Sparse Representations, Numerical Linear Algebra, and Optimization

Gitta Kutyniok (Technische Universität Berlin),
Michael Saunders (Stanford University),
Stephen Wright (University of Madison-Wisconsin),
Özgur Yilmaz (University of British Columbia)

October 5–10, 2014

1 Overview of the Field

Sparse representations of functions and data have become crucial for a wide range of applications in technology and science, such as data analysis, imaging, and the construction of efficient solvers for ODEs and PDEs. With many important recent developments in frame theory and compressed sensing, the understanding of the underlying mathematics has grown substantially, and connections to other mathematical fields have started to evolve. Examples include the use of compressed sensing techniques for model reduction in numerical linear algebra, novel optimization algorithms for compressed sensing and image processing problems, and the design of regularization functions that yield the desired type of sparsity in the solutions of reconstruction problems.

The key idea of sparsity is that functions and signals arising naturally in many contexts can be described using only a small number of significant terms in a suitable basis or frame. This observation lies at the heart of many lossy compression techniques such as JPEG or MP3. Interestingly, sparsity is useful not only for compression purposes, but also in the design of sensing and sampling techniques for capturing signals. Under suitable conditions, sparse high-dimensional signals can be recovered efficiently from what would previously have been considered a highly incomplete set of measurements. This discovery is the basis of the field of compressed sensing. It has led to a paradigm change in information theory and has caused great excitement across many areas of science and engineering.

One key to the success of compressed sensing in many application areas is the generation of data-dependent sparsifying systems, customarily termed "dictionary learning." Because of the highly complex nature of this problem, the mathematical theory is in the early stages of development, and fundamental new ideas are still needed. A key property that ensures the occurrence of sparse expansions is the *redundancy* of these systems—i.e., these are complete systems that are *linearly dependent*. This is the focus of study in frame theory, which examines various aspects of redundancy as a mathematical concept. Frame theory has already impacted the area of applied harmonic analysis and sparse approximation; yet we are only beginning to grasp a fundamental understanding of redundancy measures.

Numerical linear algebra provides the mathematical foundation for stable computational algorithms based on matrix operations, with a myriad of applications ranging from signal and image processing, optimization, through data mining and model reduction, to ODE and PDE solvers. Indeed, it can be said that numerical linear algebra is at the foundation of computational science. Not only is efficient linear algebra important in algorithms for solving compressed sensing and frame analysis problems, but the fields interact in another way: Compressed sensing principles can be used to reduce extreme size linear systems that result from real-world applications, to the point where they become tractable for classical linear algebra algorithms. The first studies in linear algebra model-reduction using compressed sensing have just appeared. There is a world of new possibilities to be explored in this area.

Efficient algorithms for solving the optimization formulations of sparse approximation problems have been an area of intense research during the past five years. Indeed, Candés and Tao's foundational discovery in compressed sensing was that a seemingly intractable formulation of the sparse reconstruction problem could under certain assumptions be replaced by an equivalent convex (and highly structured) optimization formulation. This observation, which lent theoretical support to the use of ℓ_1 and total-variation functions as devices for inducing desired structure in the solutions of optimization problems, opened the door to a burst of work on optimization algorithms that exploit the specific structure of the sparse reconstruction problems. Cross-fertilization with other areas, particularly computational statistics and machine learning, has led to an exciting body of interdisciplinary work that has great potential for continued growth and development.

2 Recent Developments and Open Problems

The subject of the workshop is rife with interesting open problems, the scope of which only seems to grow as the area is developed. We mention several themes along these lines that were addressed in the workshop.

- As the size and complexity of data sets increases, the demands on algorithms continue to grow. There is a continuing need for algorithmic innovations to tackle enormous problems efficiently; we cannot rely solely in improvements in computational hardware. This issue arose in many guises during the meeting: Herrmann's massive geophysical data sets; Balzano's real-time processing of structure-from-motion data; Willett's use of online learning techniques; and the use of randomized algorithms, hierarchical matrix representations, and new perspectives on preconditioning in linear algebra problems of extreme scale.

- The theoretical and algorithmic breakthroughs in sparse recovery problems during the past decade have been astonishing, and several important practical applications have been identified. But much remains to be done in making sparse recovery truly practical across the full range of potential applications. Issues that need to be addressed include weakening of the too-strong assumptions on incoherence and restricted isometry that underlie the effectiveness of convex relaxations, the fact that observations are quantized, the need to calibrate uncertain sensing matrices as part of the recovery process, the rapid degradation of recovery algorithms as noise levels and observational coherence increase, and the extremely large size and incompleteness of some data sets. Algorithms and formulations need to be made more robust and self-calibrating while maintaining their efficiency across broader regimes. Devices such as dimension reduction and nonconvex formulations must be employed where relevant.

- Phase retrieval—the recovery of phase information about a complex signal from measurements of its magnitude—was proposed 25 years ago as one of the original grand challenge problems in computational science. It is a central problem in crystallography, and thus in structural biology. The obvious optimization formulations of the problem are notoriously nonconvex. But we have seen this before: sparse recovery problems are intractable in their obvious formulations, but yield to powerful convex relaxations in certain regimes. Recent works have explored connections between sparse reconstruction techniques and phase retrieval, and have shown that it is possible to find global minima of nonconvex formulations of the latter problem in some circumstances. It is likely that much more can be said by bringing the full range of experience over the past decade with sparse reconstruction to bear on phase retrieval. Further successes along these lines may have an enormous impact.

- A great deal has been accomplished with convex formulations, which can be solved efficiently with optimization algorithms. But we may be reaching the limits of this methodology: It is increasingly apparent that nonconvex formulations are more powerful in many settings, and more broadly applicable. (Even in the most elementary formulations of variables selection in statisical regression, nonconvex penalties give better results that the popular ℓ_1 penalty over a wide range of problems.) It has been noted, however, that nonconvexity does not necessarily connote intractability. Some nonconvex optimization formulations have global minimizers that are easily found by standard algorithms. In others, clever presolving techniques or algorithmic innovations yield either the global minimizer or provably accurate approximations to it. We lack a great deal of understanding about this phenomenon of seemingly "tractable" nonconvex formulations. Is there some underlying convex structure, perhaps in a higher-dimensional embedding? If so, how can we recognize this structure, characterize the problems that have this structure, and exploit it in algorithms? Several talks at the meeting touched on this theme, including Strohmer's keynote, and Chandrasekaran's relaxation hierarchy for signomial polynomials.

- An enduring mystery in the current burst of activity in "big data" is the unreasonable effectiveness of deep learning (multilayer neural networks) in speech, vision, and many other applications of machine learning. At least one speaker at the meeting (Mixon) discussed an interpretation of deep learning in terms of scattering transforms. It is quite possible that frame theory can provide valuable insight into this topic, which is one of the hottest topics in data science today.

3 Presentation Highlights

The talks consisted of four longer keynote talks, with each presenting an introduction to one of four research areas: compressed sensing, frame theory, numerical linear algebra, and optimization. Twenty shorter talks presented a bouquet of topics, most of which were related to employing sparsity as prior information.

Below, we discuss the presentations according to their specific focus, each time starting with the keynote talk. Most presentations covered at least two of our four targeted areas.

3.1 Compressed Sensing

Thomas Strohmer (**Keynote Talk**) presented an introduction to compressed sensing and pointed out probable future directions for this research area. In the first part, he discussed how compressed sensing provides a methodology to solve the problem of reconstructing a sparse signal exactly from an underdetermined system of linear equations in a computationally efficient manner via convex optimization. He presented the key ingredients of compressed sensing and extensions such as matrix completion, incoherence, and various notions of sparsity. In the second part, which focussed on current challenges and possible opportunities, he discussed the chasm between discrete and continuous models; self-calibrating compressed sensing and uncertainty mitigation; and fast and provable nonconvex algorithms related to sparsity.

Laura Balzano considered low-dimensional linear subspace approximations to high-dimensional data, which have applications when missing data is inevitable because of difficulties in collecting it. She described recent results on estimating subspace projections from incomplete data, including convergence guarantees and performance of the GROUSE algorithm (Grassmannian Rank-One Update Subspace Estimation) [1]. She explained the relationship between GROUSE and an incremental SVD algorithm, and presented results for GROUSE on problems in computer vision.

Felix Herrmann focussed on the area of exploration seismology and its formulation as a PDE-constrained inverse problem. The main challenges are extremely large data sets ($\approx 10^{15}$ measurements), large sets of unknowns ($\approx 10^9$ variables), and the hyperbolic PDEs themselves. He presented two competing approaches developed by his group, respectively based on optimization and compressed sensing [18, 13]. Both approaches utilize randomized dimension-reduction techniques to reduce the number of PDE solves, but they differ in how they control the errors.

Rayan Saab focussed on compressed sensing and frame theory. He considered the question of reconstruction from quantized compressed sensing measurements—a natural question because in the digital era, an acquisition process is typically followed by quantization. He first discussed a novel approach to handle quantization of frame coefficients by a post-processing step consisting of a discrete random Johnson-Lindenstrauss embedding of the integrated bit-stream. Near-optimal approximation accuracy as a function of the number of bits used can be shown for this method, and it holds for a large class of frames including smooth frames and random frames [11]. He then showed that the same encoding scheme applied to quantized compressed sensing measurements yields near-optimal approximation accuracy as a function of the bit-rate. A different reconstruction scheme is needed but it still uses convex optimization.

Jared Tanner focussed on algorithmic aspects of compressed sensing and matrix completion. He presented a novel algorithm that balances low per iteration complexity with fast asymptotic convergence. He proved numerically that this approach has faster recovery time than any other known algorithm, both for small-scale problems and for massively parallel GPU implementations. His method, named *conjugate gradient iterative hard thresholding* [2], is based on the classical nonlinear conjugate gradient algorithm.

Vladimir Temlyakov discussed greedy sparse approximation in Banach spaces with respect to redundant dictionaries [9]. He considered the design and analysis of such greedy methods, including Lebesgue-type inequalities to measure efficiency. He introduced a generalization of the RIP concept in a Hilbert space to dictionaries in a Banach space, and analyzed the Weak Chebyshev Greedy Algorithm—a generalization of the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit).

Rachel Ward focussed on matrix completion. Former results in this area typically required that the underlying matrix satisfy a restrictive structural constraint on its row and column spaces, and the subset of elements is then sampled uniformly at random. She showed that any $n \times n$ matrix of rank r can be exactly recovered from as few as $O(nr \log^2 n)$ randomly chosen elements, provided the random choice follows a specific biased distribution based on leverage scores of the underlying matrix. She proved that this specific form of sampling is in a certain sense nearly necessary, and presented three ways the results can be used when the leverage scores are not known in advance [8].

Rebecca Willett studied the problem of tracking dynamic point processes on networks, which is not a classical compressed sensing problem but bears some resemblance. Cascading chains of interactions are a salient feature of many real-world social, biological, and financial networks. Typically, only individual events associated with network nodes can be observed, usually without knowledge of the underlying dynamic network structure. Rebecca asked the question of tracking how such events within networks influence future events. She uses techniques from online learning frameworks and a multivariate Hawkes model to encapsulate autoregressive features of observed events within a network. With no prior knowledge of the network, her method performs almost as well as would be possible with complete knowledge [10].

3.2 Frame Theory

Dustin Mixon (**Keynote Talk**) presented an introduction to frame theory— the study of overcomplete yet stable expansions. One focus is to design frames for diverse applications, each with its own evaluation criteria (tightness, symmetry, incoherence, fast transforms, or signal model representation). Applications include time-frequency analysis, compression, and machine learning.

Pete Casazza discussed his work on the problem of phase retrieval: recovery of a vector from the absolute values of its inner products against a family of measurement vectors. The problem has been well studied in mathematics and engineering. A generalization exists in engineering: recovery of a vector from measurements consisting of norms of its orthogonal projections onto a family of subspaces. Pete provided several characterizations of subspaces that yield injective measurements—hence phase retrieval is potentially possible—and through a concrete construction proved that phase retrieval in this general case can be achieved with $2M-1$ projections of arbitrary rank [3]. He also raised and provided (partial) answers to questions on the minimal number of subspaces to have injectivity and how closely this problem compares to the usual phase retrieval problem with families of measurement vectors.

3.3 Numerical Linear Algebra

Daniel Kressner (**Keynote Talk**) presented a survey about the role of sparsity and low rank in numerical linear algebra. He pointed out that one of the key drivers of developments in numerical linear algebra has traditionally been the need for solving large-scale linear systems and eigenvalue problems. These arise naturally from finite difference or finite element discretizations of partial difference equations. The matrices involved are typically sparse, suggesting Krylov subspace methods and sparse direct solvers. Other types of data-sparsity have also gained importance, such as the low-rank structure of hierarchical matrices and hierarchically semi-separable matrices. Exploiting approximate sparsity in the desired solution is another direction of research that links numerical linear algebra with other disciplines, especially optimization.

Felix Krahmer presented a randomized algorithm for approximating matrix-vector multiplication, with the goal of computing dictionary representations. The algorithm makes heavy use of approximate spherical designs, and its proof of performance is based on Johnson–Lindenstrauss projections.

Dominique Orban spoke on linear algebra for matrix-free optimization. Sequential quadratic programming, augmented Lagrangian, and interior-point methods all need to solve symmetric saddle-point linear systems, which become symmetric quasi-definite with the help of regularizaiton. He discussed different forms of the equations for computing search directions in optimization, and iterative methods that exploit their structure.

Fred Roosta-Khorasani focussed on Monte-Carlo methods for the estimation of the trace of an implicit matrix A (one that is known only via matrix-vector products). In many applications, A is symmetric positive semi-definite, and the trace is estimated by averaging quadratic forms of A with random vector realizations from a suitable probability distribution. He focussed on the Gaussian distribution and derived bounds on the number of matrix-vector products required to guarantee a probabilistic bound on the relative error, revealing direct connections between the performance of the Gaussian estimator, the rank of the matrix, and its stable-rank [16].

Martin Stoll studies problems from PDE-constrained optimization in which controls with a certain sparsity are preferred. In his approach, in addition to the classical sparsity term a directional sparsity term is added. His talk focussed on the development of preconditioners for saddle-point systems and their use within Krylov subspace solvers. Typically this requires robust approximations to the relevant Schur complement.

Zdeněk Strakoš and Jörg Liesen gave a sequence of two talks on the general topic of sparsity, local and global information in numerical solution of PDEs. They first noted that in the finite element method (FEM) for discretizing partial differential equations (PDEs), the finite-dimensional piecewise polynomial approximation subspaces are generated using locally supported basis functions that typically vanish on all but a small number of elements determining the decomposition of the domain. This leads to sparsity in the matrix representation of the discretized operator. When iterative methods (such as Krylov subspace methods) are applied to solve the resulting algebraic problem, sparsity of the discretized matrix does not automatically mean an advantage, and the matter should be considered within the context of the original infinite-dimensional mathematical model. Normally, a transformation (preconditioning) of the discretized problem is needed to assure fast convergence. Preconditioners incorporating coarse space information (such as multilevel preconditioners or domain decomposition techniques) are often efficient because they handle naturally the global exchange of information between various parts of the domain. The key point of this talk was the interplay between the local discretization and global algebraic computation. The speakers interpret algebraic preconditioning as transformation of the discretization basis and simultaneous change of the inner product in the associated function space. Moreover, they compare the distribution of the algebraic and discretization errors over the domain, and interpret the algebraic error as a possible change of the discretization basis [12, 14].

Jean-Philippe Vert spoke on the introduction of new matrix norms for structured matrix estimation. Non-smooth convex penalties are currently state-of-the-art for estimating sparse models through convex optimization procedures such as the lasso. He showed that a norm introduced by Xiao, Zhou, and Wu (2010) is an atomic norm that is optimal in a certain sense for estimating matrices with orthogonal columns. He then showed that the (k, q)-trace norm, a new convex penalty for estimating low-rank matrices with sparse factors, outperforms the L_1 norm and other norms, especially for sparse principal component analysis [15].

3.4 Optimization

Michael Friedlander (**Keynote Talk**) spoke about algorithms for sparse optimization. He described the fundamental building blocks and surveyed some of the main research directions in this area, noting that convex optimization plays a key role.

Sasha Aravkin presented a general variational framework that allows efficient algorithms for denoising problems, where a functional is minimized subject to a constraint on fitting the observed data. The framework applies to vector recovery (sparse optimization) and matrix recovery (matrix completion and robust PCA). He discussed performance aspects and noted that the efficiency of his approach relies on efficient first-order solvers. For matrix recovery, novel ideas in factorized and accelerated first-order methods were incorporated into the variational framework.

Coralia Cartis was concerned with line-search methods—an important class of algorithms for unconstrained nonconvex optimisation that rely on approximately computing a local descent direction and a step along this direction in such a way that sufficient decrease is achieved. To ensure that sufficient decrease is possible, the direction must satisfy certain requirements. In large-scale applications, meeting the requirements may be prohibitively expensive. Coralia presented global convergence rates for a line-search method based on random models and directions whose quality is ensured only with certain probability. In the convex and strongly convex case, she could improve those results. She finished by presenting a probabilistic cubic regularisation variant that allows approximate probabilistic second-order models and showed improved complexity bounds compared to probabilistic first-order methods.

Venkat Chandrasekaran focussed on signomial programs (SPs): optimization problems consisting of an objective and constraints specified by signomials (sums of exponentials of linear functionals of a decision variable). Such programs are non-convex optimization problems in general, but some instances of NP-hard problems can be reduced to SPs. He described a hierarchy of convex relaxations that provide successively tighter lower bounds on the optimal value in SPs. The approach relies on the observation that the relative entropy function provides a convex parametrization of certain sets of globally nonnegative signomials with efficiently computable nonnegativity certificates. The sequence of lower bounds converges to the global optimum for broad classes of SPs [6].

Mark Schmidt proposed the stochastic average gradient (SAG) method for optimizing the sum of a finite number of smooth convex functions. As for stochastic gradient methods, the iteration cost is independent of the number of terms in the sum. By incorporating a memory of previous gradient values, he improves the convergence rate of SAG from $O(1/k)$ to a linear convergence rate of the form $O(p^k)$ for some $p < 1$. Some good practical properties are achieved. The method supports regularization and sparse datasets, it allows an adaptive step-size and has a termination criterion, it allows mini-batches, and its performance can be improved by non-uniform sampling [17].

Ewout van den Berg presented a hybrid quasi-Newton projected-gradient method for the optimization of convex functions over a polyhedral set. He described applications such as the lasso, bound-constrained optimization, and optimization over the simplex.

4 Scientific Progress Made and Outcome of the Meeting

The meeting proved very successful. Participants from the four different communities interacted closely as planned. Many participants mentioned to us that they greatly enjoyed the excellent talks, the fruitful discussions, the inspiring atmosphere at BIRS, and the tireless support of the BIRS staff. Several asked us if another meeting of this type could be held, with invited representatives from the same four research areas.

Let us summarize the impact of our meeting.

- *Initiation of communication between the four research areas*

  A main goal of our workshop was to invite people from the four research areas of compressed sensing, frame theory, numerical linear algebra, and optimization, with sparse representations being the common factor. This BIRS workshop was a unique opportunity to initiate a fertile discussion among those representatives. The four keynote talks indeed provided excellent introductions to the respective research areas. These and the shorter talks led to many vivid debates, showing that all groups highly benefitted from being exposed to different methodologies and ideas. We can report that these discussions led to several new collaborations within the four groups.

- *Intensification of new directions in the field*

  Various new directions both theoretical and applied were presented during talks and vividly discussed afterwards. One main new direction is the utilization of randomized algorithms for numerical linear algebra problems such as matrix-vector multiplication, but related to sparsity. This direction requires a close interaction among essentially all four research areas. It is currently in its beginning stage, and far more developments can be expected in the near future. As further directions, also at the beginning stage, we especially mention a general framework for PDE solvers that balances discretization aspects and numerical linear algebra considerations such as preconditioning (Zdeněk Strakoš and Jörg Liesen). The workshop was a unique opportunity to discuss the most recent results and stimulate these directions.

- *Discussion of interactions across research areas*

  Since a main goal of the workshop was to bring together the aforementioned key areas required for sparsity methodologies in data sciences, we put a particular focus on interactions across research areas. The four keynote talks were one main approach to implement improved each participant's understanding of the other three areas. Another approach was two discussion sessions, one of which was about notational issues. It turned out that it is indeed a major obstacle for many researchers to read publications from other areas. We then assembled a collection of notational problems in the sense of notations used in the other areas with a very different meaning. This made everybody aware of the fact that notation has to be carefully thought through, especially when aiming to have a readership in a different research area.

- *Discussion of methodologies*

  Several talks made the audience aware of methods being present in other areas that might be useful in their own work as well. One example was the range of optimization approaches for sparse recovery presented by members of the optimization community such as Michael Friedlander and Ewout von den Berg, which will be highly beneficial for participants working in compressed sensing. The talk by Thomas Strohmer discussed various obstacles that are still present in compressed sensing, such as self-calibrating compressed sensing, which, as he pointed out, can only be overcome by methodologies from other areas. Also, many problems in frame theory can be regarded as problems in numerical linear algebra, yet those two communities almost never interact. The workshop permitted open discussion of methods that could assist other communities, such as the diverse preconditioners described by Daniel Kressner.

- *Introduction of young scientists*

  Several of our participants were young and very promising scientists, such as Ewout von den Berg, Martin Stoll, and even Dustin Mixon. The workshop gave them an exceptional chance to present themselves and get in contact with the leading researchers not only in their own field but also in the three other research areas, and also to broaden their horizon. After the workshop, we received excited feedbacks from this group, voicing the general opinion that this was a unique opportunity.

- *Manifestation of the future direction of the field*

  Since this workshop brought together the main leaders in the four research areas of compressed sensing, frame theory, numerical linear algebra, and optimization with sparse representations, with particular focus on the common element of sparse representations, it presented the chance to debate and perceive the future directions of this field. Intense discussions took place right after most talks, as well as during the two scheduled general discussion sessions. Interesting open problems and possible future research directions came up. We therefore expect this workshop to have a signal effect that will significantly influence the future research in this field, in particular by bringing the four research areas more closely together.

  As particular topics that came up in the discussion session we mention the following. Large-scale computations are still a major topic, as indicated in Felix Herrmann's talk. This led to a vibrant discussion of how methods can be developed to solve such problems in an efficient and robust (to noise) way. In particular, Felix met with several linear algebra people to derive an approach that will be greatly beneficial for his truly huge problems. On the theoretical side, the question arose of to what extent Hilbert space methods should be extended to the Banach space setting, to allow a more general viewpoint and to include L^p spaces in a natural way.

References

[1] L. Balzano, B. Recht, and R. Nowak, High-dimensional matched subspace detection when data are missing, *Proceedings of IEEE International Symposium on Information Theory (ISIT)* (2010), 1638–1642.

[2] J. Blanchard, J. Tanner, and K. Wei, Conjugate Gradient Iterative Hard Thresholding: observed noise stability for compressed sensing, *IEEE Transactions on Signal Processing*. Accepted (2014).

[3] J. Cahill, P.G. Casazza, J. Peterson, and L. Woodland, Phase retrieval by projections, *arXiv preprint*, arXiv:1305.6226 [math.FA]

[4] E. Candès and B. Recht, Exact matrix completion via convex optimization, *Foundations of Computational Mathematics* **9** (2009), 717–772.

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

[6] V. Chandrasekaran and P. Shah, Relative entropy relaxations for signomial optimization, *arXiv preprint*, arXiv:1409.7640 [math.OC]

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* **20** (1998), 33–61.

[8] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, Completing any low-rank matrix, provably, *arXiv preprint*, arXiv:1306.2979 [stat.ML]

[9] V. N. Temlyakov, Sparse approximation and recovery algorithms in Banach spaces, *Forum of Mathematics, Sigma*, **2** (2014), e12 doi:10.1017/fms.2014.7

[10] E.C. Hall and R. M. Willet, Tracking dynamic point processes on networks, *arXiv preprint* arXiv:1409.0031 [stat.ML]

[11] M. Iwen and R. Saab, Near-optimal encoding of finite frame expansions, *Journal of Fourier Analysis and Applications* **19** (2013), 1255–1273.

[12] J. Màlek and Z. Strakoš, Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs. *SIAM Spotlights Series*, SIAM, Philadelphia, 2014.

[13] H. Mansour, F. J. Herrmann, and Ö. Yılmaz, Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization, *Geophysics* **78** (2013), V193–V206.

[14] J. Papež, J. Liesen, and Z. Strakoš, Distribution of the discretization and algebraic error in numerical solution of partial differential equations, *Linear Algebra and Its Applications* **449** (2014), 89–114.

[15] E. Richard, G. Obozinski, and Jean-Philippe Vert, Tight convex relaxations for sparse matrix factorization, *arXiv preprint*, arXiv:1407.5158 [stat.ML]

[16] F. Roosta-Khorasani and U. Ascher, Improved bounds on sample size for implicit matrix trace estimators, *arXiv preprint*, arXiv:1308.2475 [cs.NA]

[17] M. Schmidt, N. Le Roux, and F. Bach, Minimizing finite sums with the stochastic average gradient, *arXiv preprint*, arXiv:1309.2388 [math.OC]

[18] T. van Leeuwen and F. J. Herrmann, 3D frequency-domain seismic inversion with controlled sloppiness , *SIAM Journal on Scientific Computing* **36** (2014), S192–S217.

[19] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Transactions on Signal Processing* **57** (2009), 2479–2493.