Mathematisches Forschungsinstitut Oberwolfach

Report No. 29/2012

DOI: 10.4171/OWR/2012/29

Applied Harmonic Analysis and Sparse Approximation

Organised by
Ingrid Daubechies, Durham NC
Gitta Kutyniok, Berlin
Holger Rauhut, Bonn
Thomas Strohmer, Davis

June 10th – June 16th, 2012

Abstract. Applied harmonic analysis and sparse approximation are highly active research areas with a lot of recent exciting developments. Their methods have become crucial for a wide range of applications in technology and science, such as signal and image processing. Understanding of the underlying mathematics has grown vastly. Interestingly, there are a lot of connections to other fields, such as convex optimization, probability theory and Banach space geometry. Yet, many problems in these areas remain unsolved or even unattacked. The workshop intended to bring together world leading experts in these areas, to report on recent developments, and to foster new developments and collaborations.

Mathematics Subject Classification (2000): 42-XX, 65Txx, 94Axx, 65K05, 15A52.

Introduction by the Organisers

The workshop *Applied Harmonic Analysis and Sparse Approximation* was organized by Ingrid Daubechies (Durham), Gitta Kutyniok (Berlin), Holger Rauhut (Bonn) and Thomas Strohmer (Davis). This meeting was attended by 53 participants from about 10 countries and 3 continents.

Applied Harmonic Analysis has seen enormous developments in the last three decades. Its tools and methods have turned out to be crucial in a wide range of signal and image processing applications, in numerical algorithms for the solution of operator equations, and in inverse problems. In addition, mathematical understanding of the underlying harmonic analysis has grown vastly – with research sometimes driven by the needs of various applications and its results, in turn leading to renewed interplay with applications. Sparse approximation and

compressed sensing constitute a recent development with clear roots in applied harmonic analysis as well as in other areas such as statistics and optimization. It has been the focus of an intense research activity in recent years, and – despite its young age – has become already a fairly mature mathematical discipline. This does by far not mean that all problems are solved – only that the field is past its childhood: There are still various important and sometimes deep open problems in applied harmonic analysis and sparse approximation. Compressed sensing has moreover shown intriguing connections to the fields of random matrix theory and convex optimization.

One main focus of research in applied harmonic analysis is on the development of novel structured dictionaries which exhibit specific properties – for example, optimal sparse approximation – for desired classes of mathematical objects. This has led to the recent introduction of **curvelets**, **shearlets**, and other representation systems, which have already impacted both theoretically oriented questions such as sparse expansions of Fourier integral operators and application orientation areas such as image processing. There remains, however, a wide range of fundamental open questions; for instance, the understanding and/or characterization of associated function spaces, which is essential for a mathematically satisfactory analysis of such representations systems.

One key property that ensures the occurrence of sparse expansions is the redundancy of these systems. This is the focus of the research area of **frame theory**, which studies various aspects of redundancy as a mathematical concept. Frame theory already impacted the whole area of applied harmonic analysis and sparse approximation significantly; yet, surprisingly, a fundamental understanding of redundancy measures is just at its beginning.

Sparsity has become a very important concept in recent years in applied mathematics, especially in mathematical signal and image processing. The key idea is that many types of functions and signals arising naturally in these contexts can be described using only a small number of significant terms in a suitable basis or frame, often a wavelet basis, a Gabor frame or a shearlet frame. This is essentially the reason why many lossy compression techniques such as JPEG or MP3 work so well. Quite interestingly, sparsity is useful not only for compression purposes. The new field of **compressed sensing** predicts that sparse high-dimensional signals can be recovered efficiently from what would previously have been considered highly incomplete measurements. This discovery has led to a fundamentally new approach to certain signal and image recovery problems, which can in fact be regarded as a paradigm change. Remarkably, the main constructions for good measurement matrices known so far are random, and the mathematical research in compressive sensing uses also tools, sometimes quite sophisticated, from probability theory and the geometry of Banach spaces.

The workshop featured 31 talks, thereof 9 longer overview talks. Moreover, a session of very short presentation of 3 minutes took place on Monday, which we called the *3 Minutes of Fame* (following Andy Warhol's concept of 15 minutes of fame). Every participant had the possibility to contribute to this session, and this

experiment worked out very well. In particular, it provided a quick overview on what the participants are presently working or would like to discuss with other participants.

Some highlights of the presentations include:

- **Super-resolution:** Emmanuel Candés reported on a new approach to super-resolution based on convex optimization. Quite surprisingly and in contrast to compressive sensing, this theory does not require randomness and demonstrates that one can extrapolate the high-frequency content of a signal from only low-frequency information.
- **Co-sparsity:** A new direction in sparse approximation uses the notion of cosparsity. Rather than requiring that has a sparse representation with respect to a basis or frame, one requires that Ωx is sparse, where Ω is a so-called analysis operator. This approach has advantages in certain signal processing tasks. An overview talk on this topic was given by Michael Elad, and this idea appeared also in some of the shorter talks, e.g. by Emily King and Ignace Loris.
- **Signal separation:** Very accurate bounds for the problem of separating two signals "of different nature" via convex programming were presented in the talk of Joel Tropp. This rather general theory covers several situations including that both vectors are sparse in different bases and that one is a sparse matrix and the other one a low rank matrix.
- **Phase retrieval:** Thomas Strohmer presented a new method to recover a signal x based on knowledge of only the absolute values of some coefficients $|\langle a_k, x\rangle|$ with respect to some vectors a_k. This problem arises for instance in diffraction imaging, where one measures the absolute value of the Fourier transform. The method builds on ideas from low rank matrix recovery and uses a convex optimization program to find x. First results were reported, but many open problems remain. A shorter talk on phase retrieval was presented by Radu Balan.

Further new interesting developments include the use of sparsity in flocking (Massimo Fornasier), new constructions of shearlets, results for corresponding function spaces and applications in image segmentation (Stephan Dahlke, Gabriele Steidl, Gerd Teschke; Philipp Grohs), new results around the restricted isometry property (Bernhard Bodmann; Felix Krahmer) as well as new algorithms for sparse recovery (Ignace Loris; Sergey Voronin).

The organizers would like to take the opportunity to thank MFO for providing support and a very inspiring environment for the workshop. The magic of the place (as coined by one of the participants) and the pleasant atmosphere contributed greatly to the success of the workshop.

Workshop: Applied Harmonic Analysis and Sparse Approximation

Table of Contents

Akram Aldroubi
Subspace segmentation and modeling high dimensional data 1767

Radu Balan
Reconstruction of Signals from Magnitudes of Redundant Representations 1769

Gregory Beylkin
On optimal rational approximations and their applications 1771

Holger Boche
Spaceability of Functions with Complicated Quantization and Truncation Behavior 1773

Bernhard G. Bodmann (joint with Jameson Cahill and Peter G. Casazza)
Fusion frames from matrices with the restricted isometry property 1776

Robert Calderbank
Faster than Nyquist but Slower than Tropp 1778

Emmanuel J. Candès (joint with Carlos Fernandez-Granda)
Towards a Mathematical Theory of Super-Resolution 1781

Peter G. Casazza
Spectral Tetris Constructions for Sensor Networks 1783

Stephan Dahlke, Gabriele Steidl, Gerd Teschke (joint with Sören Häuser)
Shearlet Coorbit Spaces: Traces and Embeddings in Higher Dimensions . 1784

Michael Elad
The Analysis Sparse Model - Definition, Pursuit, Dictionary Learning, and Beyond 1790

Matthew Fickus
Hyperspectral imaging, compressed sensing and chromo-tomography 1790

Massimo Fornasier
Sparse optimal control in consensus emergence 1791

Karlheinz Gröchenig (joint with Joachim Stöckler)
Gabor Frames and Totally Positive Functions 1794

Philipp Grohs (joint with Gitta Kutyniok)
On the Structure of Anisotropic Frames 1796

Mijail Guillemard
Topological Aspects in Time-Frequency Analysis 1798

Max Hügel (joint with Holger Rauhut, Thomas Strohmer)
Remote sensing via ℓ_1-minimization 1800

Emily J. King (joint with Gitta Kutyniok, Xiaosheng Zhuang)
Analysis of Inpainting via Clustered Sparsity and Microlocal Analysis . . 1802

Felix Krahmer (joint with Shahar Mendelson, Holger Rauhut)
Suprema of Chaos Processes and the Restricted Isometry Property 1806

Wang-Q Lim (joint with Wolfgang Dahmen, Gitta Kutyniok and Gerrit Welper)
Adaptive Anisotropic Refinement Scheme for Transport Dominated Problems 1808

Ignace Loris (joint with Caroline Verhoeven)
A generalization of the iterative soft-thresholding algorithm for non-separable penalties 1811

Friedrich Philipp (joint with Gitta Kutyniok, Kasso A. Okoudjou, and Elizabeth K. Tuley)
Scalable Frames 1814

Justin Romberg (joint with Ali Ahmed and Benjamin Recht)
Blind Deconvolution and Channel Protection using Random Codes 1816

Naoki Saito (joint with Yuji Nakatsukasa and Ernest Woei)
Localization Phenomena of Laplacian Eigenfunctions on Graphs & Sparse Graph Constructions 1819

Zuowei Shen (joint with Jianfeng Cai, Bin Dong, Stanley Osher)
Image restoration: total variation, wavelet frames, and beyond 1822

Joachim Stöckler (joint with Maria Charina, Mihai Putinar, Claus Scheiderer)
Algebraic geometry and tight wavelet frames 1823

Thomas Strohmer (joint with Emmanuel Candès, Vlad Voroninski, and YoninaEldar)
Phase Retrieval and Convex Optimization 1824

Arthur Szlam (joint with Karol Gregor, Yann LeCun)
Fast subspace search and applications to object recognition 1826

Joel A. Tropp (joint with Michael B. McCoy)
Sharp recovery bounds for convex deconvolution, with applications 1829

Tino Ullrich (joint with Zung Dinh)
N-Widths for High-Dimensional Sparse Approximations 1829

Sergey Voronin
Two New Algorithms Useful For the Regularization of Large Scale Inverse Problems With Sparsity Constraints 1832

Rachel Ward (joint with Deanna Needell)
Performance guarantees for total variation minimization from undersampled measurements 1835

Abstracts

Subspace segmentation and modeling high dimensional data

Akram Aldroubi

Let $\mathcal{U} = \bigcup_{i=1}^{M} S_i$ where $\{S_i \subset \mathcal{B}\}_{i=1}^{M}$ is a set of subspaces of a Hilbert space or Banach space $\mathcal{B}$. Let $\mathbf{W} = \{w_j \in \mathcal{B}\}_{j=1}^{N}$ be a set of data points drawn from $\mathcal{U}$. The goal of subspace clustering is to identify all of the subspaces that the data $\mathbf{W} = \{w_1, ..., w_N\} \in \mathcal{B}$ is drawn from and assign each data point w_i to the subspace it belongs to. The number of subspaces, their dimensions, and a basis for each subspace are to be determined even in presence of noise, missing data, and outliers. The subspace clustering or segmentation problem can be stated as follows: Then,

(1) determine the number of subspaces M,
(2) find each subspace S_i,
(3) collect the data points belonging to the same subspace into the same cluster.

The ambient space $\mathcal{B}$ maybe be finite or infinite dimensional and the subspace S_i may be infinite dimensional as well, e.g., $\mathcal{B} = L^2(\mathbb{R}^n)$ and the S_is are unknown shift invariant spaces with at most r generators [3, 5]. There are many engineering and mathematics applications in which data lives in a union of low dimensional subspaces, and where the subspace segmentation problem is central, (see e.g., [1, 4, 8, 10] and the references therein).

When M is known, the subspace segmentation problem, for both the finite and infinite dimensional space cases, can be formulated as follows:
Let $\mathcal{B}$ be a Banach space, $\mathbf{W} = \{w_1, \ldots, w_N\}$ a finite set of vectors in $\mathcal{B}$. For $i = 1, \ldots, M$, let $\mathcal{C} = C_1 \times C_2 \times \cdots \times C_M$ be the cartesian product of M family C_i of closed subspaces of $\mathcal{B}$ each containing the trivial subspace $\{0\}$. Thus, an element $\mathcal{C}$ is a sequence $\{S_1, \ldots, S_M\}$ of M subspaces of $\mathcal{B}$ with $S_i \in C_i$.

Problem 1

(i) Given a finite set $\mathbf{W} \subset \mathcal{B}$, a fixed p with $0 < p \leq \infty$, and a fixed integer $M \geq 1$, find the infimum of the expression

$$e(\mathbf{W}, \mathbf{V}) := \sum_{w \in \mathbf{W}} \min_{1 \leq j \leq M} d^p(w, S_j),$$

over $\mathbf{V} = \{S_1, \ldots, S_M\} \in \mathcal{C}$, and $d(x, y) := \|x - y\|_{\mathcal{B}}$.
(ii) Find a sequence of M-subspaces $\mathbf{V}^o = \{S_1^o, \ldots, S_M^o\} \in \mathcal{C}$ (if it exists) such that

$$e(\mathbf{W}, \mathbf{V}^o) = \inf\{e(\mathbf{W}, \mathbf{V}) : \mathbf{V} \in \mathcal{C}\}. \tag{1}$$

Note that this problem is neither linear nor convex for any p. In the presence of outliers, it is shown that $p = 1$ is a good choice [8], and a good choice for light-tailed noise is $p = 2$. The necessary and sufficient conditions for the existence of a solution when $p = 2$ and $\mathcal{B}$ is a Hilbert space can be found in [4]. However, the necessary and sufficient conditions for the existence of a solution when $p \neq 2$ is an open problem.

Many algorithms for solving Problem 1, or some special cases of the subspace segmentation problem when noise is present have been proposed. Algebraic methods, such as GPCA [11], or the RREF [1] work perfectly in noiseless environment but do not perform as well in noisy environment. However, the output to these algorithms can be set as initial conditions to convergent iterative algorithms, see e.g., [5]. It is still not known if there are a good thresholding or denoising algorithms that render the algebraic methods more performant.

In the presence of noise, the best algorithms to date are those based on spectral clustering (see e.g., [2, 12, 9]). However, some of the theoretical and quantitative issues related to these methods are still open.

References

[1] A. Aldroubi and A. Sekmen, *Reduced Row Echelon Form and Non-Linear Approximation for Subspace Segmentation and High-Dimensional Data Clustering*, preprint (2012).

[2] A. Aldroubi and A. Sekmen, *Nearness to Local Subspace Algorithm for Subspace and Motion Segmentation*, (2011), preprint.

[3] A. Aldroubi I. Krishtal, R. Tessera, and H. Wang, *Principle shift-invariant spaces with extra invariance nearest to observed data*, Collectanea Mathematica, (2011).

[4] A. Aldroubi and R. Tessera, *On the existence of optimal unions of subspaces for data modeling and clustering*, Foundation of Computational Mathematics **11** (3) (2011), 363–379.

[5] A. Aldroubi, C. Cabrelli, and U. Molter, *Optimal non-linear models for sparsity and sampling*, J. of Fourier Anal. and Applications **14** (5) (2009), 793–812.

[6] A. Aldroubi, C. Cabrelli, D. Hardin, and U. Molter, *Optimal shift-invariant spaces and their Parseval frame generators*, Appl. Comput. Harmon. Anal. (2007), 273–283.

[7] E. Elhamifar, R. Vidal, *Sparse subspace clustering*, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2790–2797.

[8] G. Lerman, T. Zhang, *Robust recovery of multiple subspaces by geometric lp minimization* **39** (2011) 2686–2715.

[9] M. Soltanolkotabi and E. J. Candès, *A geometric analysis of subspace clustering with outliers* preprint (2011).

[10] R. Vidal, *A tutorial on subspace clustering*, IEEE Signal Processing Magazine (2010).

[11] R. Vidal, Y. Ma, and S. Sastry, *Generalized principal component analysis (GPCA)*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **27** (2005) 1945–1959.

[12] T. Zhang, A. Szlam, Y. Wang, G. Lerman. Hybrid Linear Modeling via Local Best-fit Flats. Arxiv preprint, (2010).

Reconstruction of Signals from Magnitudes of Redundant Representations

Radu Balan

This paper is concerned with the question of reconstructing a vector x in a finite-dimensional real or complex Hilbert space H when only the magnitudes of the coefficients of the vector under a redundant linear map are known.

Specifically our problem is to reconstruct $x \in H$ up to a global phase factor from the magnitudes $\{|\langle x, f_k\rangle| \ , \ 1 \leq k \leq m\}$ where $\{f_1, \ldots, f_m\}$ is a frame (complete system) for H. In H we consider the following equivalence relation:

$$x, y \in H \ , \ x \sim y \text{ iff } y = zx \text{ for some scalar z with } |z| = 1 \tag{1}$$

Let $\hat{H} = H/\sim$ be the set of classes of equivalence induced by this relation. Thus $\hat{x} = \{x, -x\}$ in the real case (when $H = \mathbb{R}^n$), and $\hat{x} = \{e^{i\alpha}x, 0 \leq \alpha < 2\pi\}$ in the complex case (when $H = \mathbb{C}^n$). The analysis map induces the following nonlinear map

$$\varphi : \hat{H} \to (\mathbb{R}^+)^m \ , \ \varphi(\hat{x}) = (|\langle x, f_k\rangle|^2)_{1 \leq k \leq m} \tag{2}$$

where $\mathbb{R}^+ = \{x \ , \ x \in \mathbb{R} \ , \ x \geq 0\}$ is the set of nonnegative real numbers. The main problem is to analyze when φ is injective, and, when it is so, to find efficient algorithms for its invertibility.

A previous paper [3] described the importance of this problem to signal processing, in particular to the analysis of speech. Of particular interest is the case when the coefficients are obtained from a Windowed Fourier Transform (also known as Short-Time Fourier Transform), or an Undecimated Wavelet Transform (in audio and image signal processing). While [3] presents some necessary and sufficient conditions for reconstruction, the general problem of finding fast/efficient algorithms is still open. In [2] we describe one solution in the case of STFT coefficients.

For vectors in real Hilbert spaces, the reconstruction problem is easily shown to be equivalent to a combinatorial problem. In [4] this problem is further proved to be equivalent to a (nonconvex) optimization problem.

A different approach (which we called the "algebraic approach") was proposed in [1]. While it applies to both real and complex cases, noisless and noisy cases, the approach requires solving a linear system of size exponentially in space dimension. The algebraic approach mentioned earlier generalizes the approach in [5] where reconstruction is performed with complexity $O(n^2)$ (plus computation of the principal eigenvector for a matrix of size n). However this method requires $m = O(n^2)$ frame vectors.

Recently the authors of [6] developed a convex optimization algorithm (PhaseLift) and proved its ability to perform exact reconstruction in the absence of noise, as well as its stablity under noise conditions. In a separate paper, [7], the authors developed further a similar algorithm in the case of windowed DFT transforms.

In this paper we analyze an iterative algorithm based on regularized least-square criterion. Consider the following function

$$J : H \times H \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+ \tag{3}$$

$$J(u, v, \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle u, f_k\rangle\langle f_k, v\rangle|^2 + \lambda\|u\|^2 + \mu\|u - v\|^2 + \lambda\|v\|^2.$$

Our goal is to minimize $\|y - \varphi(u)\|^2 = J(u, u, 0, \mu)$ over u, for some (and hence any) value $\mu \in \mathbb{R}^+$. Our strategy is the following iterative process:

$$x^{t+1} = argmin_u J(u, x^t, \lambda_t, \mu_t) \tag{4}$$

for some initialization x^0 and policy of decreasing $(\lambda_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$.

In this paper we have also proved additional necessary and sufficients conditions for injectivity of the nonlinear map φ. Specifically we showed:

Theorem Given a m-set of vectors $\mathbb{F} = \{f_1, \ldots, f_m\} \subset H$ the following conditions are equivalent:

(1) For any disjoint partition of the frame set $\mathbb{F} = \mathbb{F}_1 \cup \mathbb{F}_2$, either $\mathbb{F}_1$ spans H or $\mathbb{F}_2$ spans H;

(2) For any two vectors $x, y \in H$ if $n \neq 0$ and $y \neq 0$ then

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 > 0;$$

(3) There is a positive real constant $a_0 > 0$ so that for all $x, y \in H$,

$$\sum_{k=1}^{m} |\langle x, f_k\rangle|^2 |\langle y, f_k\rangle|^2 \geq a_0\|x\|^2\|y\|^2 \tag{5}$$

(4) There is a positive real constant $a_0 > 0$ so that for all $x \in H$,

$$R(x) := \sum_{k=1}^{m} |\langle x, f_k\rangle|^2 \langle \cdot, f_k\rangle f_k \geq a_0 I \tag{6}$$

where the inequality is in the sense of quadratic forms.

Remark

1. The constants in (3) and (4) above are the same (hence the same notation).

2. In the real case (when $H = \mathbb{R}^n$), any of the above conditions is equivalent to injectivity of the nonlinear map φ.

3. In the complex case, the above four conditions are necessary but not sufficient for injectivity of the map φ.

References

[1] R. Balan, *A Nonlinear Reconstruction Algorithm from Absolute Value of Frame Coefficients for Low Redundancy Frames*, Proceedings of SampTA Conference, Marseille, France May 2009.

[2] R. Balan, *On Signal Reconstruction from Its Spectrogram*, Proceedings of the CISS Conference, Princeton NJ, May 2010.

[3] R. Balan, P. Casazza, D. Edidin, *On signal reconstruction without phase*, Appl.Comput.Harmon.Anal. **20** (2006), 345–356.
[4] R. Balan, P. Casazza, D. Edidin, *Equivalence of Reconstruction from the Absolute Value of the Frame Coefficients to a Sparse Representation Problem*, IEEE Signal.Proc.Letters, **14** (5) (2007), 341–343.
[5] R. Balan, B. Bodmann, P. Casazza, D. Edidin, *Painless reconstruction from Magnitudes of Frame Coefficients*, J.Fourier Anal.Applic., **15** (4) (2009), 488–501.
[6] E. Candés, T. Strohmer, V. Voroninski, *PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming*, to appear in Communications in Pure and Applied Mathematics (2011)
[7] E. Candés, Y. Eldar, T. Strohmer, V. Voloninski, *Phase Retrieval via Matrix Completion Problem*, preprint 2011
[8] P. Casazza, *The art of frame theory*, Taiwanese J. Math., (2) **4** (2000), 129–202.
[9] J. Finkelstein, *Pure-state informationally complete and "really" complete measurements*, Phys. Rev. A **70** (2004), no. 5, doi:10.1103/PhysRevA.70.052107

On optimal rational approximations and their applications

Gregory Beylkin

Theoretical results on optimal rational approximations (showing their efficiency in approximating functions with isolated singularities) have been around for at least half a century. On the other hand, practical construction of (near) optimal rational approximations have been limited to relatively small problems and, when applied to functions with singularities, achieved only limited accuracy. The reason is that algorithms for this purpose rely on computing small singular values and corresponding singular vectors as well as on finding roots of polynomials of high degree, both generally ill-advised numerical tasks. Recently, we developed an algorithm for computing with high relative precision con-eigenvalues of Cauchy matrices [4]. This algorithm leads to a practical approach for constructing (near) optimal rational approximations, thus making such approximations a tool in numerical analysis.

Specifically, given a rational function with n poles in the unit disk, a rational approximation with $m \ll n$ poles in the unit disk may be obtained from the mth con-eigenvector of an $n \times n$ Cauchy matrix, where the associated con-eigenvalue $\lambda_m > 0$ gives the approximation error in the L^∞ norm. Unfortunately, standard algorithms do not accurately compute small con-eigenvalues (and the associated con-eigenvectors) and, in particular, yield few or no correct digits for con-eigenvalues smaller than the machine roundoff. We have developed a fast and accurate algorithm for computing con-eigenvalues and con-eigenvectors of positive-definite Cauchy matrices, yielding even the tiniest con-eigenvalues with high relative accuracy. The algorithm computes the mth con-eigenvalue in $\mathcal{O}\left(m^2 n\right)$ operations and, since the con-eigenvalues of positive-definite Cauchy matrices decay exponentially fast, we obtain (near) optimal rational approximations in $\mathcal{O}\left(n \left(\log \delta^{-1}\right)^2\right)$ operations, where δ is the approximation error in the L^∞ norm. We provide error bounds demonstrating high relative accuracy of the computed con-eigenvalues and the high accuracy of the unit con-eigenvectors.

We also developed several applications of the new algorithm. These include constructing near optimal rational representations of large data sets (e.g. musical recordings) [2], solving Burgers' equation with a small (10^{-5}) viscosity [3], a new approach to the design of near optimal filters [1] and a new algorithm for X-ray tomography with improved resolution of singularities (in progress).

In contrast to wavelet-type approximations, near optimal rational approximations are effectively shift invariant. In order to obtain a near optimal rational approximation of a large data set, we first construct its intermediate B-spline representation. Then, by using a new rational approximation of B-splines, we arrive at a suboptimal rational approximation of the data set. We then use fast and accurate reduction algorithm to obtain a near optimal rational approximation from a suboptimal one. Our approach requires first splitting the data into large segments, which may later be merged together, if needed. We also provide a fast algorithm for evaluating these rational approximations. In particular, this allows us to interpolate the original data to any grid. One of the applications of our algorithm is the compression of audio signals.

Using the reduction algorithm, we develop a numerical calculus for rational representations of functions. Indeed, while operations such as multiplication and convolution increase the number of poles in the representation, we use the reduction algorithm to maintain an optimally small number of poles. To demonstrate the efficiency, robustness, and accuracy of our approach, we solve Burgers' equation with small viscosity ν. It is well known that its solutions exhibit moving transition regions of width $\mathcal{O}(\nu)$, so that this equation provides a stringent test for adaptive PDE solvers. We show that optimal rational approximations capture the solutions with high accuracy using a small number of poles. In particular, we solve the equation with local accuracy $\epsilon = 10^{-9}$ for viscosity as small as $\nu = 10^{-5}$.

We develop a systematic method for designing highly accurate and efficient infinite impulse response (IIR) and finite impulse response (FIR) filters given their specifications. In our approach, we first meet the specifications by constructing an IIR filter with, possibly, a large number of poles. We then construct, for any given accuracy, an optimal IIR version of such filter (with a minimal number of poles). Finally, also for any given accuracy, we convert the IIR filter to an efficient FIR filter cascade (either serial or parallel). Since in this FIR approximation the non-causal part of the IIR filter only introduces an additional delay (as a function of the desired accuracy), our IIR construction does not have to enforce causality. Thus, we obtain a simple method for constructing linear phase filters if the specifications so require. All of these procedures are accomplished via robust, fast algorithms.

Using the fact that the full Fourier series of a periodic rational function may be recovered from (sufficiently many) of its samples, we construct near optimal rational approximations of projections and demonstrate that we can improve resolution of reconstruction in X-ray tomography.

This presentation is based on the joint work with Lucas Monzón, Terry Haut, Ryan Lewis, Matt Reynolds and Anil Damle.

References

[1] G. Beylkin, R.D. Lewis, and L. Monzón. On the Design of Highly Accurate and Efficient IIR and FIR Filters. *IEEE Trans. Signal Process.*, 2011. http://dx.doi.org/10.1109/TSP.2012.2197397.

[2] A. Damle, G. Beylkin, T. S. Haut, and L. Monzón. Near optimal rational approximations of large data sets. *Appl. Comput. Harmon. Anal.*, 2012. Submitted.

[3] T. Haut, G. Beylkin, and L. Monzón. Solving Burgers' equation using optimal rational approximations. *Appl. Comput. Harmon. Anal.*, 2012. http://dx.doi.org/10.1016/j.acha.2012.03.004 (electronic).

[4] T. S. Haut and G. Beylkin. Fast and accurate con-eigenvalue algorithm for optimal rational approximations. *arXiv:1012.3196 [math.NA]*, 2011.

Spaceability of Functions with Complicated Quantization and Truncation Behavior

Holger Boche

A well-known fact about the convergence behavior of the Shannon sampling series for functions in the Paley–Wiener space $\mathcal{PW}_\pi^1$ is Brown's theorem, which states the uniform convergence on compact subsets of $\mathbb{R}$. The truncation of the finite Shannon sampling series $\sum_{k=-N}^{N} f(k)\frac{\sin(\pi(t-k))}{\pi(t-k)}$ is done in the domain of the function f because only the samples $f(k)$, $k=-N,\dots,N$ are taken into account. In contrast, it is also possible to control the truncation of the series in the codomain of f by considering only the samples $f(k)$, $k\in\mathbb{Z}$, whose absolute value is larger than or equal to some threshold $\delta>0$. This leads to the approximation process

$$(A_\delta f)(t) := \sum_{\substack{k=-\infty \\ |f(k)|\geq\delta}}^{\infty} f(k)\frac{\sin(\pi(t-k))}{\pi(t-k)}. \tag{1}$$

In general, $A_\delta f$ is only an approximation of f, and we want the function $A_\delta f$ to be close to f if δ is sufficiently small. Here, we analyze a more general approximation process

$$(A_\delta^T f)(t) := (TA_\delta f)(t) = \sum_{\substack{k=-\infty \\ |f(k)|\geq\delta}}^{\infty} f(k)h_T(t-k), \tag{2}$$

where additionally a linear time invariant (LTI) system T is applied. Clearly, (1) is a special case of (2) with T being the identity operator. Surprisingly, the approximation errors of the approximation processes (1) and (2) do not always decrease as the threshold δ tends to zero, i.e., as more and more samples are used for the approximation. Depending on the function $f\in\mathcal{PW}_\pi^1$ and the LTI system T, the approximation process $(A_\delta^T f)(t)$ can diverge unboundedly, even for fixed $t\in\mathbb{R}$, as δ goes to zero.

In order to continue the discussion, we fix some notation. Let $\hat{f}$ denote the Fourier transform of a function f, where $\hat{f}$ is to be understood in the distributional sense. For $\sigma>0$ let $\mathcal{B}_\sigma$ be the set of all entire functions f with the property that

for all $\epsilon > 0$ there exists a constant $C(\epsilon)$ with $|f(z)| \leq C(\epsilon)\exp\big((\sigma+\epsilon)|z|\big)$ for all $z \in \mathbb{C}$. The Bernstein space $\mathcal{B}_\sigma^p$ consists of all functions in $\mathcal{B}_\sigma$, whose restriction to the real line is in $L^p(\mathbb{R})$, $1 \leq p \leq \infty$. A function in $\mathcal{B}_\sigma^p$ is called bandlimited to σ. It is well known, that $\mathcal{B}_\sigma^p \subset \mathcal{B}_\sigma^s$ for $1 \leq p \leq s \leq \infty$. For $\sigma > 0$ and $1 \leq p \leq \infty$ we denote by $\mathcal{PW}_\sigma^p$ the Paley-Wiener space of functions f with a representation $f(z) = 1/(2\pi)\int_{-\sigma}^{\sigma} g(\omega)e^{iz\omega}\ \mathrm{d}\omega$, $z \in \mathbb{C}$, for some $g \in L^p[-\sigma,\sigma]$. If $f \in \mathcal{PW}_\sigma^p$ then $g(\omega) = \hat{f}(\omega)$. The norm for $\mathcal{PW}_\sigma^p$, $1 \leq p < \infty$, is given by $\|f\|_{\mathcal{PW}_\sigma^p} = (1/(2\pi)\int_{-\sigma}^{\sigma}|\hat{f}(\omega)|^p\ \mathrm{d}\omega)^{1/p}$. A subset G of a metric space X is said to be nowhere dense in X if the closure $[G]$ does not contain a non-empty open set of X. G is said to be of the first category (or meager) if G is the countable union of sets each of which is nowhere dense in X. The complement of a set of the first category is called a residual set. One property that shows the richness of residual sets is the following: The countable intersection of residual sets is always a residual set.

1. Stable LTI Systems

Since our analyses involve stable linear time-invariant (LTI) systems, we briefly review some definitions and facts. A linear system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ is called stable if the operator T is bounded, i.e., if $\|T\| = \sup_{\|f\|_{\mathcal{PW}_\pi^1}\leq 1}\|Tf\|_{\mathcal{PW}_\pi^1} < \infty$. Furthermore, it is called time-invariant if $(Tf(\cdot - a))(t) = (Tf)(t-a)$ for all $f \in \mathcal{PW}_\pi^1$ and $t, a \in \mathbb{R}$. For every stable LTI system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ there exists exactly one function $\hat{h}_T \in L^\infty[-\pi,\pi]$ such that $(Tf)(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{f}(\omega)\hat{h}_T(\omega)e^{i\omega t}\ \mathrm{d}\omega$ for all $f \in \mathcal{PW}_\pi^1$. Conversely, every function $\hat{h}_T \in L^\infty[-\pi,\pi]$ defines a stable LTI system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$. We have $h_T = T\,\mathrm{sinc}$, where $\mathrm{sinc}(t) = \sin(\pi t)/(\pi t)$ for $t \neq 0$ and $\mathrm{sinc}(t) = 1$ for $t = 0$. The operator norm of a stable LTI system T is given by $\|T\| = \|\hat{h}_T\|_\infty$. Note that $\hat{h}_T \in L^\infty[-\pi,\pi] \subset L^2[-\pi,\pi]$, and consequently $h_T \in \mathcal{PW}_\pi^2$.

For certain functions f and stable LTI systems T, we can use $\sum_{k=-N}^{N} f(k)h_T(t-k)$ to obtain an approximation of Tf. Here, we analyze the analogous approximation process (2), where the truncation is controlled in the codomain of f instead of the domain of f, for functions in $\mathcal{PW}_\pi^1$. By A_δ^T we denote the operator that maps $f \in \mathcal{PW}_\pi^1$ to $A_\delta^T f$ according to (2). If $f \in \mathcal{PW}_\pi^1$ we have $\lim_{t\to\infty} f(t) = 0$ by the Riemann-Lebesgue lemma, and it follows that the series in (2) has only finitely many summands, which implies $A_\delta^T f \in \mathcal{PW}_\pi^2 \subset \mathcal{PW}_\pi^1$. The analysis of the approximation processes (2) is difficult, because the operator $A_\delta^T \colon \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^2$ is non-linear and discontinuous.

2. Behavior for Fixed t

Let Φ be the set of all continuous, positive, and monotonically decreasing functions ϕ defined on $(0,1]$ that satisfy $\lim_{\delta\to 0}\phi(\delta) = \infty$ and $\phi(\delta) \geq 1$ for all $0 < \delta \leq 1$.

For fixed $t \in \mathbb{R}$, we want to characterize the stable LTI systems T for which the set

$$\mathcal{D}_1(T,t,\phi) := \{f \in \mathcal{PW}_\pi^1 : \limsup_{\delta\to 0} \frac{|(A_\delta^T f)(t)|}{\phi(\delta)} = \infty\}$$

is non-empty and, and in the case where $\mathcal{D}_1(T,t,\phi)$ is non-empty we are interested in structure of this set. The function $\phi \in \Phi$ is introduced in the above expression in order to describe the divergence speed of $(A_\delta^T f)(t)$.

Theorem 1. *Let T be an stable LTI system, $t \in \mathbb{R}$, and $\phi \in \Phi$. Then we have $\mathcal{D}_1(T,t,\phi) \neq \emptyset$ if and only if $\sum_{k=-\infty}^{\infty} |h_T(t-k)| = \infty$. Further, if $\sum_{k=-\infty}^{\infty} |h_T(t-k)| = \infty$ then $\mathcal{D}_1(T,t,\phi)$ is a residual set.*

Next, we want to apply Theorem 1. For the LTI system $T = Id$, where Id denotes the identity operator, we have $h_T = \text{sinc}$ and thus obtain, as a special case of (2), the sampling series (1), which is the Shannon sampling series that uses only the samples that are larger than or equal to the threshold δ. Since $\sum_{k=-\infty}^{\infty} |\text{sinc}(t-k)| = \infty$ for all $t \in \mathbb{R} \setminus \mathbb{Z}$, the next corollary is an immediate consequence of Theorem 1.

Corollary 1. *Let $t \in \mathbb{R} \setminus \mathbb{Z}$ and $\phi \in \Phi$. Then $\{f \in \mathcal{PW}_\pi^1 : \limsup_{\delta\to 0} \frac{|(A_\delta f)(t)|}{\phi(\delta)} = \infty\}$ is a residual set.*

The next corollary strengthens this assertion.

Corollary 2. *For every $\phi \in \Phi$, the set of functions $f \in \mathcal{PW}_\pi^1$, for which we have for all $t \in \mathbb{R} \setminus \mathbb{Z}$ $\limsup_{\delta\to 0} \frac{|(A_\delta f)(t)|}{\phi(\delta)} = \infty$, is a residual set.*

3. Behavior of the L^∞-Norm

Next we study the behavior of $\|A_\delta^T f\|_\infty$, i.e. the L^∞-norm of the approximation process, as the threshold δ is decreased to zero. The set of interest in this case is

$$\mathcal{D}_1^\infty(T,\phi) = \{f \in \mathcal{PW}_\pi^1 : \limsup_{\delta\to 0} \frac{\|A_\delta^T f\|_\infty}{\phi(\delta)} = \infty\}.$$

Theorem 2. *Let T be a stable LTI system and $\phi \in \Phi$. Then we have $\mathcal{D}_1^\infty(T,\phi) \neq \emptyset$ if and only if $h_T \notin \mathcal{B}_\pi^1$. Further, if $h_T \notin \mathcal{B}_\pi^1$ then $\mathcal{D}_1^\infty(T,\phi)$ is a residual set.*

Corollary 3. *Let T be a stable LTI system and $\phi \in \Phi$. Then we have $\mathcal{D}_1^\infty(T,\phi) \neq \emptyset$ if and only if $(\sum_{k=-\infty}^{\infty} |h_T(k)| = \infty$ or $\sum_{k=-\infty}^{\infty} \left|h_T\left(k+\frac{1}{2}\right)\right| = \infty)$. Moreover, if $\mathcal{D}_1^\infty(T,\phi) \neq \emptyset$ then $\mathcal{D}_1^\infty(T,\phi)$ is a residual set.*

4. Spaceability

A subset $D \subset \mathcal{PW}_\pi^1$ is called spaceable if there exists an infinite-dimensional closed linear subspace S_D of $\mathcal{PW}_\pi^1$ with $S_D \subset D$.

Theorem 3. *The set $\{f \in \mathcal{PW}_\pi^1 : \limsup_{\delta\to 0} \|A_\delta f\|_\infty = \infty\} \cup \{f \equiv 0\}$ is spaceable.*

For all $f \in S_D$, $f \not\equiv 0$, we have $\limsup_{\delta\to 0}\|A_\delta f\|_\infty = \infty$, i.e., every non-trivial function in the subspace S_D is not approximable by $A_\delta f$.

Open Problem. In general, oversampling is known to improve the convergence behavior of approximation processes. However, it is known that, for certain stable LTI systems, oversampling cannot remove the divergence of the approximation process with thresholding [2]. It is an interesting open problem to characterize the systems that can be stably approximated with approximation processes that use oversampling.

This work was partly supported by the German Research Foundation (DFG) under grant BO 1734/13-2. A journal paper that describes the results in more details is in preparation [1].

References

[1] Holger Boche and Ullrich J. Mönich, *On the Behavior of the Threshold Operator for Bandlimited Functions*, Journal of Fourier Analysis and Applications, in preparation.

[2] Holger Boche and Ullrich J. Mönich, *An impossibility result for linear signal processing under thresholding*, IEEE Transactions on Signal Processing, **58** (2010), 3, 1082–1094,

Fusion frames from matrices with the restricted isometry property

Bernhard G. Bodmann

(joint work with Jameson Cahill and Peter G. Casazza)

This talk investigates the use of matrices satisfying the restricted isometry property (RIP) in the theory of fusion frames. The design of fusion frames with desirable properties remains challenging [1, 3, 4]. We show that RIP matrices, when viewed as tight frames satisfying the restricted isometry property, give rise to nearly tight fusion frames. We also show how to replace parts of the RIP frame with orthonormal systems while maintaining the restricted isometry property.

Definition. A *frame* $\Phi = \{\phi_i\}_{i=1}^M$ for ℓ_N^2 is a spanning sequence. The *analysis operator* of Φ is $T : x \mapsto (\langle x, \phi_i\rangle)_{i=1}^M$ and the *frame operator* $S = T^*T$. The frame is called ϵ-*nearly tight* if the operator inequality $aI \le T^*T \le bI$ holds with *frame constants* $a = \frac{c}{1+\epsilon}, b = (1+\epsilon)c$ and some $c > 0$ and it is *unit norm* if $\|\phi_i\| = 1$ for all $i \in \{1, 2, \dots, M\}$.

A *Riesz (basic) sequence* $\{\phi_i\}_{i=1}^N$ is a linearly independent sequence with associated analysis operator T. It is called ϵ-Riesz if $aI \le TT^* \le bI$ with $a = \frac{1}{1+\epsilon}, b = 1+\epsilon$. A sequence $\Phi = \{\phi_i\}_{i=1}^M$ in ℓ_N^2 has ϵ-RIP for size $s \le N$ if for all $J \subset \{1, 2, \dots, M\}$, $|J| \le s$, $\{\phi_i\}_{i\in J}$ is ϵ-Riesz. An $N \times M$ matrix Φ whose columns form an ϵ-Riesz sequence is called an ϵ-RIP matrix.

We can construct Φ such that its columns can be partitioned into orthonormal systems, for example the "trivial" case of selecting random, uniformly distributed

[0]B.G.B. is in part supported by NSF grant DMS 1109545 and AFOSR grant FA9550-11-1-0245, J.C. by NSF grant DMS 1008183, and P.G.C. by NSF DMS 1008183, NSF ATD 1042701 and AFOSR grant FA9550-11-1-0245.

vectors on the unit sphere of a Hilbert space. The general case in which the orthonormal systems can have size larger than one can be interpreted as a fusion frame with subspaces that are spanned by the orthonormal systems.

A *fusion frame* $\{W_i, v_i\}_i \in K$, $W_i = \overline{W_i} \subset \mathcal{H}$, $v_i > 0$, for a Hilbert space $\mathcal{H}$ satisfies $a\|x\|^2 \leq \sum_{i\in K} v_i^2 \|P_i x\|^2 \leq b\|x\|^2 \forall x \in \mathcal{H}$ with P_i the orthogonal projection onto W_i and some $0 < a \leq b < \infty$. It is ϵ-nearly c-tight if $a = \frac{c}{1+\epsilon}, b = c(1+\epsilon)$.

The first main result addresses how coarse we can make the partition of a unit-norm tight frame into orthonormal systems.

Theorem ([2]). Let $\{\phi_i\}_{i=1}^M$ be a unit norm tight frame for ℓ_N^2 which is ϵ-RIP for sets of size s, and let $\{I_j\}_{j=1}^K$ be any partition of $\{1, 2, \ldots, M\}$ with $|I_j| \leq s$ and

$$W_j = \text{span}_{i\in I_j} \phi_i,$$

then $\{W_j, 1\}_{j=1}^K$ is a fusion frame with fusion frame bounds

$$\frac{M}{(1+\epsilon)N}, \ \frac{M(1+\epsilon)}{N}.$$

Moreover, if $L \subset \{1, 2, \ldots, K\}$ and for $j \in L$ we have $J_j \subset I_j$ with $\sum_{j=1}^K |J_j| \leq s$ then

$$\frac{1}{(1+\epsilon)^2} \sum_{j=1}^L \| \sum_{i\in J_j} a_i \varphi_i\|^2 \leq \| \sum_{j=1}^L \sum_{i\in J_j} a_i \varphi_i\|^2 \leq (1+\epsilon)^2 \sum_{j=1}^K \| \sum_{i\in J_j} a_i \varphi_i\|^2.$$

The above result on fusion frames concerns properties of *subspaces*. What can we say about spanning systems? We recall that for ϵ-Riesz $\{\phi_i\}_{i\in I_j}$ and $S_j = \sum_{i\in I_j} \phi_i \otimes \phi_i^*$, the system $\{S_j^{-1/2}|_{W_j} \phi_i\}_{i\in I_j}$ is orthonormal basis for W_j. The next main result answers the question whether we can we orthonormalize subsets of a frame in an RIP preserving way.

Theorem ([2]). Let $\{\varphi_i\}_{i=1}^M$ be a family of vectors in ℓ_N^2 having the restricted isometry property with constant $0 < \epsilon < 1$ for sets of size s. Partition $\{1, 2, \ldots, M\}$ into sets $\{I_j\}_{j=1}^K$ with $|I_j| \leq s$, and for $K_1 \leq K$, let $S_j = \sum_{i\in I_j} \phi_i \otimes \phi_i^*$ for $j \in \{1, 2, \ldots, K_1\}$, and

$$\{S_j^{-1/2} \varphi_i\}_{i\in I_j; j=1,2,\ldots,K_1} \cup \{\varphi_i\}_{i\in I_j : K_1+1 \leq j \leq K} =: \{\psi_i\}_{i=1}^M.$$

Given $J \subset \{1, 2, \ldots, M\}$ with $|J| \leq s$ and coefficients $\{a_i\}_{i\in J}$, then

$$\left[\frac{(1-\epsilon-\epsilon^2)^{1/2}}{1+\epsilon} - \epsilon\sqrt{K_1}\right]^2 \sum_{i\in J} |a_i|^2$$

$$\leq \| \sum_{i\in J} a_i \psi_i\|^2 \leq \left[(1+\epsilon)^{3/2} + \epsilon\sqrt{K_1}\right]^2 \sum_{i\in J} |a_i|^2.$$

In brief, after orthonormalizing K_1 of the subsets, lower and upper Riesz bounds for subsets of size s are

$$\left[\frac{(1-\epsilon-\epsilon^2)^{1/2}}{1+\epsilon} - \epsilon\sqrt{K_1}\right]^2$$

and

$$\left[(1+\epsilon)^{3/2}+\epsilon\sqrt{K_1}\right]^2 .$$

Corollary. If $K_1 < \frac{1}{\epsilon^2}\frac{1-\epsilon-\epsilon^2}{(1+\epsilon)^2}$ then we retain RIP. So, $K_1 = O(\epsilon^{-2})$.
Example. If $\epsilon = 0.2$, $K_1 = 13$; if $\epsilon = 0.1$, $K_1 = 73$.

References

[1] B. G. Bodmann, "Optimal linear transmission by loss-insensitive packet encoding," Appl. Comput. Harmon. Anal. **22** (2007) 274-285.
[2] B. G. Bodmann, J. Cahill and P. G. Casazza, "Fusion frames and the Restricted Isometry Property," Numerical Functional Analysis and Optimization, to appear.
[3] P. G. Casazza, G. Kutyniok, and S. Li, "Fusion Frames and Distributed Processing," Appl. Comput. Harmon. Anal. **25** (2008), 114–132.
[4] G. Kutyniok, A. Pezeshki, A.R. Calderbank, T. Liu, "Robust Dimension Reduction, Fusion Frames, and Grassmannian Packings," Appl. Comput. Harmon. Anal. **26** (2009), 64-76.

Faster than Nyquist but Slower than Tropp

Robert Calderbank

Well, who isn't? This talk described a modification to the Random Demodulator (RD) of Tropp et al. for sub-Nyquist sampling of frequency-sparse signals. It involves replacing the random waveform integral to the operation of the RD with a constrained random waveform with a slower switching rate. It was motivated by the difficulty in generating ideal fast switching waveforms. This Constrained Random Demodulator (CRD) provides a relaxation of the hardware requirements, a modest increase in addressable bandwidth, and a slight decrease in the recovery guarantees. The theoretical analysis also provides insight into how the statistics of the constrained random waveform should be chosen. Recovery of the input signal tones is improved if the power spectrum of the constrained waveform matches the distribution on the tones of the input signal (i.e., the distribution is proportional to the power spectrum).

This example shows the value of adaptation in compressed sensing and motivates an information theoretic criterion for designing measurement matrices in other applications. This criterion is maximizing the mutual information between the projected signal and the class label. Applying a recent theoretical result due to Palomar and Verdu (PV) on the gradient of mutual information, we show that this optimization problem can be solved directly using gradient descent, without requiring simplification of the objective function.

We then presented experiments where compressive sensing methods are applied to image reconstruction. We improved on standard methods and described a fast online method of projection design.

We considered two classes of results based upon random projection design. The random Gaussian Mixture Model (GMM) results shown below employ the patch-based CS construction and the learned GMM-based prior p(x). We also considered CS design in which the projections are performed directly on the entire image, rather than at the patch level. If one performs CS inversion based on traditional CS algorithms, which employ l1 and related regularization, the quality of the inversion is markedly worse than that using the proposed approach, with learned signal models p(x); we therefore do not show these results here, because they do not fit on the same scale as the results presented below. This is not surprising, as the patch-dependent signal model p(x) is much richer, and tailored to the data than is simple sparsity. To provide a fairer comparison, when performing inversion for the case in which the projections are performed directly on the entire image, we consider an underlying wavelet basis and perform inversion based on the sophisticated hidden Markov tree (HMT) wavelet model for images. This signal model p(x) could in principle also be used within the theory to design a projection matrix applicable to the entire image. However, the significant advantage of the GMM construction is that the posterior of the underlying signal may be constituted analytically, while for the HMT expensive computational methods are needed. Therefore, we only show HMT inversion results when the projection matrix is constituted at random, thereby providing a comparison of inversion quality of the GMM (patch based) and the HMT (entire image), based upon random projections.

We considered offline design of the patch-based projection matrix based upon the Renyi measure of entropy, as well as based upon mutual information. For online Renyi and Palomar-Verdu (PV) design, we do not make a simplifying single-Gaussian assumption when designing each row of M. By contrast the online PDS method uses the most probable Gaussian from the posterior to design the next projection at each step (this is therefore an approximation). The PDS method is very fast, while online PV is expensive, and therefore is shown principally for comparison (may not be done in practice, where online design must be fast).

First comparing the results based on random projections, the results based upon the (learned) patched-based GMM and based on the entire-image-based HMT are comparable in reconstruction quality. Sometimes the GMM results are slightly better, and other times the HMT results are better. However, there is no comparison with respect to computation speed. The HMT results are expensive, being based upon a Gibbs sampler. By contrast the GMM results are very fast, with the inversion analytic. The additional big advantage of the GMM representation is that it allows convenient design of patch-dependent projection matrices, which we consider next.

Each of the designed projection methods yield significant improvement relative to random, and after approximately 6 projections per patch we note that the online results are significantly better than offline design. For the first approximately 5 measurements per patch, the offline and online results are comparable; we attribute this to an inadequate number of measurements to obtain an accurate signal model, and therefore little gain manifested by adaptivity. However, after approximately

FIGURE 1. Barbara

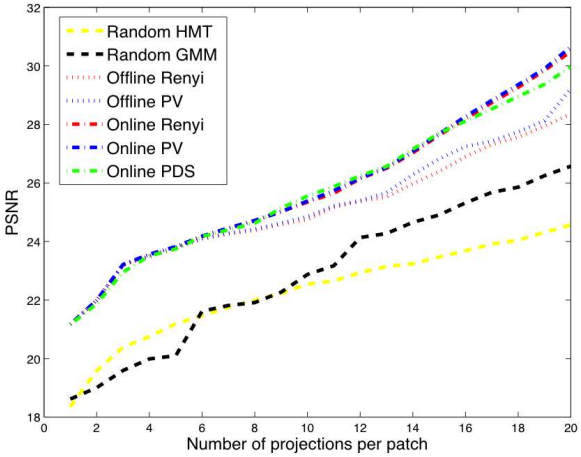


FIGURE 2. Reconstruction Results for Barbara

6 measurements per patch it appears that the posterior signal model becomes accurate, yielding advantages of adaptivity. Concerning online design, inversion quality based on the simple and fast online PDS performs quite competitively relative to the online Renyi and PV design (which do not make a simplification to a single Gaussian), despite the fact that it assumes that the patch is drawn from a single Gaussian.

Towards a Mathematical Theory of Super-Resolution

Emmanuel J. Candès
(joint work with Carlos Fernandez-Granda)

This talk develops a mathematical theory of super-resolution. Super-resolution is a word used in many different contexts; here, it is understood as the problem of recovering the fine details of an object—the high end of its spectrum—from coarse scale information only—from samples at the low end of the spectrum. In this extended abstract, we present typical results and refer to [3] for further details and references.

Suppose we have many point sources at unknown locations in $[0,1]$ and with unknown complex-valued amplitudes. Formally, our unknown signal x is of the form

$$x = \sum_j a_j \delta_{\tau_j},$$

where the amplitudes $a_j \in \mathbb{C}$ and the locations $\tau_j \in T \subset [0,1]$. We only observe Fourier samples of this object up until a frequency cut-off f_c; namely, we are given $n = 2f_c + 1$ low-frequency coefficients (Nyquist sampling)

$$y(k) = \int_0^1 e^{-i2\pi kt} x(\mathrm{d}t) = \sum_j a_j e^{-i2\pi k t_j}, \text{ where } k \in \mathbb{Z} \text{ and } |k| \le f_c. \tag{1}$$

For simplicity, we use matrix notation,

$$y = \mathcal{F}_n x$$

for (1). Informally, we have information about our object x up to a resolution limit equal to $\lambda_c = 1/f_c$ that we can understand as some sort of Rayleigh limit. Nevertheless, our work shows that one can super-resolve these point sources with infinite precision—i.e. recover the exact locations and amplitudes—by solving a simple convex optimization problem, which can be reformulated as a semidefinite program. This holds provided that the distance between sources is at least $2\lambda_c$.

To recover the signal, simply solve

$$\min \|\tilde{x}\|_{\mathrm{TV}} \quad \text{subject to} \quad \mathcal{F}_n \tilde{x} = y, \tag{2}$$

in which $\|\cdot\|_{\mathrm{TV}}$ is the classical total-variation norm and is the continuous analog of the ℓ_1 norm (e.g. with x as above, $\|x\|_{\mathrm{TV}} = \sum_j |a_j|$). Introduce the minimum distance defined as

$$\Delta(T) = \inf_{(t,t') \in T \,:\, t \neq t'} |t - t'|_\infty \qquad T \subset [0,1]. \tag{3}$$

Our main result is this:

Theorem 1 ([3])**.** *If the support T of the signal x obeys*

$$\Delta(T) \ge 2\,\lambda_c, \tag{4}$$

then the minimum total-variation norm solution is exact. For a real-valued signal x, a minimum distance of $1.87\lambda_c$ suffices.

In words, one can perfectly resolve the signal by simple convex programming provided that the spikes are sufficiently separated. This result extends to higher dimensions and other models. In one dimension for instance, it is possible to recover a piecewise smooth function by resolving the discontinuity points with infinite precision as well. The paper [3] also shows that the same result holds if the point sources are located in the plane, in 3D-space and so on.

At first sight, finding the solution to the total-variation norm problem (2) might seem quite challenging, as it requires solving an optimization problem over an infinite dimensional space. It is of course possible to approximate the solution by discretizing the support of the signal, but this could lead to an increase in complexity if the discretization step is reduced to improve precision. In this lecture, we show that (2) can be cast as a semidefinite program and demonstrate that highly accurate solutions can be found rather easily. This formulation is similar to that in [2] which concerns a related infinite dimensional convex program.

We also demonstrate that our approach is robust to noise. In particular, we develop some theoretical results explaining how the accuracy of the super-resolved signal is expected to degrade when both the noise level and the *super-resolution factor* vary. Suppose we have noisy data

$$y = \mathcal{F}_n x + w \quad \Longleftrightarrow \quad s = \mathcal{P}_n x + z$$

in which $\mathcal{P}_n = \mathcal{F}_n^* \mathcal{F}_n$ is the projection onto the first n Fourier modes; above $s = \mathcal{F}_n^* y$ and $z = \mathcal{F}_n^* w$. Assume we have bounded noise in the sense that

$$\|\mathcal{P}_n z\|_{\mathrm{TV}} \leq \delta,$$

and recover the signal by solving

$$\min \|\tilde{x}\|_{\mathrm{TV}} \quad \text{subject to} \quad \|s - \mathcal{P}_n \tilde{x}\|_{\mathrm{TV}} \leq \delta. \tag{5}$$

With noise, it is of course impossible to hope for perfect super-resolution. However, suppose we wish to resolve the signal up to a finite finer resolution $\lambda_f < \lambda_c$. It is fruitful to think of the ratio

$$\mathrm{SRF} = \frac{\lambda_c}{\lambda_f}$$

as a super-resolution factor. For instance, a value of $\mathrm{SRF} = 4$ means that we wish to extrapolate the spectrum by a factor of 4. How well then can re recover the finer features of x? Our second results is this:

Theorem 2 (C. and Fernandez Granda (2012))**.** *Suppose φ_{λ_f} is a low-pass filter that suppresses frequencies above $f = 1/\lambda_f$. Then under the same hypotheses as above, the solution to* (5) *obeys*

$$\|(\hat{x} - x) * \varphi_{\lambda_f}\|_{\mathrm{TV}} \leq C_0 \cdot \mathrm{SRF}^2 \cdot \delta,$$

where C_0 is a numerical constant.

This explains precisely how the error increases as the super-resolution factor increases.

The lecture will discuss other results and approaches to super-resolution, in particular [1, 4, 5, 6] with a special attention to the beautiful and recent work of

Kahane [7]. Finally, we will explain that when the distance between spikes falls below $\lambda_c/2$, the super-resolution problem becomes hopelessly ill posed. This is a consequence of Slepian's seminal work on prolate spheroidal functions [8] (see also [4]).

References

[1] Batenkov, D.; Yomdin, Y. Algebraic Fourier reconstruction of piecewise smooth functions. *Math. Comput.* **81** (2012), no. 277.

[2] Bhaskar, B. N.; Tang, G.; Recht, B. Atomic norm denoising with applications to line spectral estimation. Preprint.

[3] Candès, E. J.; Fernandez-Granda, C. (2012) "Towards a mathematical theory of super-resolution". To appear in *Comm. Pure Appl. Math.*

[4] Donoho, D. L. Superresolution via sparsity constraints. *SIAM J. Math. Anal.* **23** (1992), no. 5, 1309–1331.

[5] Dragotti, P.; Vetterli, M.; Blu, T. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix. *IEEE Transactions on Signal Processing* **55** (2007), no. 5, 1741–1757.

[6] Eckhoff, K. S. Accurate reconstructions of functions of finite regularity from truncated Fourier series expansions. *Math. Comput.* **64**

[7] Kahane, J. P. Analyse et synthèse harmoniques. Preprint. (1995), 671–690.

[8] Slepian, D. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *Bell System Technical Journal* **57** (1978), 1371–1430.

Spectral Tetris Constructions for Sensor Networks

Peter G. Casazza

Wireless sensor networks are an emerging technology which enables cost-effective and reliable surveillance. But there are serious problems with getting the needed information from the network. The sensors may have small batteries (which are constantly burning out), with small bandwidth. So we may need *substations* which are collecting information from the sensors and *passing it on to the next level*. We have transmission losses, quantization errors, overlapping information from the sensors and a host of other problems. Basically, the sensors represent vectors in a Hilbert space and they *sense* information by taking the inner product of the signal with the *sensor vectors*.

Fusion frames [5] (see also [6]) are a platform for dealing with the above problems but need much more development to be much more useful in this setting. In particular, fusion frames are designed to do local or distributed processing of a signal and *fusing* this information later to reconstruct the signal. They do this by projecting the signal onto the *fusion subspaces*, doing local processing and reconstructing. The idea is basically that we have a family of subspaces of a Hilbert space and by projecting a signal onto these subspaces we obtain a collection of *vector coefficients* that represent the signal uniquely. This then gives unique reconstruction from these vector coefficients.

Spectral Tetris is a recent breakthrough in the construction of (sparse) frames and fusion frames [2]. Until now, we have been living off *existence proofs* which

tell us what types of frames exist. However, in practice, we did not know how to find hardly any of these - except a few concrete constructions of specialized classes of frames. Spectral tetris is a general method for constructing equal norm tight frames [2] and some types of fusion frames [2]. It is also true that spectral tetris produces the sparsest tight frames in general [4]. Since its inception, spectral tetris has been generalized in many different directions to produce constructions of frames with arbitrary eigenvalues for the frame operator and arbitrary prescribed norms for the frame vectors - keeping in mind that there are restrictions on when these two can line up [1, 3, 3].

References

[1] R. Calderbank, P.G. Casazza, A. Heinecke, G. Kutyniok and A. Pezeshki, *Sparse fusion frames: Existence and construction*, Adv. Comput. Math. **35** (2011), 1-31.

[2] P.G. Casazza, M. Fickus, D.M. Dixon, Y. Wang and Z. Zhou, *Constructing tight fusion frames*, ACHA **30** (2011) 175–187.

[3] P.G. Casazza, A. Heinecke, and G. Kutyniok, *Optimally sparse fusion frames*, To Appear: IEEE Trans. Inform. Theory.

[4] P.G. Casazza, A. Heinecke, F. Krahmer and G. Kutyniok, *Optimally sparse frames*, Preprint.

[5] P.G. Casazza and G. Kutyniok, *Frames of Subspaces*, Contemp. Math. **345** (2004) 87–113.

[6] P.G. Casazza, G. Kutyniok and S. Li, *Fusion frames and distributed processing*, ACHA **25** (2008) 114-132.

Shearlet Coorbit Spaces: Traces and Embeddings in Higher Dimensions

Stephan Dahlke, Gabriele Steidl, Gerd Teschke

(joint work with Sören Häuser)

This papers examines structural properties of the recently developed shearlet coorbit spaces in higher dimensions. We prove embedding theorems for subspaces of shearlet coorbit spaces resembling shearlets on the cone in three dimensions into Besov spaces. The results are based on general atomic decompositions of Besov spaces. Furthermore, we establish trace results for these subspaces with respect to the coordinate planes. It turns out that in many cases these traces are contained in lower dimensional shearlet coorbit spaces.

1. Introduction

In recent years it has turned out that shearlets have the potential to retrieve directional information so that they became interesting for many applications, see [14, 17, 19]. Moreover, quite surprisingly, the shearlet transform has the outstanding property to stem from a square integrable group representation [2]. This remarkable fact provides the opportunity to design associated canonical smoothness spaces by applying the general coorbit theory derived by Feichtinger and Gröchenig [7, 8, 9, 12]. Indeed, in [3, 4] the above relationships have been clarified and new smoothness spaces, the so-called shearlet coorbit spaces, have been established. In particular, it has been shown that all the conditions needed in the

context of the coorbit space theory to obtain atomic decompositions and Banach frames can be satisfied in the shearlet setting.

2. Shearlets on $\mathbb{R}^d$

In this section, we recall basic results about the shearlet group on $\mathbb{R}^d$, $d \geq 2$, its square integrable representations and shearlet coorbit spaces from [4].

2.1. **Shearlet Group and Shearlet Transform.** For $a \in \mathbb{R}^* := \mathbb{R} \setminus \{0\}$ and $s \in \mathbb{R}^{d-1}$, let

$$A_a := \begin{pmatrix} a & 0_{d-1}^{\mathrm{T}} \\ 0_{d-1} & \operatorname{sgn}(a)|a|^{\frac{1}{d}} I_{d-1} \end{pmatrix} \quad \text{and} \quad S_s := \begin{pmatrix} 1 & s^{\mathrm{T}} \\ 0_{d-1} & I_{d-1} \end{pmatrix}$$

be the *parabolic scaling matrix* and the *shear matrix*, respectively, where $\operatorname{sgn}(a)$ denotes the sign of a. The *(full) shearlet group* $\mathbb{S}$ is defined to be the set $\mathbb{R}^* \times \mathbb{R}^{d-1} \times \mathbb{R}^d$ endowed with the group operation

$$(a, s, t)\,(a', s', t') = (aa', s + |a|^{1-1/d} s', t + S_s A_a t').$$

For the shearlet group the mapping $\pi : \mathbb{S} \to \mathcal{U}(L_2(\mathbb{R}^d))$ defined by

$$\pi(a, s, t)\,\psi(x) := |\det A_a|^{-\frac{1}{2}} \psi(A_a^{-1} S_s^{-1}(x - t)) \tag{1}$$

is a unitary representation of $\mathbb{S}$. The representation (1) is also *square integrable*, i.e., it is irreducible and there exists a nontrivial *admissible* function $\psi \in L_2(\mathbb{R}^d)$ fulfilling the *admissibility condition*

$$\int_{\mathbb{S}} |\langle f, \pi(a, s, t)\psi\rangle|^2 \, d\mu_{\mathbb{S}}(a, s, t) < \infty.$$

The function $\psi \in L_2(\mathbb{R}^d)$ is called an *admissible shearlet* and the transform $\mathcal{SH}_\psi : L_2(\mathbb{R}^d) \to L_2(\mathbb{S})$ defined by

$$\mathcal{SH}_\psi(f)(a, s, t) := \langle f, \pi(a, s, t)\psi\rangle,$$

continuous shearlet transform. It is known that there exist both band-limited and compactly supported shearlets, see [2, 5, 16, 18].

2.2. **Shearlet Coorbit Spaces.** Let w be a real-valued, continuous and submultiplicative weight on $\mathbb{S}$, i.e., $w(gh) \leq w(g)w(h)$ for all $g, h \in \mathbb{S}$. For $1 \leq p \leq \infty$, let

$$L_{p,w}(\mathbb{S}) := \{F \text{ measurable} : Fw \in L_p(\mathbb{S})\}.$$

Furthermore, we assume that the weight function w satisfies all the coorbit-theory conditions as stated in [12, Section 2.2]. A function contained in

$$\mathcal{A}_w := \{\psi \in L_2(\mathbb{R}^d) : \mathcal{SH}_\psi(\psi) = \langle \psi, \pi(\cdot)\psi\rangle \in L_{1,w}(\mathbb{S})\}$$

is called an *analyzing vector*. For an analyzing vector ψ we can consider the space

$$\mathcal{H}_{1,w} := \{f \in L_2(\mathbb{R}^d) : \mathcal{SH}_\psi(f) = \langle f, \pi(\cdot)\psi\rangle \in L_{1,w}(\mathbb{S})\},$$

with norm $\|f\|_{\mathcal{H}_{1,w}} := \|\mathcal{SH}_\psi(f)\|_{L_{1,w}(\mathbb{S})}$ and its anti-dual $\mathcal{H}_{1,w}^{\sim}$, the space of all continuous conjugate-linear functionals on $\mathcal{H}_{1,w}$. Then the inner product on

$L_2(\mathbb{R}^d) \times L_2(\mathbb{R}^d)$ extends to a sesquilinear form on $\mathcal{H}_{1,w}^{\sim} \times \mathcal{H}_{1,w}$. Therefore for $\psi \in \mathcal{H}_{1,w}$ and $f \in \mathcal{H}_{1,w}^{\sim}$ the *extended representation coefficients*

$$\mathcal{SH}_\psi(f)(a,s,t) := \langle f, \pi(a,s,t)\psi\rangle_{\mathcal{H}_{1,w}^{\sim}\times\mathcal{H}_{1,w}}$$

are well-defined.

Let m be a w-moderate weight on $\mathbb{S}$ We are interested in the following Banach spaces which are called *shearlet coorbit spaces*

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

To construct Banach frames in coorbit spaces, the following **b**etter subset $\mathcal{B}_w$ of $\mathcal{A}_w$ has to be non-empty:

$$\mathcal{B}_w := \{\psi \in L_2(\mathbb{R}^d) : \mathcal{SH}_\psi(\psi) \in \mathcal{W}^L(L_\infty, L_{1,w}(\mathbb{S}))\},$$

By Theorem [6, Theorem 1.1] there exist compactly supported shearlets $\psi \in L_2(\mathbb{R}^d)$ which are contained in $\mathcal{B}_w$ for certain weights w. The following theorem collects results about the existence of atomic decompositions and Banach frames from [3, 7].

Theorem 2. *Let $1 \le p \le \infty$ and $\psi \in \mathcal{B}_w$, $\psi \neq 0$. Then there exists a (sufficiently small) neighborhood U of e so that for any U-dense and relatively separated set $X = \{g_i = (a_i, s_i, t_i) : i \in \mathcal{I}\}$ the set $\{\pi(g_i)\psi)\}$ provides an atomic decomposition and a Banach frame for $\mathcal{SC}_{p,m}$:*

Atomic Decompositions*: If $f \in \mathcal{SC}_{p,m}$, then*

$$f = \sum_{i\in\mathcal{I}} c_i(f)\pi(g_i)\psi,$$

where the sequence of coefficients depends linearly on f and satisfies

$$\|(c_i(f))_{i\in\mathcal{I}}\|_{\ell_{p,m}} \lesssim \|f\|_{\mathcal{SC}_{p,m}}$$

with $\ell_{p,m}$ being defined by

$$\ell_{p,m} := \{c = (c_i)_{i\in\mathcal{I}} : \|c\|_{\ell_{p,m}} := \|c\, m\|_{\ell_p} < \infty\},$$

where $m = (m(g_i))_{i\in\mathcal{I}}$. Conversely, if $(c_i)_{i\in\mathcal{I}} \in \ell_{p,m}$, then $f = \sum_{i\in\mathcal{I}} c_i\pi(g_i)\psi$ is in $\mathcal{SC}_{p,m}$ and

$$\|f\|_{\mathcal{SC}_{p,m}} \lesssim \|(c_i)_{i\in\mathcal{I}}\|_{\ell_{p,m}}.$$

Banach Frames: *The set $\{\pi(g_i)\psi : i \in \mathcal{I}\}$ is a Banach frame for $\mathcal{SC}_{p,m}$ which means that*

i) $\|f\|_{\mathcal{SC}_{p,m}} \sim \|(\langle f, \pi(g_i)\psi\rangle_{\mathcal{H}_{1,w}^{\sim}\times\mathcal{H}_{1,w}})_{i\in\mathcal{I}}\|_{\ell_{p,m}}$,

ii) *there exists a bounded, linear reconstruction operator $\mathcal{R}$ from $\ell_{p,m}$ to $\mathcal{SC}_{p,m}$ such that $\mathcal{R}\left((\langle f, \pi(g_i)\psi\rangle_{\mathcal{H}_{1,w}^{\sim}\times\mathcal{H}_{1,w}})_{i\in\mathcal{I}}\right) = f$.*

3. Characterization of Besov Spaces and Coorbit Spaces

In the next section, we will show that traces of shearlet coorbit spaces onto certain hyperplanes are contained in Besov spaces or again in shearlet coorbit spaces. The proof of these trace theorems will heavily rely on the characterization

- of Besov spaces via atomic decompositions,
- of coorbit spaces via expansions of molecules.

The following subsections provide the results which will be necessary for our analysis.

3.1. Atoms in Besov Spaces.

For $\alpha > 1$, $D > 1$ and $K \in \mathbb{N}_0$, a K times differentiable function ϕ on $\mathbb{R}^d$ is called a *K-atom* if the following two conditions are fulfilled:

A1) $\operatorname{supp}\phi \subset DQ_{j,l}(\mathbb{R}^d)$ for some $l \in \mathbb{R}^d$,
where $DQ_{j,l}(\mathbb{R}^d)$ denotes the cube in $\mathbb{R}^d$ centered at $\alpha^{-j}l$ with sides parallel to the coordinate axes and side length $2\alpha^{-j}D$.

A2) $|\mathrm{D}^\gamma\phi(x)| \leq \alpha^{|\gamma|j}$ for $|\gamma| \leq K$.

Now the homogeneous Besov spaces can be characterized as follows.

Theorem 3. *Let $D > 1$ and $K \in \mathbb{N}_0$ with $K \geq 1 + \lfloor\sigma\rfloor$, $\sigma > 0$ be fixed. Let $1 \leq p \leq \infty$. Then $f \in B^\sigma_{p,q}$ if and only if it can be represented as*

$$f(x) = \sum_{j\in Z}\sum_{l\in\mathbb{Z}^d} \lambda(j,l)\phi_{j,l}(x), \tag{2}$$

where the $\phi_{j,l}$ are K-atoms with $\operatorname{supp}\phi_{j,l} \subset DQ_{j,l}(\mathbb{R}^d)$ and

$$\|f\|_{B^\sigma_{p,q}} \sim \inf\Big(\sum_{j\in\mathbb{Z}} \alpha^{j(\sigma-\frac{d}{p})q}\big(\sum_{l\in\mathbb{Z}^d}|\lambda(j,l)|^p\big)^{\frac{q}{p}}\Big)^{\frac{1}{q}}$$

where the infimum is taken over all admissible representations (2).

3.2. Molecules in Shearlet Coorbit Spaces.

Further, we will make use of the recently introduced molecules in general coorbit spaces, see [13]. We summarize the results needed from [13] for our shearlet coorbit spaces. Let $\psi \in \mathcal{B}_w$, $\psi \neq 0$ and let $X := \{g_i\}_{i\in\mathcal{I}}$ be a U-dense, relatively separated family in $\mathbb{S}$. A collection of functions $\{\phi_i\}_{i\in\mathcal{I}}$ from $L_2(\mathbb{R}^d)$ is called a *set of molecules*, if there exists an envelope function $H \in \mathcal{W}^R(L_\infty, L_{1,w}(\mathbb{S}))$ such that

$$|\mathcal{SH}_\psi(\phi_i)(g)| \leq H(g_i^{-1}g), \quad i \in \mathcal{I}.$$

This definition of the molecules does not depend on the particular choice of $\psi \in \mathcal{B}_w$. The following synthesis property was proved in [13] for general coorbit spaces.

Theorem 4. *Let $\{\phi_i\}_{i\in\mathcal{I}}$ be a set of molecules subordinated to $H \in \mathcal{W}^R(L_\infty, L_{1,w}(\mathbb{S}))$. If $(c_i)_{i\in\mathcal{I}} \in \ell_{p,m}$, $1 \leq p \leq \infty$, then $f := \sum_{i\in\mathcal{I}} c_i\phi_i \in \mathcal{SC}_{p,m}$ and $\|f\|_{\mathcal{SC}_{p,m}} \lesssim \|(c_i)_{i\in\mathcal{I}}\|_{\ell_{p,m}}$.*

4. Traces of Shearlet Coorbit Spaces

In this section, we are interested in traces of shearlet coorbit spaces. To keep the technicalities at a reasonable level, we restrict ourselves to the practically most important case of three dimensions. Moreover, we are only interested in weights

$$m(a,s,t) = m(a) := |a|^{-r}, r \geq 0$$

and use the abbreviations

$$\mathcal{SC}_{p,r} := \mathcal{SC}_{p,m} \quad \text{and} \quad \psi_{j,k,l} := \pi(a,s,t)\psi.$$

By Theorem 2, any $f \in \mathcal{SC}_{p,r}$ can be written as

$$f(x) = \sum_{j\in\mathbb{R}} \sum_{k\in\mathbb{R}^2} \sum_{l\in\mathbb{R}^3} c(j,k,l)\psi_{j,k,l}(x). \tag{3}$$

To derive reasonable trace and embedding theorems, it is necessary to introduce the following subspaces of $\mathcal{SC}_{p,r}$. For fixed $\psi \in B_w$, we denote by $\mathcal{SC}^{(\eta)}_{p,r}$, $\eta \in \{0,1\}^2$ the closed subspace of $\mathcal{SC}_{p,r}$ consisting of those functions which are representable as in (3) but with integers $|k_i| \leq \alpha^{\frac{2j}{3}}$ if $\eta_i = 1$. These subspaces resemble *shearlets on the cone*. We want to investigate the traces of functions lying in the subspaces $\mathcal{SC}^{(\eta)}_{p,r}$ with respect to the coordinate planes. For symmetry reasons we can restrict our attention to the x_1x_2-plane and to the x_2x_3-plane. We start with the latter one, where we prove that the traces are contained in Besov spaces.

Theorem 5. *Let $\mathrm{Tr}_{x_1} f$ denote the restriction of f to the x_2x_3-plane, that is, $(\mathrm{Tr}_{x_1} f)(x_2,x_3) := f(0,x_2,x_3)$. Then the embedding $\mathrm{Tr}_{x_1}(\mathcal{SC}^{(1,1)}_{p,r}(\mathbb{R}^3)) \subset B^{\sigma_1}_{p,p}(\mathbb{R}^2) + B^{\sigma_2}_{p,p}(\mathbb{R}^2)$ holds true, where $\sigma_1 + 2\lfloor\sigma_1\rfloor = 3r - \frac{21}{2} + \frac{8}{p}$ and $\sigma_2 = 3r - \frac{5}{2} + \frac{2}{p}$.*

The proof uses atomic decomposition of shearlet coorbit spaces. Let us now turn to traces on the x_1x_2-plane. In this case the shear parameter will play an additional role so that the traces will again be contained in shearlet coorbit spaces.

Theorem 6. *Let $\mathrm{Tr}_{x_3} f$ denote the restriction of f to the x_1x_2-plane, that is, $(\mathrm{Tr}_{x_3} f)(x_1,x_2) := f(x_1,x_2,0)$. Then $\mathrm{Tr}_{x_3}(\mathcal{SC}^{(0,1)}_{p,r}(\mathbb{R}^3)) \subset \mathcal{SC}_{p,r_1}(\mathbb{R}^2) + \mathcal{SC}_{p,r_2}(\mathbb{R}^2)$, where $r_1 = r - \frac{5}{6} + \frac{2}{3p}$ and $r_2 = r - \frac{1}{6}$.*

The proof uses molecules in shearlet coorbit spaces.

5. Embeddings into Besov Spaces

In this section, we prove the following embedding result of certain subspaces of shearlet coorbit spaces in three dimensions into (sums of) homogeneous Besov spaces. We like to mention that embedding results in Besov spaces have also been shown for the curvelet setting by Borup and Nielsen [1]. However, the technique used by these authors is completely different since they work in the frequency domain.

Theorem 7. *The embedding* $\mathcal{SC}_{p,r}^{(1,1)}(\mathbb{R}^3) \subset B_{p,p}^{\sigma_1}(\mathbb{R}^3) + B_{p,p}^{\sigma_2}(\mathbb{R}^3)$*, holds true, where*

$$\sigma_1 + 2\lfloor \sigma_1 \rfloor = 3r - \frac{21}{2} + \frac{9}{p} \quad \textit{and} \quad \sigma_2 - \frac{2}{3}\lfloor \sigma_2 \rfloor = r + \frac{5}{3p} + \frac{7}{6}.$$

The proof uses again atomic decomposition of shearlet coorbit spaces.

References

[1] L. Borup and M. Nielsen, *Frame decomposition of decomposition spaces*, J. Fourier Anal. Appl. 13/1, 39 - 70 (2007).

[2] 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., 6, 157 - 181 (2008).

[3] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke, *Shearlet coorbit spaces and associated Banach frames*, Appl. Comput. Harmon. Anal. 27/2, 195 - 214 (2009).

[4] S. Dahlke, G. Steidl, and G. Teschke, *The continuous shearlet transform in arbitrary space dimensions*, J. Fourier Anal. App. 16, 340 - 354 (2010).

[5] S. Dahlke, G. Steidl, and G. Teschke, *Shearlet coorbit spaces: compactly supported analyzing shearlets, traces and embeddings*, J. Fourier Anal. Appl., to appear.

[6] S. Dahlke, S. Häuser, G. Steidl, and G. Teschke, *Shearlet coorbit spaces: traces and embeddings*, Monatshefte für Mathematik, 1–18 (2012).

[7] H. G. Feichtinger and K. Gröchenig, *A unified approach to atomic decompositions via integrable group representations*, Proc. Conf. "Function Spaces and Applications", Lund 1986, Lecture Notes in Math. 1302 (1988), 52 - 73.

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

[9] H. G. Feichtinger and K. Gröchenig, *Banach spaces related to integrable group representations and their atomic decomposition II*, Monatsh. Math. 108, 129 - 148 (1989).

[10] M. Frazier and B. Jawerth, *Decomposition of Besov spaces*, Indiana University Mathematics Journal 34/4, 777 - 799 (1985).

[11] K. Gröchenig, *Foundations of Time–Frequency Analysis*, Birkhäuser, Boston, Basel, Berlin, 2001.

[12] K. Gröchenig, *Describing functions: Atomic decompositions versus frames*, Monatsh. Math. 112, 1 - 42 (1991).

[13] K. Gröchenig and M. Piotrowski. *Molecules in coorbit spaces and boundedness of operators.* Studia Math., 192(1):61 - 77, 2009.

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

[15] L.I. Hedberg and Y. Netrusov, *An axiomatic approach to function spaces, spectral synthesis, and Luzin approximation*, Memoirs of the American Math. Soc. 188, 1- 97 (2007).

[16] P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Construction of compactly supported shearlet frames*, Constructive Approximation, 1–52 (2010).

[17] G. Kutyniok and D. Labate, *Resolution of the wavefront set using continuous shearlets*, American Mathematical Society, 361(5), 2719–2754 (2009).

[18] G. Kutyniok, J. Lemvig, and W.-Q. Lim, *Compactly supported shearlets*, Approximation Theory XIII: San Antonio 2010, 163–186 (2012).

[19] G. Kutyniok and D. Labate. Shearlets: The First Five Years. Oberwolfach Report 44 (2010), 1-5.

[20] C. Schneider, *Besov spaces of positive smoothness*, PhD thesis, University of Leipzig, 2009.

[21] H. Triebel, *Function Spaces I*, Birkhäuser, Basel - Boston - Berlin, 2006

The Analysis Sparse Model - Definition, Pursuit, Dictionary Learning, and Beyond

Michael Elad

The synthesis-based sparse representation model for signals has drawn a considerable interest in the past decade. Such a model assumes that the signal of interest can be decomposed as a linear combination of a few atoms from a given dictionary. In this talk we concentrate on an alternative, analysis-based model, where an analysis operator – hereafter referred to as the "Analysis Dictionary" - multiplies the signal, leading to a sparse outcome. While the two alternative models seem to be very close and similar, they are in fact very different. In this talk we define clearly the analysis model and describe how to generate signals from it. We discuss the pursuit denoising problem that seeks the zeros of the signal with respect to the analysis dictionary given noisy measurements. Finally, we explore ideas for learning the analysis dictionary from a set of signal examples. We demonstrate this model's effectiveness in several experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis dictionary.
Joint work with Ron Rubinstein (former PhD student), Tomer Peleg (PhD student), Remi Gribonval and Sangnam Nam (INRIA, Rennes), and Mike Davies (UEdin). Relevant papers are [1, 2, 3]

References

[1] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, *The Cosparse Analysis Model and Algorithms*, to appear in Applied and Computational Harmonic Analysis (ACHA).
[2] R. Rubinstein, T. Peleg and M. Elad, *Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model*, submitted to IEEE Trans. on Signal Processing.
[3] T. Peleg and M. Elad, *Performance Guarantees of the Thresholding Algorithm for the Co-Sparse Analysis Model*, submitted to IEEE Trans. on Information Theory.

Hyperspectral imaging, compressed sensing and chromo-tomography

Matthew Fickus

Traditional hyperspectral imaging involves taking an enormous number of measurements. However, the theory of compressed sensing (CS) gives hope that far fewer measurements actually need to be taken, provided the scene is sufficiently simple. Indeed, several CS-based hyperspectral imagers have recently been proposed. Such imagers would be invaluable in situations where conventional imaging systems have insufficient temporal resolution (i.e., imaging extremely brief events). We discuss the basic principles of the proposed CS-based systems. We then discuss a competing hyperspectral imaging technology that also promises to drastically reduce the number of needed measurements: chromo-tomography, namely applying ideas from tomography to color.

Sparse optimal control in consensus emergence

Massimo Fornasier

We consider the dynamical system provided by the Cucker-Smale model of *consensus emergence*

$$\begin{cases} \dot{x}_i = v_i \in \mathbb{R}^d \\ \dot{v}_i = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} \frac{v_j - v_i}{(1+\|x_j - x_i\|^2)^\beta} \in \mathbb{R}^d \end{cases},$$

where $\beta > 0$ governs the rate of communication between particles. Denote $a(t) := a_\beta(t) = \frac{1}{(1+t^2)^\beta}$. In compact matrix form

$$\begin{cases} \dot{x} = v \\ \dot{v} = -L_x v \end{cases}$$

where L_x is the Laplacian of the matrix[1] $(a(\|x_j - x_i\|)/N)_{i,j=1}^N$ and depends on x. Without loss of generality $\bar{v} = 0$ and $\bar{x}(t) = \bar{x}(0) = \frac{1}{N}\sum_{i=1}^N x_i(0)$. Whenever the rate of communication is sufficiently strong or, in other words, β is sufficiently small, the system converges to a consensus in the parameter v independently of the initial condition.

Theorem 1 (Cucker-Smale, Ha-Tadmor, Carrillo-F.-Rosado-Toscani)**.** *Let $(x(t), v(t)) \in C^1([0,+\infty), \mathbb{R}^{2d\times N})$ be the solution of the Cucker-Smale system. We denote*

$$\mathcal{V}(t) = \max_{i=1,\dots N} \|v_i(t)\|, \quad \mathcal{V}_0 = \mathcal{V}(0).$$

If $0 < \beta < \frac{1}{2}$ then

$$\mathcal{V}(t) \le \mathcal{V}_0 e^{-a(2\bar{\mathcal{X}})t} \to 0, t \to \infty, \quad \exists \bar{\mathcal{X}} > 0.$$

Actually one has $\mathcal{V}(t) \to 0$ also for $\beta = 1/2$.

When $\beta > 1/2$ then one has convergence to consensus only for specific initial configurations which are representing certain concentration properties of the system. Consider the symmetric bilinear form

$$B(u,v) = \frac{1}{2N^2}\sum_{i,j}\langle u_i - u_j, v_i - v_j\rangle = \frac{1}{N}\sum_{i=1}^N \langle u_i, v_i\rangle - \langle \bar{u}, \bar{v}\rangle,$$

and

$$X(t) = B(x(t), x(t)), \quad V(t) = B(v(t), v(t)).$$

Then the following result gives sufficient conditions for consensus.

[1]The *Laplacian* L of A is given by $L = D - A$, with $D = \mathrm{diag}(d_1, \dots, d_N)$ and $d_k = \sum_{j=1}^N a_{kj}$

Theorem 2 (Ha-Ha-Kim)**.** *Let* $(x_0, v_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ *be such that* $X_0 = B(x_0, x_0)$ *and* $V_0 = B(v_0, v_0)$ *satisfy*

$$\sqrt{N} \int_{\sqrt{NX_0}}^{\infty} a(\sqrt{2}r)dr > \sqrt{V_0}\,.$$

Then the solution with initial data (x_0, v_0) *tends to consensus.*

If $\beta > 1/2$ then the latter consensus condition is *not* satisfied by all $(x_0, v_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$. In particular there are counterexamples to consensus emergence. In such cases where neither $\beta \leq 1/2$ nor the sufficient concentration properties for consensus emergence are satisfied, we wonder whether one can "help" the system to achieve consensus by external intervention. We consider admissible controls, given by measurable functions $u = (u_1, \ldots, u_N) : [0, +\infty) \to \mathbb{R}^N$ such that $\sum_{i=1}^N \|u_i(t)\| \leq M$ for every $t > 0$, for a given constant M:

$$\begin{cases} \dot{x}_i = v_i \\ \dot{v}_i = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N} a(\|x_j - x_i\|)(v_j - v_i) + u_i \end{cases}$$

for $i = 1, \ldots, N$, and $x_i \in \mathbb{R}^d$, $v_i \in \mathbb{R}^d$. The aim is then to find admissible controls steering the system to the consensus region.

Proposition 1 (Caponigro-Fornasier-Piccoli-Trélat). For every initial condition $(x_0, v_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ and $M > 0$ there exist $T > 0$ and $u : [0, T] \to (\mathbb{R}^d)^N$, with $\sum_{i=1}^N \|u_i(t)\| \leq M$ for every $t \in [0, T]$ such that the associated solution tends to consensus.

Proof. Consider a solution of the system with initial data (x_0, v_0) associated with a feedback control $u = -\alpha(v - \bar{v})$, with $0 < \alpha \leq M/(N\sqrt{B(v_0, v_0)})$. Then

$$\begin{aligned} \frac{d}{dt}V(t) &= \frac{d}{dt}B(v(t), v(t)) \\ &= -2B(L_x v(t), v(t)) + 2B(u(t), v(t)) \\ &\leq 2B(u(t), v(t)) = -2\alpha B(v - \bar{v}, v - \bar{v}) = -2\alpha V(t). \end{aligned}$$

Therefore $V(t) \leq e^{-2\alpha t}V(0)$ and $V(t)$ tends to 0 exponentially fast as $t \to \infty$. Moreover $\sum_{i=1}^N \|u_i\| \leq M$. □

However the strategy described in the proof above requires to activate a control on *every* agent, and we wonder whether more economical choices are possible. In particular we wish to make

$$\begin{aligned} \frac{d}{dt}V(t) &= \frac{d}{dt}B(v(t), v(t)) \\ &= -2B(L_x v(t), v(t)) + 2B(u(t), v(t)) \end{aligned}$$

the smallest possible and use the *minimal amount of intervention*. This leads to minimize $B(u(t), v(t))$ with additional *sparsity constraints* on u. We propose then the following greedy strategy.

Theorem 3 (Caponigro-Fornasier-Piccoli-Trélat)**.** *For every initial condition* $(x_0, v_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ *and* $M > 0$ *there exist* $T > 0$ *and a* sparse *control* $u : [0,T] \to (\mathbb{R}^d)^N$*, with* $\sum_{i=1}^N \|u_i(t)\| \leq M$ *for every* $t \in [0,T]$ *such that the associated solution tends to consensus. More precisely, we can choose adaptively the control law explicitly as one of the solutions of the variational problem*

$$\min B(v,u) + \gamma(x) \sum_{i=1}^N \|u_i\| \quad \text{subject to } \sum_{i=1}^N \|u_i\| \leq M\,,$$

where

$$\gamma(x) = \sqrt{N} \int_{\sqrt{NB(x,x)}}^{\infty} a(\sqrt{2}r)dr.$$

The control $u(t)$ *is a sparse vector with at most one nonzero coordinate, i.e.,* $u_i(t) \neq 0$ *for a unique* $i \in \{1,\dots,N\}$ *and* $u_j(t) = 0$ *for* $j \neq i$ *for almost every* $t \in [0,T]$*.*

Denote $v_\perp = v - \bar{v}$. Let us construct explicitly the control law from the variational principle described above. If $B(v,v) \leq \gamma(x)^2$, then $\|v_{\perp_i}\| \leq \gamma(x)$ for every $i = 1,\dots,N$ and

$$u_1 = \cdots = u_N = 0 \Rightarrow \text{reached flocking region.}$$

If $B(v,v) > \gamma(x)^2$ then there exists $i \in \{1,\dots,N\}$ such that

$$\|v_{\perp_i}\| > \gamma(x) \quad \text{and} \quad \|v_{\perp_i}\| \geq \|v_{\perp_j}\| \quad \text{for every } j = 1,\dots,N.$$

Therefore we can choose $i \in \{1,\dots,N\}$ satisfying it, and a control law

$$u_i = -M \frac{v_{\perp_i}}{\|v_{\perp_i}\|}, \quad \text{and} \quad u_j = 0, \quad \text{for every } j \neq i.$$

Hence the control acts instantaneously on the most "stubborn" agent. This choice of the control makes $V(t) = B(v(t),v(t))$ vanishing in finite time, hence there exists T such that $B(v(t),v(t)) \leq \gamma(x)^2$, $t \geq T$, and it maximizes the rate of convergence among all the possible feedback control strategies. *This can be viewed as a mathematical description of the general principle for which a policy maker should consider more favorable, in order to obtain consensus, to intervene with stronger actions on the fewest possible instantaneous optimal leaders than trying to control more agents, with minor strength.*

References

[1] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat, Sparse stabilization and optimal control of the Cucker–Smale model *preprint*, 2012

[2] F. Cucker and S. Smale. On the mathematics of emergence. *Jpn. J. Math.*, 2(1):197–227, 2007.

[3] S.-Y. Ha, T. Ha, and J.-H. Kim. Emergent behavior of a Cucker-Smale type particle model with nonlinear velocity couplings. *IEEE Trans. Automat. Control*, 55(7):1679–1683, 2010.

Gabor Frames and Totally Positive Functions

Karlheinz Gröchenig
(joint work with Joachim Stöckler)

The fundamental problem of Gabor analysis is to determine triples (g, α, β) consisting of an L^2-function g and lattice parameters $\alpha, \beta > 0$, such that the set of functions $\mathcal{G}(g, \alpha, \beta) = \{e^{2\pi i \beta l t} g(t - \alpha k) : k, l \in \mathbb{Z}\}$ constitutes a frame for $L^2(\mathbb{R})$. Thus the fundamental problem is to determine the set (the *frame set*)

$$\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}^2_+ : \mathcal{G}(g, \alpha, \beta) \text{ is a frame } \}\,. \tag{1}$$

Under mild conditions, precisely, if g is in the Feichtinger algebra M^1, then the set $\mathcal{F}(g)$ is open in $\mathbb{R}^2_+$. Furthermore, if $g \in M^1$, then $\mathcal{F}(g)$ contains a neighborhood U of 0 in $\mathbb{R}^2_+$. Much effort has been spent to improve the analytic estimates and make this neighborhood as large as possible, see e.g., [1, 6]. The fundamental density theorem asserts that $\mathcal{F}(g)$ is always a subset of $\{(\alpha, \beta) \in \mathbb{R}^2_+ : \alpha\beta \leq 1\}$ [2]. If $g \in M^1$, then a subtle version of the uncertainty principle, the so-called Balian-Low theorem, states that $\mathcal{F}(g) \subseteq \{(\alpha, \beta) : \alpha\beta < 1\}$. This means that $\{(\alpha, \beta) : \alpha\beta \leq 1\}$ is the maximal set that can occur as a frame set $\mathcal{F}(g)$.

Until now, the catalogue of windows g for which $\mathcal{F}(g)$ is completely known, consists of the following functions: if g is either the Gaussian $g(t) = e^{-\pi t^2}$, the hyperbolic secant $g(t) = (e^t + e^{-t})^{-1}$, the exponential function $e^{-|t|}$, then $\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}^2_+ : \alpha\beta < 1\}$; if g is the one-sided exponential function $g(t) = e^{-t}\chi_{\mathbb{R}^+}(t)$, then $\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}^2_+ : \alpha\beta \leq 1\}$. In addition, the dilates of these functions and their Fourier transforms, $g(t) = (1 + 2\pi i t)^{-1}$ and $g(t) = (1 + 4\pi^2 t^2)^{-1}$, also have the same frame set. The case of the Gaussian was solved independently by Lyubarski [5] and Seip [8] in 1990 with methods from complex analysis. The case of the hyperbolic secant can be reduced to the Gaussian with a trick of Janssen and Strohmer, the case of the exponential functions is due to Janssen. We note that in all these cases the necessary density condition $\alpha\beta < 1$ (or $\alpha\beta \leq 1$) is also sufficient for $\mathcal{G}(g, \alpha, \beta)$ to generate a frame.

The example of the Gaussian lead Daubechies to conjecture that

$$\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}^2_+ : \alpha\beta < 1\}$$

whenever g is a positive function in L^1 with positive Fourier transform in L^1 [1, p. 981]. This conjecture was disproved by Janssen [4].

In our talk we treated a modification of Daubechies' conjecture and explained that the frame set of a large class of functions is indeed the maximal set $\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}^2_+ : \alpha\beta < 1\}$.

The main observation is that all functions above — the Gaussian, the hyperbolic secant, and the exponential functions — are *totally positive functions*. This means that for every two sets of increasing real numbers $x_1 < x_2 < \cdots < x_N$ and $y_1 < y_2 < \cdots < y_N$, $N \in \mathbb{N}$, the determinant of the matrix $[g(x_j - y_k)]_{1 \leq j,k \leq N}$ is non-negative.

Indeed, for a large class of totally positive functions to be defined in (2) the frame set $\mathcal{F}(g)$ can be determined completely. Based on Schoenberg's classification of totally positive functions [7], f is called totally positive of finite type M, if its Fourier transform factors as

$$\hat{f}(\xi) = C \prod_{j=1}^{M} (1 + 2\pi i \delta_j \xi)^{-1}, \tag{2}$$

for $M \in \mathbb{N}, C > 0$, and $\delta_k \in \mathbb{R}$.

Theorem 1. *Assume that $g \in L^2(\mathbb{R})$ is a totally positive function of finite type ≥ 2. Then $\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}_+^2 : \alpha\beta < 1\}$. In other words, $\mathcal{G}(g, \alpha, \beta)$ is a frame, if and only if $\alpha\beta < 1$.*

Our theorem increases the number of functions with known frame set from six to uncountable. Among the examples of totally positive functions of finite type are the two-sided exponential $e^{-|t|}$ (already known), the truncated power functions $g(t) = e^{-t} t^r \chi_{\mathbb{R}_+}$ for $r \in \mathbb{N}$, the function $g(t) = (e^{-at} - e^{-bt})\chi_{\mathbb{R}^+}(t)$ for $a, b > 0$, or the asymmetric exponential $g(t) = e^{at}\chi_{\mathbb{R}^+}(-t) + e^{-bt}\chi_{\mathbb{R}^+}(t)$, and the convolutions of totally positive functions of finite type.

Using a partial fraction decomposition of (2), one can obtain explicit formulas for totally positive functions of finite type and their Zak transform.

To compare with Daubechies' original conjecture, we note that every totally positive and even function possesses a positive Fourier transform. Theorem 1 yields a large class of functions for which Daubechies' conjecture is indeed true. Furthermore, Theorem 1 suggests the modified conjecture that the frame set of every continuous totally positive function is $\mathcal{F}(g) = \{(\alpha, \beta) \in \mathbb{R}_+^2 : \alpha\beta < 1\}$.

As a corollary of the theorem one obtains results about Gabor frames on $\ell^2(\mathbb{Z})$ (corresponding to discrete signals) and on $L^2(\mathbb{T})$ (corresponding to periodic signals) by sampling and periodizing.

This is a report on work in progress. The first results are available in the preprint [3].

References

[1] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory*, 36(5):961–1005, 1990.

[2] K. Gröchenig. *Foundations of time-frequency analysis*. Birkhäuser Boston Inc., Boston, MA, 2001.

[3] K. Gröchenig, J. Stöckler. Gabor Frames and Totally Positive Functions. Preprint. ArXive ???

[4] A. J. E. M. Janssen. Some counterexamples in the theory of Weyl-Heisenberg frames. *IEEE Trans. Inform. Theory*, 42(2):621–623, 1996.

[5] Y. I. Lyubarskiĭ. Frames in the Bargmann space of entire functions. In *Entire and subharmonic functions*, pages 167–180. Amer. Math. Soc., Providence, RI, 1992.

[6] A. Ron and Z. Shen. Weyl–Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$. *Duke Math. J.*, 89(2):237–282, 1997.

[7] I. J. Schoenberg. On Pólya frequency functions, I. The totally positive functions and their Laplace transforms. *J. Analyse Math.*, 1: 331–374, 1951.

[8] K. Seip. Density theorems for sampling and interpolation in the Bargmann-Fock space. I. *J. Reine Angew. Math.*, 429:91–106, 1992.

On the Structure of Anisotropic Frames

Philipp Grohs
(joint work with Gitta Kutyniok)

In recent years the field of geometric multiscale analysis has seen a flurry of activity, specifically in the construction of representation systems which are capable of efficiently encoding anisotropic structures such as edges in images.

This effort has brought forward a whole zoo of different constructions (among which we mention curvelets and shearlets) all possessing similar properties. This latter fact is usually proven by a rather complicated case-by-case study for each seperate system.

The motivation of [2] is the question whether such a cumbersome repetition of proofs is really necessary or if a higher-level viewpoint could be useful for a unification and better understanding of these previous results.

This has led to the introduction of *parabolic molecules* in [2] which can be defined as certain systems of functions $(m_\lambda)_{\lambda\in\Lambda}$ each one having frequency support in parabolic wedges associated to an orientation and spatial support in rectangles with parabolic aspect ratio

$$\text{length} \sim \text{width}^2.$$

Each index λ is associated with a scale $s_\lambda \in \mathbb{R}_+$, a direction $\theta_\lambda \in (-\pi, \pi]$ and a location $x_\lambda \in \mathbb{R}^2$ which can be formalized by introducing a so-called *parametrization* $\Phi_\Lambda : \Lambda \to \mathbb{P} := \mathbb{R}_+ \times (-\pi, \pi] \times \mathbb{R}^2,\ \lambda \mapsto (s_\lambda, \theta_\lambda, x_\lambda)$.

The formal definition of parabolic molecules is as follows: We shall write $\langle x\rangle := (1+x^2)^{1/2}$, $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ the rotation matrix of angle θ and $D_a := \mathrm{diag}(a, \sqrt{a})$ for the anisotropic dilation matrix associated with $a > 0$.

Definition 1. Let Φ_Λ be a parametrization. A family $(m_\lambda)_{\lambda\in\Lambda}$ of functions $m_\lambda \in L_2(\mathbb{R}^2)$ is called a *family of parabolic molecules* of order (R, M, N_1, N_2) if it can be written as

$$m_\lambda(x) = 2^{3s_\lambda/4} a^{(\lambda)} \left(D_{2^{s_\lambda}} R_{\theta_\lambda} (x - x_\lambda)\right)$$

such that

$$\left|\partial^\beta \hat{a}^{(\lambda)}(\xi)\right| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2}$$

for all $|\beta| \le R$. The implicit constants are uniform over $\lambda \in \Lambda$.

The two main findings of [2] are that

(i) All known curvelet or shearlet systems in the literature are systems of parabolic molecules.
(ii) Any two systems of parabolic molecules possess equivalent approximation properties.

To show (ii) we consider a specific *index distance* $\omega : \mathbb{P} \times \mathbb{P} \to \mathbb{R}_+$ and prove that the cross Gramian matrix of any two systems of parabolic molecules possesses strong off-diagonal decay in terms of ω. More specifically we show the following

Theorem 2. *Let $(m_\lambda)_{\lambda\in\Lambda}$, $(p_\mu)_{\mu\in M}$ be two systems of parabolic molecules of order (R, M, N_1, N_2) with*

$$R \geq 2N, \quad M > 4N - \frac{5}{4}, \quad N_1 \geq 2N + \frac{3}{4}, \quad N_2 \geq 2N.$$

Then

$$|\langle m_\lambda, p_\mu\rangle| \lesssim \omega\left((s_\lambda, \theta_\lambda, x_\lambda), (s_\mu, \theta_\mu, x_\mu)\right)^{-N}.$$

Using this "almost orthogonality" property between different systems of parabolic molecules one can transfer approximation results between two such systems. For instance one can readily establish the celebrated approximation results of [1] for any system of parabolic molecules.

Theorem 3. *Assume that $(m_\lambda)_{\lambda\in\Lambda}$ is a system of parabolic molecules of order (R, M, N_1, N_2) such that*

(1) *$(m_\lambda)_{\lambda\in\Lambda}$ constitutes a frame for $L_2(\mathbb{R}^2)$,*
(2) *Λ is k-admissible for all $k > 2$ (this is in particular true for the shearlet parametrization),*
(3) *it holds that*

$$R \geq 6, \quad M > 12 - \frac{5}{4}, \quad N_1 \geq 6 + \frac{3}{4}, \quad N_2 \geq 6.$$

Then the frame $(m_\lambda)_{\lambda\in\Lambda}$ possesses an almost best N-term approximation rate of order $N^{-1+\varepsilon}$, $\varepsilon > 0$ arbitrary for cartoon images (i.e. compactly supported, bivariate, piecewise C^2 functions with C^2 curve singularities).

In particular this includes results from [3, 4].

References

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

[2] P. Grohs and G. Kutyniok, *Parabolic Molecules*, submitted (2012). Available as arXiv preprint arXiv:1206.1958v1.

[3] K. Guo and D. Labate, *Optimally sparse multidimensional representation using shearlets*, SIAM J. Math. Analy., 39:298-318, (2007).

[4] G. Kutyniok and W.-Q. Lim, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory, 163(11):1564-1589, (2011).

Topological Aspects in Time-Frequency Analysis

Mijail Guillemard

We consider the problem of studying the components f_ℓ of a signal $f = \sum_\ell f_\ell$ using a combination of topological methods with time-frequency analysis. The basic illustration for our setting is the field of polyphonic audio analysis, where various sounds f_ℓ, originated from different simultaneous events, are recorded in a single structure f. For many applications, extracting from f the components f_ℓ can be a difficult task, and a direct usage of modern harmonic analysis tools such as wavelets, Gabor analysis, or frame theory can be insufficient. Our objective is to study f using geometrical and topological properties of a set $X_f = \{x_i\}_i$ defined by collecting local information of f as in the context of time-frequency analysis.

In **time-frequency analysis** [4] a signal f is analyzed by considering a partitioning in segments $x_b = fg_b$, for g a window function, and $g_b(t) = g(t-b)$. This procedure is a local analysis strategy that can be abstractly described using a locally compact group $\mathcal{G}$ acting in a Hilbert space $\mathcal{H}$. The space $\mathcal{H}$ is the space of signals, and the action of $\mathcal{G}$ is required to be an irreducible and square integrable group representation, $\pi : G \to \mathcal{U}(\mathcal{H})$, defined as a homomorphism between G and $\mathcal{U}(\mathcal{H})$, the space of unitary operators in $\mathcal{H}$. With this representation, the *voice transform* is constructed as $V_\psi : \mathcal{H} \to L^2(\mathcal{G})$, with $V_\psi f(x) = \langle f, \pi(x)(\psi)\rangle$ for $f \in \mathcal{H}, x \in G$, and ψ a particular (so called admissible) element of $\mathcal{H}$. Gabor and wavelet transforms are basic examples where ψ is a Gaussian window function and a wavelet, respectively.

In **dimensionality reduction and manifold learning** [8], the objective is to study a *point cloud data* (or *dataset*), defined as a finite family of vectors $X = \{x_i\}_{i=1}^m \subset \mathbb{R}^n$ in a n-dimensional Euclidean space. We assume that X is sampled from $\mathcal{M}$, a (low dimensional) *space*, seen as a submanifold of $\mathbb{R}^n$ or it can be also considered as the geometrical realization of a topological space (e.g. simplicial complex) in $\mathbb{R}^n$. We have then, $X \subset \mathcal{M} \subset \mathbb{R}^n$ with $p := \dim(\mathcal{M}) < n$. Another concept is a homeomorphic copy of $\mathcal{M}$, denoted by Ω, embedded in a low dimensional space $\mathbb{R}^d$ (with $d < n$), together with a homeomorphism $\mathcal{A} : \Omega \to \mathcal{M} \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^d$. The main advantage of using Ω instead of $\mathcal{M}$ in concrete applications, is the low-dimensionality of the environment $\mathbb{R}^d$.

Persistent Homology [1, 3] is a subfield of **computational topology**, and it provides an efficient strategy for computing topological properties of a dataset $X = \{x_i\}_{i=1}^m \subset \mathbb{R}^n$. In the context of topological analysis, homology theories are fundamental tools, and these can be described as functors from an adequate category (e.g. topological spaces) to a given category of algebraic structures (e.g. abelian groups). Simplicial homology is the basic example where the concept of abstract simplicial complex is the object to study. When considering simplicial homology for the analysis of a raw dataset $X = \{x_i\}_{i=1}^m$, a basic problem is that simplicial complex structures need to be constructed. Persistent homology provides a useful strategy for this problem by efficiently constructing a multiscale topological overview of a point cloud data. The fundamental idea is to construct a family of simplicial complexes by considering the spaces $\mathbb{X}_\epsilon = \cup_{i=1}^m B(x_i, \epsilon)$, where

a ball $B(x_i, \epsilon)$ of radius $\epsilon > 0$ is centered around each point x_i of the dataset X. The efficient algorithmic procedures of this framework is a major milestone in the application of algebraic topology in computational settings.

Noncommutative geometry [7] provides powerful concepts for studying geometrical and topological spaces X using algebraic structures based on C^*-algebras. A basic component in this framework is the concept of noncommutative quotients, which uses noncommutative C^*-algebras for studying "bad" quotients spaces $X = Y/\sim$ for an equivalence relation $\sim$ in Y. In the general framework of noncommutative geometry, the equivalence relation $\sim$, is replaced with a groupoid, which is a powerful generalization of the notion of an equivalence relation, groups, group actions, etc. In our work, we attempt to use the topological counterpart of this framework, noncommutative topology, in order to analyze a time-frequency space X_f related to a signal f.

Our basic proposal [5, 6] is to consider a *functional cloud* $X_f = \{x_i\}_{i=1}^k$ (that we also denote by $M_f = \{x_i\}_i$ for non necessarily discrete cases) defined as a quotient space $M^G_{V_\psi f} = F_{V_\psi f}/G$ for $F_{V_\psi f}$ the graph of $V_\psi f|_{\text{supp}V_\psi f}$, a voice transform of f, and G a groupoid. Our objective is to use noncommutative C^*-algebras, and the concept of Morita equivalence $\overset{m}{\sim}$, as a new type of analysis layer in signal processing. Given $f = \sum_{\ell=1}^k f_\ell$, our goal is to study the space $M^G_{V_\psi f} = F_{V_\psi f}/G$ via $C_0(M^G_{V_\psi f}) \overset{m}{\sim} \mathcal{A} \rtimes_{lt,r} G$ for $\mathcal{A} = \left\{[h_{ij}] \in M_k(C_0(F_{V_\psi f})), h_{ij} \in C_0(U_i \cap U_j)\right\}$ a noncommutative C^*-algebra, with $\{U_\ell\}_{\ell=1}^k$ a covering of $F_{V_\psi f}$ describing time-frequency information of the components f_ℓ. Another objective of our framework is to design computational algorithms based on our developments of signal analysis via C^*-algebras. Here, our proposal is to combine the framework of persistent homology, with tools from *AF-algebras* which are an important family of C^*-algebras, particularly useful for studying finite simplicial complexes [7]. We remark that another important strategy for constructing a C^*-algebra from a simplicial complex has been proposed in [2].

References

[1] G. Carlsson. Topology and data. *American Mathematical Society*, 46(2):255–308, 2009.

[2] J. Cuntz. Noncommutative simplicial complexes and the Baum−Connes conjecture. *Geometric and Functional Analysis*, 12(2):307–329, 2002.

[3] H. Edelsbrunner and J. Harer. Persistent homology - a survey. In *Surveys on discrete and computational geometry: twenty years later: AMS-IMS-SIAM Joint Summer Research Conference, June 18-22, 2006, Snowbird, Utah*, volume 453, page 257. American Mathematical Society, 2008.

[4] H. G. Feichtinger and K. Gröchenig. A unified approach to atomic decompositions via integrable group representations. *Lect. Notes in Math.*, 1302:52–73, 1988.

[5] M. Guillemard. *Some Geometrical and Topological Aspects of Dimensionality Reduction in Signal Analysis*. PhD thesis, Hamburg University, 2011.

[6] M. Guillemard and A. Iske. On groupoid C^*-algebras, persistent homology and time-frequency analysis. *Preprint*, 2011.

[7] G. Landi. *An Introduction to Noncommutative Spaces and their Geometries*. Springer, Heidelberg, 1997.

[8] J.A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.

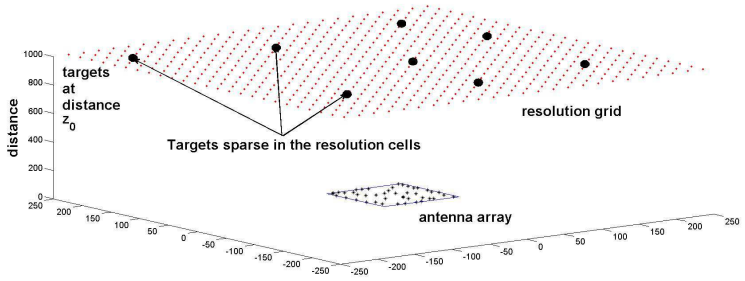


FIGURE 1. The targets at distance z_0 distributed sparsely in the target domain.

Remote sensing via ℓ_1-minimization

MAX HÜGEL
(joint work with Holger Rauhut, Thomas Strohmer)

We analyze a compressive sensing scheme arising from radar and similar applications. The aim is to detect the locations and reflectivities of remote targets by sending probing signals from an antenna array and recording the reflected signals. Standard techniques, such as matched field processing or time reversal methods work well for the detection of very few, well separated targets. However, when the number of targets increases and/or some targets are adjacent to each other, these methods run into severe problems. Moreover, these methods have major difficulties when the dynamic range between the reflectivities of the targets is large. Following [3], we propose a compressive sensing based approach to the inverse scattering problem to overcome the ill-posedness by utilizing sparsity of the target scene. In fact, sparsity is a natural assumption as the targets typically occupy only a small fraction of the overall region of interest. As common in compressive sensing [1, 4] randomness is used and in this setup it is realized by placing the antennas at random locations on a square.

Suppose an array of n transducers is located in the square $[0, B]^2$. The spatial part of a wave of wavelength $\lambda > 0$ emitted from some point source $b \in [0, B]^2$ and recorded at another point $r \in \mathbb{R}^3$ is given by the Green's function G of the Helmholtz equation,

$$G(r, b) := \frac{\exp\left(\frac{2\pi i}{\lambda} \|r - b\|_2\right)}{4\pi \|r - b\|_2}. \tag{1}$$

Our aim is to image targets which are at distance $z_0 > 0$. We make the idealizing assumption that the targets are on a discretized grid of meshsize $d_0 > 0$ in the domain $T := [-L, L]^2 \times \{z_0\}$, where $L > 0$ determines the size of the target domain, see 1 for an illustration.

In order to simplify the analysis we approximate the Green function under the far field assumption $z_0 \gg B + L$ by

$$G(r,b) \approx \widetilde{G}(r,b) := \frac{\exp\left(\frac{\pi i}{\lambda z_0}\|(x,y)-(\xi,\eta)\|_2^2\right)}{4\pi z_0}. \tag{2}$$

Choosing the meshsize d_0 such that the aperture condition

$$\rho := \frac{d_0 B}{\lambda z_0} \in \mathbb{N}, \tag{3}$$

is met, then the normalized system of functions

$$\widehat{G}(b,r_\ell) := 4\pi z_0 \widetilde{G}(b,r_\ell), \quad b \in [0,B], \ell \in [N],$$

satisfies the orthonormality relation

$$\frac{1}{B^2}\int_{[0,B]^2} \widehat{G}(b,r_\ell)\overline{\widehat{G}(b,r_m)}db = \delta_{\ell m}.$$

Assume we have a vector $(x_j)_{j\in[N]} \in \mathbb{C}^N$ of reflectivities on the resolution grid. We sample n antenna positions $b_1,\dots,b_n \in [0,B]^2$ independently at random according to the uniform distribution on $[0,B]^2$. If antenna element $b_j \in [0,B]^2$ transmits and $b_k \in [0,B]^2$ receives, then we model the echo y_{jk} with the Born approximation

$$y_{jk} = \sum_{\ell=1}^{N} \widehat{G}(b_j,r_\ell)\widehat{G}(r_\ell,b_k)x_\ell, \quad (j,k) \in [n]^2. \tag{4}$$

We study the transmit-receive mode where one antenna element transmits at a time and the whole aperture receives the echo. Then the appropriately scaled sensing matrix $A \in \mathbb{C}^{n^2\times N}$ is given entrywise by

$$A_{(j,k),\ell} := \widehat{G}(b_j,r_\ell)\widehat{G}(r_\ell,b_k), \quad (j,k) \in [n]^2, \ell \in [N], \tag{5}$$

and $y = Ax$ by (4). Due to the randomness in the b_k, $k \in [n]$, the matrix A is then a (structured) random matrix with coupled rows and columns. In many scenarios the number of targets is small compared to the grid size, which naturally leads to sparsity in the vector $x \in \mathbb{C}^N$ of reflectivities, $\|x\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s$.

As common in compressive sensing we study reconstruction of x from noisy measurements $y = Ax$ via ℓ_1-minimization,

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \quad \text{subject to } Az = y.$$

The main result of [5] stated next gives an estimate of the required number of antennas ensuring perfect recovery in a nonuniform setting.

Theorem 1. *Let $x \in \mathbb{C}^N$ with $\|x\|_0 \leq s$ be a fixed vector and $A \in \mathbb{C}^{n^2\times N}$ be a draw of the random scattering matrix. If, for $\varepsilon > 0$,*

$$n^2 \geq Cs\log^2(N/\varepsilon) \tag{6}$$

then with probability at least $1-\varepsilon$, ℓ_1-minimization recovers x from $y = Ax$. The constant $C > 0$ is universal.

This result can be made stable under noise on the measurements y and under passing from sparse to approximately sparse vectors, see [5] for details. Moreover, note that the number of measurements is n^2 in this context. The theorem is nonuniform in the sense that the theorem holds for a fixed vector and does not imply that a single draw of the random matrix is able to recover all sparse vectors simultaneously. Such a result would be implied by an estimate of the so-called restricted isometry property [4, 2], but it presently remains an open problem to establish this property for our scattering matrix.

The proof of the theorem relies on an estimate of the conditioning of a single column submatrix of A corresponding to the support of x, which is achieved via decoupling techniques and the non-commutative Bernstein inequality. Moreover, combinatorial estimates similar to [1, 6] are pursued in order to construct a dual-certificate ensuring perfect recovery as well as approximate recovery in the noisy case. A new general criterion for recovery via ℓ_1-minimization in the noisy case is provided in [5] for this purposes. The paper [5] also contains numerical experiments illustrating the theoretical findings.

References

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

[2] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[3] A. Fannjiang, P. Yan, and T. Strohmer. Compressed remote sensing of sparse objects. *SIAM J. Imag. Sci.*, 3(3):596–618, 2010.

[4] M. Fornasier and H. Rauhut. Compressive sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[5] M. Hügel, H. Rauhut and T. Strohmer. Remote sensing via ℓ_1-minimization. Preprint Arxiv:1205.1366, 2012.

[6] H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comput. Harmon. Anal.*, 22(1):16–42, 2007.

Analysis of Inpainting via Clustered Sparsity and Microlocal Analysis

Emily J. King

(joint work with Gitta Kutyniok, Xiaosheng Zhuang)

Land development and geographic features many times prevent optimal sensor placement when collecting seismic data [HFH10, HH08]. Photographs may be scratched. Blocks of videos may be missing due to imperfect steaming. These examples of missing data arise in very different situations but may be formulated as similar mathematical problems. *Inpainting* is the term for the data recovery problem applied to images or videos. In what follows, we set up a general theoretical framework for a certain approach to data recovery which may be analyzed by a novel concept called cluster coherence, and we use this framework to prove results about the success of particular systems, namely shearlets, in inpainting.

One common approach to data recovery is using variational methods [BBCSV, BBS01, BSCB00, CS02]. However, over the last few years, methods based on sparsity have become more popular and have proven themselves heuristically and theoretically to be powerful in a large variety of situations. Based on the apparent success of the numerical algorithms which inpaint based on ℓ_1-minimization of analysis coefficients in [ESQD05, CCS10, DJLSX], we focus on this approach which we shall specifically define below. Detailed exposition may be found in the papers [KKZ11, KKZ12]. Additionally, this analysis-side inpainting may be viewed as a discretization of the variational methods [CDOS12].

We now define what is meant by analysis-side inpainting. A collection of vectors $\Phi = \{\varphi_i\}_{i\in I}$ in a separable Hilbert space $\mathcal{H}$ forms a *Parseval frame* for $\mathcal{H}$ if for all $x \in \mathcal{H}$,

$$\sum_{i\in I} |\langle x, \varphi_i\rangle|^2 = \|x\|^2.$$

With a slight abuse of notation, given a Parseval frame Φ, we also use Φ to denote the *synthesis operator*

$$\Phi : \ell_2(I) \to \mathcal{H}, \quad \Phi(\{c_i\}_{i\in I}) = \sum_{i\in I} c_i\varphi_i.$$

With this notation, Φ^* is called the *analysis operator*. Assume that the complete signal x_0 lies in some Hilbert space $\mathcal{H}$ and that $\mathcal{H} = \mathcal{H}_M \oplus \mathcal{H}_K$ with respective orthogonal projections P_M and P_K. The part of x_0 that is known to us is $P_K x_0$, and we would like to recover x_0. We solve the following minimization problem:

$$\min_x \|\Phi^* x\|_1 \text{ s.t. } P_K x^0 = P_K x. \tag{1}$$

The sequence $\Phi^* x$ is called the sequence of *analysis coefficients* and optimizing this problem is *analysis-side inpainting*.

One would not expect any arbitrary Parseval frame to inpaint successfully. This brings us to the definition of two different measurements which will allow us to analyze how successful a particular Φ is when used in (1). That is, if $x^\star$ solves (1), how small is $\|x - x^\star\|_2$? The following two concepts were first introduced in [DK12] to analyze problems of geometric separation and are modifications of *sparsity* and *mutual coherence* which take into account the geometry of the Parseval frame.

Definition 1. Fix $\delta > 0$. Given a Hilbert space $\mathcal{H}$ with a Parseval frame Φ with index set I, $x \in \mathcal{H}$ is *δ-relatively sparse in Φ (with respect to Λ)* if

$$\|(\Phi^* x)_\lambda : \lambda \in I\backslash\Lambda\|_1 \le \delta.$$

Now let $\Phi_1 = \{\varphi_{1i}\}_{i\in I}$ and $\Phi_2 = \{\varphi_{2j}\}_{j\in J}$ lie in a Hilbert space $\mathcal{H}$ and let $\Lambda \subseteq I$. Then the *cluster coherence* $\mu_c(\Lambda, \Phi_1; \Phi_2)$ of Φ_1 and Φ_2 with respect to Λ is defined by

$$\mu_c(\Lambda, \Phi_1; \Phi_2) = \max_{j\in J} \sum_{i\in\Lambda} |\langle \varphi_{1i}, \varphi_{2j}\rangle|.$$

With these ideas, we can characterize the success of (1).

Theorem 2. *Fix $\delta > 0$ and suppose that x^0 is δ-relatively sparse in Φ. Let $x^\star$ solve (1). Then*

$$\|x^\star - x^0\|_2 \leq \frac{2\delta}{1 - 2\mu_c(\Lambda, P_M\Phi; \Phi)}.$$

A similar result holds when noise is incorporated into the theory. Furthermore, a similar result holds with ℓ_1-minimization is replaced with a simple thresholding of the analysis coefficients. The set Λ here plays an auxiliary role. Any Λ which yields δ-relative sparsity may be used. Even though in general, increasing the size of Λ decreases the δ that may be used, if Φ is a sparsifying dictionary, then a carefully chosen small Λ will yield a small δ. Decreasing the size of Λ increases the value of $1 - 2\mu_c(\Lambda, P_M\Phi; \Phi)$, thus decreasing the error of $x^\star$ in approximating x^0.

We may now apply Theorem 2 to specific cases. While wavelets handle 1-dimensional data well, they do not capture directional information in higher dimensions. This has lead a number of researchers to consider alternate systems. *Shearlets*, introduced in [GKL06], are one such system. Shearlets are the optimal sparsifying systems for images governed by anisotropic structures and have the advantage that they provide a unified concept of the continuum and the digital [KL12, KL11].

Definition 3. The *cone-adapted shearlet system* $\mathcal{SH}(\phi, \psi, \tilde{\psi})$ generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of

$$\begin{aligned}
&\{\phi(\cdot - m) : m \in \mathbb{Z}^2\},\\
&\{2^{3j/4}\psi(S_k A_{2^j} \cdot - m) \quad : \quad j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}, \text{ and}\\
&\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j} \cdot - m) \quad : \quad j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.
\end{aligned}$$

where $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and $S_\ell = \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix}$.

We would also like to show that shearlets are good tools for inpainting. Thus, the Parseval frames we will use in (1) are either wavelets W_j or cone-adapted shearlets S_j. The signal to be recovered x^0 is a linear singularity in $\mathbb{R}^2$ filtered by sub bands $w\mathcal{L}_j = w\mathcal{L} * F_j$. We consider the mask

$$M_h = \{(x_1, x_2) \in \mathbb{R}^2 : |x_1| \leq h\} \quad h > 0.$$

Then P_M is the product of an $L^2(\mathbb{R}^2)$ function with the indicator function of M_h. We vary h with j, denoted h_j, and also denote x_K with $f_j = P_K w\mathcal{L}_j$. The positive results in the following theorem also hold for thresholding.

Theorem 4.

- *For $h_j = o(2^{-j})$ (this is critical in thresholding case) as $j \to \infty$,*

$$\frac{\|W_j - w\mathcal{L}_j\|_2}{\|w\mathcal{L}_j\|_2} \to 0, \quad j \to \infty.$$

- *For* $h_j = o(2^{-j/2})$ *as* $j \to \infty$,

$$\frac{\|S_j - w\mathcal{L}_j\|_2}{\|w\mathcal{L}_j\|_2} \to 0, \quad j \to \infty.$$

This shows that shearlets are able to inpaint larger gaps than wavelets are. Future directions of research would be to implement shearlet inpainting, to prove theoretical results for simultaneously geometrically separating and inpainting using analysis-side minimization, to find the critical bound for the size of the gap that wavelets can asymptotically inpaint using ℓ_1 minimization (the known critical bound is for thresholding), and to apply the framework of parabolic molecules [GK12].

References

[BBCSV] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, *Filling-in by joint interpolation of vector fields and gray levels*, IEEE Trans. Image Process. **10** (2001), no. 8, 1200–1211.

[BBS01] M. Bertalmio, A.L. Bertozzi, and G. Sapiro, *Navier-stokes, fluid dynamics, and image and video inpainting*, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001), IEEE, 2001, pp. I–355–I362.

[BSCB00] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, Proceedings of SIGGRAPH 2000, New Orleans, July 2000, pp. 417–424.

[CCS10] Jian-Feng Cai, Raymond H. Cha, and Zuowei Shen, *Simultaneous cartoon and texture inpainting*, Inverse Probl. Imag. **4** (2010), no. 3, 379 – 395.

[CDOS12] Jian-Feng Cai, Bin Dong, Stanley Osher, and Zuowei Shen, *Image restoration: Total variation, wavelet frames, and beyond*, J. Amer. Math. Soc. (2012), To appear.

[CS02] Tony F. Chan and Jianhong Shen, *Mathematical models for local nontexture inpaintings*, SIAM J. Appl. Math. **62** (2001/02), no. 3, 1019–1043 (electronic). MR 1897733 (2003f:65110)

[DJLSX] Bin Dong, Hui Ji, Jia Li, Zuowei Shen, and Yuhong Xu, *Wavelet frame based blind image inpainting*, Appl. Comput. Harmon. Anal. **32** (2012), no. 2, 268–279.

[DK12] D. L. Donoho and G. Kutyniok, *Microlocal analysis of the geometric separation problem*, Comm. Pure Appl. Math. (2012), Accepted.

[ESQD05] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal. **19** (2005), no. 3, 340–358. MR 2186449 (2007h:94007)

[GK12] P. Grohs and G. Kutyniok, *Parabolic molecules*, preprint. (2012).

[GKL06] K. Guo, G. Kutyniok, and D. Labate, *Sparse multidimensional representations using anisotropic dilation und shear operators*, Wavelets und Splines (G. Chen und M. J. Lai, ed.), Nashboro Press, Nashville, TN, 2006, pp. 189–201.

[HFH10] Gilles Hennenfent, Lloyd Fenelon, and Felix J. Herrmann, *Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach*, Geophysics **75** (2010), no. 6, WB203–WB210.

[HH08] F. J. Herrmann and G. Hennenfent, *Non-parametric seismic data recovery with curvelet frames*, Geophys. J. Int. **173** (2008), 233–248.

[KKZ11] Emily J. King, Gitta Kutyniok, and Xiaosheng Zhuang, *Analysis of data separation and recovery problems using clustered sparsity*, SPIE Proceedings: Wavelets and Sparsity XIV **8138** (2011).

[KKZ12] ———, *Analysis of inpainting via clustered sparsity and microlocal analysis*, Submitted, 2012.

[KL11] Gitta Kutyniok and Wang-Q. Lim, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory **163** (2011), 1564–1589.

[KL12] Gitta Kutyniok and Demetrio Labate (eds.), *Shearlets: Multiscale analysis for multivariate data*, Applied and Numerical Harmonic Analysis, Birkhäuser, 2012.

Suprema of Chaos Processes and the Restricted Isometry Property

Felix Krahmer

(joint work with Shahar Mendelson, Holger Rauhut)

The theory of compressed sensing considers the following problem: Let $A \in \mathbb{C}^{m\times n}$ and let $x \in \mathbb{C}^n$ be s-sparse, i.e., $x_i = 0$ for all but s indices i. One seeks to recover x uniquely and efficiently from linear measurements $y = Ax$, although $m \ll n$. A sufficient condition to ensure that this is possible is the Restricted Isometry Property (RIP).

A matrix $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$ has this property for sparsity $s < n$ and level δ if the restricted isometry constant δ_s, defined as the smallest number such that

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{\Phi x}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{x}\|_2^2 \quad \text{for all } s\text{-sparse } \boldsymbol{x},$$

is small [1]. For example, $\delta_{2s} \leq 0.49$ is known to be sufficient (see, e.g., [5]) to ensure that ℓ_1-minimization

$$\min \|\boldsymbol{z}\|_1 \quad \text{subject to } \boldsymbol{\Phi z} = \boldsymbol{y},$$

reconstructs every s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{\Phi x}$.

In practice, structure is an additional requirement on the measurement matrix $\boldsymbol{\Phi}$, both to accelerate computations using fast matrix-vector multiplies and to model specific applications. An example for a matrix structure that arose in models for specific applications and also possesses fast matrix multiplication properties is a partial circulant matrix, describing a subsampled convolution with a fixed vector. Potential applications such matrices include system identification, radar and cameras with coded aperature.

More precisely, the (circular) convolution of two vectors $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^n$ is defined by

$$(\boldsymbol{z} * \boldsymbol{x})_j := \sum_{k=1}^{n} z_{j\ominus k} x_k, \quad j = 1, \ldots, n,$$

where $j \ominus k = j - k \mod n$ is the cyclic subtraction; the circulant matrix $\boldsymbol{H} = \boldsymbol{H_z} \in \mathbb{C}^{n\times n}$ associated with $\boldsymbol{z}$ is given by $\boldsymbol{Hx} = \boldsymbol{z} * \boldsymbol{x}$ and has entries $H_{jk} = z_{j\ominus k}$. Moreover, we let $\Omega \subset \{1, \ldots, n\}$ be an arbitrary (fixed) set of cardinality m, and denote by $\boldsymbol{R}_\Omega : \mathbb{C}^n \to \mathbb{C}^m$ the operator that restricts a vector $\boldsymbol{x} \in \mathbb{C}^n$ to its entries in Ω. Then our object of study, the corresponding partial random circulant matrix, is the (normalized) combination of a circulant matrix and such a projection.

We discuss the restricted isometry property for such matrices, where the vector $\boldsymbol{z}$ is chosen at random. This randomized setup is in line with the observation that the most efficient matrix constructions which are known to have the restricted isometry property are all random matrices. While corresponding results also hold for the more general case of vectors with independent subgaussian entries [4],

we will, for simplicity, focus on Rademacher vectors. Here a Rademacher vector $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ is a random vector with independent entries distributed according to $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$. Then the associated partial random circulant matrix is given by $\boldsymbol{\Phi} = m^{-1/2}\boldsymbol{R}_\Omega \boldsymbol{H}_{\boldsymbol{\varepsilon}} \in \mathbb{R}^{m\times n}$ and acts on vectors $\boldsymbol{x} \in \mathbb{C}^n$ via

$$\boldsymbol{\Phi}\boldsymbol{x} = \frac{1}{\sqrt{m}}\boldsymbol{R}_\Omega(\boldsymbol{\varepsilon} * \boldsymbol{x}).$$

We show the restricted isometry property of $\boldsymbol{\Phi}$ in a near-optimal parameter regime:

Theorem 1. *Let $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ be a draw of a partial random circulant matrix generated by a Rademacher vector $\boldsymbol{\varepsilon}$. If*

$$m \geq c\delta^{-2} s\,(\log^2 s)(\log^2 n),$$

then with probability at least $1 - n^{-(\log n)(\log^2 s)}$, the restricted isometry constant of $\boldsymbol{\Phi}$ satisfies $\delta_s \leq \delta$. The constant $c > 0$ is universal.

These results improve the best previously known estimates for a partial random circulant matrix [6], namely that $m \geq C_\delta(s \log n)^{3/2}$ is a sufficient condition for achieving $\delta_s \leq \delta$ with high probability (see also [3] for an earlier work on this problem).

For the proof, note that one can rewrite the restricted isometry constant of a partial circulant matrix A based on a Rademacher vector as

$$\begin{aligned}
\delta_s(A) &= \sup_{\substack{x\in S^{n-1} \\ |\operatorname{supp} x|\leq s}} \left| \|Ax\|_2^2 - 1 \right| \\
&= \sup_{\substack{x\in S^{n-1} \\ |\operatorname{supp} x|\leq s}} \left| \|P_\Omega x * \epsilon\|_2^2 - 1 \right| \\
&= \sup_{\substack{x\in S^{n-1} \\ |\operatorname{supp} x|\leq s}} \left| \|V_x\epsilon\|_2^2 - \mathbb{E}\|V_x\epsilon\|_2^2 \right|,
\end{aligned}$$

where V_x is defined through $V_x y := P_\Omega x * y$. This motivates the study of *suprema of chaos processes* of the form

$$\sup_{A\in\mathcal{A}} \left| \|A\epsilon\|_2^2 - \mathbb{E}\|A\epsilon\|_2^2 \right|,$$

where $\mathcal{A}$ is a set of matrices. The key ingredients of the proof of Theorem 1 are moment bounds for such suprema. The formulation of these bounds uses the generic chaining methodology of Talagrand [7]. In particular, the following definition will be crucial to formulate the result.

Definition 2 ([7]). For a metric space (T, d), an *admissible sequence* of T is a collection of subsets of T, $\{T_s : s \geq 0\}$, such that for every $s \geq 1$, $|T_s| \leq 2^{2^s}$ and $|T_0| = 1$. For $\beta \geq 1$, define the γ_β functional by

$$\gamma_\beta(T, d) = \inf \sup_{t\in T} \sum_{s=0}^{\infty} 2^{s/\beta} d(t, T_s),$$

where the infimum is taken with respect to all admissible sequences of T.

Furthermore, denote by $d_F(\mathcal{A})$ the radius of $\mathcal{A}$ in the Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$, i.e., $d_F(\mathcal{A}) = \sup_{A\in\mathcal{A}} \|A\|_F$. In terms of these concepts, the bound of the first moment reads as follows.

Theorem 3 ([4])**.** *Let $m, n \in \mathbb{N}$, let ϵ be a Rademacher vector of length n, and let $\mathcal{A} \subset \mathbb{R}^{m\times n}$ be a symmetric set of matrices. Then one has*

$$\mathbb{E} \sup_{A\in\mathcal{A}} \left| \|A\epsilon\|_2^2 - \mathbb{E}\|A\epsilon\|_2^2 \right| \lesssim d_F(\mathcal{A})\gamma_2(\mathcal{A}, \|\cdot\|_{2\to 2}) + (\gamma_2(\mathcal{A}, \|\cdot\|_{2\to 2}))^2 .$$

From such moment bounds, Theorem 1 is deduced using a Dudley-type entropy integral (cf. [2]).

References

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

[2] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis, **1** (1967), 290–330.

[3] J. Haupt, W. Bajwa, G. Raz, and R. D. Nowak, *Toeplitz compressed sensing matrices with applications to sparse channel estimation*, IEEE Trans. Inform. Theory **56** (2010), 5862–5875.

[4] F. Krahmer, S. Mendelson, and H. Rauhut, *Suprema of chaos processes and the restricted isometry property*, preprint (2012).

[5] Q. Mo and S. Li, *New bounds on the restricted isometry constant δ_{2k}*, Appl. Comput. Harmon. Anal. **31** (2011), 460–468.

[6] H. Rauhut, J. K. Romberg, and J. A. Tropp, *Restricted isometries for partial random circulant matrices*, Appl. Comput. Harmon. Anal. **32** (2012), 242–254.

[7] M. Talagrand, *The Generic Chaining*, Springer Monographs in Mathematics, Springer-Verlag (2005).

Adaptive Anisotropic Refinement Scheme for Transport Dominated Problems

Wang-Q Lim

(joint work with Wolfgang Dahmen, Gitta Kutyniok and Gerrit Welper)

1. Introduction

Solutions of hyperbolic conservation laws or more generally of transport dominated equations are governed by anisotropic features such as singularities concentrated on lower dimensional embedded manifolds. In this case, solutions typically exhibit strong anisotropic features such as shear layers or shock fronts. The metric entropy of compact sets of such functions suggest that *'good'* approximation methods should give rise to distortion rates that cannot be achieved by isotropic refinements corresponding to classical wavelet bases. Since the stability of discretizations based on anisotropic mesh refinements is not a straightforward matter, an interesting alternative is offered by recent developments centering on directional

representation systems like curvelets (see [1]) or shearlets (see [4]). Best N-term approximations from such systems are known to resolve wave front sets - the prototype of anisotropic singularities - at a nearly optimal rate [1, 6] when compared with the metric entropy of such classes. However, the currently known directional representation systems do generally not form Riesz bases but merely frames for a specific function space, namely for $L^2(\mathbb{R}^d), d = 2, 3$, [1, 6]. Adaptive discretizations of operator equations using such representation systems must therefore take this fact into account.

Recently, a general framework for L^2–stable variational formulations of linear unsymmetric operators, taking first order transport equations in bounded domains, has been introduced in [2]. In this framework, the trial space X indeed can be chosen as $L^2(\mathbb{R}^d)$, in which the variational solution is sought and the proposed stability concept is based on perturbations of certain *'ideal'* test spaces in Petrov-Galerkin formulations. Our aim in this report is to develop adaptive refinement scheme providing nearly optimal approximation rate for linear transport equations using directional representation system – especially shearlets – based on this L^2–stable variational formulation.

2. Linear transport equations

In this section we briefly discuss model problems that we will consider. Throughout this report, we will only consider 2 dimensional bounded domain $D = [0, 1]^2$ with boundary $\Gamma = \partial D$ but our scheme which will be introduced here can be generalized to more general settings – for instance, a bounded, polyhedral domain in 2D or 3D. First, we assume that velocity field $\overrightarrow{b}(x), x \in D$ is differentiable, i.e. $\overrightarrow{b} \in C^1(\overline{D})^2$. Likewise, $c \in C^0(\overline{D})$ will serve as the reaction term in the *first order transport equation*

$$\begin{aligned} Au := \overrightarrow{b} \cdot \nabla u + cu &= f \text{ in } D, \\ u &= g \text{ on } \Gamma_-. \end{aligned} \tag{1}$$

Here, Γ_- is the *inflow boundary* defined by

$$\Gamma_- = \{x \in \Gamma : \overrightarrow{b}(x) \cdot \overrightarrow{n}(x) < 0\},$$

where $\overrightarrow{n}(x)$ is the exterior unit normal vector for $x \in \Gamma$.

3. Adaptive Refinement Scheme using Shearlets

The anisotropic structures can be distinguished by location and direction which indicates that our way of analyzing and representing the data should capture not only location, but also directional information. This observation inspires various constructions of directional representation systems to achieve sparse approximations of data governed by anisotropic features. Among those representation systems, shearlets were introduced as a means to optimal sparsely encode anisotropic singularities of multivariate data while allowing compactly supported analyzing elements [6]. In contrast to curvelets, shearlets treat the continuum and digital setting uniformly, thereby enabling faithful numerical realizations [4]. We first

briefly review the customarily employed notion of (cone-adapted) shearlets. For this, for $j \geq 0, k \in \mathbb{Z}$, let

$$A_j = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\lfloor j/2 \rfloor} \end{pmatrix}, \; S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad M_c = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix},$$

where $c = (c_1, c_2)$ and c_1, c_2 are some positive constants. The definition of (cone-adapted) discrete shearlets (in 2D) can now be phrased as follows.

Definition 1. Let $c = (c_1, c_2) \in (\mathbb{R}_+)^2$. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ the *(cone-adapted) discrete shearlet system* $SH(\phi, \psi, \tilde{\psi}; c)$ is defined by

$$SH(\phi, \psi, \tilde{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \tilde{\Psi}(\tilde{\psi}; c),$$

where

$$\Phi(\phi; c_1) = \{\phi(\cdot - m) : m \in c_1 \mathbb{Z}^2\}$$
$$\Psi(\psi; c) = \{2^{\frac{3}{4}j} \psi(S_k A_j \cdot - m) : j \geq 0, -2^{\lceil j/2 \rceil} \leq k \leq 2^{\lceil j/2 \rceil}, m \in M_c \mathbb{Z}^2\},$$

and each element in $\tilde{\Psi}(\tilde{\psi}; c)$ is defined from element in $\Psi(\psi; c)$ by switching the order of variables x_1 and x_2. If $SH(\phi, \psi, \tilde{\psi}; c)$ is a frame for $L^2(\mathbb{R}^2)$, we refer to ϕ as a *scaling function* and ψ and $\tilde{\psi}$ as *shearlets*.

In [6], decay conditions on compactly supported shearlet generators ψ and $\tilde{\psi}$ were derived which ensure that the associated (cone-adapted) shearlet system $SH(\phi, \psi, \tilde{\psi}; c)$ forms a frame for $L^2(\mathbb{R}^2)$. Further, it was shown that compactly supported shearlets provide almost optimally sparse approximations for functions which are C^2 smooth apart from C^2 singularity curves. This indicates that the anisotropic structures of piecewise smooth functions – in particular, singularities of a solution of linear transport equation (1) – can be efficiently encoded using shearlets. However, until now there has been little attempt to utilize directional representation systems for adaptive solvers for solving PDEs. In fact, in addition to this sparse approximation property, there are additional desiderata in order to use directional representation systems in the adaptive scheme as follows.

- *Adaptation to bounded domain.* Since solutions are typically just defined on a bounded domain, basis elements which are supported on such a domain, in particular, whose regularity is adapted to this domain, are required. This will also allow for naturally incorporating boundary conditions.
- *Piecewise polynomial basis elements.* It is desirable to be able to compute the entries of a mass matrix efficiently.

In [3], new construction of shearlet system, so called *Alpert shearlets*, satisfying all the desirable properties above as well as optimal sparse approximation property has been introduced. Based on L^2–stable variational formulation in [2], we have developed a novel adaptive scheme, where we use *Alpert shearlets* to approximate solutions of linear transport equations given as (1) – see [3]. Various numerical test results in [3] show that our scheme approximates solutions of (1) at almost

optimal rate. The main advantage of the adaptive scheme is that there is hierarchical structure associated with *Alpert shearlet* system with respect to scale and shear parameters j and k. Indeed this provides very efficient adaptive ansiotropic refinements for approximating the solutions. Also, 2D shearlet system we described in this report can be naturally extended to higher dimensions – for instance, we refer to [6] for shearlets in 3D. Therefore, there is a strong potential of the shearlet based adaptive scheme for higher dimensional problems although we only considered 2D problems in [3].

References

[1] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure and Appl. Math. **56** (2004), 216–266.

[2] W. Dahmen, C. Huang, C. Schwab, and G.Welper, *Adaptive Petrov-Galerkin methods for first order transport equations*, IGPM Report **321**, RWTH Aachen, 2011.

[3] W. Dahmen, G. Kutyniok, W.-Q Lim and G. Welper *Shearlet Adaptive Petrov-Galerkin Method*, preprint.

[4] G. Kutyniok and D. Labate, eds., *Shearlets: Multiscale Analysis for Multivariate Data*, Birkhäuser Boston, 2012.

[5] G. Kutyniok, J. Lemvig, and W.-Q Lim. *Optimally Sparse Approximations of 3D Functions by Compactly Supported Shearlet Frames*, SIAM J. Math. Anal., to appear.

[6] P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Construction of compactly supported shearlet frames*, Constr. Approx., **35** (2012), 21–72.

[7] G. Kutyniok and W.-Q Lim, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory, **163** (2011), 1564–1589.

A generalization of the iterative soft-thresholding algorithm for non-separable penalties

Ignace Loris

(joint work with Caroline Verhoeven)

The present research report focuses on the efficient use of sparse representations for the regularization and solution of ill-posed problems. In particular, we discuss simple iterative algorithms for the minimization of certain convex functionals encountered in this area.

Sparsity refers to the property that data can be modeled using only few significant coefficients, without loss of important information. In the context of this report it is used as a way of imposing a priori information on the solution of an ill-posed inverse problem. A popular technique for recovering sparse vectors x from noisy data y is found by using a penalty term of type $\|x\|_1 = \sum_i |x_i|$ in combination with a least squares data misfit term:

$$\min_x \tfrac{1}{2}\|Kx - y\|_2^2 + \lambda\|x\|_1. \tag{1}$$

Here y are noisy data, x is the unknown quantity and K is a known linear operator.

The above functional is non-differentiable, which makes numerical minimization more difficult. A simple iterative method for finding the minimizer of (1) is the

so-called iterative soft-thresholding algorithm (IST) [1, 2]:

$$x^{n+1} = \mathcal{S}_\lambda[x^n + K^T(y - Kx^n)] \qquad x^0 = \text{arbitrary and } \|K\| < \sqrt{2}. \tag{2}$$

$\mathcal{S}_\lambda$ acts componentwise and $\mathcal{S}_\lambda(u) = u - \lambda\,\mathrm{sgn}(u)$ when $|u| \geq \lambda$ and $\mathcal{S}_\lambda(u) = 0$ for $|u| \leq \lambda$. However, convergence of this algorithm can be slow and an faster version has been proposed in [3].

Another, more general, minimization problem that is often encountered in sparse regularization of ill-posed inverse problems is:

$$\min_x \tfrac{1}{2}\|Kx - y\|_2^2 + \lambda\|Ax\|_1. \tag{3}$$

Here a second linear operator A mixes the variables also in the penalty term. In [4] we have developed a generalized iterative soft-thresholding algorithm that converges to a minimizer of this functional. It takes the form:

$$\left\{\begin{array}{lcll} \bar{x}^{n+1} &=& x^n + K^T(y - Kx^n) - A^T w^n & \\ w^{n+1} &=& \mathcal{P}_\lambda(w^n + A\bar{x}^{n+1}) & \\ x^{n+1} &=& x^n + K^T(y - Kx^n) - A^T w^{n+1} & x^0, w^0 = \text{arbitrary}, \end{array}\right. \tag{4}$$

where $\mathcal{P}_\lambda = \mathrm{Id} - \mathcal{S}_\lambda$ is the projection on the ℓ_∞-ball of radius λ.

Just as the original IST algorithm (2), the 'generalized IST' algorithm (4) is an explicit algorithm that uses only matrix-vector multiplications and a simple convex projection operator. Convergence is proven in a finite dimensional setting when $\|K\| < \sqrt{2}$ and $\|A\| < 1$. We also derived a $\mathcal{O}(N^{-1})$ bound on the functional for the sequence of averages $\sum_{n=1}^N x^n/N$. The algorithm reduces to the IST algorithm (2) in the special case $A = 1$ (or A orthogonal). Other algorithms exist as well.

The algorithm (4) can be used to solve 'total variation' (TV) penalized problems popular in image restoration. In that case the operator A in (3) is simply a local differencing operator ('gradient'). Another application is found in structured sparsity where sparsity is expressed in terms of grouped variables. In this case the operator A is used for grouping variables and the projection $\mathcal{P}_\lambda$ takes on a different (but still completely explicit) expression.

In some applications it is necessary to find the minimizer of functional (3) under additional linear constraints $Bx = b$, i.e. one is interested in the problem:

$$\min_{Bx=b} \tfrac{1}{2}\|Kx - y\|_2^2 + \lambda\|Ax\|_1. \tag{5}$$

In [5], we showed that the algorithm

$$\left\{\begin{array}{lcll} \bar{v}^{n+1} &=& v^n - (Bx^n - b) & \\ \bar{x}^{n+1} &=& x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^T w^n & \\ w^{n+1} &=& \mathcal{P}_\lambda(w^n + A\bar{x}^{n+1}) & \\ x^{n+1} &=& x^n + K^T(y - Kx^n) + B^T\bar{v}^{n+1} - A^T w^{n+1} & \\ v^{n+1} &=& v^n - \frac{1}{\alpha}(Bx^{n+1} - b) & x^0, w^0, v^0 = \text{arbitrary}. \end{array}\right. \tag{6}$$

converges to a minimizer of this problem if $\|\frac{1}{2}K^TK + B^TB\| < 1$, $\|AA^T\| < 1$ and $\alpha > \frac{1}{2}$.

Using similar techniques an iterative algorithm for the constrained problem

$$\min_{\|Kx-y\|\leq\epsilon} \|Ax\|_1 \tag{7}$$

was derived:

$$\begin{cases} \bar{x}^{n+1} &= x^n - K^T(v^n + (v^n - v^{n-1})/\theta) - A^T w^n \\ w^{n+1} &= P_\mu\left(w^n + A\bar{x}^{n+1}\right) \\ x^{n+1} &= x^n - K^T(v^n + (v^n - v^{n-1})/\theta) - A^T w^{n+1} \\ v^{n+1} &= (1-\theta)v^n + \theta\mathcal{T}_{y,\epsilon}\left(v^n + Kx^{n+1}\right) \qquad x^0, w^0, v^0 = \text{arbitrary}, \end{cases} \tag{8}$$

and convergence was proven [6] for $0 < \theta \leq 1$, $\mu > 0$, $\|K\| < 1$ and $\|A\| < 1$. Here $\mathcal{T}_{y,\epsilon} = \mathrm{Id} - \mathcal{Q}_{y,\epsilon}$ where $\mathcal{Q}_{y,\epsilon}$ is the projection on the ℓ_2-ball of radius ϵ around y. Its performance was compared numerically with a number of other algorithms and illustrated on generalized TV constrained inversions.

All the above algorithms can be generalized to handle an arbitrary convex function $H(Ax)$ instead of $\|Ax\|_1$ as long as the proximity operator of H is known. Notice that the proximity operator of $H(A\cdot)$ does not need to be known.

A current theme of research is the comparison of the generalized IST algorithm of [4] with the block-iterative algorithm of [7] for TV regularized medical imaging. Both algorithms appear to show about the same speed of convergence. Although the latter algorithm is block-iterative, it does approximate the (non-differentiable) TV function by a differentiable penalty. The generalized IST algorithm uses techniques of convex optimization to avoid such an approximation. An interesting direction for future research is therefore the development of a block-iterative algorithm for the minimization of a TV penalized least functional, that does not approximate the TV function, but that would use the same proximal operators as the algorithm (4). As a first step, we will look at the problem of finding $\arg\min_{x,Kx=y} \|Ax\|_1$ by a block iterative method. The plan is to later generalize this to $\arg\min_{x,\|Kx-y\|\leq\epsilon} \|Ax\|_1$ using a block-iterative algorithm.

References

[1] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications On Pure And Applied Mathematics*, 57(11):1413–1457, November 2004.

[2] Patrick L. Combettes and Valerie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, January 2005.

[3] Amir Beck and Marc Teboulle. A fast iterative shrinkage-threshold algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2:183–202, 2009.

[4] Ignace Loris and Caroline Verhoeven. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Problems*, 27:125007, 2011.

[5] Ignace Loris and Caroline Verhoeven. An iterative algorithm for sparse and constrained recovery with applications to divergence-free current reconstructions in magneto-encephalography. *Computational Optimization and Applications*, 2012. Special Issue on Optimization methods for inverse problems in imaging. Accepted.

[6] Ignace Loris and Caroline Verhoeven. Iterative algorithms for total variation-like reconstructions in seismic tomography. *International Journal on Geomathematics*, 2012. Accepted..

[7] Michel Defrise, Christian Vanhove, and Xuan Liu. An algorithm for total variation regularization in high-dimensional linear problems. *Inverse Problems*, 27(6):065002, 2011.

Scalable Frames

Friedrich Philipp
(joint work with Gitta Kutyniok, Kasso A. Okoudjou, and Elizabeth K. Tuley)

A frame for a separable Hilbert space $\mathcal{H}$ is a system of vectors $\Phi = (\varphi_j)_{j\in J} \subset \mathcal{H}$, $J \subset \mathbb{N}$, which satisfies a stability condition

$$A\|x\|^2 \leq \sum_{j\in J} |\langle x, \varphi_j\rangle|^2 \leq B\|x\|^2, \quad x \in \mathcal{H}, \tag{1}$$

where $B \geq A > 0$. Thus, frames are typically redundant, yet stable systems and have therefore established themselves as a standard tool in applied mathematics, computer science, and engineering. The stability of a frame can be measured by the ratio B/A of the optimal frame bounds in (1) which is exactly the square of the condition number of the so-called analysis operator $T_\Phi = (\langle \cdot, \varphi_j\rangle)_{j\in J}$ of the frame. A frame is thus optimally conditioned if it is a *Parseval frame*, i.e. if $A = B = 1$.

In applications it is often desirable to construct a Parseval frame from a given frame. One possibility of doing so is to compute the Parseval frame $(S_\Phi^{-1/2}\varphi_j)_{j\in J}$, where $S_\Phi = T_\Phi^* T_\Phi$ is the frame operator of Φ. But the problem of inverting the frame operator might be ill-conditioned, so that this method is of rather theoretical interest. The simplest way to modify a frame is just to multiply its frame vectors with scalars. Therefore, we call a frame *scalable* if such an operation leads to a Parseval frame. If the scalars used are all larger than some $\delta > 0$, we say that the frame is *strictly scalable*. Scalability can also be expressed in operator-theoretic terms:

Theorem 1. *A frame $\Phi = (\varphi_j)_{j\in J}$ for $\mathcal{H}$ is (strictly) scalable if and only if there exists a non-negative (strictly positive, respectively) self-adjoint diagonal operator D in $\ell^2(J)$ such that the operator DT_Φ, mapping from $\mathcal{H}$ to $\ell^2(J)$, is isometric.*

Hence, scalability is closely connected with the preconditioning problem in numerical linear algebra (see, e.g., [2]). But in the latter case the problem is to find the smallest possible condition number of DT_Φ where D runs through the diagonal operators or matrices. Here, we ask for necessary and sufficient conditions for optimal preconditioning, i.e. for which frames there exists a diagonal operator such that DT_Φ has exactly condition number one. The first of our two presented characterizations is the following.

Theorem 2. *Let $\Phi = (\varphi_j)_{j\in J}$ be a frame for $\mathcal{H}$. Then the following statements are equivalent.*

(i) *The frame Φ is strictly scalable.*
(ii) *There exist a Hilbert space $\mathcal{K}$ and an injective bounded operator $L : \mathcal{K} \to \ell^2(J)$ with closed range such that $TT^* + LL^*$ is a strictly positive bounded diagonal operator in $\ell^2(J)$.*

(iii) *There exist a Hilbert space* $\mathcal{K}$ *and a frame* $\Psi = (\psi_j)_{j\in J}$ *for* $\mathcal{K}$ *such that the vectors* $(\varphi_j, \psi_j) \in \mathcal{H} \oplus \mathcal{K}$ *form an orthogonal basis of* $\mathcal{H} \oplus \mathcal{K}$.

In the case when $\mathcal{H} = \mathbb{K}^N$ (where $\mathbb{K}$ stands for either $\mathbb{C}$ or $\mathbb{R}$) and $J = \{1, \dots, M\}$, $M \geq N$, we have $\ell^2(J) = \mathbb{K}^M$, and $\mathcal{K}$ in Theorem 2 can be chosen as $\mathbb{K}^{M-N}$. In particular, a basis of $\mathbb{K}^N$ is strictly scalable if and only if it is an orthogonal basis. If there is only one frame vector more than dimensions, i.e. $M = N + 1$, Theorem 2 leads to a simple scalability test (see [1, Corollary 2.9]).

In the real finite-dimensional case (i.e. $\mathcal{H} = \mathbb{R}^N$ and $J = \{1, \dots, M\}$, $M \geq N$) it is natural to ask for the geometric nature of scalable frames. Our following second characterization theorem will lead to such a geometric interpretation.

Theorem 3. *Let* $\Phi = (\varphi_j)_{j=1}^M \subset \mathbb{R}^N \setminus \{0\}$ *be a frame for* $\mathbb{R}^N$. *Then the following statements are equivalent.*

(i) Φ *is not scalable.*
(ii) *There exists a symmetric matrix* $Y \in \mathbb{R}^{N\times N}$ *with* $\operatorname{tr}(Y) = 0$ *such that* $\varphi_j^T Y \varphi_j > 0$ *for all* $j = 1, \dots, M$.

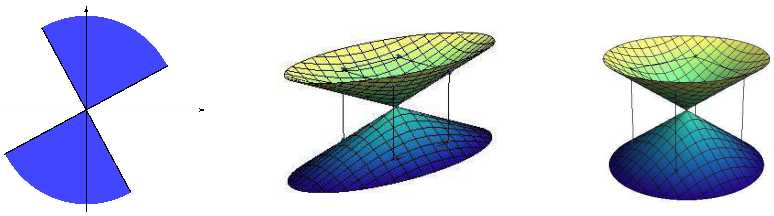
Before providing the geometric interpretation of scalability, let us first draw the following corollary which is a direct consequence of Theorem 3.

Corollary 4. *The set of non-scalable frames is open in the following sense: Given a non-scalable frame* $(\varphi_j)_{j=1}^M \subset \mathbb{R}^N \setminus \{0\}$, *then there exists* $\varepsilon > 0$ *such that each system of vectors* $(\psi_j)_{j=1}^M \subset \mathbb{R}^N$ *with* $\|\varphi_j - \psi_j\| < \varepsilon$ *for all* $j = 1, \dots, M$ *is a non-scalable frame.*

For the convenience of the reader, here we only consider the geometry of scalable frames in $\mathbb{R}^2$ or $\mathbb{R}^3$. First, note that, by Theorem 3, the frame $\Phi = (\varphi_j)_{j=1}^M$ in $\mathbb{R}^N$ ($N \in \{2, 3\}$) is non-scalable if and only if a rotated and reflected version $(\psi_j)_{j=1}^M$ of Φ is contained in a set of the form

$$\left\{ x \in \mathbb{R}^N : \sum_{i=1}^N c_i x_i^2 > 0 \right\},$$

where $\sum_{i=1}^N c_i = 0$. For each N this is a cone with the 2^N points $(\pm 1, \dots, \pm 1)^T \in \mathbb{R}^N$ on its boundary. In the special case $N = 2$ it is just the 45°-rotation of the set $\{(x, y)^T : xy > 0\}$. Hence, *a frame in* $\mathbb{R}^2$ *is non-scalable if and only if all its vectors lie in a rotation of the quadrant cone* $\{(x, y)^T : xy > 0\}$. In the 3D-case it is easily seen that the quadrant cone must be replaced by a range of special elliptical cones. For a detailed description (also for general N) see [1].



(a) (b) (c)

FIGURE 1. (a) shows a sample region of vectors of a non-scalable frame in $\mathbb{R}^2$. (b) and (c) show examples of sets which determine sample regions in $\mathbb{R}^3$.

REFERENCES

[1] G. Kutyniok, K.A. Okoudjou, F. Philipp, E.K. Tuley, *Scalable frames*, arXiv:1204.1880v3, Preprint.

[2] K. Chen, Matrix preconditioning techniques and applications, Cambridge Monographs on Applied and Computational Mathematics 19. Cambridge: Cambridge University Press xxiii, 2005.

Blind Deconvolution and Channel Protection using Random Codes

JUSTIN ROMBERG

(joint work with Ali Ahmed and Benjamin Recht)

We consider the problem of recovering two unknown vectors, $\boldsymbol{w}$ and $\boldsymbol{x}$, of length L from their circular convolution. We make the structural assumption that the two vectors are members known subspaces, one with dimension N and one with dimension K. Although the observed convolution is nonlinear in both $\boldsymbol{w}$ and $\boldsymbol{x}$, it is linear in the rank-1 matrix formed by their outer product $\boldsymbol{w}\boldsymbol{x}^*$. This observation allows us to recast the deconvolution problem as low-rank matrix recovery problem from linear measurements, whose natural convex relaxation is a nuclear norm minimization program.

We prove the effectiveness of this relaxation by showing that for "generic" signals, the program can deconvolve $\boldsymbol{w}$ and $\boldsymbol{x}$ exactly when the maximum of N and K is almost on the order of L. That is, we show that if $\boldsymbol{x}$ is drawn from a random subspace of dimension N, and $\boldsymbol{h}$ is a vector in a subspace of dimension K whose basis vectors are "spread out" in the frequency domain, then nuclear norm minimization recovers $\boldsymbol{w}\boldsymbol{x}^*$ without error.

We discuss this result in the context of blind channel estimation in communications. If we have a message of length N which we code using a random $L \times N$ coding matrix, and the encoded message travels through an unknown linear time-invariant channel of maximum length K, then the receiver can recover both the channel response and the message when $L \gtrsim N + K$, to within constant and log factors.

The problem is formalized as follows. A message vector $\boldsymbol{m} \in \mathbb{R}^N$ is encoded through an $L \times N$ encoding matrix $\boldsymbol{C}$ — we will take $\boldsymbol{C}$ to be a random matrix, and will discuss how it is generated below. The protected message $\boldsymbol{x} = \boldsymbol{Cm}$ travels through a channel whose impulse response is $\boldsymbol{h} \in \mathbb{R}^K$ (and so $\boldsymbol{w}$ is simply $\boldsymbol{h}$ zero-padded to length L). The receiver observes

$$\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{Cm}. \tag{1}$$

We can write the observations in matrix-vector form as

$$\boldsymbol{y} = \boldsymbol{HCm},$$

where $\boldsymbol{H} = \operatorname{circ}(\begin{bmatrix} h(0) & h(1) & \cdots & h(K) & 0 & \cdots 0 \end{bmatrix})$ is the $L \times L$ circulant matrix generated from h after it has been zero-padded. We will find it convenient to treat this problem in the Fourier domain; taking the discrete Fourier transform of $\boldsymbol{y}$ yields

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{H}}\boldsymbol{FCm}, \tag{2}$$

where $\hat{\boldsymbol{H}} = \operatorname{diag}(\begin{bmatrix} \hat{h}(1) & \cdots & \hat{h}(L) \end{bmatrix})$ is the L-point discrete Fourier transform (DFT) of $\boldsymbol{h}$ and $\boldsymbol{F}$ is the (normalized) $L \times L$ DFT matrix,

$$\boldsymbol{F}(\omega, \ell) = \frac{1}{\sqrt{L}} \mathrm{e}^{-\mathrm{j}2\pi(\omega-1)(\ell-1)/L}.$$

Since we are "blind" to the channel $\boldsymbol{h}$, the inverse problem (2) is *nonlinear* — the observations in $\hat{\boldsymbol{y}}$ contain scalar products between the entries in $\boldsymbol{m}$ and the entries in $\boldsymbol{h}$. But by rewriting (2) carefully, we can view $\hat{\boldsymbol{y}}$ as a set of linear measurements of the $K \times N$ rank-1 matrix $\boldsymbol{X}_0 = \boldsymbol{hm}^*$. We do this by breaking apart the action of the channel $\boldsymbol{H}$ on each component of $\boldsymbol{Cm}$. The coded message $\boldsymbol{Cm}$ is of course a linear combination of the columns of $\boldsymbol{C}$; we can interpret $\boldsymbol{y}$ in this lights as the first column of $\boldsymbol{C}$ weighted by $m(1)$ convolved with $\boldsymbol{h}$ plus the second column of $\boldsymbol{C}$ weighted by $m(2)$ convolved with $\boldsymbol{h}$, etc. With $\boldsymbol{C}_n$ as the nth column of $\boldsymbol{C}$, we can write this in matrix form as

$$\begin{aligned} \boldsymbol{y} &= m(1)(\boldsymbol{C}_1 * \boldsymbol{h}) + m(2)(\boldsymbol{C}_2 * \boldsymbol{h}) + \cdots + m(N)(\boldsymbol{C}_N * \boldsymbol{h}) \\ &= \begin{bmatrix} \operatorname{toep}(\boldsymbol{C}_1) & \operatorname{toep}(\boldsymbol{C}_2) & \cdots & \operatorname{toep}(\boldsymbol{C}_N) \end{bmatrix} \begin{bmatrix} m(1)\boldsymbol{h} \\ m(2)\boldsymbol{h} \\ \vdots \\ m(N)\boldsymbol{h} \end{bmatrix}, \end{aligned}$$

where $\operatorname{toep}(\boldsymbol{C}_n)$ is the $L \times K$ matrix whose columns are the first K circular shifts of $\boldsymbol{C}_n$. We will find it convenient to write this system in the Fourier domain; since $\boldsymbol{F}$ is orthogonal, recovering $\hat{\boldsymbol{y}} = \boldsymbol{Fy}$ is the same as recovering $\boldsymbol{y}$. Using the fact that

$$\operatorname{toep}(\boldsymbol{C}_n) = \boldsymbol{F}^* \boldsymbol{D}_n \tilde{\boldsymbol{F}},$$

where $\tilde{\boldsymbol{F}}$ consists of the first K columns of $\boldsymbol{F}$ and $\boldsymbol{D}_n$ is an $L \times L$ diagonal matrix whose diagonal entries are the DFT of $\boldsymbol{C}_n$: $\boldsymbol{D}_n = \operatorname{diag}\{\sqrt{L}\boldsymbol{F}\boldsymbol{C}_n\}$. The frequency-domain observations $\hat{\boldsymbol{y}}$ are now given by

$$\hat{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{D}_1\tilde{\boldsymbol{F}} & \boldsymbol{D}_2\tilde{\boldsymbol{F}} & \cdots & \boldsymbol{D}_N\tilde{\boldsymbol{F}} \end{bmatrix} \begin{bmatrix} m(1)\boldsymbol{h} \\ m(2)\boldsymbol{h} \\ \vdots \\ m(N)\boldsymbol{h} \end{bmatrix}. \tag{3}$$

The form of (3) makes it clear that $\hat{\boldsymbol{y}}$ can be interpreted as a set of *linear* measurements of the rank-1 matrix $\boldsymbol{h}\boldsymbol{m}^*$,

$$\hat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{h}\boldsymbol{m}^*), \tag{4}$$

where $\mathcal{A}$ maps $K \times N$ matrices to $\mathbb{R}^L$. Recovering $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$ is then the same as recovering both $\boldsymbol{h}$ and $\boldsymbol{m}$ up to a multiplicative factor.

By making the matrix $\boldsymbol{C}$ random, we can interpret (4) as random linear measurements of a low-rank matrix. The structure of our measurements (given in (3)), however, is completely different than anything which has appeared in the literature to date. We will generate the coding matrix in the Fourier domain by taking the $d_\ell(n)$ to be independent Steinhaus random variables. To preserve symmetry in the Fourier domain (in order to keep our code book real-valued), we take

$$\begin{aligned} d_\ell(n) &= \mathrm{e}^{\mathrm{j}\theta(\ell,n)}, \\ \theta(\ell,n) &\sim \text{Uniform}[0,2\pi), \quad \ell = 1,\dots,(L+1)/2, \\ \theta(\ell,n) &= -\theta(L-\ell+1,n), \quad \ell = (L+3)/2,\dots,L. \end{aligned}$$

(We assume here that L is odd for simplicity; extending the construction to even L is straightforward.) This method of generating the codes is chosen purely for mathematical convenience — it will achieve the desired result while keeping the analysis as simple as possible. It is very likely that codes generated using different models (simply taking the entries of $\boldsymbol{C}$ to be independent subgaussian random variables, for example) can be shown to be equally effective.

Given $\boldsymbol{y}$ (and hence $\hat{\boldsymbol{y}}$), we recover $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$ using the convex program

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{X}\|_* \\ \text{subject to} & \hat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{X}) \end{array}. \tag{5}$$

Our main result gives a sufficient condition on the code length L that allows perfect recovery of both the message $\boldsymbol{m}$ and the channel $\boldsymbol{h}$ from the observation of their convolution $\boldsymbol{y}$ by solving (5). The code is generated at random, and is effective for arbitrary, fixed $\boldsymbol{h}$ and $\boldsymbol{m}$ with high probability. Along with the message length N and the channel length K, the bound depends on the *channel coherence*

$$\mu_h^2 = L \cdot \max_{1 \le \ell \le L} |\langle \boldsymbol{h}, \boldsymbol{f}_\ell \rangle|^2.$$

The parameter μ_h roughly quantifies how "flat" the channel is in the frequency domain. If the distribution of the energy in $\boldsymbol{h}$ is spread more or less evenly across

the entire band, then μ_h^2 will be a small constant. Note that it is always the case that $1 \leq \mu_h^2 \leq K$.

Theorem 1. *Let $\boldsymbol{m} \in \mathbb{R}^N$ be a message vector, and $\boldsymbol{h} \in \mathbb{R}^K$ be a channel vector with channel coherence μ_h. Let $\hat{\boldsymbol{y}} \in \mathbb{R}^L$ be the nonlinear observations of $\boldsymbol{m}$ and $\boldsymbol{h}$ as in* (3)*. Then there is a constant C such that if*

$$L \;\geq\; C \max\left(K, \mu_h^2 N\right) \log^2(KN), \tag{6}$$

then with probability $1 - O(N^{-1})$, $\boldsymbol{h}$ and $\boldsymbol{m}$ are recoverable from $\hat{\boldsymbol{y}}$: the $K \times N$ matrix $\boldsymbol{X}_0 = \boldsymbol{h}\boldsymbol{m}^*$ is the unique solution to (5)*.*

Proof of this theorem can be found in the preprint [1].

References

[1] A. Ahmed, B. Recht, and J. Romberg, *Multipath Channel using Random Codes*, Preprint, June 2012, available at `http://bit.ly/L6SBdP`.

Localization Phenomena of Laplacian Eigenfunctions on Graphs & Sparse Graph Constructions

Naoki Saito

(joint work with Yuji Nakatsukasa and Ernest Woei)

The first part of the talk, we described our current understanding on the phase transition phenomenon of the graph Laplacian eigenfunctions constructed on a certain type of unweighted trees [9], which we previously observed through our numerical experiments conducted on dendritic trees of mouse's retinal ganglion cells [11, 12]. The eigenvalue distribution for such a tree is a smooth bell-shaped curve starting from the eigenvalue 0 up to 4. Then, at the eigenvalue 4, there is a sudden jump. Interestingly, the eigenfunctions corresponding to the eigenvalues below 4 are *semi-global* oscillations (like Fourier modes) over the entire tree or one of the branches; on the other hand, those corresponding to the eigenvalues above 4 are much more *localized* and *concentrated* (like wavelets) around junctions/branching vertices.

For a special class of trees called *starlike trees*, each of which has only one vertex whose degree is larger than 2, we obtained a complete understanding of such phase transition phenomenon:

Theorem 1 ([12, 9])**.** *A starlike tree has exactly one graph Laplacian eigenvalue greater than or equal to* 4*. The equality holds if and only if the starlike tree is $K_{1,3}$, which is also known as a* claw*. Moreover, let $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_n)^\mathsf{T}$, where ϕ_j is the value of the eigenfunction corresponding to the largest eigenvalue at the vertex v_j, $j = 1, \ldots, n$. Then, the absolute value of this eigenfunction at the central vertex v_1 cannot be exceeded by those at the other vertices, i.e.,*

$$|\phi_1| > |\phi_j|, \quad j = 2, \ldots, n.$$

The dendritic trees of neurons, of course, are more complicated than starlike trees. However, we managed to prove the following theorem for general unweighted graphs.

Theorem 2 ([9]). *For any unweighted* graph G *of finite volume, we have*

$$0 \le m_G([4,\infty)) \le \#\{j \in [1,n] \,|\, d(v_j) \gneqq 2\},$$

where $m_G(I)$ is the number of the Laplacian eigenvalues of G belonging to the interval $I \subset \mathbb{R}$. Moreover, each eigenfunction corresponding to $\lambda \ge 4$ has its largest component (in the absolute value) on the vertices whose degree are larger than 2.

Moreover, we also proved the decay of the eigenfunction magnitudes along a branching path for a general graph.

Theorem 3 ([9]). *Suppose that an unweighted graph G has a branch consisting of a path of length k, say, $\{v_{i_1}, \dots, v_{i_k}\}$ with v_{i_k} being the leaf of that branch. Then for any $\lambda > 4$, the corresponding eigenfunction $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_n)^{\mathsf{T}}$ satisfies*

$$|\phi_{i_{j+1}}| \le \gamma |\phi_{i_j}| \quad \text{for } j = 1, 2, \dots, k-1, \quad \gamma := 2/(\lambda - 2) < 1.$$

Hence $|\phi_{i_j}| \le \gamma^{j-1} |\phi_{i_1}|$ for $j = 1, \dots, k$, that is, the eigenfunction along the branch decays exponentially with the rate γ.

For the proofs of the above theorems, the Geršgorin Circle Theorem played a crucial role [9].

We have also identified a unique class of trees that can have an eigenvalue exactly equal to 4 as follows:

Corollary 4 ([9] based on Theorem 4 of Guo [6]). *A tree has an eigenvalue exactly equal to 4 iff it consists of vertex disjoint copies of $K_{1,3}$.*

There are still many open problems related to the graph Laplacian eigenvalue 4 and the eigenfunction localizations. For example, for a finite lattice graph (or regular grids) in $\mathbb{R}^d$ of size n^d, we know that:

$$m_G(4) \begin{cases} = 0 & \text{if } d = 1; \\ = n - 1 & \text{if } d = 2; \\ \ge \binom{d}{2}(n-1) & \text{if } d \ge 3. \end{cases}$$

However, it seems quite difficult to figure out the exact multiplicity of the eigenvalue 4 for $d \ge 3$. For example, in the case of $d = 3$, for most n, we have $m_G(4) = 3(n-1)$; however, if n is a multiple of 6, then $m_G(4) = 3n$. Moreover, if $n = 15$, then $m_G(4) = 3(n+1)$. There seems to be an intimate relationship with the analytic number theory and the uncertainty principle [14], which we would like to investigate in the near future.

We also know that unexpected things can happen if we deal with *weighted* graphs. For example, for a 1D lattice graph (or a path graph) with uniform weights, there is no eigenfunction localization phenomenon, i.e., the eigenfunctions are simply cosine functions with increasing frequency. As soon as one introduces

uneven edge weights in such a simple 1D lattice graph, the wavelet-like localized eigenfunctions start to emerge [9]. We are currently analyzing this phenomenon using the similar techniques we used to prove the above theorems.

Finally, the second part of the talk, we briefly discussed the importance of constructing a *sparse* graph with a given set of input vectors as vertices. A sparse graph here means that a connected graph with a relatively small number of edges. Compared to complete or dense graphs, such sparse graphs often provide: 1) computational efficiency for processing data defined or measured at vertices; and 2) better performance of the tasks at hand (e.g., classification, regression, segmentation, etc.). In particular, we reviewed two recent methods of such sparse graph constructions: one is the method based on the ℓ^2-minimization of the product of the graph Laplacian matrix and the transposed input data matrix proposed by Daitch, Kelner, and Spielman [4]; and the other is the so-called ℓ^1-graph that uses ℓ^1-sparse approximation of input vectors proposed by Cheng, et al. [1]. The DKS method requires convex quadratic programming and may not be too robust for outliers and noise unless their modified approach called an α-soft graph that allows αn vertices with low weighted degree is used. However, the determination of this parameter $\alpha \in (0, 1)$ may not be trivial. On the other hand, the ℓ^1-graph seems quite robust against outliers and noise, but the computational burden looks severe. We are currently comparing these methods in terms of computational costs and performances using several standard datasets. We also commented some recent effort of constructing wavelet-like transforms on graphs (see, e.g., [3, 13, 8, 7, 5, 2, 10]), and emphasized the importance of transferring conventional Fourier and wavelet analysis tools developed for the standard Euclidean domains (intervals, cubes, and regular lattices) to more general graph settings.

References

[1] H. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, *Learning with ℓ^1-graph for image analysis*, IEEE Trans. Image Process., **19** (2010), 858–866.

[2] R. R. Coifman and M. Gavish, *Harmonic analysis of digital data bases*, in Wavelets and Multiscale Analysis: Theory and Applications (J. Cohen and A. Zayed, eds.), Applied and Numerical Harmonic Analysis, Boston, MA, 2011, Birkhäuser.

[3] R. R. Coifman and M. Maggioni, *Diffusion wavelets*, Applied and Computational Harmonic Analysis, **21** (2006), 53–94.

[4] S. I. Daitch, J. A. Kelner, and D. A. Spielman, *Fitting a graph to vector data*, in Proc. 26th Intern. Conf. Machine Learning (L. Bottou and M. L. Littman, eds.), Omnipress, 2009, 201–208.

[5] M. Gavish, B. Nadler, and R. R. Coifman, *Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning*, in Proc. 27th Intern. Conf. Machine Learning (J. Fürnkranz and T. Joachims, eds.), Haifa, Israel, 2010, Omnipress, 367–374.

[6] J.-M. Guo, *The kth Laplacian eigenvalue of a tree*, J. Graph Theor., **54** (2006), 51–57.

[7] D. K. Hammond, P. Vandergheynst, and R. Gribonval, *Wavelets on graphs via spectral graph theory*, Applied and Computational Harmonic Analysis, **30** (2011), 129–150.

[8] M. Jansen, G. P. Nason, and B. W. Silverman, *Multiscale methods for data on graphs and irregular multidimensional situations*, J. R. Stat. Soc. Ser. B, Stat. Methodol., **71** (2008), 97–125.
[9] Y. Nakatsukasa, N. Saito, and E. Woei, *Mysteries around the graph Laplacian eigenvalue* 4, tech. rep., Dept. Math., Univ. California, Davis, 2012. Submitted for publication.
[10] R. M. Rustamov, *Average interpolating wavelets on point clouds and graphs.* arXiv:1110.2227v1 [math.FA], 2011.
[11] N. Saito and E. Woei, *Analysis of neuronal dendrite patterns using eigenvalues of graph Laplacians*, JSIAM Letters, **1** (2009), 13–16. Invited paper.
[12] ———, *On the phase transition phenomenon of graph Laplacian eigenfunctions on trees*, RIMS Kôkyûroku, **1743** (2011), 77–90.
[13] A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. Bremer, Jr., *Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions*, in Wavelets XI, Proc. SPIE **5914** (M. Papadakis, A. F. Laine, and M. A. Unser, eds.), 2005. Paper # 59141D.
[14] T. Tao, *An uncertainty principle for cyclic group*, Math. Res. Lett., **12** (2005), 121–127.

Image restoration: total variation, wavelet frames, and beyond

Zuowei Shen
(joint work with Jianfeng Cai, Bin Dong, Stanley Osher)

This talk is based on the recent paper [1].
The variational techniques (e.g., the total variation based method [3]) are well established and effective for image restoration, as well as many other applications, while wavelet frame based approach is relatively new and came from a different school (see e.g. [4]) . This paper is designed to establish a connection between these two major approaches for image restoration. The main result of this paper shows that when spline wavelet frames of [2] are used, a special model of a wavelet frame method, called the analysis based approach, can be viewed as a discrete approximation at a given resolution to variational methods. A convergence analysis as image resolution increases is given in terms of objective functionals and their approximate minimizers. This analysis goes beyond the establishment of the connections between these two approaches, since it leads to new understandings for the both approaches. First, it provides geometric interpretations to the wavelet frame based approach as well as its solutions. On the other hand, for any given variational model, wavelet frame based approaches provide various and flexible discretizations which immediately lead to fast numerical algorithms for both the wavelet frame based approaches and the corresponding variational model. Furthermore, the built-in multiresolution structure of wavelet frames can be utilized to adaptively choose proper differential operators in different regions of a given image according to the order of the singularity of the underlying solutions. This is important when multiple orders of differential operators are used in various models that generalize the total variation based method. These observations will enable us to design new methods according to the problems at hand, hence, lead to wider applications of both the variational and wavelet frame based approaches.

Links of wavelet frame based approaches to some more general variational methods developed recently will also be discussed.

References

[1] J. Cai,B. Dong, S. Osher, and Z. Shen *Image restoration: total variation, wavelet frames, and beyond*, Journal of the American Mathematical Society **xx** (2012), to appear.

[2] A. Ron and Z. Shen *The Analysis of the Analysis Operator*, Journal of Functional Analysis, **148** (1997), 407–448.

[3] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, **60** (1992), 259–268.

[4] Z. Shen, *Wavelet frames and image restorations*, Proceedings of the International Congress of Mathematicians, Hyderabad, India, **IV** (2010), 2834–2863.

Algebraic geometry and tight wavelet frames

Joachim Stöckler

(joint work with Maria Charina, Mihai Putinar, Claus Scheiderer)

Recent results from real algebraic geometry and the theory of moments are related in a novel framework in [1] to the existence of multivariate tight wavelet frames. The construction of multivariate tight frames by the so-called Unitary Extension Principle [3] and succeeding results [2] require the existence of multivariate trigonometric polynomials $Q_1, \ldots, Q_N$, such that a given nonnegative trigonometric polynomial f has a sum-of-squares decomposition

$$f = \sum_{j=1}^{N} Q_j^* Q_j.$$

Scheiderer's local-global principles of real algebraic geometry [4] confirm the existence of such a decomposition in the two-dimensional case, and provide sufficient conditions based on the Hessian of f at all zeros of f for higher dimensions. Moreover, two methods for the construction of Q_j, $1 \le j \le N$, are presented. The first construction uses tools from semi-definite programming. The second construction employs the representation of non-negative trigonometric polynomials as transfer functions of linear systems.

References

[1] M. Charina, M. Putinar, C. Scheiderer, J. Stöckler, *A real algebra perspective on multivariate tight wavelet frames*, arXiv:1202.3596 (February 2012).

[2] M. J. Lai, J. Stöckler, *Construction of multivariate compactly supported tight wavelet frames*, Appl. Comp. Harmonic Anal. **21** (2006), 324–348.

[3] A. Ron, Z. Shen, *Affine systems in $L_2(R^d)$: the analysis of the analysis operator*, J. Func. Anal. **148** (1997), 408–447.

[4] C. Scheiderer, *Sums of squares on real algebraic surfaces*, Manuscripta Math. **119** (2006), 395–410.

Phase Retrieval and Convex Optimization

Thomas Strohmer
(joint work with Emmanuel Candès, Vlad Voroninski, and YoninaEldar)

In many applications, one would like to acquire information about an object but it is impossible or very difficult to measure and record the phase of the signal. The problem is then to reconstruct the object from intensity measurements only. A problem of this kind that has attracted a considerable amount of attention over the last hundred years or so, is of course that of recovering a signal or image from the intensity measurements of its Fourier transform as in X-ray crystallography. As is well-known, such phase retrieval problems are notoriously difficult to solve numerically.

We develop a novel methodology for phase retrieval based on a rigorous and flexible numerical framework. Whereas most of the existing methods seek to overcome nonuniqueness by imposing additional constraints on the signal, we pursue a different direction by assuming no constraints at all on the signal. There are two main components to our approach.

- *Multiple structured illuminations.* We suggest collecting several diffraction patterns providing 'different views' of the sample or specimen. This can be accomplished in a number of ways: for instance, by modulating the light beam falling onto the sample or by placing a mask right after the sample. Taking multiple diffraction patterns usually yields uniqueness.

  A variety of methods have been proposed to carry out these multiple measurements; depending on the particular application, these may include the use of various gratings and/or of masks, the rotation of the axial position of the sample, and the use of defocusing implemented in a spatial light modulator. Other approaches include *ptychography*, an exciting field of research, where one records several diffraction patterns from overlapping areas of the sample.
- *Formulation of phase recovery as a matrix completion problem.* We suggest (1) lifting up the problem of recovering a vector from quadratic constraints into that of a recovering of a rank-one matrix from affine constraints, and (2) relaxing the combinatorial problem into a convenient convex program. Since the lifting step is fundamental to our approach, we will refer to the proposed numerical framework as *PhaseLift*. The price we pay for trading the nonconvex quadratic constraints into convex constraints is that we must deal with a highly underdetermined problem. However, recent advances in the areas of compressive sensing and matrix completion have shown that such convex approximations are often exact.

A significant aspect of our methodology is that our systematic optimization framework offers a principled way of dealing with noise, and makes it easy to handle various statistical noise models. This is important because in practice, measurements are always noisy. In fact, our framework can be understood as an elaborate regularized maximum likelihood method. Lastly, our framework can also include

a priori knowledge about the signal that can be formulated or relaxed as convex constraints.

Formally, suppose $x \in \mathbb{C}^n$ is a discrete signal and that we are given information about the squared modulus of the inner product between the signal and some vectors z_i, namely,

$$b_i = |\langle x, z_i \rangle|^2, \quad i = 1, \ldots, m.$$

We would like to know $\langle x, z_i \rangle$ and record both phase and magnitude information but can only record the magnitude; in other words, phase information is lost.

Letting $\mathcal{A}$ be the linear transformation

$$\begin{array}{ccc} \mathcal{H}^{n\times n} & \mapsto & \mathbb{R}^m \\ X & \mapsto & \{z_i^* X z_i\}_{1\le i\le m} \end{array}$$

which maps Hermitian matrices into real-valued vectors, one can express the data collection $b_i = |\langle x, z_i \rangle|^2$ as

$$b = \mathcal{A}(xx^*).$$

For reference, the adjoint operator $\mathcal{A}^*$ maps real-valued inputs into Hermitian matrices, and is given by

$$\begin{array}{ccc} \mathbb{R}^m & \mapsto & \mathcal{H}^{n\times n} \\ y & \mapsto & \sum_i y_i\, z_i z_i^*. \end{array}$$

As observed in [3, 2] the phase retrieval problem can be cast as the matrix recovery problem

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0. \end{array}$$

Indeed, we know that a rank-one solution exists so the optimal X has rank at most one. We then factorize the solution as xx^* in order to obtain solutions to the phase-retrieval problem. This gives x up to multiplication by a unit-normed scalar. This is all we can hope for since if x is a solution to the phase retrieval problem, then cx for any scalar $c \in \mathbb{C}$ obeying $|c| = 1$ is also solution.[1]

Rank minimization is in general NP hard, and we propose, instead, solving a trace-norm relaxation. Formally, we suggest solving

$$\begin{array}{ll} \text{minimize} & \operatorname{trace}(X) \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0. \end{array} \tag{1}$$

If the solution has rank one, we factorize it as above to recover our signal. This method which lifts up the problem of vector recovery from quadratic constraints into that of recovering a rank-one matrix from affine constraints via semidefinite programming is known under the name of *PhaseLift* [2].

Our main result is that the convex program recovers x exactly (up to global phase) provided the number m of magnitude measurements is on the order of $n \log n$.

[1]When the solution is unique up to multiplication by such a scalar, we shall say that unicity holds up to global phase.

Theorem 1. *Consider an arbitrary signal x in $\mathbb{C}^n$ and suppose that the number of measurements obeys $m \geq c_0\, n \log n$, where c_0 is a sufficiently large constant. Then in both the real and complex cases, the solution to the trace-minimization program is exact with high probability in the sense that* (1) *has a unique solution obeying*

$$\hat{X} = xx^*.$$

This holds with probability at least $1 - 3e^{-\gamma \frac{m}{n}}$, where γ is a positive absolute constant.

Expressed differently, the theorem above establishes a rigorous equivalence between a class of phase retrieval problems and a class of semidefinite programs. We refer to [1] for the proof as well as a theorem establishing robustness vis a vis additive noise. Details about multiple structured illuminations as well as numerical algorithms can be found in [2].

References

[1] E. J. Candes, T. Strohmer and V. Voroninski. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Communications on Pure and Applied Mathematics*, to appear.

[2] E. J. Candes, Y. Eldar, T. Strohmer and V. Voroninski. Phase retrieval via matrix completion. *Preprint*, 2011.

[3] A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. Technical report, Stanford University, 2010.

Fast subspace search and applications to object recognition

Arthur Szlam

(joint work with Karol Gregor, Yann LeCun)

Sparse modeling [3, 4] has proven to be a useful framework for signal processing. Each point from a dataset consisting of vectors in a Euclidean space is represented by a vector with only a few nonzero coefficients. Sparse modeling has lead to state of the art algorithms in image denoising, inpainting, supervised learning, and of particular interest here, object recognition. The systems described in [1, 2, 5, 6, 7, 8] use sparse coding as an integral element. Since the coding is done densely in an image with relatively large dictionaries, this is a computationally expensive part of the recognition system, and a barrier to real time application.

The main contribution described in this extended abstract of our paper[9] is a fast approximate algorithm for finding sparse representations; we use this algorithm to build a system with near state of the art recognition performance that runs in real time. During inference the algorithm uses a tree to assign an input to a group of allowed dictionary elements and then finds the corresponding coefficient values using a cached pseudoinverse. We give an algorithm for learning the tree, the dictionary and the dictionary element assignment.

One standard formulation of sparse coding is to consider N d-dimensional real vectors $X = \{x_1, \ldots, x_N\}$ and represent them using N K-dimensional real vectors $Z = \{z_1, \ldots, z_N\}$ using a $k \times d$ dictionary matrix W by solving

$$\text{argmin}_{Z,W} \sum_k ||Wz_k - x_k||^2, \text{ s.t. } ||z_k||_0 \leq q, \tag{1}$$

where $||\cdot||_0$ measures the number of nonzero elements of a vector; each input vector x is thus represented as a vector z with at most q nonzero coefficients. While this problem is not convex, and in fact the problem in the Z variable is NP-hard, there exist algorithms for solving both the problem in Z (e.g. Orthogonal Matching Pursuit, OMP [10]) and the problem in both variables (e.g. K-SVD [4]) that work well in many practical situations.

A simple form of structured sparsity is given by specifying a list of L allowable active sets of coefficients (coefficients that are allowed to be nonzero), and some function $g : \mathbb{R}^d \mapsto \{1, ..., L\}$ associating to each x to one of the L configurations. An example of this is the output of many subspace clustering algorithms. There each of the L configurations corresponds to the subspace spanned by a basis B_j, $j = 1, \ldots, L$. We can treat the sparse coding dictionary W as concatenation of the bases $W = [B_1 \ldots B_L]$. If each B is of size q then the allowed active sets of coefficients are $\{1, ..., q\}$, $\{q + 1, ..., 2q\}$, etc. This sort of method is used in object recognition in [6].

We discusses learning the L configurations as well as the dictionary. We introduce a LLoyd-like algorithm [15] that alternates between updating the dictionary, updating the assignments of each data point to the groups, and updating the dictionary elements associated to a group via simultaneous OMP (SOMP) [16].

At inference time, we need a fast method for determining which group an x belongs to. This is computationally expensive if there is a large number of groups and one needs check the projection onto each group. However, by specializing the Lloyd type algorithm to the case when each group is composed of a union of (perhaps only one) leaves of a binary decision tree, we will build a fast inference scheme into the learned dictionary. The key idea is that by using SOMP, we can learn which leaves should use which dictionary elements as we train the dictionary. To code an input, we march it down the tree until we arrive at the appropriate leaf. In addition to the decision vectors and thresholds, we will store a lookup table with the active set of each leaf as learned above, and the pseudoinverse of the columns of W corresponding to that active set. Thus after following x down the tree we need only make one matrix multiplication to get the coefficients.

Finally, we would like use these algorithms to build an accurate real time recognition system. We focus on a particular architecture studied in [1, 2, 6, 8]. We use this pipeline with two modifications. First we write our own fast implementation of the SIFT descriptor. Second we use our fast algorithm for the sparse coding step. The resulting system achieves nearly the same performance as exact sparse coding calculation but processes 321×481 size images at the rate of 20 frames per second on a laptop computer with a quad core cpu.

There is much left to understand here. Hashing algorithms empirically work well in the sense of finding almost as near as the nearest one when there are not too many subspaces. On the other hand, when the number of subspaces is combinatorially large, as in Equation (1), they do not perform very well in the mean square error sense. However, even though in that case the l_2 error is not nearly optimal, they still perform well as a component of the object recognition system. It would be nice to understand rigorously under what conditions we can expect good mean square error results, and while it is too much, perhaps, to hope for rigor, better understanding of what is actually important for the recognition system would be useful.

References

[1] S. Lazebnik, C.S., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR'06. (2006)
[2] Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: CVPR'09. (2009)
[3] Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381** (1996) 607–609
[4] Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54** (2006) 4311–4322
[5] Kavukcuoglu, K., Ranzato, M., LeCun, Y.: Fast inference in sparse coding algorithms with applications to object recognition. Technical Report CBLL-TR-2008-12-01, Computational and Biological Learning Lab, Courant Institute, NYU (2008)
[6] J. Yang, K.Y., Huang., T.: Efficient highly overcomplete sparse coding using a mixture model. In: European Conference on Computer Vision. (2010)
[7] Boureau, Y., Bach, F., LeCun, Y., Ponce, J.: Learning mid-level features for recognition. In: Proc. International Conference on Computer Vision and Pattern Recognition (CVPR'10), IEEE (2010)
[8] Boureau, Y., La Roux, N., Bach, F., Ponce, J., LeCun, Y.: Ask the locals: multi-way local pooling for image recognition. In: International Conference on Computer Vision. (2011)
[9] Szlam, A., Gregor, K., LeCun, Y.: Fast approximations to structured sparse coding and applications to object classification. In: European Conference on Computer Vision. (2012)
[10] Pati, Y.C., Rezaiifar, R., Rezaiifar, Y.C.P.R., Krishnaprasad, P.S.: Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In: Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers. (1993) 40–44
[11] Jenatton, R., Mairal, J., Obozinski, G., Bach, F.: Proximal methods for sparse hierarchical dictionary learning. In: International Conference on Machine Learning (ICML). (2010)
[12] Kim, S., Xing, E.P.: Tree-guided group lasso for multi-task regression with structured sparsity. In: ICML. (2010) 543–550
[13] Jacob, L., Obozinski, G., Vert, J.P.: Group lasso with overlap and graph lasso. In: Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09, New York, NY, USA, ACM (2009) 433–440
[14] Baraniuk, R.G., Cevher, V., Duarte, M.F., Hegde, C.: Model-Based Compressive Sensing. (2009)
[15] Lloyd, S.P.: Least squares quantization in pcm. IEEE Transactions on Information Theory **28** (1982) 129–137
[16] Gilbert, A.C., Strauss, M.J., Tropp, J.A.: Simultaneous Sparse Approximation via Greedy Pursuit. IEEE Trans. Acoust. Speech Signal Process. **5** (2005) 721–724

Sharp recovery bounds for convex deconvolution, with applications

Joel A. Tropp
(joint work with Michael B. McCoy)

Deconvolution refers to the challenge of identifying two structured signals given only the sum of the two signals and prior information about their structures. A standard example is the problem of separating a signal that is sparse with respect to one basis from a signal that is sparse with respect to a second basis. Another familiar case is the problem of decomposing an observed matrix into a low-rank matrix plus a sparse matrix. This paper describes and analyzes a framework, based on convex optimization, for solving these deconvolution problems and many others.

This work introduces a randomized signal model which ensures that the two structures are incoherent, i.e., generically oriented. For an observation from this model, the calculus of spherical integral geometry provides an exact formula that describes when the optimization problem will succeed (or fail) to deconvolve the two constituent signals with high probability. This approach identifies a summary statistic that reflects the complexity of a particular signal. The difficulty of separating two structured, incoherent signals depends only on the total complexity of the two structures.

Some applications include (i) deconvolving two signals that are sparse in mutually incoherent bases; (ii) decoding spread-spectrum transmissions in the presence of impulsive errors; and (iii) removing sparse corruptions from a low-rank matrix. In each case, the theoretical analysis of the convex deconvolution method closely matches its empirical behavior.

N-Widths for High-Dimensional Sparse Approximations

Tino Ullrich
(joint work with Zung Dinh)

In recent decades, there has been increasing interest in solving problems that involve functions depending on a large number d of variables. These problems arise from many applications in mathematical finance, chemistry, physics, especially quantum mechanics, and meteorology. It is not surprising that these problems can almost never be solved analytically such that one is interested in a proper framework and efficient numerical methods for an approximate treatment. Classical methods suffer the so-called "curse of dimensionality" coined by Bellmann in 1957. In fact, the computation time typically grows exponentially in the dimension d, and the problems become intractable already for mild dimensions without further assumptions on the model. A classical model, widely studied in literature, is to impose certain smoothness conditions on the function to be approximated; in particular, it is assumed that mixed derivatives are bounded.

In the present talk, we discuss linear sparse hyperbolic cross grid approximations and the well-known Kolmogorov n-widths in isotropic Sobolev space H^γ, $\gamma \in \mathbb{R}$, of

periodic multivariate function classes with both, mixed and isotropic smoothness properties in high-dimensional settings. In particular, if W is a class of d-variate functions and n represents the dimension of the linear approximation space, both parameters n and d play the same essential role for the asymptotic estimates of the n-widths $d_n(W, X)$.

If X is a normed space and W a subset in X then the Kolmogorov n-width $d_n(W, X)$ is given by

$$d_n(W, X) := \inf_{L_n} \sup_{f \in W} \inf_{g \in L_n} \|f - g\|_X,$$

where the outer inf is taken over all linear manifolds L_n in X of dimension at most n. It represents a characterization of the optimal worst case error from a llinear subspace of order n. In the talk we are interested in measuring the approximation error in H^γ, therefore we can assume X to be a Hilbert space H.

For the unit balls U^α and $U^{\alpha\mathbf{1}}$ of the periodic d-variate isotropic Sobolev space H^α and the space $H^{\alpha\mathbf{1}}$ with mixed smoothness $\alpha > 0$, the following well-known estimates hold true

$$A(\alpha, d)n^{-\alpha/d} \;\le\; d_n(U^\alpha, L_2) \;\le\; A'(\alpha, d)n^{-\alpha/d}, \tag{1}$$

and

$$B(\alpha, d)n^{-\alpha}(\log n)^{\alpha(d-1)} \;\le\; d_n(U^{\alpha\mathbf{1}}, L_2) \;\le\; B'(\alpha, d)n^{-\alpha}(\log n)^{\alpha(d-1)}. \tag{2}$$

Here $A(\alpha, d)$, $A'(\alpha, d)$, $B(\alpha, d)$, $B'(\alpha, d)$ are certain constants which are usually not computed explicitly. The inequalities (1) are a direct generalization of the first result on n-widths proven by Kolmogorov [3] where the exact values of n-widths were obtained for the univariate case. The inequalities (2) were proven by Babenko already in 1960, where a linear approximation on hyperbolic cross spaces of trigonometric polynomial is used. These estimates are quite satisfactory if d the number of variables is small and fixed. In high-dimensional settings, i.e., if d is large, it turns out that the smoothness of the isotropic Sobolev class U^α is not a proper concept. Indeed, in (1) the curse of dimensionality occurs since here $n_\varepsilon \ge C(\alpha, d)\varepsilon^{-d/\alpha}$. In this talk, we extend and refine existing estimates. In particular, we give the lower and upper bounds for constants $B(\alpha, d)$, $B'(\alpha, d)$ in (2) with regards to α, d. In fact, we are concerned with measuring the approximation error in the isotropic smoothness space H^γ. This is motivated by Galerkin's method and estimates of the error in the energy norm.

The curse of dimensionality is not sufficiently clarified unless "constants" such as $B(\alpha, d)$, $B'(\alpha, d)$ in (2) for d_n are not completely determined. We are interested, so far possible, in explicitly determining these constants. We computed $d_n(U, H^\gamma)$ and $n_\varepsilon(U, H^\gamma)$ where U is the unit ball $U^{\alpha,\beta}$ in $H^{\alpha,\beta}$. Here $H^{\alpha,\beta}$ is a hybrid type space combining mixed smoothness α and isotropic smoothness β.

It seems that smoothness is not enough for ridding the curse of dimensionality. However, by imposing some additional restrictions on functions in $U^{\alpha,\beta}$ this is possible. In fact, $U^{\alpha,\beta}_\nu$ is the set of all functions $f \in U^{\alpha,\beta}_\nu$ actually depending on at most ν (unknown) variables by formally being a d-variate function. For this function class, the curse of dimensionality is broken. We establish sharp lower and

upper bounds in an explicit form of all relevant components depending on α, β, γ and d, n, ν. This includes the case (2) and its modifications when $\alpha > \gamma = \beta = 0$. In contrast to [2] we also obtain lower bounds and prove therefore that sparse hyperbolic cross approximation is optimal in terms of Kolmogorov n-widths. For the case $\alpha > \gamma - \beta > 0$, we prove that the optimized sparse grid spaces from [2] are optimal for $d_n(U_*^{\alpha,\beta}, H^\gamma)$. Moreover, the modifications given are optimal for $d_n(U^{\alpha,\beta}, H^\gamma)$ and $d_n(U_\nu^{\alpha,\beta}, H^\gamma)$. In the case $\alpha > \gamma - \beta = 0$, we prove that classical hyperbolic cross spaces (see, e.g., [5]) and their modifications are optimal for $d_n(U_*^{\alpha,\beta}, H^\gamma)$, $d_n(U^{\alpha,\beta}, H^\gamma)$ and $d_n(U_\nu^{\alpha,\beta}, H^\gamma)$. We obtain for instance the relations

$$\begin{aligned} \frac{1}{2^{\rho+3\delta}}\nu^\delta \left(1 + \frac{d}{\nu(2^{\rho/\delta}-1)}\right)^{\delta\nu} n^{-\delta} \;&\le\; d_n(U_\nu^{\alpha,\beta}, H^\gamma) \\ &\le\; \left(\frac{\alpha}{\delta}\right)^\delta 2^{2\rho+\delta}\nu^\delta \left(1 + \frac{d}{2^{\rho/\delta}-1}\right)^{\delta\nu} n^{-\delta}, \end{aligned}$$

if $n \ge \frac{\alpha}{\delta}\nu 2^{\nu(2\alpha/\delta+1)}(1 + d/(2^{\rho/\delta}-1))^\nu$, where $\delta := \alpha + \beta - \gamma$ and $\rho := \gamma - \beta$. On the other hand, what concerns the curse of dimensionality, we show negative results for the class $U^{\alpha,\beta}$ in H^γ. A corresponding result for the ε-dimension n_ε states that the number $n_\varepsilon(U_\nu^{\alpha,\beta}, H^\gamma)$ is bounded polynomially in d and ε^{-1} from above. As a consequence, according to [4, (2.3)], we obtain that the problem is polynomially tractable. In addition, the case $\gamma = \beta$, which contains the classical situation with $U_\nu^{\alpha\mathbf{1}}$ instead of $U^{\alpha\mathbf{1}}$ in (2), gives as well the polynomial tractability.

Let us mention the relation to the results of Novak and Woźniakowski on weighted tensor product problems with finite order weights [4, 5.3]. Their approach also limits the number ν of active variables in a function via a finite order weight sequence (of order ν). However, since here in most cases neither the spaces $H^{\alpha,\beta}$ of the functions to be approximated, nor the space H^γ, where the approximation error is measured, are tensor product spaces of univariate ones, our results are not included in [4, Theorem 5.8].

References

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

[2] M. Griebel and S. Knapek, *Optimized general sparse grid approximation spaces for operator equations*, Math. Comp., **78**,268 (2009):2223–2257.

[3] A.N. Kolmogorov, *Über die beste Annäherung von Funktionen einer Funktionklasse*, Ann. Math., **37** (1936), 107–111.

[4] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems, Volume I: Linear Information*, EMS Tracts in Mathematics, Vol. 6., Eur. Math. Soc. Publ. House, Zürich 2008.

[5] V. N. Temlyakov, *Approximation of Periodic Functions*, Computational Mathematics and Analysis Series. Nova Science Publishers Inc., Commack, NY, 1993.

Two New Algorithms Useful For the Regularization of Large Scale Inverse Problems With Sparsity Constraints

Sergey Voronin

We introduced two new algorithms useful for the regularization of large scale inverse problems which possess sparsely representable solutions. These algorithms are designed to minimize a generalized sparsity promoting functional and are based on a reweighted two norm technique, which makes the non-smooth terms differentiable. They are simple to implement and parallelize and exhibit promising numerical performance.

The motivation for these algorithms lies in an application in Geotomography. This inverse problem, like many others from the physical sciences, has several key characteristics which impact the type of algorithms that may be used to find reconstructions. We have a very large under-determined linear system with a noisy data vector and a badly conditioned matrix, from which we would like to find a solution in the least squares sense. Additional constraints must be imposed to make the problem well posed, which is done via regularization. It turns out that a sparse solution may be found under the action of a transform (for example a Wavelet transform) and we would like to look for such solutions. These are the settings under which we design our schemes. For reasons of size, we prefer simple, easy to implement and easy to parallelize methods which do not require significantly more than matrix vector and matrix transpose vector multiplications.

Two typical constraints are often enforced to give sparse solutions: the minimization of the ℓ_0 and ℓ_1 norms. In the context of noisy problems, this corresponds to solving the problems: $\min ||x||_0$ s.t. $||Ax - b||_2 \leq \epsilon$ and $\min ||x||_1$ s.t. $||Ax - b||_2 \leq \epsilon$ which for some constant τ corresponds to minimizing the functionals:

$$G(x) = ||Ax - b||_2^2 + 2\tau||x||_0 \quad \text{and} \quad F(x) = ||Ax - b||_2^2 + 2\tau||x||_1$$

which we call the ℓ_0 and ℓ_1 functionals below. When x is expected to be sparse only under a certain transform, we can replace $||x||_1$ above with $||Wx||_1$ where W denotes the desired transform. For the first functional (ℓ_0), a wide variety of methods from compressed sensing exist. However, the ℓ_0 norm which counts the number of nonzero elements, is highly non-convex and difficult to deal with numerically, because of the existence of local minima. Additionally, these methods often rely on the matrix satisfying the Restricted Isometry Property, which amounts to satisfying a condition of the form:

$$(1 - \sigma_s)||y||_2^2 \leq ||A_s y||_2^2 \leq (1 + \sigma_s)||y||_2^2$$

for every s-columned submatrix A_s of A for some small constant σ_s and sparse vectors y. This is problematic for large matrices which are not well conditioned, since any such matrix A_s would likely have a nonzero null space that contains some sparse vectors. The end result is that methods based on the ℓ_0 norm are difficult to use for problems with our requirements.

For this reason, an attractive option is ℓ_1 optimization, which involves a convex functional with a single global minimum. The main difficulty here is treating the

non-smooth term $||x||_1 = \sum_{k=1}^{N} |x_k|$. Subgradient methods then lead to the soft-thresholding operation which is a component wise nonlinear operator defined via the minimization problem: $\mathbb{S}_\tau(b) = \arg\min_x ||x-b||^2 + 2\tau||x||_1$. A simple scheme called the Iterative Soft Thresholding Algorithm (ISTA) can then be used to minimize the ℓ_1 functional. The scheme can be derived from a so called majorization-minimization approach, where instead of minimizing the original function, we minimize the function which majorizes it, resulting in a simpler problem and a very straightforward algorithm:

$$x^{n+1} = \arg\min_x ||Ax-b||_2^2 + ||x-x^n||_2^2 - ||A(x-x^n)||_2^2 + 2\tau||x||_1$$
$$\implies x^{n+1} = \mathbb{S}_\tau(x^n + A^T b - A^T A x^n)$$

The advantage of the above scheme is its convergence for any matrix with spectral norm less than one: $||A||_2 < 1$ and any initial guess x^0. The disadvantage, however, is that the convergence of this algorithm is known to be very slow. In particular, it was shown in [1] that:

$$F(x^n) - F(x^*) \leq \frac{C||x^0 - x^*||_2^2}{2n}$$

where $x^* = \lim_{n\to\infty} x^n$. This motivates the search for different algorithms, which are of about the same complexity but which converge faster. One such answer is given by the FISTA scheme which employs a simple trick from the work of Nesterov and can be implemented simply as:

$$z^n = x^{n-1} + \frac{t_{n-1}-1}{t_n}(x^{n-1} - x^{n-2}) \quad , \quad t_n \in \mathbb{R}$$
$$x^{n+1} = \mathbb{S}_\tau(z^n + A^T b - A^T A z^n)$$

and has a significantly faster rate of convergence:

$$F(x^n) - F(x^*) \leq \frac{C_2||x^0 - x^*||_2^2}{(n+1)^2}$$

In this work we searched for other schemes based on a different approach. Instead of treating the non-smooth term of the ℓ_1 functional directly and using the soft thresholding operator, we have replaced the non-smooth portion by a smooth approximation. One such approximation is possible, simply by convoluting the absolute value function with a smooth function having a shrinking support, such as a narrow Gaussian. This simple approach results in a smooth approximation to the ℓ_1 functional but is rather crude and does not give good numerical results. In this work we instead use a reweighted approach that gets more accurate as the iterations progress:

$$||x||_1 = \sum_{k=1}^{N} |x_k| \approx \sum_{k=1}^{N} w_k^n x_k^2$$

where we use the weight $w_k^n = \frac{1}{\sqrt{(x_k^n)^2+(\epsilon_n)^2}}$ where the parameter $\epsilon_n \to 0$. The trick to proving convergence turns out to be in picking the parameter ϵ_n in a suitable way so that the limiting points of the iteration satisfy the required optimality conditions. Using a majorization-minimization approach similar to that of ISTA, we can define the algorithm via the minimization:

$$x^{n+1} = \arg\min_x ||Ax-b||_2^2 + ||x-x^n||_2^2 - ||A(x-x^n)||_2^2 + 2\tau \sum_{k=1}^N w_k^n x_k^2$$

$$\implies x_k^{n+1} = \frac{1}{1+\tau w_k^n}\left(x_k^n + A^T b - A^T A x_k^n\right)$$

where the ϵ_n are defined simply by $\epsilon_n = \min(\epsilon_{n-1}, \frac{1}{N}\sqrt{||x^n - x^{n-1}||} + \alpha^n)$ for some small $\alpha > 0$ and some initial $\epsilon_0 = 1$. For this to work we prove that $||x^n - x^{n-1}|| \to 0$. Numerically, this algorithm exhibits performance similar to that of the FISTA scheme. We show that with minor modifications for the weights, the scheme can be extended to minimize a more general functional:

$$||Ax-b||_2^2 + 2\sum_{k=1}^N \lambda_k |x_k|^{q_k} \quad , \quad 1 \le q_k \le 2$$

Finally, we propose a scheme that is more complicated at each step, yet more powerful. Instead of the above minimization, we define simply:

$$x^{n+1} = \arg\min_x ||Ax-b||_2^2 + 2\tau \sum_{k=1}^N w_k^n x_k^2 = \arg\min_x ||Ax-b||_2^2 + 2\tau ||D_n x||_2^2$$

Both terms of the above are quadratic and can be differentiated to yield:

$$x^{n+1} = (A^T A + \tau D_n^2)^{-1} A^T b$$

where D_n is a diagonal matrix containing the elements $\sqrt{w_k^n}$ where w_k^n are defined as before. The difficulty in the proofs lies again in picking the right subsequence ϵ_n, which is more challenging in this case, since the estimate $||x^n - x^{n-1}|| \to 0$ does not (at least readily) come about. At each iteration, the method requires a linear solve, which can be done, for instance, using some variant of a conjugate gradient method. At this step, the solution of the previous iteration can be used as a warm start for the linear solve so that at later iterations very few inner linear solve iterations are required. With this scheme, the number of outer iterations is significantly smaller and the total runtime can be reduced. Finally, by taking more general weights, the scheme can be adopted to minimize the generalized functional above, which is useful in applications. The proofs assume that the linear solve at each step is carried to completion. Both methods are simple to use and attractive, especially for larger problems, since they involve only simple matrix vector operations. In both cases we prove that all limit points of a particular subsequence of the iterates satisfy the optimality conditions for our generalized sparsity promoting functional, without any requirements on the matrix A except for its spectral norm to be less than one.

Acknowledgement

I would like to express many thanks to my Ph.D. advisor, Prof. Ingrid Daubechies for her help with various aspects of this work. I would like to also thank Prof. Holger Rauhut, for hosting my visit and giving me the initial idea to pursue the linear solve based scheme stated above, in Bonn, Germany.

References

[1] Beck, A. and Teboulle, M., A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM Imaging Science, 2: 183-202, doi: 10.1137/080716542

[2] Daubechies, I. and Defrise, M. and De Mol, C. (2004), An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics, 57: 1413-1457. doi: 10.1002/cpa.20042

[3] Daubechies, I., DeVore, R., Fornasier, M. and Gunturk, C. S. (2010), Iteratively reweighted least squares minimization for sparse recovery. Communications on Pure and Applied Mathematics, 63: 1-38. doi: 10.1002/cpa.20303

[4] Simons, F., Loris, I., Nolet, G., Daubechies, I., Voronin, S., Judd, S., Vetter, P., Charlety, J., Vonesch, C. (2011), Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. Geoph. J. Int., doi: 10.1111/j.1365-246X.2011.05190.x

Performance guarantees for total variation minimization from undersampled measurements

Rachel Ward

(joint work with Deanna Needell)

Compressed sensing (CS) is a new signal processing methodology where signals are acquired in compressed form as undersampled linear measurements. The applications of CS are abundant, ranging from radar and error correction to many areas of image processing. The underlying assumption that makes such acquisition and reconstruction possible is that most natural signals are *sparse* or *compressible*. We say that a signal $\boldsymbol{x} \in \mathbb{C}^p$ is s-sparse when

$$\|\boldsymbol{x}\|_0 \stackrel{\text{def}}{=} |\text{supp}(\boldsymbol{x})| \leq s \ll p. \tag{1}$$

Compressible signals are those which are well-approximated by sparse signals. In the CS framework, we acquire $m \ll p$ nonadaptive linear measurements of the form $\boldsymbol{y} = M(\boldsymbol{x}) + \boldsymbol{\xi}$, where $M : \mathbb{C}^p \to \mathbb{C}^m$ is an appropriate linear operator and $\boldsymbol{\xi}$ is vector modeling additive noise. The theory of CS ensures that under suitable assumptions on the measurement operator M, a sufficiently compressible signal can be accurately approximated by the signal of minimal ℓ_1-norm consistent with the measurements,

$$(L_1) \qquad \hat{\boldsymbol{x}} = \underset{\boldsymbol{w}}{\text{argmin}} \|\boldsymbol{w}\|_1 \quad \text{such that} \quad \|M(\boldsymbol{w}) - \boldsymbol{y}\|_2 \leq \varepsilon,$$

where ε bounds the noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$. The program (L_1) may be cast as a second order cone program (SOCP) and can be solved efficiently using standard convex programming methods.

To guarantee robust recovery of compressible signals via (L_1), Candès and Tao introduced in [1] the restricted isometry property (RIP) for a measurement operator M.

Definition 1. A linear operator $M : \mathbb{C}^p \to \mathbb{C}^m$ is said to have the restricted isometry property (RIP) of order $s \in \mathbb{N}$ and level $\delta \in (0,1)$ if

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \leq \|M(\boldsymbol{x})\|_2^2 \leq (1+\delta)\|\boldsymbol{x}\|_2^2 \qquad \text{for all } s\text{-sparse } \boldsymbol{x} \in \mathbb{C}^p. \tag{2}$$

For example, a matrix whose entries are independent and identical (i.i.d.) realizations of a properly-normalized subgaussian random variable will have the RIP with probability exceeding $1 - e^{-cm}$ once $m \approx \delta^{-2} s \log(p/s)$.

Candès, Romberg, and Tao [2] showed that when the measurement operator M has the RIP of order $O(s)$ and sufficiently small constant δ, the program (L_1) recovers an estimation $\hat{\boldsymbol{x}}$ to $\boldsymbol{x}$ that satisfies the error bound

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq C \left(\frac{\|\boldsymbol{x} - \boldsymbol{x_s}\|_1}{\sqrt{s}} + \varepsilon \right), \tag{3}$$

where $\boldsymbol{x_s}$ denotes the best s-sparse approximation to the signal $\boldsymbol{x}$. Using properties about Gel'fand widths of the ℓ_1 ball due to Kashin and Garnaev–Gluskin, this is the optimal minimax reconstruction rate from $m \approx s \log(p/s)$ nonadaptive linear measurements. Due to the rotational-invariance of an RIP matrix with randomized column signs [4], a completely analogous theory holds for signals that are compressible with respect to a known orthonormal basis or tight frame $\boldsymbol{D}$ by replacing $\boldsymbol{w}$ with $\boldsymbol{D}^*\boldsymbol{w}$ inside the ℓ_1-norm of the minimization problem (L_1) [5].

0.1. **Imaging with CS.** Natural images are highly compressible with respect to their gradient representation. A typical grayscale digital image, regarded as a signal $\boldsymbol{x} \in \mathbb{C}^{N^2}$, consists primarily of slowly-varying pixel intensities, with large jumps in intensity occurring only along edges. Figure 1 illustrates the gradient sparsity of a representative image $\boldsymbol{x} \in \mathbb{C}^{N^2}$ along with its discrete directional derivatives,

$$\boldsymbol{x}_u : \mathbb{C}^{N\times N} \to \mathbb{C}^{(N-1)\times N}, \qquad (\boldsymbol{x}_u)_{j,k} = \boldsymbol{x}_{j,k+1} - \boldsymbol{x}_{j,k} \tag{4}$$

$$\boldsymbol{x}_v : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times(N-1)}, \qquad (\boldsymbol{x}_v)_{j+1,k} = \boldsymbol{x}_{j,k+1} - \boldsymbol{x}_{j,k} \tag{5}$$

The discrete gradient transform $\boldsymbol{\nabla} : \mathbb{C}^{N^2} \to \mathbb{C}^{N\times N\times 2}$ is defined in terms of the directional derivatives,

$$\big(\boldsymbol{\nabla}[\boldsymbol{x}]\big)_{j,k} \stackrel{\text{def}}{=} \begin{cases} \big((\boldsymbol{x}_u)_{j,k}, (\boldsymbol{x}_v)_{j,k}\big), & 1 \leq j \leq N-1, \quad 1 \leq k \leq N-1 \\ \big(0, (\boldsymbol{x}_v)_{j,k}\big), & j = N, \quad 1 \leq k \leq N-1 \\ \big((\boldsymbol{x}_u)_{j,k}, 0\big), & k = N, \quad 1 \leq j \leq N-1 \\ \big(0,0\big), & j = k = N \end{cases}$$

FIGURE 1. An image, along with its horizontal and vertical discrete directional derivatives.

The ℓ_1-norm of the discrete gradient defines a seminorm for the space $\mathbb{C}^{N^2}$, often referred to as the *total variation* seminorm and denoted by

$$|\boldsymbol{x}|_{TV} \stackrel{\text{def}}{=} \|\boldsymbol{\nabla}[\boldsymbol{x}]\|_1. \tag{6}$$

Due to the gradient sparsity of natural images, it should not be surprising that the total variation minimization program

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{z}}{\operatorname{argmin}} |\boldsymbol{z}|_{TV} \quad \text{such that} \quad \|M(\boldsymbol{z}) - \boldsymbol{y}\|_2 \leq \varepsilon. \tag{TV}$$

is often used for image reconstruction, in setting of compressed sensing and more broadly in imaging applications such as denoising, deblurring, and inpainting. While (TV) is similar to the ℓ_1-minimization program (L_1), the RIP-based theoretical guarantees for (L_1) do not directly translate to recovery guarantees for (TV) because the gradient map $\boldsymbol{z} \to \boldsymbol{\nabla}[\boldsymbol{z}]$ is not well-conditioned. In fact, viewed as an invertible operator over mean-zero images, the condition number of the gradient map grows linearly with the signal side-length N.

0.2. **Our results.** We provide the first recovery guarantees for (TV) in the compressed sensing set-up. In the setting of two-dimensional images $\boldsymbol{x} \in \mathbb{C}^{N^2}$, we show that the gradient map is well-conditioned when restricted to signals lying in null space of a matrix with the restricted isometry property:

Theorem 1. There are choices of linear operators $M : \mathbb{C}^{N^2} \to \mathbb{C}^m$ with $m \approx s \log(N^2/s)$ for which the following holds for any image $\boldsymbol{x} \in \mathbb{C}^{N^2}$: Given noisy measurements $\boldsymbol{y} = M(\boldsymbol{x}) + \boldsymbol{\xi}$ with noise level $\|\boldsymbol{\xi}\|_2 \leq \varepsilon$, the total-variation minimizing signal

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{z}}{\operatorname{argmin}} |\boldsymbol{z}|_{TV} \quad \text{such that} \quad \|M(\boldsymbol{z}) - \boldsymbol{y}\|_2 \leq \varepsilon \tag{7}$$

satisfies the error bound

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq C \log(N^2/s) \Big(\frac{\|\boldsymbol{\nabla}[\boldsymbol{x}] - \boldsymbol{\nabla}[\boldsymbol{x}]_s\|_1}{\sqrt{s}} + \varepsilon \Big). \tag{8}$$

For details, see [6]. In words, the total-variation minimizer estimates $\boldsymbol{x}$ to within a factor of the noise level and best s-term approximation of its gradient. The bound in (8) is optimal up to the logarithmic factor $\log(N^2/s)$.

We also extend Theorem 1 to multidimensional signals $\boldsymbol{x} \in \mathbb{C}^{N^d}$ of arbitrary dimension $d \geq 2$, detailed in [7]. We show that the signal $\hat{\boldsymbol{x}} \in \mathbb{C}^{N^d}$ of minimal (d-dimensional) total variation seminorm consistent with $m \approx sd\log(N^d)$ appropriately-chosen linear measurements $\mathbf{y} = M(\boldsymbol{x}) + \xi$ will approximate $\boldsymbol{x}$ to within a factor of the noise level and the best s-term approximation to the (d-dimensional) discrete gradient of $\boldsymbol{x}$, modulo a single logarithmic factor in the signal dimension N^d. In particular, our results provide guarantees on total variation minimization in reconstructing three-dimensional digital movies, which represent sequences of gradually-changing images and thus have compressible (three-dimensional) discrete gradient.

0.3. **Open problems.** Our proof rests on the Sobolev inequalities for random subspaces from [7], which follow from bounds of Cohen, Dahmen, Daubechies, and DeVore in [3] on the compressibility of wavelet representations for functions of bounded variation. Unfortunately these bounds, and hence our results for total variation, do not hold in dimension $d = 1$; guarantees on the fidelity of total variation minimization in the one-dimensional setting remains an interesting open problem. It also remains open whether the sub-optimal logarithmic factor in Theorem 1 can be removed. Finally, guarantees for total variation from *structured random* measurements such as partial rows from the Discrete Fourier matrix would be of practical and theoretical interest.

References

[1] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, **51**:4203–4215, 2005.

[2] *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Applied Math., **59**8, 1207–1223,2006.

[3] A. Cohen, W. Dahmen, I. Daubechies and R. DeVore, *Harmonic analysis of the space BV*, Revistaua Matematica Iberoamericana,**19**, 1: 235–263, 2003.

[4] F. Krahmer and R. Ward, *New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property*, **43**, 3: 1269–1281, 2011.

[5] E. Candès and Y. Eldar and D. Needell and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Hamon. A.,**31**,1:59–73,2011.

[6] D. Needell and R. Ward, *Stable image reconstruction using total variation minimization*, Submitted, 2012.

[7] D. Needell and R. Ward, *Performance guarantees for total variation minimization from undersampled measurements*, Submitted, 2012.

Reporter: Holger Rauhut

Participants

Prof. Dr. Daniel Abreu
Departamento de Matematica
Universidade de Coimbra
Apartado 3008
3000 Coimbra
PORTUGAL

Prof. Dr. Akram Aldroubi
Department of Mathematics
Vanderbilt University
1326 Stevenson Center
Nashville , TN 37240
USA

Prof. Dr. Radu Balan
Department of Mathematics
University of Maryland
College Park , MD 20742-4015
USA

Prof. Dr. Gregory Beylkin
Dept. of Applied Mathematics
University of Colorado at Boulder
Campus Box 526
Boulder , CO 80309-0526
USA

Prof. Dr. Holger Boche
Lehrstuhl für Theoretische
Informationstechnik
Technische Universität München
Theresienstr. 90/IV (LTI)
80333 München

Prof. Dr. Bernhard G. Bodmann
Department of Mathematics
University of Houston
Houston TX 77204-3008
USA

Prof. Dr. A. Robert Calderbank
Pratt School of Engineering
Duke University
Durham , NC 27708
USA

Prof. Dr. Emmanuel Candes
Department of Statistics
Stanford University
Sequoia Hall
Stanford , CA 94305-4065
USA

Prof. Dr. Peter G. Casazza
Department of Mathematics
University of Missouri-Columbia
202 Mathematical Science Bldg.
Columbia , MO 65211-4100
USA

Prof. Dr. Stephan Dahlke
FB Mathematik & Informatik
Philipps-Universität Marburg
Hans-Meerwein-Strasse (Lahnbg.)
35032 Marburg

Prof. Dr. Ingrid Daubechies
Department of Mathematics
Duke University
P.O.Box 90320
Durham , NC 27708-0320
USA

Prof. Dr. Christine De Mol
Department of Mathematics
Universite Libre de Bruxelles
CP 217 Campus Plaine
Bd. du Triomphe
B-1050 Bruxelles

Dr. Shai Dekel
School of Mathematical Sciences
Tel Aviv University
Ramat Aviv
Tel Aviv 69978
ISRAEL

Prof. Dr. Michael Elad
Computer Science Department
TECHNION
Israel Institute of Technology
Haifa 32000
ISRAEL

Prof. Dr. Matthew Fickus
Department of Mathematics & Statistics
Air Force Institute of Technology
2950 Hobson Way
Kettering , OH 45433
USA

Prof. Dr. Massimo Fornasier
Zentrum Mathematik
TU München
85747 Garching

Prof. Dr. Simon Foucart
Department of Mathematics
Drexel University
Korman Center 269
3141 Chestnut St.
Philadelphia , PA 19104
USA

Prof. Dr. Karlheinz Gröchenig
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Prof. Dr. Philipp Grohs
Seminar for Applied Mathematics
ETH Zürich
Rämistr. 101
CH-8092 Zürich

Dr. C. Sinan Güntürk
Courant Institute of
Mathematical Sciences
New York University
251, Mercer Street
New York , NY 10012-1110
USA

Dr. Mijail Guillemard
Institut für Mathematik
Sekr. MA 4-1
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Prof. Dr. Bin Han
Department of Mathematical and
Statistical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1
CANADA

Max Hügel
Mathematisches Institut
Universität Bonn
Endenicher Allee 60
53115 Bonn

Dr. Emily King
Institut für Mathematik
Sekr. MA 4-1
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Jun.-Prof. Dr. Felix Krahmer
Institut für Numerische
und Angewandte Mathematik
Georg-August-Universität Göttingen
Lotzestr. 16-18
37083 Göttingen

Dr. Ilya Krishtal
Department of Mathematics
Northern Illinois University
DeKalb , IL 60115
USA

Prof. Dr. Stefan Kunis
Institut für Mathematik
Universität Osnabrück
Albrechtstrasse 28a
49069 Osnabrück

Prof. Dr. Gitta Kutyniok
Institut für Mathematik
Sekr. MA 4-1
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Prof. Dr. Demetrio Labate
Department of Mathematics
University of Houston
Houston TX 77204-3008
USA

Dr. Wang-Q. Lim
Fachbereich Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Prof. Dr. Ignace Loris
Department of Mathematics
Universite Libre de Bruxelles
CP 217 Campus Plaine
Bd. du Triomphe
B-1050 Bruxelles

Ines Melzer
FB Mathematik/Informatik
Universität Osnabrück
Albrechtstr. 28a
49076 Osnabrück

Prof. Dr. Götz Pfander
International University Bremen
School of Engineering and Science
Postfach 750561
28725 Bremen

Dr. Friedrich Philipp
Fachbereich Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Prof. Dr. Holger Rauhut
Hausdorff Center for Mathematics
Institute for Numerical Simulation
Endenicher Allee 60
53115 Bonn

Prof. Dr. Justin Romberg
School of Electrical and
Computer Engineering
Georgia Institute of Technology
777 Atlantic Drive NW
Atlanta , GA 30332-0250
USA

Dr. Jose Luis Romero
Fakultät für Mathematik
Universität Wien
Nordbergstr. 15
A-1090 Wien

Prof. Dr. Naoki Saito
Department of Mathematics
University of California, Davis
1, Shields Avenue
Davis , CA 95616-8633
USA

Dr. Karin Schnass
Johann Radon Institute
Austrian Academy of Sciences
Altenberger Straße 69
A-4040 Linz

Prof. Dr. Zuowei Shen
Department of Mathematics
National University of Singapore
2 Science Drive
Singapore 117 543
SINGAPORE

Carlos Sing-Long
Institute for Computational and
Mathematical Engineering
Stanford University
475 Via Ortega
Stanford , CA 94305-4042
USA

Prof. Dr. Gabriele Steidl
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Prof. Dr. Joachim Stöckler
Institut für Angewandte Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund

Zeljka Stojanac
Hausdorff Center for Mathematics
Institute for Numerical Simulation
Endenicher Allee 60
53115 Bonn

Prof. Dr. Thomas Strohmer
Department of Mathematics
University of California, Davis
1, Shields Avenue
Davis , CA 95616-8633
USA

Prof. Dr. Arthur Szlam
New York University
Courant Institute of Math. Sciences
251 Mercer Street
New York , NY 10012
USA

Prof. Dr. Gerd Teschke
Hochschule Neubrandenburg
University of Applied Sciences
Brodaer Straße 2
17033 Neubrandenburg

Prof. Dr. Joel A. Tropp
Annenberg Center, CMS
MC 305-16
California Institute of Technology
1200 E. California Blvd.
Pasadena , CA 91125-5000
USA

Dr. Tino Ullrich
Hausdorff Center for Mathematics
Institute for Numerical Simulation
Endenicher Allee 60
53115 Bonn

Sergey Voronin
Department of Mathematics
Princeton University
609 Fine Hall
Washington Road
Princeton , NJ 08544
USA

Dr. Jan Vybiral
Fachbereich Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin

Prof. Rachel Ward
University of Texas at Austin
Department of Mathematics
2515 Speedway
Austin , TX 78712
USA

Prof. Dr. Özgür Yilmaz
Department of Mathematics
University of British Columbia
121-1984 Mathematics Road
Vancouver BC V6T 1Z2
CANADA