,h}\|\phi\|_m.\end{aligned}$$

By weak continuity, Poincare inequality and scaling argument,

$$\int_F [\partial_h^\alpha v_h]\phi = \mathcal{O}(h)|\partial_h^\alpha v_h|_{1,h,T\cup T'}\|\phi\|_{1,T\cup T'}, \quad F = T \cap T'.$$

By Lemma 3, it is then easy to see that $\{M_h^m\}$ is a consistent approximation of H^m with following finite element approximation error estimate $\|u - u_h\|_{m,h} = \mathcal{O}(h)$.

The new FE class (which extends trivially to include the case $m = 0$) is the lowest order possible elements. Their d.o.f. are well defined on $H^m(\Omega)$ for all m and n, while the d.o.f. of most other finite elements for $H^m(\Omega)$ are not well-defined on $H^m(\Omega)$ (except for $n = 1$). Given any $n > m \geq 1$, we have the inclusion property $M_h^m = \text{span}\{\partial_h^{e_1} M_h^{m+1}, \partial_h^{e_2} M_h^{m+1}, \cdots, \partial_h^{e_n} M_h^{m+1}\}$.

Locally supported basis functions can be constructed easily. In particular, we have given all the details for $m = 1, 2$ and 3 (see [9]).

The new class of FE elements is the only known family of approximation spaces of $H^m(\Omega)$ in R^n that are universally constructed ($0 \leq m \leq n$). It has many nice properties and everything fits perfectly well. The "most peculiar" Morley element seems now "most natural".

Applications to higher order PDEs. Higher (4th or 6-th) order PDEs arise in certain applications, such as, a Cahn-Hilliard equation modelling the spinodal decomposition and coarsening phenomena in binary alloys, a 6th order phase field simulation of the morphological evolution of a strained epitaxial thin film on a compliant substrate, and a 6th equation on the oxidation of silicon in superconductor devices.

Commonly used methods for higher order problems are to reduce them to the 2nd order systems. But the reduction requires caution (see [9]). The high order PDEs can be discretized directly by our new elements without too much difficulty.

References

[1] P. Alfeld, *A trivariate Clough-Tocher scheme for tetrahedral data*, Comput. Aided Geom. Design, **1** (1984),169-181.

[2] J. Aubin, *Approximation of elliptic boundary value problems*, Wiley-Interscience, 1972,

[3] G. Baker, *Finite element methods for elliptic equations using nonconforming elements*, v. 31, 137 (1977), 45-59.

[4] O. Davydov, *Stable local bases for multivariants spline spaces*, J Approx Theory, **111** (2001), 267-297.

[5] M. J. Lai and A. LeMehaute, *A new kind of trivariate C^1 macro-element*, Advances in Computational Mathematics, **21** (2004), 273-292.

[6] L. L. Schumaker, *On super splines and finite elements*, SIAM J Numer Anal, **26** (1989), 997-1005.

[7] F. Stummel, *The generalized patch test*, SIAM J Numer Analysis, **16** (1979), 449-471.

[8] M. Wang and J. Xu, *The Morley element for fourth order elliptic equations in any dimensions*, Numer Math, **103** (2006), 155-169.

[9] M. Wang and J. Xu, *Minimal finite element spaces for 2m-th order partial differential equations in R^n*, Research Report, **29** (2006), School of Mathematical Sciences and Institute of Mathematics, Peking University.

Reporter: Angela Kunoth

Participants

Prof. Dr. Paco Arandiga
Departament de Matematica Aplicada
Universitat de Valencia
Dr. Moliner 50
E-46100 Burjassot

Prof. Dr. Peter G. Binev
Dept. of Mathematics
University of South Carolina
Columbia , SC 29208
USA

Prof. Dr. Folkmar A. Bornemann
Zentrum Mathematik
Technische Universität München
85747 Garching bei München

Prof. Dr. Dietrich Braess
Fakultät für Mathematik
Ruhr-Universität Bochum
44780 Bochum

Jürgen Braun
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Kolja Brix
Institut für Geometrie und
Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen

Dr. Martin Campos Pinto
I.R.M.A.
Universite Louis Pasteur
7, rue Rene Descartes
F-67084 Strasbourg -Cedex

Prof. Dr. Claudio Canuto
Dipartimento di Matematica
Politecnico di Torino
Corso Duca degli Abruzzi, 24
I-10129 Torino

Prof. Dr. Carsten Carstensen
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

Prof. Dr. Albert Cohen
Laboratoire Jacques-Louis Lions
Universite Pierre et Marie Curie
175 rue du Chevaleret
F-75013 Paris

Gabriela Constantinescu
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Prof. Dr. Stephan Dahlke
FB Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Strasse (Lahnbg)
35032 Marburg

Prof. Dr. Wolfgang Dahmen
Institut für Geometrie und
Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen

Prof. Dr. Ingrid Daubechies
Program in Applied & Computational Mathematics
Princeton University
218 Fine Hall - Washington Road
Princeton , NJ 08544-1000
USA

Prof. Dr. Ronald A. DeVore
Dept. of Mathematics
University of South Carolina
Columbia , SC 29208
USA

Prof. Dr. Rosa Donat
Departament de Matematica Aplicada
Universitat de Valencia
Dr. Moliner 50
E-46100 Burjassot

Prof. Dr. Hans Georg Feichtinger
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Prof. Dr. Anna C. Gilbert
Department of Mathematics
University of Michigan
2074 East Hall
530 Church Street
Ann Arbor , MI 48109-1043
USA

Prof. Dr. Karlheinz Gröchenig
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Dr. C. Sinan Güntürk
Courant Institute of Mathematical Sciences
New York University
251, Mercer Street
New York , NY 10012-1110
USA

Prof. Dr. Wolfgang Hackbusch
Max-Planck-Institut für Mathematik in den Naturwissenschaften
Inselstr. 22 - 26
04103 Leipzig

Prof. Dr. Helmut Harbrecht
Institut für Numerische Simulation
Universität Bonn
Wegelerstr. 6
53115 Bonn

Prof. Dr. Reinhard Hochmuth
FB 17 - Mathematik/Informatik -
Universität Kassel
Heinrich-Plett-Str. 40
34132 Kassel

Prof. Dr. Palle E.T. Jorgensen
Dept. of Mathematics
University of Iowa
Iowa City , IA 52242-1466
USA

Prof. Dr. Gerard Kerkyacharian
U. F. R. de Sciences Economiques
Universite de Paris X - Nanterre
200, Avenue de la Republique
F-92001 Nanterre Cedex

Prof. Dr. Angela Kunoth
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Priv.-Doz. Dr. Gitta Kutyniok
Department of Mathematics
Princeton University
609 Fine Hall
Washington Road
Princeton , NJ 08544
USA

Dr. Franz Luef
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Prof. Dr. Peter Maaß
Universität Bremen
FB3 Mathematik und Informatik
Zentrum für Technomathematik
MZH 2250
28344 Bremen

Priv.-Doz. Dr. Siegfried Müller
Institut für Geometrie und
Praktische Mathematik
RWTH Aachen
Templergraben 55
52062 Aachen

Prof. Dr. Peter Oswald
School of Engineering and Science
Jacobs University Bremen
Postfach 750561
28725 Bremen

Roland Pabel
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstr. 6
53115 Bonn

Prof. Dr. Guergana Petrova
Department of Mathematics
Texas A & M University
College Station , TX 77843-3368
USA

Prof. Dr. Pencho P. Petrushev
Dept. of Mathematics
University of South Carolina
Columbia , SC 29208
USA

Prof. Dr. Dominique Picard
Laboratoire de Probabilites et
Modeles Aleatoires
Universite Paris VII
175 rue du Chevaleret
F-75013 Paris Cedex

Prof. Dr. Bojan Popov
Department of Mathematics
Texas A & M University
College Station , TX 77843-3368
USA

Dr. Holger Rauhut
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Fernando Rodriguez-Gonzalez
Centre de Mathematiques
Ecole Polytechnique
Plateau de Palaiseau
F-91128 Palaiseau Cedex

Prof. Dr. Reinhold Schneider
Lehrstuhl für Scientific Computing
Universitäts-Hochhaus 9. Stock
Christian-Albrechts-Platz 4
24098 Kiel

Prof. Dr. Christoph Schwab
Seminar für Angewandte Mathematik
ETH-Zentrum
Rämistr. 101
CH-8092 Zürich

Prof. Dr. Rob Stevenson
Department of Mathematics
Utrecht University
P.O.Box 80.010
NL-3508 TA Utrecht

Prof. Dr. Endre Süli
Computing Laboratory
Oxford University
Wolfson Building
Parks Road
GB-Oxford OX1 3QD

Prof. Dr. Anita Tabacco
Dipartimento di Matematica
Politecnico di Torino
Corso Duca degli Abruzzi, 24
I-10129 Torino

Prof. Dr. Eitan Tadmor
Center for Scientific Computation
and Mathematical Modeling (CSCAMM)
CSIC Bldg # 406
Paint Branch Dr.
College Park MD 20742-3289
USA

Prof. Dr. Vladimir N. Temlyakov
Dept. of Mathematics
University of South Carolina
Columbia , SC 29208
USA

Prof. Dr. Karsten Urban
Institut für Numerische Mathematik
Universität Ulm
Helmholtzstr. 18
89069 Ulm

Rachel Ward
Department of Mathematics
Princeton University
Fine Hall
Washington Road
Princeton , NJ 08544-1000
USA

Prof. Dr. Przemek Wojtaszczyk
Instytut Matematyki
Stosowanej i Mechaniki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
POLAND

Prof. Dr. Jinchao Xu
Department of Mathematics
Pennsylvania State University
University Park , PA 16802
USA

Dionisio Felix Yanez Avendano
Facultad de Matematicas
Universidad de Valencia
Avda. Dr. Moliner 50
E-46100 Burjassot Valencia

Prof. Dr. Harry Yserentant
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin