Mathematisches Forschungsinstitut Oberwolfach

Report No. 18/2014

DOI: 10.4171/OWR/2014/18

Mini-Workshop: Mathematical Physics meets Sparse Recovery

Organised by
David Gross, Freiburg
Felix Krahmer, Göttingen
Rachel Ward, Austin
Andreas Winter, Bellaterra

13 April – 19 April 2014

Abstract. In recent years, there have been several fruitful interchanges of methods between the fields of sparse and low-rank recovery on the one hand and quantum information theory on the other hand. One way to understand this seemingly surprising coincidence is that the analysis of vector- and matrix-valued randomized constructions plays an important role in both fields. An example is the realization that certain matrix-valued large deviation bounds can be employed to substantially simplify and generalize the analysis of low-rank matrix recovery schemes.

In this workshop, the participants worked to identify and collaborate on further mathematical problems that are being researched in parallel by the two communities. Topics that have been discussed include

- Tools for the analysis of vector- and matrix-valued randomized constructions and their application to phase retrieval problems
- Conversely, tools for *de*-randomizing such protocols, based, e.g., on spherical designs.
- Uncertainty relations, e.g., for the task of lower-bounding the number of measurements required for signal identification.
- Time-frequency methods (known as phase-space methods in physics).
- Matrix- and tensor norms: computational tools, complexity, relaxations and their application to tensor recovery.

Mathematics Subject Classification (2010): 15xx, 15A83, 15A42, 47A52, 46N50.

Introduction by the Organisers

The mini-workshop *Mathematical Physics meets Sparse Recovery*, organised by David Gross (Freiburg), Felix Krahmer (Göttingen), Rachel Ward (Austin), and Andreas Winter (Bellaterra) successfully initiated intensive communication between researchers in two different areas. Among the participants there were, on the one hand, nine researchers focusing on mathematical signal processing, in particular on sparse recovery and related questions, and, on the other hand, seven mathematical physicists who have contributed to quantum information theory. The connection between these fields first started to arise with the works of David Gross, one of the workshop organizers, who demonstrated that certain problems in quantum information theory are closely related to problems in low-rank matrix recovery [1]. So tools developed in mathematical physics can help solve mathematical signal processing problems. Following up on his results, a collaboration between David Gross, his PhD student Richard Küng, also a workshop participant, and Felix Krahmer, another organizer, started [4], strengthening the connection between the fields. This collaboration laid the foundation for the mini-workshop.

From their collaboration experience, they realized that one of the biggest obstacles to an intensified collaboration between the fields was the often quite different mathematical language and notation used in the two fields, it was decided that a substantial part of the workshop should be devoted to communicating to the respective other group the way of thinking and the terminology employed. Another substantial part was to be devoted to communicating the general research goals to find collaboration projects, and the third intended pillar was organized discussions in small groups.

The workshop started by two introductory talks of about 90 minutes each by participants representing the sparse recovery community. Dustin Mixon gave a general overview of mathematical signal processing problems on a very broad level and Holger Rauhut continued by providing a more technical introduction to the methods currently used. The goal of these two talks was to communicate the type of problems arising in the signal processing community to those participants from quantum information theory who were experts on relevant methods while not having applied them to mathematical signal processing before. Reciprocating, Andreas Winter gave a similar 90 minute talk introducing the problems and mathematical methods at the center of quantum information theory.

After these introductions, the workshop proceeded as follows: days were kicked off by a one hour talk requested the previous day. After that, the participants were divided into collections of small, mixed groups that started working on specific questions. Topics that received significant attention by those groups are as follows:

Tight large-deviation bounds for sums of random matrices. This theory was essential to set up the connection between the two communities. Originally developed by A. Winter to treat quantum information problems, large-deviation bounds for sums of random matrices were introduced into the theory of sparse and low-rank recovery by D. Gross and later greatly refined at the hands of J. Tropp. Joel Tropp presented the state of the art and the most pressing problems as he sees the theory. A concrete problem – removing a sometimes spurious logarithmic dimension factor in front of the generally exponentially small large deviation probability – created particular interest. While the problem was not solved during the workshop, collaborations have been initiated as a result.

A conjectured "matrix arithmetic-geometric mean inequality." Recently, a noncommutative extension of the well-known arithmetic-geometric mean inequality for vectors was proposed by B. Recht and C. Ré [2], and this conjecture was communicated to the workshop participants by Felix Krahmer and Rachel Ward. Such an inequality is of interest in machine learning and signal processing as it would give theoretical justification to the observed effect that *without-replacement* sampling schemes outperform *with-replacement* sampling schemes in randomized sequential optimization algorithms. Although this inequality has been verified in certain special cases, neither a general proof nor a counter-example has been found. While the conjecture was not resolved during the workshop, Marius Junge made initial discoveries towards finding a counter-example to the conjecture using free probability and recent techniques he had devised for related matrix problems.

The nascent theory of tensor recovery. After having treated sparse vectors and low-rank matrices, the signal analysis community has recently turned their attention to the theory of learning low-rank tensors from underdetermined measurements. Tensor problems likewise appear in quantum mechanics, where many-body wave functions are just elements in large tensor spaces. After extremely well-received talks by Y.-K. Liu about applications in natural language processing and by Z. Stojanac on recent ideas for tensor norm relaxations based on theta bodies, several discussions ensued. A mixed group tried to devise measurements that are incoherent w.r.t. all low-rank tensors – generalizing previous such constructions that the quantum community routinely uses for matrices. The prospect of proving results for completely symmetric tensors was discussed. V. Cevher introduced the physicists to dual smoothing techniques that might speed up numerical methods for computing tensor norms. There was an extensive discussion as to whether tests for quantum separability based on symmetric extensions could give rise to well-performing convex proxies for tensor rank.

Phase space and time-frequency methods. Inspired by an earlier talk by G. Pfander on time-frquency analysis and sparse recovery, D. Gross presented unpublished work on phase space support-rank uncertainty relations and their role in proving lower bounds to low-rank matrix recovery problems. A collaboration between Gross and Pfander was initiated with the goal to generalize these results from discrete to continuous Gabor systems.

Stability of PhaseLift. The ill-defined inverse problem of retrieving an unknown complex vector from "amplitude" measurements – i.e. linear measurements that are ignorant towards complex phases – has received considerable attention in the field of mathematical signal processing over the last few years. Moreover, problems of this type are also of considerable interest for doing quantum state tomography – an important subfield of quantum information theory [5]. During the course of the workshop, H. Rauhut and R. Küng started to tackle the important problem of proving stability guarantees for PhaseLift in the presence of noise for certain measurement setups – most notably random coded diffraction patterns [6]. Although the problem could not be fully solved during the workshop, partial results have been obtained and yet another ongoing collaboration (between H. Rauhut, R. Küng and D. Gross) arose. In this context, also a connection in the other direction arose, namely F. Krahmer and Y.-K. Liu started a collaboration with the goal to apply mathematical tools developed and applied in the context of phase retrieval problems to problems in quantum information.

Statistical trade-offs in low-rank recovery. It had been observed numerically [5] that the performance of low-rank based estimators for quantum state tomography displays the following behavior: Assume the total number of experimental samples taken is constant. There is the freedom to use these samples to either estimate a few linear functions of the unknown low-rank matrix to a high precision, or many distinct functions more coarsely. Maybe surprisingly, it turns out that the performance of the estimator is largely independent of that choice. This is of relevance to physical experiments – but the physicists failed to find a theoretical explanation for this behavior. M. Gutta suggested an approach based on asymptotic properties of the maximum likelihood estimator, while R. Saab proposed to use results about the behavior of right-inverses of near-isometric embeddings to attack the problem. Both ideas seem promising, and a collaboration between the aforementioned researchers, D. Gross and A. Winter to further pursue these questions has been initiated.

To conclude, this mini-workshop was by all accounts a great success. Through this meeting, certain mathematical language and notation barriers were overcome, and each community became aware of new problems and common interests among the other community, as well as new applications to problems and techniques already known. At the same time, several new collaborations formed between researchers in mathematical signal processing and quantum information theory, and concrete results to open problems have already been established. We expect a number of papers to come out of this meeting in the coming years.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, "US Junior Oberwolfach Fellows".

References

[1] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Transactions on Information Theory **57** (2011), 1548–1566.

[2] B. Recht, C. Ré, *Beneath the valley of the noncommutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences*, arXiv preprint (2012).

[3] M. Muster, *Computing other invariants of topological spaces of dimension three*, Topology **32** (1990), 120–140.

[4] D. Gross, F. Krahmer, R. Kueng, *A partial derandomization of PhaseLift using spherical designs*, to appear in JFAA (2014).

[5] S. Flammia, D. Gross, Y.K. Liu, J. Eisert *Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators*, New Journal of Physics, **14(9)**, 095022 (2012).

[6] D. Gross, F. Krahmer, R. Kueng, *Improved recovery guarantees for phase retrieval from coded diffraction patterns*, preprint arXiv:1402.6286 (2014).

Mini-Workshop: Mathematical Physics meets Sparse Recovery

Table of Contents

Abstracts

Compressed sensing: Variations on a theme

Dustin G. Mixon

Compressed sensing has been an exciting subject of research over the last decade, and the purpose of this talk was to provide a brief overview of the subject. First, we considered certain related topics (namely image compression and denoising) which led up to the rise of compressed sensing. In particular, wavelets provide a useful model for images, as natural images tend to be approximated by linear combinations of particularly few wavelets. This sparsity model has enabled JPEG2000 to provide particularly efficient image compression with negligible distortion. Additionally, this model has been leveraged to remove random noise from natural images [20].

Considering natural images enjoy such a useful model, one may ask whether the model can be leveraged to decrease the number of measurements necessary to completely determine an image. For example, an MRI scan might require up to 2 hours of exposure time, and then the image might be compressed with JPEG2000 after the fact, meaning most of the measurements can be effectively ignored. So is it possible to simply measure the important parts of the image and not waste time in the image acquisition process? This is the main idea underlying *compressed sensing*, as introduced by Candès, Romberg and Tao [16] and Donoho [19].

In compressed sensing, one is faced with finding the sparsest vector (i.e., the one with the fewest nonzero entries) which solves an underdetermined linear system. Solving this problem is generally NP-hard [26], but one may convexify the problem by instead finding the vector with the smallest ℓ_1 norm. Perhaps surprisingly, this solves the original problem for "most" underdetermined linear systems, as the only requirement is for the set of solutions to the linear system to be slanted properly relative to the ℓ_1 ball. Indeed, ℓ_1 minimization has been used for applications such as deconvolution [25, 29], regression [30], and sparse approximation [17]. However, compressed sensing is different from these prior applications of ℓ_1 minimization because one may choose the "sensing" matrix in the underdetermined linear system so as to guarantee ℓ_1 recovery of the sparsest vector.

There are many different types of guarantees for compressed sensing. For the scenario in which the sensing matrix is drawn randomly every time a sparse vector is to be measured, one may ask for ℓ_1 recovery with high probability, and this non-uniform guarantee is characterized by a phase transition which has been completely resolved in [3]. We may also ask for a uniform guarantee: that the sensing matrix is fixed and must allow for ℓ_1 recovery of all sufficiently sparse vectors simultaneously; this is characterized by the so-called *null space property*, as introduced in [18]. Finally, one may ask for a uniform guarantee with stability, meaning the presumed sparse vector may actually have small entries in place of zero entries, and also one may add a small amount of unknown noise to the measurements, and yet ℓ_1 minimization will produce a correspondingly close approximation of the true "nearly sparse" vector. One popular condition on the sensing matrix which implies such a guarantee is the *restricted isometry property*, and a version of this guarantee is proved in [9]. There are several known families of random matrices which satisfy the restricted isometry property with high probability [6, 24, 27, 28].

To date, there are a few open problems in compressed sensing. First, while there are several random constructions of matrices satisfying the restricted isometry property, explicit constructions are notoriously terrible by comparison [5, 8]. This problem of "finding hay in a haystack" is common in combinatorics; for example, there is currently no explicit n-vertex graph which contains neither a clique nor an independent set of size $2\log_2 n$, they such graphs are known to exist by the probabilistic method. Second, there are many applications in which the sensing matrix is plagued with coherent column vectors, thereby preventing sparse recovery of vectors which are supported on the corresponding entries. However, there might be other types of recovery guarantees which may be proved in this case; there has been success along these lines, for example, in the case where the columns form an oversampled discrete Fourier transform matrix [12]. Finally, compressed sensing primarily focuses on signals which are sparse in an orthonormal basis (such as wavelets), but for some applications, signals are instead sparse in an overcomplete dictionary, and the theory for this case is almost nonexistent at the moment [10].

To motivate the next part of my talk, I introduced the Netflix problem. In 2006, Netflix offered a US $1 million prize to improve its movie rating prediction algorithm. Here, the idea is that movie ratings can be organized in a matrix with rows indexed by movies and columns indexed by users. However, since most users have yet to view most movies, Netflix does not have access to most of the entries of this matrix. Fortunately, if we had all of the entries of this matrix, we can assume that principal component analysis would describe the columns of this matrix as essentially lying in a low-dimensional subspace. As such, one might attempt to fill in the blanks by assuming the desired matrix has low rank. This is the motivation behind *low-rank matrix completion*.

For this problem, one is inclined to find the matrix of minimal rank given the linear measurements available. This is strikingly similar to the problem of compressed sensing, except now we seek to minimize the number of nonzero singular values. As such, the natural relaxation to consider is minimizing the nuclear norm, i.e., the sum of the singular values. Indeed, when completing a low-rank matrix from randomly reported entries, this relaxed optimization is effective provided the low-rank matrix we intend to recover is not localized at any particular entry [21].

When the linear measurements of the matrix are not entries, but linear combinations of the entries, the problem goes by a different name: *low-rank matrix recovery*. One important application of this problem is *phase retrieval*. For this inverse problem, one seeks to recover an input vector from the entrywise absolute value of the output of a known linear operator. By squaring these absolute values, each can be identified as a linear combination of entries of the input vector's outer product [7]. As such, one can hope to recover this rank-1 outer product by low-rank matrix recovery [11]. To date, there are various guarantees for phase retrieval via low-rank matrix recovery [13, 14, 15, 22].

Along the lines of low-rank matrix recovery, there are a few open problems which remain. First, a prominent application of phase retrieval via low-rank matrix recovery is X-ray crystallography, in which the measurements are masked Fourier transforms. Each mask corresponds to an exposure of a small object that one would like to image, but the object is so small that each exposure contributes to its destruction; as such, one would like to image the object with as few exposures as possible. Taking n to denote the dimensionality of the image, it is shown in [14] that $\log^4 n$ exposures suffice to image the object. Recently, [22] showed that $\log^2 n$ exposures suffice. Both of these results provide a non-uniform guarantee, and they do not prove stability. On the other hand, [4] provides a uniform guarantee for recovery from $\log n$ exposures, but using a different recovery method (i.e., not convex optimization). For both recovery methods, [15] and [1] prove stability for more general measurement ensembles. However, it remains to guarantee stability for masked Fourier transforms. It would also be interesting to find guarantees for other important measurement ensembles, such as the short-time Fourier transform. Finally, low-rank matrix recovery is invariably solved using semidefinite programming, which is rather slow in general, and so it is desirable to find speedups for particular instances (such as the instances corresponding to phase retrieval).

The last part of my talk discussed future directions related to compressed sensing. First, given a convolution of a sparse function with a function of rapid decay, one can expect to determine the two functions, especially if the nonzero entries of the sparse function are sufficiently separated. This problem is called *blind deconvolution*. In general, if you are given a convolution of two functions, each belonging to a known signal class, then you might be able to recover both functions. Taking inspiration from phase retrieval, notice that each entry of the convolution can be expressed as a linear combination of the entries of the outer product of the two functions. As such, one can hope to recover this rank-1 outer product from the convolution using low-rank matrix completion [2]. A related problem is *calibration*, in which you wish to find the sparsest solution to an underdetermined linear system, but you only know the matrix up to a parameterized family; if the parameterization is linear, then each entry of the output vector is a linear combination of entries from the outer product of the parameter vector and the sparse vector. Recently, there has been a lot of success applying this sort of *bilinear compressed sensing* to deblur images from multiple blurred exposures [23]. This would be an interesting direction for the community to pursue.

References

[1] B. Alexeev, A. S. Bandeira, M. Fickus, D. G. Mixon, *Phase retrieval with polarization*, SIAM J. Imaging Sci. **7** (2014), 35–66.

[2] A. Ahmed, B. Recht, J. Romberg, *Blind deconvolution using convex programming*, IEEE Trans. Inform. Theory **60** (2014), 1711–1732.

[3] D. Amelunxen, M. Lotz, M. B. McCoy, J. A. Tropp, *Living on the edge: Phase transitions in convex programs with random data*, Available online: arXiv:1303.6672

[4] A. S. Bandeira, Y. Chen, D. G. Mixon, *Phase retrieval from power spectra of masked signals*, to appear in Inform. Inference.

[5] A. S. Bandeira, M. Fickus, D. G. Mixon, P. Wong, *The road to deterministic matrices with the restricted isometry property*, J. Fourier Anal. Appl. **19** (2013), 1123–1149.

[6] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx. **28** (2008), 253–263.

[7] R. Balan, B. G. Bodmann, P. G. Casazza, D. Edidin, *Painless reconstruction from magnitudes of frame coefficients*, J. Fourier Anal. Appl. **15** (2009), 488–501.

[8] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, D. Kutzarova, *Explicit constructions of RIP matrices and related problems*, Duke Math. J. **159** (2011) 145–185.

[9] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Acad. Sci. Paris, Ser. I **346** (2008), 589–592.

[10] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Harmon. Anal. **31** (2011), 59–73.

[11] E. J. Candès, Y. C. Eldar, T. Strohmer, V. Voroninski, *Phase retrieval via matrix completion*, SIAM J. Imaging Sci. **6** (2013), 199–225.

[12] E. J. Candès, C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Comm. Pure Appl. Math. **67** (2014), 906–956.

[13] E. J. Candès, X. Li, *Solving quadratic equations via PhaseLift when there are about as many equations as unknowns*, Found. Comput. Math. (2012), 1–10.

[14] E. J. Candès, X. Li, M. Soltanolkotabi, *Phase retrieval from coded diffraction patterns*, Available online: arXiv:1310.3240

[15] E. J. Candès, T. Strohmer, V. Voroninski, *PhaseLift: Exact and stable recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math. **66** (2013), 1241–1274.

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

[17] S. S. Chen, D. L. Donoho, M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **20** (1998), 33–61.

[18] A. Cohen, W. Dahmen, R. DeVore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc. **22** (2009), 211–231.

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

[20] D. L. Donoho, J. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), 425–455.

[21] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Inform. Theory **57** (2011), 1548–1566.

[22] D. Gross, F. Krahmer, R. Kueng, *Improved recovery guarantees for phase retrieval from coded diffraction patterns*, Available online: arXiv:1402.6286

[23] A. Ito, A. C. Sankaranarayanan, A. Veeraraghavan, R. G. Baraniuk, *BlurBurst: Removing blur due to camera shake using multiple images*, preprint.

[24] F. Krahmer, S. Mendelson, H. Rauhut, *Suprema of chaos processes and the restricted isometry property*, to appear in Comm. Pure Appl. Math.

[25] S. Levy, P. K. Fullagar, *Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution*, Geophys. **46** (1981), 1235–1243.

[26] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput. **24** (1995), 227–234.

[27] H. Rauhut, *Compressive sensing and structured random matrices*, Radon Series Comp. Appl. Math. **9** (2010), 1–92.

[28] M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), 1025–1045.

[29] H. L. Taylor, S. C. Banks, J. F. McCoy, *Deconvolution with the ℓ_1 norm*, Geophys. **44** (1979), 39–52.

[30] R. Tibshirani, *Regression skrinkage and selection via the lasso*, J. R. Statist. Soc. B **58** (1996), 267–288.

Compressive Sensing and Low Rank Matrix Recovery

Holger Rauhut

Compressive sensing and low rank matrix recovery (matrix completion) aim at recovering objects (usually signals) of small complexity within a high-dimensional ambient space from a small number of linear measurements [12, 8, 9, 5]. In standard compressive sensing, small complexity is modeled via sparsity, i.e., it is assumed that the vectors under consideration have a sparse representation in terms of a suitable basis, or at least are well-approximated by such an expansion. An extension of compressive sensing assumes that a matrix to be recovered is of low rank [10, 3, 13]. These fields have gained substantial attention in recent years with a lot of activity in applied mathematics, physics, electrical engineering and computer science. On the one hand, this theory leads to fundamentally new approaches for certain signal processing applications and on the other hand, the involved mathematics is highly nontrivial and triggered many activities in areas such as convex optimization, harmonic analysis and random matrix theory.

In mathematical terms, given measurements

$$y = Ax$$

of a vector $x \in \mathbb{R}^N$ (or a matrix) we would like to reconstruct x in the underdetermined case that $A \in \mathbb{R}^{m\times N}$ with $m \ll N$. Obviously without further information, reconstruction is impossible, but the small complexity assumption (sparsity or low rank) may help out. While the naive approaches of ℓ_0-minimization and rank minimization are NP-hard in general [12], tractable alternatives including ℓ_1-minimization and nuclear norm minimization have been introduced. Usually, random measurement matrices A are considered in this context. In fact, a typical result in compressive sensing [6, 12, 18] states that

$$m \geq Cs\ln(N/s)$$

Gaussian random measurements are sufficient for stable recovery of s-sparse vectors via ℓ_1-minimization with high probability. Similarly,

$$m \geq Cr(n_1 + n_2)$$

Gaussian random measurements are sufficient for recovery of $n_1 \times n_2$ matrices of rank r via nuclear norm minimization with high probability [2, 10].

In practice, structure is often required in the measurement process, which leads to the study of structured random matrices in this context [23, 17], [12, Chapter 12]. Similarly guarantees as just outlined hold, for instance, for random partial Fourier matrices and extensions [5, 6, 30, 21, 23, 27, 28], partial random circulant matrices (modeling subsampled random convolutions) [22, 29, 24, 16], time-frequency structured random matrices [19, 20, 16] and scattering matrices arising in radar [15].

Recently, an interesting application of low rank matrix recovery in phase retrieval problem has triggered a high research activity [1, 4, 14]. In fact, the phase retrieval problem can be recast as recovering a rank one matrix.

Extensions to recovery of low rank tensor are currently also under investigations, but it is much harder than in the matrix case to derive theoretical recovery guarantees. Preliminary results can be found in [26, 25] and in the references therein.

The reader is referred to the overview papers [7, 11, 23] and books [9, 12] for further information on compressive sensing and low rank recovery.

References

[1] E. Candès, X. Li, M. Soltanolkotabi. *Phase retrieval from masked Fourier transforms.* Preprint (2013).

[2] E. J. Candès, Y. Plan, *Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements*, IEEE Trans. Inform. Theory **57(4)** (2011), 2342–2359.

[3] E. J Candès, B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math. **9** (2009), 717-772.

[4] E. Candès, T. Strohmer, and V. Voroninski. *PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming.* Comm. Pure Appl. Math. (to appear).

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

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

[7] E. J. Candès, M. Wakin. *An introduction to compressive sampling.* IEEE Signal Processing Magazine **25(2)** (2008), 21–30.

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

[9] Y. C. Eldar, G. Kutyniok (eds.), *Compressed Sensing - Theory and Applications*, Cambridge Univ. Press (2012).

[10] B. Rechtm, M. Fazel, P. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev. **52(3)** (2010), 471-501.

[11] M. Fornasier, H. Rauhut. *Compressive sensing.* In: Handbook of Mathematical Methods in Imaging (O. Scherzer, ed.), Springer (2011), 187–228.

[12] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser (2013).

[13] D. Gross. *Recovering low-rank matrices from few coefficients in any basis.* IEEE Trans. Inform. Theory **57(3)** (2011), 1548–1566.

[14] D. Gross, F. Krahmer, and R. Kueng. *A partial derandomization of PhaseLift using spherical designs.* J. Fourier Anal. Appl. (to appear).

[15] M. Hügel, H. Rauhut, and T. Strohmer. *Remote sensing via l1-minimization.* Found. Comput. Math. **14** (2014), 115–150.

[16] F. Krahmer, S. Mendelson, H. Rauhut. *Suprema of chaos processes and the restricted isometry property.* Comm. Pure Appl. Math. (to appear).

[17] F. Krahmer and H. Rauhut. *Structured random measurements in signal processing.* Preprint (2014).

[18] S. Mendelson, A. Pajor, and N. Tomczak Jaegermann. *Uniform uncertainty principle for Bernoulli and subgaussian ensembles.* Constr. Approx. **28(3)** (2009), 277–289.

[19] H. Rauhut, G. E. Pfander. *Sparsity in time-frequency representations.* J. Fourier Anal. Appl. **16(2)** (2010), 233–260.

[20] G. E. Pfander, H. Rauhut, J.A. Tropp. *The restricted isometry property for time-frequency structured random matrices.* Probab. Theory Relat. Fields **156** (2013), 707–737.

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

[22] H. Rauhut. *Circulant and Toeplitz matrices in compressed sensing*, Proc. SPARS'09 (2009).

[23] H. Rauhut, Holger. *Compressive sensing and structured random matrices* In: Theoretical foundations and numerical methods for sparse recovery (M. Fornasier, ed.), Radon Series Comp. Appl. Math. **9** (2010), deGruyter, 1–92.

[24] H. Rauhut, J. K. Romberg, J. A. Tropp. *Restricted isometries for partial random circulant matrices.* Appl. Comput. Harmon. Anal. **32(2)** (2012), 242–254.

[25] H. Rauhut, R. Schneider, Ž. Stojanac. *Tensor completion in hierarchical tensor formats.* Preprint (2014).

[26] H. Rauhut, R. Schneider, Ž. Stojanac. *Low rank tensor recovery via iterative hard thresholding.* Proc. SampTA2013 (2013).

[27] H. Rauhut, R. Ward. *Sparse Legendre expansions via l_1-minimization.* J. Approx. Theory **164(5)** (2012), 517–533.

[28] H. Rauhut and R. Ward. Interpolation via weighted l_1 minimization. Preprint (2013).

[29] J. K. Romberg. *Compressive sensing by random convolution.* SIAM J. Imaging Sci. **2(4)** (2009), 1098–1128.

[30] M. Rudelson, R. Vershynin. *On sparse reconstruction from Fourier and Gaussian measurements.* Comm. Pure Appl. Math. **61** (2008), 1025–1045.

A Spectral Algorithm for Latent Dirichlet Allocation

Yi-Kai Liu

(joint work with Animashree Anandkumar, Dean P. Foster, Daniel Hsu, Sham M. Kakade)

Document topic modeling is a popular task in machine learning and natural language processing, where one is given a large collection of documents, and one wants to learn a small number of topics that describe them. This task can be formally defined in different ways. One approach is to assume that each document was generated by first choosing a random topic h from some distribution, and then choosing random words $x_1, x_2, x_3, \dots$ independently from some distribution that depends on h (so that $x_1, x_2, x_3, \dots$ are conditionally independent given h). Latent Dirichlet allocation (LDA) is one model that has this structure.

In [1], we presented a spectral algorithm for learning the LDA model. The algorithm works by estimating the second- and third-order moments of the word distribution. The second-order moment is used to compute a whitening transformation, which exposes the low-rank structure in the third-order moment tensor. The third-order moment tensor is then randomly projected down to a matrix, which reveals the parameters of the LDA model. This algorithm compares favorably with existing methods based on expectation-maximization (EM).

Subsequent work [2] shows that this algorithm can be generalized to learn a large class of "multi-view" latent variable models. These algorithms can be simply described in terms of symmetric orthogonal tensor decompositions, which can be computed using the tensor power method. One open question is whether these methods can be made more robust to the effects of noisy data, statistical uncertainty, and incorrect modeling assumptions.

References

[1] A. Anandkumar, D. P. Foster, D. Hsu, S. M. Kakade and Y.-K. Liu, "A Spectral Algorithm for Latent Dirichlet Allocation," *Advances in Neural Information Processing Systems (NIPS)*, pp.926-934 (2012).
[2] Anima Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, Matus Telgarsky, "Tensor decompositions for learning latent variable models," ArXiv:1210.7559.

Theta bodies

Željka Stojanac
(joint work with Holger Rauhut)

We are interested in low-rank tensor recovery via small number of measurements. We consider a generalization of the matrix singular value decomposition to tensors called canonical decomposition (CP-decomposition) and the corresponding notion of rank and norm (tensor nuclear norm). However, the CP-decomposition and therefore its norm are in general NP-hard to compute. To overcome this difficulty we suggest closed convex relaxations of the tensor nuclear norm called theta bodies (introduced first by Lovász [5]) which can be computed via semidefinite programming. In the following, we will introduce theta bodies and show on an example of relaxation of the nuclear norm of 2×2 matrices how to build the corresponding semidefinite program.

A central problem in optimization is to find the maximum value of a linear function over a set $S \in \mathbb{R}^n$, i.e., solving

$$\max_{\mathbf{x}} \langle \mathbf{c}, \mathbf{x} \rangle \quad \text{s.t.} \quad \mathbf{x} \in S$$

which is equivalent to

$$\max_{\mathbf{x}} \langle \mathbf{c}, \mathbf{x} \rangle \quad \text{s.t.} \quad \mathbf{x} \in \text{cl}\,(\text{conv}\,(S)),$$

where $\text{cl}\,(\text{conv}\,(S))$ denotes the closure of the convex hull of the set S. For example, in linear programming the set S is a polyhedron $S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$.

We are interested in the case where S is a real algebraic set, i.e., a set of all real solutions to a finite set of polynomials. To be more precise, if a polynomial ideal $\mathcal{I}$ is generated by a finite set of polynomials, $\mathcal{I} = \langle f_1, f_2, \ldots, f_m \rangle$, the set S is the real algebraic variety of the ideal $\mathcal{I}$, i.e., $S = \nu_{\mathbb{R}}(\mathcal{I}) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0, \forall\, f \in \mathcal{I}\} = \{\mathbf{x} \in \mathbb{R}^n : f_i(\mathbf{x}) = 0, \forall\, i = 1, \ldots, m\}$.

The closure of the convex hull of an arbitrary set $S \subset \mathbb{R}^n$ and therefore also the closure of the convex hull of $S = \nu_{\mathbb{R}}(\mathcal{I})$ can be obtained as

$$\text{cl}\,(\text{conv}\,(\nu_{\mathbb{R}}(\mathcal{I}))) = \cap \left\{\mathbf{x} \in \mathbb{R}^n : \ell(\mathbf{x}) \geq 0 \text{ for all } \ell \text{ affine s.t. } \ell|_{\nu_{\mathbb{R}}(\mathcal{I})} \geq 0\right\}.$$

However, already checking for a single polynomial $\ell \in \mathbb{R}[\mathbf{x}]$ whether it is nonnegative on a set $\nu_{\mathbb{R}}(\mathcal{I})$ can be a difficult task. A natural idea is to relax the condition $\ell|_{\nu_{\mathbb{R}}(\mathcal{I})} \geq 0$ into something easier to check. One possibility to obtain the hierarchy of the convex relaxations is restricting only to affine polynomials which are k-sos mod $\mathcal{I}$, i.e. to polynomials that can be written as

$$\ell(\mathbf{x}) = \sigma(\mathbf{x}) + g(\mathbf{x}), \qquad \text{where } \sigma \in \Sigma_{2k},\ g \in \mathcal{I},$$

with Σ_{2k} denoting the sum of squares (sos) polynomials of degree at most $2k$ in $\mathbb{R}[\mathbf{x}]$. To be more precise,

$$\sigma \in \Sigma_{2k} \quad \text{if} \quad \exists\, h_1, h_2, \ldots, h_t \text{ with } \deg(h_1), \ldots \deg(h_t) \leq k \text{ s.t. } \sigma(\mathbf{x}) = \sum_{i=1}^{t} h_i^2(\mathbf{x}).$$

These relaxations are called theta bodies [1, 4]. In particular, for a fixed $k \in \mathbb{N}$ the k-th theta body of $\mathcal{I}$ is defined as

$$\mathrm{TH}_k(\mathcal{I}) = \{\mathbf{x} \in \mathbb{R}^n : \ell(\mathbf{x}) \geq 0, \text{ for all } \ell \text{ affine and } k\text{-sos mod } \mathcal{I}\}.$$

Thus, the theta bodies of $\mathcal{I}$ form a hierarchy of closed convex relaxations [4]

$$\mathrm{TH}_1(\mathbf{x}) \supseteq \mathrm{TH}_2(\mathbf{x}) \supseteq \cdots \supseteq \mathrm{TH}_k(\mathbf{x}) \supseteq \mathrm{TH}_{k+1}(\mathbf{x}) \cdots \supseteq \mathrm{cl}\left(\mathrm{conv}\left(\nu_{\mathbb{R}}(\mathcal{I})\right)\right).$$

Since we are not restricting to a particular basis of the ideal, we have to be able to do the computations at the level of the ideal $\mathcal{I}$ and therefore it is essential to find the Groebner basis of the ideal $\mathcal{I}$ with respect to some monomial ordering. The good monomial orderings will be the one that respect the degree [1], for example graded lexicographic [3] or graded reverse lexicographic ordering [3] which is used in the latter.

In the following, we will illustrate how to build a semidefinite relaxation of the unit nuclear norm ball for 2×2 matrices. We will work with polynomials in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{11}, x_{12}, x_{21}, x_{22}]$, where a matrix $\mathbf{X} \in \mathbb{R}^{2\times 2}$ is of the form

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}.$$

First, we define an ideal $\mathcal{J}$ such that its algebraic variety $\nu_{\mathbb{R}}(\mathcal{J})$ is the set of all rank one unit norm matrices so that $\mathrm{cl}\left(\mathrm{conv}\left(\nu_{\mathbb{R}}(\mathcal{J})\right)\right)$ is the unit nuclear norm ball. Notice that the ideal

$$\mathcal{J} = \left\langle x_{12}x_{21} - x_{11}x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \right\rangle$$

satisfies the desired property.

Secondly, we define a Groebner basis of the ideal $\mathcal{I}$ with respect to an appropriate ordering. In our example, the generators of the ideal form a Groebner basis of the ideal $\mathcal{J}$ with respect to the graded reverse lexicographic ordering.

Next we find the appropriate basis $B = \{f_0 + \mathcal{I}, f_1 + \mathcal{I}, \ldots\}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. For simplicity, we look only at standard monomial bases of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$. In addition, we assume that each element in the basis $B = \{f_i + \mathcal{I}\}$ of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ is represented by the polynomial whose degree equals the degree of its equivalence class, and that B is ordered so that $\deg(f_i + \mathcal{I}) \leq \deg(f_{i+1} + \mathcal{I})$. The ordered subset of the basis B of degree at most k is denoted by B_k. The basis of $\mathbb{R}[\mathbf{x}]/\mathcal{I}$ we need to find is the so called Θ-basis which satisfies the following two properties

(1) $B_1 = \{1 + \mathcal{I}, x_1 + \mathcal{I}, \ldots, x_n + \mathcal{I}\}$

(2) if $\deg(f_i + \mathcal{I}), \deg(f_j + \mathcal{I}) \leq k$ then $f_i f_j + \mathcal{I}$ is in the $\mathbb{R}$-span of B_{2k}.

For simplicity, in our example we are interested only in the first theta body $\mathrm{TH}_1(\mathcal{J})$ and therefore we need to find only the subsets B_1 and B_2 of the basis B

$$B_1 = \{1 + \mathcal{J}, x_{11} + \mathcal{J}, x_{12} + \mathcal{J}, x_{21} + \mathcal{J}, x_{22} + \mathcal{J}\}$$
$$\hat{B}_2 = \left\{x_{11}x_{12} + \mathcal{J}, x_{11}x_{22} + \mathcal{J}, x_{12}^2 + \mathcal{J}, x_{12}x_{22} + \mathcal{J}, x_{21}^2 + \mathcal{J}, x_{21}x_{22} + \mathcal{J}, x_{22}^2 + \mathcal{J}\right\},$$

where $B_2 = B_1 \cup \hat{B}_2$.

The next step consists in computing the so-called combinatorial moment matrix $\mathbf{M}_{B_k}(\mathbf{y})$. We first need to define a vector $[\mathbf{x}]_{B_k}$ which contains all the elements from the set B_k in order. Then one forms a matrix $\mathbf{X}_{B_k} = [\mathbf{x}]_{B_k}[\mathbf{x}]_{B_k}^T$ whose (i, j)-th entry is of the form $[\mathbf{X}_{B_k}]_{i,j} = f_i f_j + \mathcal{I}$. Since $\deg(f_i + \mathcal{I}), \deg(f_j + \mathcal{I}) \leq k$, by the second property of the theta basis $f_i f_j + \mathcal{I}$ is in the $\mathbb{R}$-span of B_{2k}. Thus, for all i, j there exist unique coefficients $\lambda_{i,j}^l \in \mathbb{R}$ such that

$$[\mathbf{X}_{B_k}]_{i,j} = f_i f_j + \mathcal{I} = \sum_{f_l + \mathcal{I} \in B_{2k}} \lambda_{i,j}^l \left(f_l + \mathcal{I}\right).$$

To obtain the combinatorial moment matrix one linearizes the elements of B_{2k}

$$[\mathbf{M}_{B_{2k}}(\mathbf{y})]_{i,j} = \sum_{f_l + \mathcal{I} \in B_{2k}} \lambda_{i,j}^l \mathbf{y}_l.$$

In our example, the combinatorial moment matrix is of the form

$$\mathbf{M}_{B_2}(\mathbf{y}) = \begin{pmatrix} y_0 & y_1 & y_2 & y_3 & y_4 \\ y_1 & -y_5 - y_6 - y_7 + 1 & y_8 & y_9 & y_{10} \\ y_2 & y_8 & y_5 & y_{10} & y_{11} \\ y_3 & y_9 & y_{10} & y_6 & y_{12} \\ y_4 & y_{10} & y_{11} & y_{12} & y_7 \end{pmatrix},$$

with the following table where in the first row the monomial f represents the $f + \mathcal{J}$ element of the basis B

1	x_{11}	x_{12}	x_{21}	x_{22}	x_{12}^2	x_{21}^2	x_{22}^2	$x_{11}x_{12}$	$x_{11}x_{21}$	$x_{11}x_{22}$	$x_{12}x_{22}$	$x_{21}x_{22}$
y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

Finally, we are ready to define the relaxations. The $\mathrm{TH}_k(\mathcal{I})$ is the closure of

$$Q_{B_k}(\mathcal{I}) = \pi_{\mathbb{R}^n}\left\{\mathbf{y} \in \mathbb{R}^{B_{2k}} : \mathbf{M}_{B_k}(\mathbf{y}) \succeq 0, y_0 = 1\right\},$$

where $\pi_{\mathbb{R}^n}$ is a projection on the variables $y_1, y_2, \ldots, y_n$, see [4, 1].

In our example, the first theta body is of the form

$$\mathrm{TH}_1(\mathcal{J}) = \mathrm{cl}\left(Q_{B_1}(\mathcal{J})\right) = \pi_{\mathbb{R}^4}\left\{\mathbf{y} \in \mathbb{R}^{B_2} : \mathbf{M}_{B_2}(\mathbf{y}) \succeq 0, y_0 = 1\right\}.$$

So far, we are able to prove that TH_1 coincides with the nuclear norm unit ball for 2×2 matrices. The generalization of this fact to arbitrary sized matrices is still open, see [6].

For 3rd order tensors we use a similar idea, i.e., we define the ideal $\mathcal{I}$ such that its algebraic variety $\nu_{\mathbb{R}}(\mathcal{I})$ contains all rank one unit norm tensors. In this scenario, minimization of the norm induced by $\mathrm{TH}_1(\mathcal{I})$ under an affine constraint seems to work well experimentally for low-rank tensor recovery. However, theoretical guarantees still remain an open question, see [6].

References

[1] G. Blekherman, P. A. Parillo, R. R. Thomas, *Semidefinite Optimization and Convex Algebraic Geometry*, SIAM (2013).

[2] V. Chandrasekaran, B. Recht, P. A. Parillo, A. S. Willsky, *The Convex Geometry of Linear Inverse Problems*, Foundations of Computational Mathematics **12(6)** (2012), 805–849.

[3] D. A. Cox, J. Little, D. O'Shea, *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Springer (1991).

[4] J. Gouveia, P. A. Parillo, R. R. Thomas, *Theta Bodies for Polynomial Ideals*, SIAM Journal on Optimization **20(4)** (2010), 2097–2118.

[5] L. Lovász, *Semidefinite programs and combinatorial optimization*, In Recent advances in algorithms and combinatorics **11** of CMS Books Math./Ouvrages Math. SMC (2003), Springer, New York, 137–194.

[6] H. Rauhut, Ž. Stojanac, *Low rank tensor recovery via convex optimization*, In preparation, 2014.

SIC-POVMs vs WSSUS: Quantum Information Theory meets Channel Estimation

Götz E. Pfander, Pavel Zheltov

Gabor frames play a central part in time-frequency analysis. Aside from being used in science and engineering to analyze and synthesize signals with respect to components localized in time and frequency, finite dimensional Gabor frames, that is, Gabor frames on finite Abelian groups, have been employed in mathematical physics to construct mutually unbiased basis and symmetrically information complete positive operator valued measures, referred to as SIC-POVMs. Below, we describe some recent findings on the geometry of Gabor frames and describe an engineering motivated application of SIC-POVMs in the realm of the estimation problem for so-called "wide sense stationary with uncorrelated scatterers", that is, WSSUS, channels.

For G being a finite Abelian group[1], and $\mathbb{C}^G = \{x : G \longrightarrow \mathbb{C}\}$, that is, $\mathbb{C}^G$ is a $|G|$-dimensional vector space with vector entries indexed by elements in the group G, we define unitary *translation operators* $T_k : \mathbb{C}^G \longrightarrow \mathbb{C}^G$, $k \in G$, by

$$T_k x\,(n) = x(n-k), \quad n \in G.$$

A *modulation operator* on $\mathbb{C}^G$ is pointwise multiplication with a *character* $\xi \in \widehat{G}$ on G, that is, with a group homomorphism ξ mapping G into the multiplicative group $S^1 = \{z \in \mathbb{C} : \ |z| = 1\}$,

$$M_\xi x\,(n) = \xi(n)\, x(n), \quad n \in G.$$

Combining translation and modulation, we obtain the *time-frequency shift operator* $\pi(\lambda)$, $\lambda = (k, \xi) \in G \times \widehat{G}$, that is,

$$\pi(\lambda) : \mathbb{C}^G \longrightarrow \mathbb{C}^G, \quad x \mapsto \pi(\lambda)x = \pi(k,\xi)x = M_\xi T_k x = \xi(\cdot)\, x(\cdot - k)\,.$$

[1]Recall that any finite Abelian group is a product of cyclic groups.

A *Gabor system* is then given by

$$(\varphi, \Lambda) = \{\pi(\lambda)\varphi\}_{\lambda\in\Lambda}\,,$$

where Λ is a subgroup (or just a subset) of $G \times \widehat{G}$. If (φ, Λ) spans $\mathbb{C}^G$, it forms a so-called *Gabor frame*. Gabor frames (φ, Λ) on subgroups $\Lambda \subseteq G \times \widehat{G}$ share a number of remarkable properties based on the fact that $\pi : G \times \widehat{G} \longrightarrow \mathcal{L}(\mathbb{C}^G, \mathbb{C}^G)$, $\lambda \mapsto \pi(\lambda)$, is a so-called projective representation [4].[2] For example, using the *adjoint subgroup*

$$\Lambda^\circ = \{\mu \in G \times \widehat{G} :\ \pi(\lambda)\pi(\mu) = \pi(\mu)\pi(\lambda) \ \text{ for all } \lambda \in \Lambda\}$$

of a group $\Lambda \subseteq G \times \widehat{G}$ we can formulate the following classical result from Gabor analysis attributed to Wexler–Raz and Ron–Shen [5, 4, 13].

Theorem 1. *Let Λ be a subgroup of $G \times \widehat{G}$. Then (φ, Λ) is a frame for $\mathbb{C}^G$ if and only if (φ, Λ°) is a linear independent set; and (φ, Λ) is a tight frame for $\mathbb{C}^G$ if and only if (φ, Λ°) is an orthogonal set.*

As a consequence, having $\Lambda = G \times \widehat{G}$, hence, $\Lambda^\circ = \{0\} \subseteq G \times \widehat{G}$ and $\varphi \neq 0$ implies that (φ, Λ) is a $\|\varphi\||G|$-tight frame.

In the recent development of a sampling theory for operators, the question of constructing Gabor systems $(\varphi, G \times \widehat{G})$ in general linear position has been relevant [9, 11]. Indeed, with the linear map (and a corresponding matrix) V_φ given by

$$V_\varphi : \mathbb{C}^G \longrightarrow \mathbb{C}^{G\times\widehat{G}}, \quad V_\varphi x(\lambda) = \langle x, \pi(\lambda)\varphi\rangle,$$

the following uncertainty principles were established [7, 8]. In the following, $\|u\|_0$ denotes the number of nonzero entries in a vector u.

Theorem 2. *If G is a cyclic group, then for almost every $\varphi \in \mathbb{C}^G$, the rows of V_φ are in general linear position, and, consequently, $\|V_\varphi x\|_0 \geq |G|^2 - |G| + 1$ for every $x \in \mathbb{C}^G$. If the order of G is prime, then for almost all $\varphi \in \mathbb{C}^G$ we have that all minors of V_φ are nonzero and $\|V_\varphi x\|_0 + \|x\|_0 \geq |G|^2 + 1$ for all $x \in \mathbb{C}^G$.*

The result does not extend to non-cyclic groups. An open question in this realm is the following. For G cyclic does there exist $\varphi \in \mathbb{C}^G$ with set equality[3]

$$\big\{(\|x\|_0, \|V_\varphi x\|_0),\ x \in \mathbb{C}^G\big\} = \big\{(\|x\|_0, |G|^2 - |G| + \|\widehat{x}\|_0),\ x \in \mathbb{C}^G\big\}\,?$$

As alluded to above, Theorem 2 has an application in terms of communications engineering and radar [10, 6]. Indeed, time-varying communications channels are frequently modeled as linear combinations of time-frequency shift operators, each representing a point scatterer. The task at hand is to design a sounding signal φ so that the respective channel response $H\varphi$ allows one to decipher H. In case that

[2] It is, in fact, up to isomorphisms, the only irreducible faithful projective representation of $G \times \widehat{G}$ on $\mathbb{C}^G$.

[3] $\widehat{x}$ denotes the Fourier transform of x.

we have *a priori* knowledge of the time-frequency shifts Λ caused by the paths the signals travel, then we aim to identify an operator from the class

$$\mathcal{H}_\Lambda = \Big\{ \sum_{\lambda\in\Lambda} c_\lambda \pi(\lambda),\ c_\lambda \in \mathbb{C} \Big\}, \quad \Lambda \subseteq G \times \hat{G}.$$

In case that Λ is not known but small, then, we aim to identify the class of operators

$$\mathcal{H}_s = \Big\{ \sum_{\lambda\in\Lambda} c_\lambda \pi(\lambda), c_\lambda \in \mathbb{C}\,,\ \Lambda \in G \times \hat{G} \ \text{ with } \ |\Lambda| \leq s \Big\}.$$

It is easy to observe that Theorem 2 implies that if G is cyclic, then there exists φ such that the map

$$\Phi : \mathcal{H}_\Lambda \longrightarrow \mathbb{C}^G, \quad H \mapsto H\varphi = \big(\sum_{\lambda\in\Lambda} \eta_\lambda \pi(\lambda) \big)\varphi = \sum_{\lambda\in\Lambda} \eta_\lambda\, \pi(\lambda)\varphi = V_\varphi^* \eta,$$

is injective for all Λ with $|\Lambda| \leq |G|$, that is, $V_\varphi^*|_\Lambda$ is injective. Similarly, we have that Φ acts injectively on $\mathcal{H}_s$ if $2s \leq |G|$. In applications, though, injectivity is not sufficient and we require that $V_\varphi^*|_\Lambda$ is well conditioned, that is, we require the existence of $B_{\varphi,\Lambda}$ and $A_{\varphi,\Lambda}$ with

$$A_{\varphi,\Lambda}\|\eta\|_2 = A_{\varphi,\Lambda}\|H\|_{HS} \leq \|V_\varphi^*\eta\|_2 = \|\Phi H\|_2 = \|H\varphi\|_2 \leq B_{\varphi,\Lambda}\|\eta\|_2, \quad H \in \mathcal{H}_\Lambda,$$

and $B_{\varphi,\Lambda}/A_{\varphi,\Lambda}$ small. Here, $\|H\|_{HS}$ denotes the Hilbert-Schmidt norm of H and the first equality follows from the fact that $\{\pi(\lambda)\}_{\lambda\in\Lambda}$ forms an orthonormal system with respect to the Hilbert-Schmidt inner product, see, for example, the overview article [9]. While for generic Λ vectors φ are known that guarantee injectivity, only in case that Λ is a subgroup, we can easily obtain φ with $B_{\varphi,\Lambda} = A_{\varphi,\Lambda}$. For example, with $\delta_0 \in \mathbb{C}^G$ denoting the vector which is 1 at 0 and 0 else, we have $(\delta_0, G \times \{0\})$ is an orthonormal basis and hence $A_{\delta_0, G\times\{0\}} = 1 = B_{\delta_0, G\times\{0\}}$.

In some applications, the channel is assumed to have stochastic components. In this situation, the coefficients $\boldsymbol{\eta}_\lambda$, $\lambda \in \Lambda$, are random variables and we can ask whether a stochastic operator $\boldsymbol{H}$ supported on Λ can be determined from its stochastic response $\boldsymbol{H}\varphi$ to a deterministic input signal φ.

Here, we shall address a slightly refined question. We assume that the components of $\boldsymbol{\eta}$ are zero mean and attempt to recover only the covariance $R_{\boldsymbol{\eta}} = \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^*] \in \mathbb{C}^{(G\times\widehat{G})\times(G\times\widehat{G})}$ from only the output covariance $R_{\boldsymbol{H}\varphi} = \mathbb{E}[\boldsymbol{H}\varphi(\boldsymbol{H}\varphi)^*] \in \mathbb{C}^{G\times G}$. A dimension counting argument shows that *a priori* this problem is ill-posed. Hence, we introduce a sparsity prior and consider

$$sto\mathcal{H}_\Gamma = \Big\{ \sum_{\lambda\in G\times\widehat{G}} c_\lambda \pi(\lambda),\ \mathbb{E}[\boldsymbol{\eta}_\lambda \overline{\boldsymbol{\eta}_\mu}] \neq 0 \text{ only for } (\lambda,\mu) \in \Gamma \Big\},\ \ \Gamma \subseteq (G \times \widehat{G}) \times (G \times \widehat{G})\,,$$

respectively

$$sto\mathcal{H}_s = \Big\{ \sum_{\lambda\in G\times\widehat{G}} c_\lambda \pi(\lambda),\ \mathbb{E}[\boldsymbol{\eta}_\lambda \overline{\boldsymbol{\eta}_\mu}] \neq 0 \text{ for at most } s\ (\lambda,\mu) \in (G \times \widehat{G}) \times (G \times \widehat{G}) \Big\}.$$

A trivial computation yields

$$\begin{aligned} R_{\boldsymbol{H}\varphi}(m,n) &= \mathbb{E}[\boldsymbol{H}\varphi(m)\overline{\boldsymbol{H}\varphi(n)}] \\ &= \mathbb{E}[\sum_{\lambda\in G\times\widehat{G}} \boldsymbol{\eta}(\lambda)\,\pi(\lambda)\varphi(m) \sum_{\mu\in G\times\widehat{G}} \overline{\boldsymbol{\eta}(\mu)}\,\overline{\pi(\mu)\varphi(n)}] \\ &= \sum_{\lambda\in G\times\widehat{G}} \sum_{\mu\in G\times\widehat{G}} \pi(\lambda)\varphi(m)\overline{\pi(\mu)\varphi(n)}\mathbb{E}[\boldsymbol{\eta}(\lambda)\,\overline{\boldsymbol{\eta}(\mu)}], \end{aligned}$$

and for $\boldsymbol{H} \in sto\mathcal{H}_\Gamma$, in bra-ket notation,

$$R_{\boldsymbol{H}\varphi} = \sum_{(\lambda,\mu)\in\Gamma} R_{\boldsymbol{\eta}}(\lambda,\mu)\ |\pi(\mu)\varphi\rangle\langle\pi(\lambda)\varphi|$$

Now, $R_{\boldsymbol{\eta}}$ with $\boldsymbol{H} \in sto\mathcal{H}_\Gamma$ can be recovered from $R_{\boldsymbol{H}\varphi}$ if and only if the rank one operators $\{|\pi(\mu)\varphi\rangle\langle\pi(\lambda)\varphi|\}_{(\lambda,\mu)\in\Gamma}$ are linearly independent [12, 14]. Unfortunately, Theorem 2 does not generalize to this setting, namely, if $|G| > 3$, there exists Γ of cardinality $3|G|+1 < |G\times\widehat{G}|$ where linear independence is not achieved.

The set $\Gamma_{\text{diag}} = \{(\lambda,\lambda),\ \lambda \in G\times\widehat{G}\} \subseteq \mathbb{C}^{(G\times\widehat{G})\times(G\times\widehat{G})}$ is of particular importance in communications engineering and radar. Many stochastic channels in communications engineering (or targets in radar) are assumed to have the Wide Sense Stationary with Uniform Scattering (WSSUS) property [3], that is, that the random variables η_λ are zero mean with

$$\mathbb{E}[\boldsymbol{\eta}_\lambda\overline{\boldsymbol{\eta}_\mu}] = \delta(\lambda-\mu)\,C(\lambda), \quad \lambda,\mu \in G\times\widehat{G}.$$

The function $C(\lambda) \geq 0$ is called the *scattering function* of the target. It represents the variances of all individual scatterers. The goal of determining $C(\lambda)$ corresponds to identifying $sto\mathcal{H}_{\Gamma_{\text{diag}}}$. Respective results in [17] are summarized as follows.

Theorem 3. *For G cyclic, $\varphi \neq 0$, the matrix with columns $\{|\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|\}_{\lambda\in G\times\widehat{G}}$ is invertible. Its condition number is $\max_{\lambda\in G\times\widehat{G}} |V_\varphi\varphi(\lambda)|/\min_{\lambda\in G\times\widehat{G}} |V_\varphi\varphi(\lambda)|$ which is lower bounded by $\sqrt{|G|+1}$. The lower bound is achieved if $|V_\varphi\varphi(\lambda)|$, $\lambda \neq 0$, is constant, that is if $(\varphi, G\times\widehat{G})$ is an equiangular frame.*

Note that a vector φ with $(\varphi, G\times\widehat{G})$ being an equiangular frame is generally referred to as fiducial vector in the literature, and if φ is fiducial, then $\{|\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|\}_{\lambda\in G\times\widehat{G}}$ is referred to as symmetrically information complete positive operator valued measure [2, 1, 16]. "Symmetric" refers to the fact that

$$\big\langle |\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|,\ |\pi(\mu)\varphi\rangle\langle\pi(\mu)\varphi| \big\rangle_{HS} = |\langle\pi(\lambda)\varphi, \pi(\mu)\varphi\rangle|^2$$

is constant for $\mu \neq \lambda$; "information complete" indicates that $\{|\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|\}_{\lambda\in G\times\widehat{G}}$ is complete in the space of Hilbert-Schmidt operators on $\mathbb{C}^G$, and "positive operator valued measure" describes, in short, that

$$\sum_{\lambda\in G\times\widehat{G}} |\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi| = \text{Identity}.$$

where here and in the following we assume that $\|\varphi\|_2 = 1/\sqrt{G}$. This property allows us to compute the biorthogonal basis of $\{|\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|\}_{\lambda\in G\times\widehat{G}}$ which leads to the following scattering function reconstruction formula and respective channel estimation procedures [17].

Corollary 4. *Let $\varphi \in \mathbb{C}^G$ be fiducial. Then for any WSSUS channel $\boldsymbol{H}$ we have*

$$C(\lambda) = R_{\boldsymbol{\eta}}(\lambda,\lambda) = \big\langle R_{\boldsymbol{H}\varphi},\ (|G|+1)|\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi| - \text{Identity}\big\rangle_{HS}$$
$$= (|G|+1)\langle\pi(\lambda)\varphi|R_{\boldsymbol{H}\varphi}|\pi(\lambda)\varphi\rangle - \text{trace} R_{\boldsymbol{H}\varphi}.$$

Confirmation of the so-called Zauner's conjecture would imply the existence of a fiducial vector whenever G is cyclic [16]. While the general case is still open, analytic solutions for G cyclic of order $|G| = 1, 2, \ldots, 15, 19, 24, 35, 48$, and numerical solutions for G cyclic with $|G| \leq 67$ are known [15].

The discussion above shows that WSSUS channel identification can be seen as a problem dual to the problem of determining a density matrix of a quantum state. In the latter, we attempt to reconstruct a matrix ρ from the inner products $\big\langle\rho,\ |\pi(\lambda)\varphi\rangle\langle\pi(\lambda)\varphi|\big\rangle_{HS}$, so the role of basis and dual basis are interchanged.

References

[1] D.M. Appleby, H.B.Dang, and C.A. Fuchs. Symmetric informationally-complete quantum states as analogues to orthonormal bases and minimum-uncertainty states. *Entropy*, 16(3):1484–1492, 2014.

[2] S. Colin, J. Corbett, T. Durt, and D. Gross. About sic povms and discrete wigner distributions. *Journal of Optics B: Quantum and Semiclassical Optics*, 7:S778, 2005.

[3] G. Durisi, V.I. Morgenshtern, H. Bölcskei, U.G. Schuster, and S. Shamai (Shitz). *Information theory of underspread WSSUS channels*, pages 65–116. 2011.

[4] H. G. Feichtinger, W. Kozek, and F. Luef. Gabor analysis over finite abelian groups. *Appl. Comput. Harmon. Anal.*, 26(2):230–248, 2009.

[5] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA, 2001.

[6] M.A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *Signal Processing, IEEE Transactions on*, 57(6):2275–2284, June 2009.

[7] J. Lawrence, G. E. Pfander, and D. F. Walnut. Linear independence of Gabor systems in finite dimensional vector spaces. *J. Fourier Anal. Appl.*, 11(6):715–726, 2005.

[8] R.-D. Malikiosis. A note on Gabor frames in finite dimensions. Preprint.

[9] G.E. Pfander. Sampling of operators. *J. Four. Anal. Appl.*, 19(3):612, 2013.

[10] G.E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.

[11] G.E. Pfander and D. Walnut. Sampling and reconstruction of operators. Preprint.

[12] G.E. Pfander and P. Zheltov. Sampling of stochastic operators. *Information Theory, IEEE Transactions on.* To appear.

[13] G.E. Pfander. Gabor frames in finite dimensions. In Peter G. Casazza and Gitta Kutyniok, editors, *Finite Frames*, Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2012.

[14] G.E. Pfander and P. Zheltov. Identification of stochastic operators. *Applied and Computational Harmonic Analysis*, 36(2):256 – 279, 2014.

[15] A.J. Scott and M. Grassl. Symmetric informationally complete positive-operator-valued measures: A new computer study. *Journal of Mathematical Physics*, 51(4), 2010.

[16] G. Zauner. *Quantum Designs Foundations of a Non-Commutative Theory of Designs.* PhD thesis, University of Vienna, Austria, 1999.

[17] G.E. Pfander and P. Zheltov. Estimation of overspread scattering functions. Preprint.

Estimating group transformations via convex relaxation

Afonso S. Bandeira
(joint work with Moses Charikar, Yutong Chen, Amit Singer, and Andy Zhu)

Let X be a space of objects and $\mathcal{G}$ be a group of transformations acting on X. Suppose we have n measurements of the form

$$y_i = g_i \cdot x + \epsilon_i,$$

for $i = 1, \dots, n$, where x is a fixed but unknown element of X, $g_1, \dots, g_n$ are unknown elements of $\mathcal{G}$, and ϵ_i are independent noise terms. We refer to the statistical estimation problem of the n group elements as the *multi-reference alignment problem.*

Of particular interest is the case of alignment of $L-$dimensional signals, where $\mathcal{G} = \mathbb{Z}/L\mathbb{Z}$ is the cyclic group of shifts (see [2]). In this case one observes noisy shifted copies of an unknown signal $x \in \mathbb{R}^L$ and is tasked with recovering the shifts (up to a global shift).

The challenge in obtaining the maximum likelihood estimator (MLE) is that the parameter space is non-convex and is exponentially large in n. We consider a convex relaxation using semidefinite programming (SDP), considered in [2] for the particular case of shifted signals. This relaxation is numerically shown to be tight with high probability for a wide range of parameters, that is, the SDP recovers the MLE with high probability. The tightness of this SDP is only understood in very specific instances [1] and understanding this phenomenon in more generality remains an interesting open problem [1, 4].

This approach is preferable to approaches that compare pairs of observations y_i, y_j in order to estimates the pairwise group transformation $g_i g_j^{-1}$ and, from these, obtain estimates for the group elements g_i's [1, 3, 5, 7] as the MLE incorporates the likelihood of all possible pairwise group transformations $g_i g_j^{-1}$ and not simply which one is the most likely. Moreover, as opposed to most iterative heuristics like expectation maximization or simulated annealing this approach does not risk getting trapped in local optima. Moreover, when the SDP relaxation is tight, which seems to be often the case (and tightness is easily checked), one is guaranteed to have found the true MLE. Indeed, numerical tests suggest that the MLE based approach outperforms existing approaches.

Besides alignment of signals, this method has applications in many problems including the problem, in theoretical chemistry, of determining the minimum energy positions of a set of atoms, the shape matching problem in computer graphics [6], and the cryo-electron microscopy problem in molecular imaging. In the context of cryo-electron microscopy a variation is considered, where the measurement model is of the form

$$y_i = P\left(g_i \cdot x\right) + \epsilon_i,$$

where P is a tomographic projection operator.

References

[1] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer. Linear inverse problems on Erdős-Rényi graphs: Information-theoretic limits and efficient recovery. *to appear in IEEE International Symposium on Information Theory (ISIT2014)*.

[2] A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu. Multireference alignment using semidefinite programming. *5th Innovations in Theoretical Computer Science (ITCS 2014)*, 2014.

[3] A. S. Bandeira, C. Kennedy, and A. Singer. Approximating the little Grothendieck problem over the orthogonal and unitary groups. *Available online at arXiv:1308.5207 [cs.DS]*, 2013.

[4] A. S. Bandeira, Y. Khoo, and A. Singer. Open problem: Tightness of maximum likelihood semidefinite relaxations. *Conference on Learning Theory (COLT 2014) Open Problem, to appear*, 2014.

[5] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger inequality for the graph connection Laplacian. *SIAM J. Matrix Anal. Appl.*, 34(4):1611–1630, 2013.

[6] Q.-X. Huang and L. Guibas, *Consistent shape maps via semidefinite programming*, Computer Graphics Forum **32** (2013), no. 5, 177–186.

[7] A. Singer, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. Comput. Harmon. Anal. **30** (2011), no. 1, 20 – 36.

Composite self-concordant minimization

Volkan Cevher

We describe a variable metric framework for minimizing the sum of a self-concordant function and a possibly non-smooth convex function, endowed with an easily computable proximal operator. We theoretically establish the convergence of our framework without relying on the usual Lipschitz gradient assumption on the smooth part. An important highlight of our work is a new set of analytic step-size selection and correction procedures based on the structure of the problem. We describe concrete algorithmic instances of our framework for quantum (process) tomography applications and demonstrate them numerically on both synthetic and real data.

Reporter: Afonso S. Bandeira

Participants

Afonso S. Bandeira
Department of Mathematics
Princeton University
Fine Hall
Princeton, NJ 08544-1000
UNITED STATES

Prof. Dr. Volkan Cevher
Ecole Polytechnique Fédérale de Lausanne
EPFL STI IEL LIONS
ELE 233 (Bâtiment ELE)
Station 11
1015 Lausanne
SWITZERLAND

Prof. Dr. David Groß
Physikalisches Institut
Universität Freiburg
79104 Freiburg
GERMANY

Prof. Dr. Madalin Guta
School of Mathematical Sciences
The University of Nottingham
University Park
Nottingham NG7 2RD
UNITED KINGDOM

Prof. Dr. Marius Junge
Dept. of Mathematics, University of Illinois at Urbana-Champaign
273 Altgeld Hall MC-382
Urbana, IL 61801-2975
UNITED STATES

Jun.-Prof. Dr. Felix Krahmer
Institut für Numerische
und Angewandte Mathematik
Georg-August-Universität Göttingen
Lotzestr. 16-18
37083 Göttingen
GERMANY

Richard Küng
Physikalisches Institut
Universität Freiburg
79104 Freiburg
GERMANY

Dr. Yi-Kai Liu
National Institute of Standards and Technology
100 Bureau Drive, Stop 8910
Gaithersburg MD 20899-8910
UNITED STATES

Dr. Dustin Mixon
Air Force Institute of Technology
2950 Hobson Way
Wright OH 45433
UNITED STATES

Prof. Dr. Götz Pfander
International University Bremen
School of Engineering and Science
Postfach 750561
28725 Bremen
GERMANY

Prof. Dr. Holger Rauhut
Lehrstuhl für Mathematik C (Analysis)
RWTH Aachen
Templergraben 55
52062 Aachen
GERMANY

Prof. Dr. Rayan Saab
Department of Mathematics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0112
UNITED STATES

Zeljka Stojanac
Lehrstuhl C für Mathematik
RWTH Aachen
52056 Aachen
GERMANY

Prof. Dr. Joel A. Tropp
Annenberg Center, CMS
MC 305-16
California Institute of Technology
1200 E. California Blvd.
Pasadena, CA 91125-5000
UNITED STATES

Prof. Rachel Ward
University of Texas at Austin
Department of Mathematics
2515 Speedway
Austin, TX 78712
UNITED STATES

Prof. Dr. Andreas Winter
Departament de Física
Àrea de Física Teòrica
Edifici C
Campus de la UAB
08193 Bellaterra (Barcelona)
SPAIN

Prof. Dr. Michael M. Wolf
Zentrum Mathematik, M5
TU München
Boltzmannstr. 3
85748 Garching b. München
GERMANY