Mathematisches Forschungsinstitut Oberwolfach

Report No. 24/2011

DOI: 10.4171/OWR/2011/24

Random Matrices, Geometric Functional Analysis and Algorithms

Organised by
Michel Ledoux, University of Toulouse, France
Mark Rudelson, University of Michigan, USA
Gideon Schechtman, Weizmann Institute, Israel

May 1st – May 7th, 2011

Abstract. The workshop gathered close to 50 participants on the topics of random matrix theory, high dimensional convex geometry and probabilistic methods in theoretical computer science. It favored cooperation between researchers in these interlaced areas by providing an interacting working atmosphere most appreciated by the participants. We hope this workshop will advance the developments at these interfaces further.

Mathematics Subject Classification (2000): 15A52, 46B07, 68W20.

Introduction by the Organisers

The workshop was successful in bringing together experts in three related fields: random matrix theory, geometric functional analysis and theoretical computer science. It was the opportunity to encourage further cooperation between people within and across these areas. 28 talks were delivered during the conference. Special efforts were made by the speakers from the three different communities to make their works accessible by the others in order to favor exchanges and discussions.

Among the specific areas discussed during the workshop, random matrix developments took an important part, as one of the hot topics of the current research. R. Vershynin presented new important results on invertibility of symmetric matrices based on Littlewood-Offord problems for quadratic forms. His talk emphasized the connection between the areas covered by the workshop, with problems coming from the random matrix theory, and methods coming from geometric functional

analysis. This number-theoretic aspect was further developed by K. Costello linking probabilistic and arithmetic properties of homogenous polynomials. V. Vu and J. Yin presented some of the most striking recent achievements on universality of the eigenvalue spacing distribution (works of T. Tao and V. Vu on one side and L. Erdös, H. T. Yau and collaborators on the other). Sharp bounds for singular values for matrices in log-concave ensembles with applications to approximate reconstruction was another highlight (talks by R. Latała, A. Litvak and N. Tomczak-Jaegermann). This series of works applies delicate geometric properties of log-concave measures to random matrices, and finds further application in signal reconstruction. M. Krishnapour presented a joint work with A. Guionnet and O. Zeitouni proving a long-standing single ring conjecture. This conjecture asserted that the empirical spectra of a unitary invariant ensemble of matrices converges to a measure, whose support is one ring, regardless of the potential. Other talks on random matrices included polynomials, tail bounds on sums of random matrices, heavy tail models, log-gases, etc. Connections between random matrix theory and quantum information theory, free probability and statistics completed the picture.

Recent developments in high dimensional convex geometry included results on tight embeddings in non-Euclidean spaces. Concentration inequalities and sharp bounds on log-concave measures with geometric applications to log-concave ensembles were presented by O. Guédon and R. Latała. B. Klartag presented a result on the vector in subspace problem which gives an application of convex geometry to the computer science area of communication complexity.

Advances in theoretical computer science are intimately related to both random matrices and high dimensional convex geometry. A talk of N. Srivastava described how a solution of a computer science problem of graph sparsification led to an unexpected improvement of several old results in geometric functional analysis. Algorithms and complexity theory figured in the talk of S. Khot on games with strong soundness, and in the talk of the structure of Y. Xiao on local optima. A talk by A. Barvinok discussed arithmetic properties of random matrices with prescribed row and column sums. Such matrices, called random contingency tables, appear in statistical analysis of large data arrays.

Workshop: Random Matrices, Geometric Functional Analysis and Algorithms

Table of Contents

Abstracts

Covariance estimation and invertibility of random matrices

Roman Vershynin

Estimation of covariance matrices is a basic problem in statistics and its applications. Consider a mean zero random vector X valued in $\mathbb{R}^n$. The covariance matrix of X is the $n \times n$ positive semidefinite matrix

$$\Sigma = \mathbb{E}\, XX^T.$$

Our goal is to estimate Σ from a sample $X_1, \dots, X_N$ taken from the same distribution as X. A classical unbiased estimator for Σ is the sample covariance matrix

$$\Sigma_N = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T.$$

A basic question is to determine the minimal sample size N which guarantees that Σ is accurately estimated by Σ_N. More precisely, for a given accuracy $\varepsilon > 0$ we are interested in the minimal $N = N(n, \varepsilon)$ so that

$$\mathbb{E}\, \|\Sigma_N - \Sigma\| \le \varepsilon \|\Sigma\|$$

where $\|\cdot\|$ denotes the spectral (operator) norm.

It follows from Rudelson's theorem [5] that $N = O(n \log n)$ for general distributions supported in a ball of radius $O(\sqrt{n})$, see [8, Section 4.3]. In general the logarithmic oversampling is needed. It is an open problem to describe the distributions for which no logarithmic oversampling is needed, i.e. for which $N = O(n)$. Such are sub-gaussian distributions (this follows from a standard covering argument [8, Section 4.3]) and even sub-exponential distributions (this is a recent result of Adamczak, Litvak, Pajor and Tomczak [1] that answered Kannan-Lovasz-Simonovits question). A conjecture is that the logarithmic oversampling is almost never needed; in particular that $N = O(n)$ for distributions with $2 + \varepsilon$ moments.

If the distribution has some structure then it may happen that $N = o(n)$, which is the regime preferred in modern applications. Sparsity can manifest itself differently. For example, suppose the distribution is approximately k-dimensional where $k = o(n)$, i.e. the covariance matrix Σ has effective rank k. In this case, one can deduce from Rudelson's theorem that $N = O(k \log n)$, so for low-dimensional distributions one has the desired bound $N = o(n)$, see [8, Section 4.3]. A different form of structure appears if the covariance matrix Σ is sparse, i.e. most of the coordinates are uncorrelated. Levina and Vershynin [4] proved that if Σ has k non-zeros per row then $N = O(k \log^6 n)$, so one again may have the desired bound $N = o(n)$.

The second part of this talk addresses the invertibility problem of random matrices. At the heart of random matrix theory lies the realization that the spectrum of a random matrix H tends to stabilize as the dimensions of H grow to infinity.

This phenomenon is captured by the limit laws of random matrix theory, in particular by Wigner's semicircle law, the circular law, and Marchenko-Pastur law. One can think of these laws as relatives of the central limit theorem, although the way the random entries of H determine the spectrum is more complicated than the sum of the entries studied in the central limit theorem.

The limit laws offer us a clear global and asymptotic picture of the spectrum of H. In the last few years, a considerable progress was made on the more difficult local and non-asymptotic regimes. In the non-asymptotic regime, the dimensions of H are fixed rather than grow to infinity. In the local regime, one zooms in on a small part of the spectrum of H, ideally until one sees individual eigenvalues. As an important example, suppose one zooms in on zero. The location of the eigenvalue nearest zero determines the invertibility properties of H, i.e. the probability that a random matrix H is non-singular, and the typical value of the spectral norm of the inverse of H. The invertibility properties determine in turn whether the matrix H is well conditioned, which is a matter of importance in numerical analysis.

We report on the recent progress on the invertibility problem for general Wigner matrices H, which are symmetric $n \times n$ matrices with iid above-diagonal entries that have zero mean, unit variance and sub-gaussian moments (the latter can be relaxed). A result in [9] shows that the eigenvalues $\lambda_k(H)$ satisfy for each $z \in \mathbb{R}$ and $\varepsilon \geq 0$ that

$$\mathbb{P}\Big\{ \min_k |\lambda_k(H) - z| \leq \varepsilon n^{-1/2} \Big\} \leq C\varepsilon^{1/9} + 2e^{-n^c}.$$

where $C, c > 0$ are constants that depend only on the subgaussian moments of the entries of H.

This result shows that H is singular with an exponentially small probability, that $\|H^{-1}\| = O(n)$ and hence the condition number of H is linear in the dimension n, and that the spectrum of H is fully delocalized – the eigenvalues are spread out, they do not tend to stick to any particular point at scales comparable to their average gap. This result improves upon the polynomial singularity bound $O(n^{-1/8+\varepsilon})$ due to Costello, Tao and Vu [2], and it generalizes, up to constant factors, previous results for distributions whose first few moments match the moments of the normal distribution (due to the universality results of Tao and Vu [6, 7]) and for continuous distributions in the bulk of the spectrum (due to Erdös, Schlein and Yau [3]).

References

[1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. **23** (2010), 535–561.

[2] K. Costello, T. Tao, V. Vu, *Random symmetric matrices are almost surely non-singular*, Duke Math. J. **135** (2006), 395–413.

[3] L. Erdös, B. Schlein, H.-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not., to appear.

[4] E. Levina, R. Vershynin, *Partial estimation of covariance matrices*, Probability Theory and Related Fields, to appear. DOI: 10.1007/s00440-011-0349-4.

[5] M. Rudelson, *Random vectors in the isotropic position*, J. Funct. Anal. **164** (1999), 60–72.

[6] T. Tao, V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math., to appear.
[7] T. Tao, V. Vu, *Random matrices: Universality of local eigenvalue statistics up to the edge*, Comm. Math. Phys., to appear.
[8] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*. In: Compressed Sensing: Theory and Applications, eds. Yonina Eldar and Gitta Kutyniok. Cambridge University Press, to appear (2010).
[9] R. Vershynin, *Invertibility of symmetric random matrices*, submitted (2010).

The Vector in Subspace Problem

Bo'az Klartag
(joint work with Oded Regev)

Suppose $A \subseteq S^{n-1}$ is a fixed measurable set and let $E \subset \mathbb{R}^n$ be a random k-dimensional subspace. We consider the random variable

$$\sigma_E(A \cap E)$$

where σ_E denoted the uniform probability measure on the sphere $S^{n-1} \cap E$. We show that the random variable $\sigma_E(A \cap E)$ is rather concentrated, even when the set A is quite small. Denote $R = \log(1/\sigma_{n-1}(A))$, where σ_{n-1} is the uniform probability measure on the sphere S^{n-1}. We show that for any $0 \leq t \leq 1$,

$$\mathbb{P}\left\{ \left| \frac{\sigma_E(A \cap E)}{\sigma_{n-1}(A)} - 1 \right| \geq t \right\} \leq C \exp(-ckt^2/R^2)$$

where $C, c > 0$ are universal constants. The estimate is in general sharp for $k \leq 9n/10$ (better estimates exists when k is very close to n).

We view this inequality as a "sampling inequality". Suppose we have a subset of the sphere that, say, occupies only $\exp(-n^{1/3})$-fraction of the sphere. We would like to estimate its exact measure, with an error of at most 1%. Our only access to the set is via its intersection with a random $n/2$-dimensional subspace $E \subset \mathbb{R}^n$, and we use the measure of the intersection as an unbiased estimator for the measure of the set. The inequality above states that this is a pretty accurate estimator, in high dimension: The chances of an error of more than 1% is very small, only $C \exp(-cn^{1/3})$.

The main application of the inequality we present is related to the field of *Quantum Communication*. Specifically, a decade ago Ran Raz presented a (partial) function for which there is a quantum protocol communicating only O(log n) qubits, but for which any classical (randomized, bounded-error) protocol requires poly(n) bits of communication. That quantum protocol requires two rounds of communication. Ever since Raz's paper it was open whether the same exponential separation can be achieved with a quantum protocol that uses only one round of communication. We settle this question in the affirmative, using our sampling inequality as a central technical tool.

Partial transposition of random matrices

Guillaume Aubrun

In the recent years, very fruitful connexions were discovered between Random Matrix Theory and Quantum Information Theory. The most prominent example is Hastings's use of the probabilistic method to obtain counterexamples to a long-standing problem: the additivity conjecture [2].

A basic object in Quantum Information Theory is a quantum state. We stick to the finite-dimensional case, and we consider a quantum state to be simply a positive operator with trace one. We denote by $D(\mathbf{C}^n)$ the set of states on $\mathbf{C}^n$. There is a natural family of probability measures on $D(\mathbf{C}^n)$ which is obtained by normalizing (Gaussian) Wishart matrices. Let G be a $n \times p$ random matrix with i.i.d. entries with distribution $N_{\mathbf{C}}(0,1)$, and define

$$\rho = \frac{GG^\dagger}{\operatorname{tr} GG^\dagger},$$

which is a random quantum state on $\mathbf{C}^n$. We call $\mu_{n,p}$ the distribution of ρ. When $p \geq n$, the measure $\mu_{n,p}$ has a simple density with respect to the Lebesgue measure on the hyperplane of trace one operators:

$$d\mu_{n,p}(\rho) \propto (\det \rho)^{p-n} \mathbf{1}_{\{\rho \geq 0\}} d\rho.$$

This family of random states has also a physical interpretation in terms of open quantum systems. The statistical repartition of their eigenvalues is described by the Marčenko–Pastur distribution, in the asymptotic limit when n, p go to infinity with $\lim p/n = \alpha \in (0, \infty)$.

Usually in Quantum Information Theory, we are interested in the case when the Hilbert space carries an extra tensor product structure, which corresponds to the description of a shared quantum system. We consider for simplicity the case of states on $\mathbf{C}^d \otimes \mathbf{C}^d$ (which we identify with $\mathbf{C}^n$ for $n = d^2$). If we think of states as matrices, this corresponds to adding a block structure.

An fundamental operation which plays an important role in Quantum Information Theory is the partial transposition. The partial transposition ρ^Γ of a state $\rho \in D(\mathbf{C}^d \otimes \mathbf{C}^d)$ is defined as

$$\rho^\Gamma = (\mathrm{Id} \otimes T)(\rho),$$

where T is the usual transposition on $d \times d$ matrices. In the block-matrices representation, this corresponds to applying the usual transposition inside each block. A state ρ is called PPT (Positive Partial Transpose) if ρ^Γ is a positive operator. This notion is useful since non-PPT states are necessary entangled (the converse is false beyond very small dimensions), and this is a very practical (polynomial time) criterion, while deciding whether a state is entangled or not is computationnally hard.

It is therefore relevant to ask what the spectrum of the partial transposition of random states looks like, and especially whether it is entirely positive. We answer both questions: the spectral distribution of ρ^Γ is given by a *non-centered*

semicircular distribution, and there is a phase transition from PPT to non-PPT when the Wishart parameter α equals 4. Here are more precise statements.

Theorem 1. Let ρ be a random state on $\mathbf{C}^d \otimes \mathbf{C}^d$, distributed according to the measure $\mu_{d^2,p}$. Assume that d, p go to infinity and that the ratio p/d^2 tends to a limit $\alpha \in (0, \infty)$. Then the empirical spectral distribution of $d^2\rho^\Gamma$ approaches a semicircular distribution with mean 1 and variance $1/\alpha$, supported in the interval $[1 - 2/\sqrt{\alpha}, 1 + 2/\sqrt{\alpha}]$.

Theorem 2. Under the same hypotheses, the largest eigenvalue of $d^2\rho^\Gamma$ approaches $1 + 2/\sqrt{\alpha}$ and the smallest eigenvalue approaches $1 - 2/\sqrt{\alpha}$.

Both proofs are based on the moments method. See [1].

References

[1] G. Aubrun, *Partial transposition of random states and non-centered semicircular distributions*, preprint

[2] M. B. Hastings, Superadditivity of communication capacity using entangled inputs, *Nature Physics* **5**, 255 (2009).

Random matrices with group-theoretic linear structure

Mark Meckes

The most obvious way to devise a model of a random matrix is simply to choose each of the matrix entries independently. Since the work of Wigner in the 1950s, a great deal of attention has focused on random matrices lying in the linear subspaces of real symmetric or complex Hermitian matrices; to generate such random matrices one may pick all the above-diagonal entries independently. In a 1999 survey paper [2], Bai raised the possibility of investigating random matrices which lie in other important linear subspaces of matrix space, in particular mentioning Toeplitz, Hankel, and Markov matrices. Several years later Bryc, Dembo, and Jiang [4] and Hammond and Miller [5] proved that (appropriately normalized) random real symmetric Toeplitz matrices have a limiting spectral distribution (LSD); [4] proved the same about the other two models and determined the order of the spectral norm in the Markov case. Meckes [7] and Adamczak [1] determined the order of the spectral norm of random Toeplitz and Hankel matrices.

Beginning with Bose and Mitra [3], a number of authors considered random circulant matrices, which are a further restriction of random Toeplitz matrices. Bose and Mitra showed that real symmetric random circulant matrices have an LSD which is normal. Meckes [8] showed that without the symmetry constraint, random circulant matrices have a *complex normal* LSD.

Here we propose to investigate a different random matrix model which generalizes circulant matrices. Let G be a finite abelian group. (The case of nonabelian G will be the subject of future study.) Let $\{Y_g \mid g \in G\}$ be random variables and define a random matrix $X \in \mathbb{C}^{G \times G}$ by $X_{g,h} = Y_{gh^{-1}}$. A matrix of this form

is sometimes called a G-circulant matrix; when G is a cyclic group one obtains a classical circulant matrix. The action of such a matrix is the same as convolution with the vector $(Y_g)_{g\in G}$. Thus this may equally well be thought of as a model of a random convolution operator on G, and its spectrum may be studied by means of Fourier analysis on G, which turns this into a question about the distribution of values of a random Fourier series on G. Then in particular the order of the spectral norm of X follows as a special case of results of Marcus and Pisier [6].

The new results about LSDs are more sensitive than in previously studied contexts to details both of the structure of G and whether the Y_g are real or (say) complex with uncorrelated real and imaginary parts, and the real symmetric and complex Hermitian cases may even exhibit different behavior from each other. In general, the LSD is a mixture of two Gaussian distributions, either both real or one real and one complex. The parameters of the mixture depend on whether we are in the complex or real setting, and on the fraction of elements of G which are of order 2. (The strange behaviors that may arise in this model do not appear for classical circulant matrices precisely because large cyclic groups have a negligible fraction of such elements.)

References

[1] R. Adamczak, *A few remarks on the operator norm of random Toeplitz matrices*, J. Theoret. Probab. **23** (2010), no. 1, 85-108.

[2] Z. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611-677.

[3] A. Bose and J. Mitra, *Limiting spectral distribution of a special circulant*, Statist. Probab. Lett. **60** (2002), no. 1, 111-120.

[4] W. Bryc, A. Dembo, and T. Jiang, *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, Ann. Probab. **34** (2006), no. 1, 1-38.

[5] C. Hammond and S. Miller, *Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices*, J. Theoret. Probab. **18** (2005), no. 3, 537-566.

[6] M. Marcus and G. Pisier, *Random Fourier Series with Applications to Harmonic Analysis*, Annals of Mathematics Studies, 101. Princeton University Press, Princeton, N.J., 1981.

[7] M. Meckes, *On the spectral norm of a random Toeplitz matrix*, Electron. Comm. Probab. **12** (2007), 315-325

[8] M. Meckes, *Some results on random circulant matrices*, pp. 213–223 of *High Dimensional Probability V: The Luminy Volume*, IMS Collections, 5. Institute of Mathematical Statistics, Beachwood, OH, 2009.

Probability inequalities for sums of random matrices

Joel A. Tropp

1. Overview

Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent, self-adjoint random matrices with dimension $d \times d$. Our goal is to provide bounds for the probability

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_{k=1}^{n} \boldsymbol{X}_k\right) \geq t\right\}. \tag{1}$$

The symbol $\lambda_{\max}$ denotes the (algebraically) maximum eigenvalue of a self-adjoint matrix. We wish to harness properties of the individual summands to obtain information about the behavior of the sum. The approch here leads to simple estimates that are relatively general and easy to use in applied settings. The cost is that the results are not quite sharp for every example.

This research begins with the observation that controlling (1) resembles the classical problem of developing tail bounds for a sum of independent real random variables. There are some compelling analogies between self-adjoint matrices and real numbers that suggest it may be possible to extend classical techniques to the matrix setting. Indeed, this dream can be realized.

In a notable paper [1], Ahlswede and Winter show that elements from the Laplace transform technique generalize to the matrix setting. Further work in this direction includes [10, 4, 8, 9]. These techniques are closely related to noncommutative moment inequalities [7, 3, 5] and their applications in random matrix theory [11, 12].

2. The Matrix Laplace Transform Method

To begin, we show how Bernstein's Laplace transform technique extends to the matrix setting. The basic idea is due to Ahlswede–Winter [1], but we follow Oliveira [9] in this presentation. Fix a positive number θ. Observe that

$$\begin{aligned}\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_k \boldsymbol{X}_k\right) \geq t\right\} &= \mathbb{P}\left\{\exp\left\{\lambda_{\max}\left(\sum\nolimits_k \theta\boldsymbol{X}_k\right)\right\} \geq \mathrm{e}^{\theta t}\right\} \\ &\leq \mathrm{e}^{-\theta t}\cdot\mathbb{E}\exp\left\{\lambda_{\max}\left(\sum\nolimits_k \theta\boldsymbol{X}_k\right)\right\} \\ &= \mathrm{e}^{-\theta t}\cdot\mathbb{E}\,\lambda_{\max}\left(\exp\left\{\sum\nolimits_k \theta\boldsymbol{X}_k\right\}\right) \\ &< \mathrm{e}^{-\theta t}\cdot\mathbb{E}\operatorname{tr}\exp\left\{\sum\nolimits_k \theta\boldsymbol{X}_k\right\}. \qquad (2)\end{aligned}$$

The first identity uses the positive homogeneity of the eigenvalue map; the second relation is Markov's inequality; the third line is the spectral mapping theorem; and the last part holds because the exponential of a self-adjoint matrix is positive definite.

At this point, previous authors interpreted the quantity

$$\mathbb{E}\operatorname{tr}\exp\left\{\sum\nolimits_k \theta\boldsymbol{X}_k\right\}$$

as a matrix extension of the classical moment generating function (mgf). They attempted to generalize the fact that the mgf of an independent sum is the product of the mgfs of the summands. Roughly, the hope seemed to be that

$$\text{⟩⟩}\quad \mathbb{E}\operatorname{tr}\exp\{\textstyle\sum_k \theta\boldsymbol{X}_k\} = \operatorname{tr}\prod_k \mathbb{E}\,\mathrm{e}^{\theta\boldsymbol{X}_k}. \quad\text{⟨⟨}$$

This ostensible identity fails completely. Why? In the matrix setting, it is generally not true that $\mathrm{e}^{\boldsymbol{X}+\boldsymbol{Y}} \neq \mathrm{e}^{\boldsymbol{X}}\mathrm{e}^{\boldsymbol{Y}}$. The Golden–Thompson inequality [2, Ch. IX] can be used as a limited substitute: $\operatorname{tr}\mathrm{e}^{\boldsymbol{X}+\boldsymbol{Y}} \leq \operatorname{tr}\mathrm{e}^{\boldsymbol{X}}\mathrm{e}^{\boldsymbol{Y}}$. But the obvious extension to three matrices is false: $\operatorname{tr}\mathrm{e}^{\boldsymbol{X}+\boldsymbol{Y}+\boldsymbol{Z}} \not\leq \operatorname{tr}\mathrm{e}^{\boldsymbol{X}}\mathrm{e}^{\boldsymbol{Y}}\mathrm{e}^{\boldsymbol{Z}}$. On reflection, it becomes clear

that results like this cannot be true because the trace of a product of three positive matrices can be a negative number. In the past, researchers have circumvented this problem using some clever iterative procedures.

Nevertheless, we need a new idea if we want to find the natural extension the classical approach. The key observation is that we should try to **extend the additivity rule for cumulants**. To do so, we need more tools. The following result is one of the crown jewels of matrix analysis.

Theorem 1 (Lieb [6]). *Let $\boldsymbol{H}$ be a self-adjoint matrix. Then the map*

$$\boldsymbol{A} \longmapsto \operatorname{tr} \exp\{\boldsymbol{H} + \log \boldsymbol{A}\}$$

is concave on the positive-definite cone.

We apply Lieb's theorem through the following simple corollary.

Corollary 2 (Tropp 2010). *Let $\boldsymbol{H}$ be a fixed self-adjoint matrix, and let $\boldsymbol{X}$ be a random self-adjoint matrix. Then*

$$\mathbb{E} \operatorname{tr} \exp\{\boldsymbol{H} + \boldsymbol{X}\} \leq \operatorname{tr} \exp\left\{\boldsymbol{H} + \log \mathbb{E}\, \mathrm{e}^{\boldsymbol{X}}\right\}.$$

When we apply the corollary iteratively, we obtain the following inequality in our setting.

(3)
$$\operatorname{tr} \exp\left\{\log \mathbb{E} \exp\left\{\sum\nolimits_k \theta \boldsymbol{X}_k\right\}\right\} = \mathbb{E} \operatorname{tr} \exp\left\{\sum\nolimits_k \theta \boldsymbol{X}_k\right\} \leq \operatorname{tr} \exp\left\{\sum\nolimits_k \log \mathbb{E}\, \mathrm{e}^{\theta \boldsymbol{X}_k}\right\}.$$

The bound (3) states that the cumulant generating function (cgf) of a sum of independent random matrices is controlled by the sum of the cgfs of the individual matrices. Introducing (3) into (2), we reach

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_k \boldsymbol{X}_k\right) \geq t\right\} \leq \inf_{\theta > 0}\left[\mathrm{e}^{-\theta t} \cdot \operatorname{tr} \exp\left\{\sum\nolimits_k \log \mathbb{E}\, \mathrm{e}^{\theta \boldsymbol{X}_k}\right\}\right]. \tag{4}$$

The latter inequality is the natural matrix extension of the classical Laplace transform approach.

3. Example: Matrix Rademacher series

The simplest application of (4) concerns Rademacher series with matrix coefficients. Let $\{\boldsymbol{A}_k\}$ be a finite sequence of fixed, self-adjoint matrices with dimension d. Let $\{\varepsilon_k\}$ be a sequence of independent Rademacher random variables. We claim that

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_k \varepsilon_k \boldsymbol{A}_k\right) \geq t\right\} \leq d \cdot \mathrm{e}^{-t^2/2\sigma^2} \quad \text{where} \quad \sigma^2 = \left\|\sum\nolimits_k \boldsymbol{A}_k^2\right\|. \tag{5}$$

The symbol $\|\cdot\|$ denotes the spectral norm, or Hilbert space operator norm, of a matrix. A related calculation, which we omit, yields

$$\mathbb{E}\, \lambda_{\max}\left(\sum\nolimits_k \varepsilon_k \boldsymbol{A}_k\right) \leq \sigma \cdot \sqrt{2 \log d}.$$

For every example, this bound on the expectation is sharp up to the square-root log factor.

The inequality (5) has some interesting relations to earlier results. An alternative proof uses sharp noncommutative Khintchine inequalities [3] to bound the matrix mgf. In comparison, the approach described by Ahlswede and Winter [1] leads to the weaker inequality

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_k \varepsilon_k \boldsymbol{A}_k\right) \geq t\right\} \leq d \cdot \mathrm{e}^{-t^2/2\rho^2} \quad \text{where} \quad \rho^2 = \sum\nolimits_k \left\|\boldsymbol{A}_k^2\right\|.$$

The latter estimate also follows from Tomczak-Jaegermann's moment bounds [13] for Rademacher series in the Schatten classes.

To establish the claim (5), we need to study the cgf of a fixed matrix modulated by a Rademacher variable. Note that

$$\log \mathbb{E}\, \mathrm{e}^{\varepsilon\theta\boldsymbol{A}} = \log\cosh(\theta\boldsymbol{A}) \preccurlyeq \frac{\theta^2}{2}\boldsymbol{A}^2.$$

The semidefinite relation follows from the scalar inequality $\log\cosh(x) \leq x^2/2$. Introduce this estimate (with appropriate justifications!) into the tail bound (4) to reach

$$\begin{aligned}
\mathbb{P}\left\{\lambda_{\max}\left(\sum\nolimits_k \varepsilon_k \boldsymbol{A}_k\right) \geq t\right\} &\leq \inf_{\theta>0}\ \mathrm{e}^{-\theta t} \cdot \operatorname{tr}\exp\left\{\frac{\theta^2}{2}\sum\nolimits_k \boldsymbol{A}_k^2\right\} \\
&\leq \inf_{\theta>0}\ \mathrm{e}^{-\theta t} \cdot \exp\left\{\frac{\theta^2}{2} \cdot \lambda_{\max}\left(\sum\nolimits_k \boldsymbol{A}_k^2\right)\right\} \\
&= \inf_{\theta>0}\ \mathrm{e}^{-\theta t} \cdot \mathrm{e}^{\theta^2\sigma^2/2}.
\end{aligned}$$

Optimize with respect to θ to complete the proof of (5).

Finally, let us mention that these ideas can be extended to study rectangular matrices. Consider a finite sequence $\{\boldsymbol{B}_k\}$ of fixed $d_1 \times d_2$ matrices. Then

$$\mathbb{P}\left\{\left\|\sum\nolimits_k \varepsilon_k \boldsymbol{B}_k\right\| \geq t\right\} \leq (d_1+d_2)\cdot \mathrm{e}^{-t^2/2\sigma^2} \quad \text{where} \quad \sigma^2 = \left\|\sum\nolimits_k \boldsymbol{B}_k\boldsymbol{B}_k^*\right\| \vee \left\|\sum\nolimits_k \boldsymbol{B}_k^*\boldsymbol{B}_k\right\|.$$

Remarkably, this estimate follows immediately from (5) by applying that result to the self-adjoint matrices

$$\boldsymbol{A}_k = \begin{bmatrix} \mathbf{0} & \boldsymbol{B}_k \\ \boldsymbol{B}_k^* & \mathbf{0} \end{bmatrix}.$$

See [14] for further details.

Acknowledgments

This work has been supported in part by ONR awards N00014-08-1-0883 and N00014-11-1-0025, AFOSR award FA9550-09-1-0643, and a Sloan Fellowship.

References

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, Mar. 2002.

[2] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, Berlin, 1997.

[3] A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Ann.*, 319:1–16, 2001.

[4] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, Oct. 2009. To appear. Available at `arXiv:0910.1879`.
[5] M. Junge and Q. Xu. On the best constants in some non-commutative martingale inequalities. *Bull. London Math. Soc.*, 37:243–253, 2005.
[6] E. H. Lieb. Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.*, 11:267–288, 1973.
[7] F. Lust-Piquard. Inégalités de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Math. Acad. Sci. Paris*, 303(7):289–292, 1986.
[8] R. I. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Available at `arXiv:0911.0600`, Feb. 2010.
[9] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Elect. Comm. Probab.*, 15:203–212, 2010.
[10] B. Recht. Simpler approach to matrix completion. *J. Mach. Learn. Res.*, Oct. 2009. To appear. Available at `http://pages.cs.wisc.edu/~brecht/papers/09.Recht.ImprovedMC.pdf`.
[11] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164:60–72, 1999.
[12] M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *J. Assoc. Comput. Mach.*, 54(4):Article 21, 19 pp., Jul. 2007. (electronic).
[13] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes S_p $(1 \leq p < \infty)$. *Studia Math.*, 50:163–182, 1974.
[14] J. A. Tropp. User-friendly tail bounds for sums of random matrices. Available at `arXiv:1004.4389`, Apr. 2010.

Concentration inequalities for log-concave measures

Olivier Guédon
(joint work with Emanuel Milman)

I have presented several questions from the high dimensional geometry of convex bodies or log-concave measures. Let X be a random vector uniformly distributed on a convex body K in $\mathbb{R}^n$ or more generally distributing according to a log-concave density of probability f where log-concavity means that for any $\lambda \in (0,1)$ and any $x, y \in \mathbb{R}^n$, $f((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda} f(y)^{\lambda}$. This random vector is said to be isotropic if $\mathbb{E}X \otimes X = \mathrm{Id}$. In all this report, X will denote an isotropic log-concave random vector in $\mathbb{R}^n$ and its density with respect to the Lebesgue measure in $\mathbb{R}^n$ will be denoted by f. Every letters C, c will denote universal constants which occurrence may change from line to line.

In [12], Kannan, Lovasz and Simonovits asked about a variant of the isoperimetric inequality for an isotropic convex body, that is about the worse constant in Cheeger inequality where the measure is uniformly distributed on an isotropic convex body. More generally, in a log-concave setting, what is h such that for any $S \subset \mathbb{R}^n$,

$$\mathbb{P}^+(S) \geq h^{1/2} \mathbb{P}(X \in S)(1 - \mathbb{P}(X \in S)) \tag{1}$$

where the unit Euclidean ball is denoted by B_2^n and

$$\mathbb{P}^+(S) = \liminf_{\varepsilon \to 0} \frac{\mathbb{P}(X \in S + \varepsilon B_2^n) - \mathbb{P}(X \in S)}{\varepsilon}$$

They conjectured that up to a universal constant, the extremal sets in this inequality are the half-spaces, hence h is a universal constant.

In the middle of the nineties, Ball asked the following : does $|X|_2$ concentrate around its expectation significantly smaller than the trivial bound suggested by $\mathbb{V}\mathrm{ar}|X|_2 \leq \mathbb{E}|X|_2^2$? More precisely, in [1], they asked if there exists a sequence ε_n going to zero when n goes to infinity such that for any log-concave isotropic random vector X,

$$\mathbb{P}(||X|_2 - \sqrt{n}| \geq \varepsilon_n \sqrt{n}) \leq \varepsilon_n ? \tag{2}$$

This question was motivated by the fact that a positive answer to this question implies a central limit theorem for log-concave measures, a question independently asked in [4] and [1].

A stronger version of this conjecture was put forth by Bobkov and Koldobsky in [3] : can we prove that

$$\sqrt{\mathbb{V}\mathrm{ar}|X|_2} \leq C \quad \text{or} \quad \sigma_f^2 = \frac{\mathbb{V}\mathrm{ar}|X|_2^2}{\mathbb{E}|X|_2^2} = \frac{\mathbb{V}\mathrm{ar}|X|_2^2}{n} \leq C \quad ? \tag{3}$$

Another important problem from the high dimensional geometry of convex bodies is the hyperplane conjecture [18]: does there exist a universal constant c such that for any convex body $K \subset \mathbb{R}^n$ of volume 1, you may find a direction $\theta \in S^{n-1}$ such that $\mathrm{Vol}_{n-1}(K \cap \theta^{\perp}) \geq c$? Equivalently, for any log-concave isotropic density f in $\mathbb{R}^n$,

$$f(0)^{1/n} \leq C \quad ? \tag{4}$$

It is well known [16] that (1) implies that for any smooth function F,

$$\mathbb{V}\mathrm{ar}F(X) \leq \frac{C}{h} \mathbb{E}|\nabla F(X)|_2^2$$

In the talk, I have presented some of the recent results around these questions. I will try to describe some of them, not in chronological order. Milman in [17] proved that (1) with h being a universal constant is equivalent to proving that for any 1-Lipschitz function F, $\mathbb{V}\mathrm{ar}F(X) \leq C$. It is then easy to observe that (3) is just a particular case of (1) with $F(X) = |X|_2$ or $F(X) = |X|_2^2$. In their original paper, Kannan, Lovasz and Simonovits [12] proved that (1) holds true with $1/h \leq C\mathbb{E}|X|_2 \leq C\sqrt{n}$ while Bobkov [2] improved it by showing that $1/h \leq C\left(\mathbb{V}\mathrm{ar}|X|_2^2\right)^{1/4}$. Moreover, it is well known [16] that (1) implies that for every $t > 0$,

$$\mathbb{P}\left(\left||X|_2 - \sqrt{n}\right| \geq t\sqrt{n}\right) \leq C \exp(-C\, t\, \sqrt{n}).$$

Ball announced in his series of lectures during the semester Phenomena in High Dimensions organized at the Institut Henri Poincaré in 2006 that a positive answer to (1) with h being a universal constant implies the hyperplane conjecture (4). In [5], Eldan and Klartag proved a stronger version:

$$\sup f(0)^{1/n} \leq C \sup \sigma_f$$

where the suprema are taken over f being isotropic log-concave probability densities in $\mathbb{R}^n$. This means that a positive answer to (3) implies a positive answer to

(4). Fleury in [6] observed that a positive answer to (1) with h being a universal constant implies that

$$\forall\ 2 \leq p \leq c\sqrt{n}, \quad (\mathbb{E}|X|_2^p)^{1/p} \left(1 + \frac{c\,p}{n}\right) \left(\mathbb{E}|X|_2^2\right)^{1/2}$$

and proved for example that this inequality holds true when X is uniformly distributed on any isotropic generalized Orlicz ball.

A main breakthrough in this type of question was made by Paouris [19] where he proved a large deviation result. For every $t \geq C$, we have

$$\mathbb{P}(|X|_2 \geq t\sqrt{n}) \leq \exp(-C\,t\,\sqrt{n}). \tag{5}$$

While positive answers to (1) with h being a universal constant are given by Sodin [20] when X is uniformly distributed on B_p^n for $1 \leq p \leq 2$, by Latała and Wojtaszczyk [15] for $p \geq 2$, by Huet [11] when X is uniformly distributed on a body of revolution. Fleury [8] studied the case when X is uniformly distributed on a random Gaussian polytope and proved such type of results in average.

In a breakthrough paper, Klartag [13] proved the conjecture of Anttila, Ball and Perissinaki (2) with ε_n being decreasing like negative power of $\log n$. Independently, following Paouris' approach, Fleury, Guédon and Paouris [9] proved also the result with the same type of dependence for ε_n. Slightly after, Klartag [14] proved the result with polynomial dependence in the dimension for ε_n. Precisely, he proves that for every $t \in (0,1)$,

$$\mathbb{P}\left(\left||X|_2 - \sqrt{n}\right| \geq t\sqrt{n}\right) \leq C\exp(-C\,t^{3.33}n^{0.33})$$

which implies that $\sqrt{\mathbb{V}\mathrm{ar}|X|_2} \leq Cn^{0.41}$ Recently in [7], Fleury improved Klartag's thin-shell estimate to $\sqrt{\mathbb{V}\mathrm{ar}|X|} \leq Cn^{\frac{1}{2}-\frac{1}{8}}$ by obtaining the following deviation estimates:

$$\mathbb{P}(|X| \geq (1+t)\sqrt{n}) \leq C\exp(-cn^{\frac{1}{4}}t^2) \quad \forall t \in [0,1]\ ;$$

$$\mathbb{P}(|X| \leq (1-t)\sqrt{n}) \leq C\exp(-cn^{\frac{1}{8}}t) \quad \forall t \in [0,1]\ .$$

Note, however, that when $t = 1/2$, Fleury's positive and negative large-deviation estimates are both inferior to those of Klartag, and so in the mesoscopic scale $t = n^{-\delta}$ ($\delta > 0$ small), Klartag's estimates still outperform Fleury's (and Paouris' ones are inapplicable). In addition, note that both Klartag and Fleury's estimates do not seem to improve under a ψ_α condition, contrary to the ones of Paouris [19]. Recall that X (and its density) is said to be "ψ_α with constant b_α" if:

$$(\mathbb{E}|\langle X, y\rangle|^p)^{1/p} \leq b_\alpha p^{1/\alpha} \left(\mathbb{E}|\langle X, y\rangle|^2\right)^{1/2} \quad \forall p \geq 2 \quad \forall y \in \mathbb{R}^n\ .$$

All of this suggests that one might hope for a concentration estimate which recovers Paouris' sharp positive large-deviation estimate (5), improves if X is ψ_α, improves the best-known thin-shell estimate of Fleury, improves the best-known mesoscopic-deviation estimate of Klartag, interpolates continuously between all scales of t (bulk, mesoscopic, large-deviation). The aim of our work is to provide precisely such an estimate.

Theorem 1. *Let X denote an isotropic random vector in $\mathbb{R}^n$ with log-concave density, which is in addition ψ_α ($\alpha \in [1,2]$) with constant b_α. Then:*

$$\mathbb{P}(\big||X|_2 - \sqrt{n}\big| \geq t\sqrt{n}) \leq C \exp\left(-c \frac{n^{\alpha/2}}{b_\alpha^\alpha} \min(t^{2+\alpha}, t)\right) \quad \forall t \geq 0 \ .$$

In particular, we obtain the following thin-shell estimate:

$$\sqrt{\mathbb{V}ar(|X|_2)} \leq C n^{\frac{1}{2}} n^{-\frac{\alpha}{2(2+\alpha)}} b_\alpha^{\frac{\alpha}{(2+\alpha)}} .$$

I refer to [10] for more details and general formulation of the result, allowing an application of a linear transformation to X.

References

[1] M. Anttila, K. Ball, and I. Perissinaki, *The central limit problem for convex bodies*, Trans. Amer. Math. Soc. 355(12) (2003), 4723–4735.

[2] S. Bobkov. *On isoperimetric constants for log-concave probability distributions*, In *Geometric aspects of functional analysis, Israel Seminar 2004-2005*, volume 1910 of Lecture Notes in Math., pages 81–88. Springer, Berlin, 2007.

[3] S. G. Bobkov and A. Koldobsky. *On the central limit property of convex bodies*, In *Geometric aspects of functional analysis*, volume 1807 of Lecture Notes in Math., pages 44–52. Springer, Berlin, 2003.

[4] U. Brehm and J. Voigt. *Asymptotics of cross sections for convex bodies*, Beiträge Algebra Geom. 41 (2000), no. 2, 437–454.

[5] R. Eldan and B. Klartag. *Approximately gaussian marginals and the hyperplane conjecture*, To appear in Contemporary Math., Proceedings of a workshop on "Concentration, Functional Inequalities and Isoperimetry".

[6] B. Fleury. *Between Paouris concentration inequality and variance conjecture*, Ann. Inst. Henri Poincaré Probab. Stat., 46(2) (2010), 299–312.

[7] B. Fleury. *Concentration in a thin euclidean shell for log-concave measures* J. Func. Anal., 259 (2010), 832–841.

[8] B. Fleury. *Poincaré inequality in mean value for Gaussian polytopes* to appear in Prob. Theory Relat. Fields (2011).

[9] B. Fleury, O. Guédon, and G. Paouris. *A stability result for mean width of l_p-centroid bodies*, Advances in Mathematics, 214(2) (2007), 865–877.

[10] O. Guédon and E. Milman, *Interpolating Thin-Shell and Sharp Large-Deviation Estimates For Isotropic Log-Concave Measures*, to appear in Geom. Funct. Anal. (2011).

[11] N. Huet. *Spectral gap for some invariant log-concave probability measures*, Mathematika 57 (2011), no. 1, 51–62.

[12] R. Kannan, L. Lovász, and M. Simonovits. *Isoperimetric problems for convex bodies and a localization lemma*, Discrete Comput. Geom., 13(3-4) (1995) 541–559.

[13] B. Klartag. *A central limit theorem for convex sets*, Invent. Math., 168 (2007) 91–131.

[14] B. Klartag. *Power-law estimates for the central limit theorem for convex sets*, J. Funct. Anal., 245 (2007) 284–310.

[15] Latała, R. and Wojtaszczyk, J. O. *On the infimum convolution inequality*, Studia Mathematica, 189-2, (2008), 147–187.

[16] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[17] E. Milman. *On the role of convexity in isoperimetry, spectral-gap and concentration*, Invent. Math., 177(1) (2009) 1–43.

[18] V. D. Milman and A. Pajor. *Isotropic position and interia ellipsoids and zonoids of the unit ball of a normed n-dimensional space*, In Geometric Aspects of Functional Analysis, volume 1376 of Lecture Notes in Mathematics, pages 64–104. Springer-Verlag, 1987-1988.

[19] G. Paouris. *Concentration of mass on convex bodies*, Geom. Funct. Anal., 16(5) (2006), 1021–1049.
[20] S. Sodin *An isoperimetric inequality on the l_p balls*, Ann. Inst. Henri Poincaré Probab. Stat. 44 (2008), no. 2, 362–373.

Random matrices: The Universality phenomenon and the Four Moment theorem

Van Vu

Random matrix theory is a central topic in probability and mathematical physics, with many connections to various areas such that statistics, number theory, combinatorics, numerical analysis and theoretical computer science.

The main goal of random matrix theory is to derive limiting laws for the eigenvalues and eigenvectors. For the sake of presentation, in most of this survey we restrict ourself to Wigner matrices, although the results hold for more general models such as symmetric real matrices and sample covariance matrices [6, 7, 8].

Definition 1 (Wigner matrices)**.** *Let n be a large number. A* Wigner Hermitian matrix *(of size n) is defined as a random Hermitian $n \times n$ matrix M_n with upper triangular complex entries $\zeta_{ij} := \xi_{ij} + \sqrt{-1}\tau_{ij}$ $(1 \leq i < j \leq n)$ and diagonal real entries ξ_{ii} $(1 \leq i \leq n)$ where*

- *For $1 \leq i < j \leq n$, ξ_{ij}, τ_{ij} are iid copies of a real random variable ξ with mean zero and variance $1/2$.*
- *For $1 \leq i \leq n$, ξ_{ii} are iid copies of a real random variable $\tilde{\xi}$ with mean zero and variance 1.*
- *$\xi, \tilde{\xi}$ have exponential decay, i.e., there are constants C, C' such that $\mathbb{P}(|\xi| \geq t^C) \leq \exp(-t), \mathbb{P}(|\tilde{\xi}| \geq t^C) \leq \exp(-t)$, for all $t \geq C'$.*

We refer to $\xi, \tilde{\xi}$ as the atom distributions *of M_n, and ξ_{ij}, τ_{ij} as the* atom variables*. We refer to the matrix $W_n := \frac{1}{\sqrt{n}} M_n$ as the* coarse-scale normalized Wigner Hermitian matrix*, and $A_n := \sqrt{n} M_n$ as the* fine-scale normalized Wigner Hermitian matrix*.*

Example 2. *An important special case of a Wigner Hermitian matrix is the* gaussian unitary ensemble *(GUE), in which $\xi, \tilde{\xi}$ are gaussian random variables with mean zero and variance $1/2$, 1 respectively.*

Given an $n \times n$ Hermitian matrix A, we denote its n eigenvalues as

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A),$$

and write $\lambda(A) := (\lambda_1(A), \ldots, \lambda_n(A))$.

A cornerstone of this theory is the *Wigner semicircular law*. Denote by ρ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases} \tag{1}$$

Theorem 3 (Semi-circular law)**.** *Let M_n be a Wigner Hermitian matrix. Then for any real number x,*

$$\lim_{n\to\infty} \frac{1}{n}|\{1 \le i \le n : \lambda_i(W_n) \le x\}| = \int_{-2}^{x} \rho_{sc}(y)\ dy$$

in the sense of probability (and also in the almost sure sense, if the M_n are all minors of the same infinite Wigner Hermitian matrix), where we use $|I|$ to denote the cardinality of a finite set I.

Example 4. *Wigner*[10] *proved this theorem for special ensembles. The general version above is due to Pastur (see* [1] *for a detailed discussion). The semi-circular law in fact holds under substantially more general hypotheses than those given in Definition 1, but we will not discuss this matter further here.*

Theorem 3 addressed the global behavior of the eigenvalues. One of the main open problems in the field is to understand the local behavior. For instance, what can we say about the limiting law of an individual eigenvalue λ_i or that of the gap $\lambda_{i+1} - \lambda_i$, for some $1 \le i \le n$.

It has been generally believed (and in many cases explicitly conjectured; see [4, page 9] for an example)) that the local statistics (such as the above limiting distributions) are *universal*, in the sense that the limiting laws do not depend on the distribution of the atom variables (at the entries of the matrix). This phenomenon was motivated by similar phenomena in physics, such as the same laws of thermodynamics, which should emerge no matter what the details of atomic interaction and have been discussed in numerous books and surveys (see [4, 2, 3]).

It is clear that if one is able to prove the universality of a limiting law, then the problem of determining this law reduces to computing it for a specific model. This is usually doable in the GUE model, thanks to the availability of an explicit formula for the joint distribution of the eigenvalues (Ginibre's formula) and the fact that GUE is unitary invariance.

Recently, Tao and Vu [6, 7, 8] proved the following theorem, which roughly states that the joint distribution of any set of k eigenvalues (for any fixed k) depends only on the first four moment of the atom variable.

Theorem 5 (Four Moment Theorem)**.** *For any small positive constant c, there is a small positive constant c_0 such that for integer $k \ge 1$ the following holds. Let $M_n = (\zeta_{ij})_{1\le i,j\le n}$ and $M'_n = (\zeta'_{ij})_{1\le i,j\le n}$ be two random matrices where the atom distributions have exponential decay. Assume furthermore that for any $1 \le i < j \le n$, ζ_{ij} and ζ'_{ij} match to order 4 (their first four mixed moments are equal) and for any $1 \le i \le n$, ζ_{ii} and ζ'_{ii} match to order 2. Set $A_n := \sqrt{n}M_n$ and $A'_n := \sqrt{n}M'_n$, and let $G : \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$|\nabla^j G(x)| \le n^{c_0} \tag{2}$$

for all $0 \le j \le 5$ and $x \in \mathbb{R}^k$. Then for any $1 \le i_1 < i_2 \cdots < i_k \le n$, and for n sufficiently large we have

$$|\,\mathbb{E}(G(\lambda_{i_1}(A_n), \ldots, \lambda_{i_k}(A_n))) - \mathbb{E}(G(\lambda_{i_1}(A'_n), \ldots, \lambda_{i_k}(A'_n)))| \le n^{-c_0}. \tag{3}$$

The statement shows that one cannot tell the two distributions apart by looking at any "nice" test function. For a detailed discussion of this result (including many refinements, extensions and applications) we refer to [6, 5, 9].

Acknowledgements. V. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

References

[1] Z. D. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
[2] P. Deift, Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.
[3] P. Deift, Universality for mathematical and physical systems. *International Congress of Mathematicians* Vol. I, 125–152, Eur. Math. Soc., Zürich, 2007.
[4] M.L. Mehta, Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York, NY, 1967.
[5] T. Tao and V. Vu, The Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices, *submitted*.
[6] T. Tao and V. Vu, Random matrices: Universality of the local eigenvalue statistics, *Acta Mathematica* 206 (2011), 127-204.
[7] T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics up to the edge, *Comm. Math. Phys.* 298 (2010), no. 2, 549572.
[8] T. Tao, V. Vu, Random covariance matrices: university of local statistics of eigenvalues, *to appear in Annals of Probability*.
[9] T. Tao and V. Vu, Random matrices: Universal properties of eigenvectors, *to appear in Random Matrices: Theory and Applications*.
[10] P. Wigner, On the distribution of the roots of certain symmetric matrices, *The Annals of Mathematics* **67** (1958) 325-327.

Convergence of the norm of a polynomial in independent Wigner matrices

Greg W. Anderson

As part of a larger operator-theoretic investigation initiated in [11], it was shown in [10] that there are for N sufficiently large no eigenvalues outside an ϵ-neighborhood of the support of the limiting spectral distribution of a self-adjoint noncommutative polynomial in independent GUE(N) matrices. From this it follows trivially that the norm of a noncommutative polynomial in independent GUE(N) matrices (self-adjoint or not) converges almost surely. In the preprint [1] we have proved the analogous statements for Wigner matrices, forswearing the Poincaré inequality for Gaussian random variables used in [10] to do so. Now the archetype for all results of the form "no eigenvalues outside the support..." is the result of [3], which was proved under stringent fourth moment hypotheses. The paper [3] gave us hope for and guidance toward substitutes for the Poincaré inequality. The method we ultimately developed for analyzing polynomials in Wigner matrices combines ideas of [10] and [3] with some new tricks to handle

correction terms. Needless to say, the classical results of [6] and [4] about single Wigner matrices also play key roles.

Many generalizations of the random matrix results of [10] and [11] have already appeared. In [15] results in the GOE and GSE cases were obtained. These results bring to light correction terms of a sort one must handle as soon as one goes beyond the GUE case. In [5], a generalization to non-Gaussian distributions satisfying Poincaré inequalities was obtained. This result clarifies the central role of the Poincaré inequality in the arguments of [10]. In [12], a generalization involving polynomials in Gaussian Wigner matrices and deterministic matrices with convergent joint law was obtained. This result in particular provides rectangular generalizations of the random matrix results of [10].

All the works listed above exploit two extraordinarily powerful ideas from [10], namely, (i) a counterintuitively "backwards" way of estimating the error of approximate solutions of the Schwinger-Dyson equation and (ii) the linearization trick. We refine and simplify both ideas in our work. Both ideas are of permanent value and deserve wide popularization.

The paper [10] and all works following upon it presuppose and heavily exploit the well-known fact that the limiting spectral distribution of a self-adjoint polynomial in independent Wigner matrices equals the law of the corresponding polynomial in free semicircular variables. Many special cases, refinements, generalizations and proofs of this result exist in the literature. The original insight is due to Voiculescu; see, for example, [16]. We mention also [2, Chap. 5, Sec. 4] and the recent preprint [13] in order to give two more recent references. (And yet another proof emerges as a byproduct of our work.) More is true, namely, the support of the limiting spectral distribution of a polynomial in independent Wigner matrices equals the spectrum of a certain self-adjoint operator on Boltzmann-Fock space. The latter fact one can either dig out of [16, Thm. 2.6.2] or derive from scratch by elementary arguments at the level, say, of [14]. We have outlined the elementary arguments in [1] for the reader's convenience. We are grateful to K. Dykema for communicating them to us. The spectral representation of the support is an important tool for us since in effect it converts the analytic problem of controlling a support into an algebraic problem of controlling a spectrum.

In [4], convergence of the largest eigenvalue of a Wigner matrix was established under stringent fourth moment hypotheses. In [1], by exploiting a truncation strategy more or less the same as in [3], we specialize our general results to obtain a "polynomialization" of the results of [4]. Thus, although it may seem at first glance that the model studied in [1] has excessively generous moment assumptions, in fact the theory does touch the fourth moment boundary.

In many papers equations more or less the same as the Schwinger-Dyson equation albeit under different names have been studied and exploited for probabilistic purposes. The papers [7], [8] and [9] have been especially significant influences. But note that we do not solve the Schwinger-Dyson equation by an iterative procedure as do the cited authors. Rather, more in keeping with the approach of [10],

we harvest the solutions we need fully formed from Boltzmann-Fock space, thus painlessly gaining control of domains of definition.

There has recently been much progress on universality for (single) Wigner matrices, as the talks of Vu and of Yin at the conference made plain. It is overwhelmingly likely that universality holds in the polynomial case. We believe polynomial universality is an excellent direction for future research because many if not quite all the necessary tools seem already to exist. We hope our results can serve as a point of departure for investigation of edge-universality in the polynomial case.

References

[1] G. Anderson, *Convergence of the norm of the largest singular value of a polynomial in independent Wigner matrices*, arXiv:1103.4825.

[2] G. Anderson, A. Guionnet and O. Zeitouni, *An Introduction to Random Matrices,* Cambridge studies in advanced mathematics 118, Cambridge University Press, Cambridge 2010.

[3] Z. D. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices,* Annals Probab. **26** (1998), 316–345.

[4] Z. D. Bai and Y. Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix.* Annals Probab. **16** (1988), 1729–1741.

[5] M. Capitaine and C. Donati-Martin, *Strong asymptotic freeness for Wigner and Wishart matrices,* Indiana Univ. Math. J., **56** (2007), 767–803.

[6] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices,* Combinatorica **1** (1981), 233–241.

[7] W. Hachem, P. Loubaton, J. Najim, *Deterministic equivalents for certain functionals of large random matrices,* Ann. Appl. Probab. **17** (2007), no. 3, 875–930.

[8] W. Hachem, P. Loubaton, J. Najim, *A CLT for information-theoretic statistics of gram random matrices with a given variance profile,* Ann. Appl. Probab. **18** (2008), no. 6, 2071–2130.

[9] J. Helton, R. Rashidi Far, R. Speicher, *Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints,* Int. Math. Res. Not. IMRN **2007**, no. 22, Art. ID rnm086.

[10] U. Haagerup, H. Schultz and S. Thorbjørnsen, *A random matrix approach to the lack of projections in* $C^*_{\mathrm{red}}(\mathbb{F}_2)$., Adv. Math. **204** (2006), 1–83.

[11] U. Haagerup and S. Thorbjørnsen, *A new application of random matrices:* $\mathrm{Ext}(C^*(\mathbb{F}_2))$ *is not a group,* Annals Math. **162** (2005), 711–775.

[12] C. Male, *The norm of polynomials in large random and deterministic matrices,* arXiv:1004.4155v4, 3 Mar 2011.

[13] M. W. Meckes and S. J. Szarek, *Concentration for noncommutative polynomials in random matrices,* arXiv:1101.1923v1, 10 Jan 2011.

[14] G. J. Murphy, C^**-algebras and operator theory,* Academic Press, San Diego, CA, 1990.

[15] H. Schultz, *Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases,* Probab. Theory Related Fields **131** (2005), 261–309.

[16] D. V. Voiculescu, K. J. Dykema, A. Nica, *Free random variables. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.* CRM Monograph Series, 1. American Mathematical Society, Providence, RI, 1992.

Spectrum of non-hermitian heavy tailed random matrices

Charles Bordenave
(joint work with Pietro Caputo and Djalil Chafaï)

The *eigenvalues* of an $n \times n$ complex matrix M are the roots in $\mathbb{C}$ of its characteristic polynomial. We label them $\lambda_1(M), \dots, \lambda_n(M)$ so that $|\lambda_1(M)| \geq \cdots \geq |\lambda_n(M)| \geq 0$. We also denote by $s_1(M) \geq \cdots \geq s_n(M)$ the *singular values* of M, defined for every $1 \leq k \leq n$ by $s_k(M) := \lambda_k(\sqrt{MM^*})$ where $M^* = \overline{M}^\top$ is the conjugate transpose of M. We define the empirical spectral measure and the empirical singular values measure as

$$\mu_M = \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(M)} \quad \text{and} \quad \nu_M = \frac{1}{n}\sum_{k=1}^{n} \delta_{s_k(M)}.$$

Let $(X_{ij})_{i,j\geq 1}$ be i.i.d. complex random variables with cumulative distribution function F. Consider the matrix $X = (X_{ij})_{1\leq i,j\leq n}$. If F has finite positive variance σ^2, then Marchenko and Pastur [2] have proved that a.s. (almost surely)

$$\nu_{\frac{1}{\sqrt{n}}X} \underset{n\to\infty}{\rightsquigarrow} \mathcal{Q}_\sigma \tag{1}$$

where $\rightsquigarrow$ denotes the weak convergence of probability measures and the probability measure $\mathcal{Q}_\sigma$ is the quarter-circular law with Lebesgue density

$$x \mapsto \frac{1}{\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\mathbf{1}_{[\mathbf{0},\mathbf{2}\sigma]}(\mathbf{x}). \tag{2}$$

A proof of (1) is based on a classical approach for Hermitian random matrices with bounded second moment: truncation, centralization, recursion on the resolvent, and cubic equation for the limiting Cauchy-Stieltjes transform.

Girko's famous circular law theorem states under the same assumptions that a.s.

$$\mu_{\frac{1}{\sqrt{n}}X} \underset{n\to\infty}{\rightsquigarrow} \mathcal{U}_\sigma \tag{3}$$

where $\mathcal{U}_\sigma$ is the uniform law on the disc $\{z \in \mathbb{C}; |z| \leq \sigma\}$. This statement was established through a long sequence of partial results, the general case (3) being finally obtained by Tao and Vu [3] by using Girko's Hermitization with logarithmic potentials and uniform integrability, the convergence (1), and polynomial bounds on the extremal singular values.

We have investigated what happens when F does not have a finite second moment and considered the hypothesis:

$$\mathbf{P}(|X_{11}| \geq t) \sim_{t\to\infty} c\, t^{-\alpha} \text{ for some } c > 0.$$

For every $n \geq 1$, let us define the i.i.d. $n \times n$ complex matrix $A = A_n$ by

$$A_{ij} := a_n^{-1} X_{ij} \tag{4}$$

for every $1 \leq i, j \leq n$ and $a_n \sim_{n\to\infty} c^{1/\alpha} n^{1/\alpha}$. Belinschi, Dembo and Guionnet have proved that a.s.

$$\nu_A \underset{n\to\infty}{\rightsquigarrow} \nu_\alpha.$$

where ν_α is a symmetric probability measure on $\mathbb{R}$ and has a continuous density. Its density at $x = 0$ is known and its tail $\nu_\alpha((t, \infty))$ is asymptotically equivalent to $t^{-\alpha}$.

We have proved a heavy tailed version of Girko's circular law theorem (3). More precisely, there exists a probability measure μ_α on $\mathbb{C}$ depending only on α such that a.s.

$$\mu_A \underset{n\to\infty}{\rightsquigarrow} \mu_\alpha.$$

This probability distribution μ_α is isotropic and has a continuous density. Its density at $z = 0$ equals

$$\frac{\Gamma(1 + 2/\alpha)^2 \Gamma(1 + \alpha/2)^{2/\alpha}}{\pi \Gamma(1 - \alpha/2)^{2/\alpha}}.$$

Furthermore, up to a multiplicative constant, the density of μ_α is equivalent to

$$|z|^{2(\alpha-1)} e^{-\frac{\alpha}{2}|z|^\alpha} \text{ as } |z| \to \infty.$$

The above formula reveals a striking contrast between μ_α and ν_α. The limiting law of the eigenvalues μ_α has a stretched exponential tail while the limiting law ν_α of the singular values is heavy tailed with power exponent α, see [1]. This does not contradict the identity $\prod_{k=1}^n |\lambda_k(A)| = \prod_{k=1}^n s_k(A)$, but it does indicate that A is typically far from being a normal matrix.

References

[1] S. Belinschi, A. Dembo, and A. Guionnet, *Spectral measure of heavy tailed band and covariance random matrices*, Comm. Math. Phys. **289** (2009), no. 3, 1023–1055.

[2] V.A. Marchenko and L.A. Pastur, *The distribution of eigenvalues in sertain sets of random matrices*, Mat. Sb. **72** (1967), 507–536.

[3] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law*, with an appendix by Manjunath Krishnapur. Ann. Probab. **38** (2010), no. 5, 2023–2065.

A Two Prover One Round Game with Strong Soundness

Subhash Khot

(joint work with Muli Safra)

Brief Description

It is well-known that for many NP-hard problems, even computing approximate solutions is computationally hard. A hard instance of 2-Prover-1-Round Game is a starting point for many of the inapproximability results and constructions of probabilistically checkable proofs (PCPs), e.g. [1, 5, 8, 9]. A 2P1R Game has a parameter R that denotes the number of different answers each prover may give

on a fixed question. The PCP Theorem [7, 4, 3] combined with Raz's Parallel Repetition Theorem [14] gives[1]:

Theorem 1. *There exists an absolute constant $\gamma > 0$ such that for all large constant R, it is NP-hard to distinguish whether the value of a 2P1R Game with R answers is 1 (called completeness parameter) or at most $\frac{1}{R^\gamma}$ (called the soundness parameter).*

In this paper, we investigate the trade-off between the number of answers R and the soundness parameter. Given the central nature of 2P1R Games, we believe this is a natural pursuit. It is easy to see that if the completeness is (close to) 1, then the soundness must be at least $\Omega(\frac{1}{R})$, since the provers may give a random answer and succeed with probability $\Omega(\frac{1}{R})$. The exponent γ in the above theorem is unspecified in Raz's paper (and the subsequent works of Holenstein [10] and Rao [13]) and even if one were to compute it, it would presumably be very tiny.[2] The main result in this paper is that the above theorem holds essentially with $\gamma = \frac{1}{6}$, albeit with imperfect completeness.

Theorem 2. (Main Theorem) *For any fixed prime $q \geq 5$ and constant $\zeta > 0$, it is NP-hard to distinguish whether a 2P1R Game with $R = q^6$ answers has value at least $1 - \zeta$ or at most $\frac{4}{q}$.*

The exponent γ does play a role in some inapproximability results. For instance, Arora et al [2] show that the Quadratic Programming Problem is inapproximable within factor $(\log n)^\gamma$. This is the problem of maximizing a quadratic form $\sum_{i,j=1}^{n} a_{ij} x_i x_j$ over all vectors $\|x\|_\infty \leq 1$ and known to be approximable within factor $O(\log n)$ [11, 12, 6] (the diagonal entries of the quadratic form are assumed to be zero; the problem becomes rather meaningless otherwise). Using Theorem 2 with super-constant setting of parameter q, we obtain the following result. In fact this application was our original motivation.

Theorem 3. *Unless $NP \subseteq DTIME(2^{poly(\log n)})$, no polynomial time algorithm can approximate the Quadratic Programming Problem within factor $(\log n)^{1/6-o(1)}$.*

One technical contribution of the paper, perhaps more interesting for future research, is an essentially black-box method to translate a *codeword test* for Hadamard code (i.e. a linearity test) to a *consistency test*, leading to a full PCP construction.

Acknowledgement. Research supported by NSF CAREER grant CCF-0833228, NSF Expeditions grant CCF-0832795, NSF Waterman Award and BSF grant 2008059.

[1] The result holds for games with the *projection* property. In this paper, all games considered are projection games. For a projection game, the number of answers for the two provers may be different; R denotes the larger of the two numbers.

[2] If the value of a game is $1-\alpha$, then the value of the k-wise repeated game is at most $(1-\alpha^p)^{ck}$ for some absolute constants c and p. We have improvements $p = 32, 3$ and for projection games $p = 2$ from [14, 10, 13] respectively. However, c still remains unspecified and hence the exponent γ remains unspecified in Theorem 1.

References

[1] S. Arora, L. Babai, J. Stern, and E. Sweedyk. The hardness of approximate optima in lattices, codes and systems of linear equations. *Journal of Computer and Systems Sciences*, 54:317–331, 1997.

[2] S. Arora, E. Berger, E. Hazan, G. Kindler, and M. Safra. On non-approximability for quadratic programs. In *Proc. Annual IEEE Symposium on Foundations of Computer Science*, pages 206–215, 2005.

[3] S. Arora, C. Lund, R. Motawani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[4] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[5] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability. *Electronic Colloquium on Computational Complexity, Technical Report TR95-024*, 1995.

[6] M. Charikar and A. Wirth. Maximizing quadratic programs: Extending Grothendieck's inequality. In *Proc. Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60, 2004.

[7] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[8] J. Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[9] J. Hastad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[10] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proc. ACM Symposium on the Theory of Computing*, pages 411–419, 2007.

[11] A. Megretski. Relaxation of quadratic programs in operator theory and system analysis. *In Systems, Approximation, Singular Integral Operators, and Related Topics (Bordeaux, 2000)*, pages 365–392, 2001.

[12] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.

[13] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proc. ACM Symposium on the Theory of Computing*, pages 1–10, 2008.

[14] R. Raz. A parallel repetition theorem. *SIAM J. of Computing*, 27(3):763–803, 1998.

Spectral Sparsification

Nikhil Srivastava

(joint work with Joshua Batson and Daniel Spielman)

Suppose G is a graph. Is there a graph H which approximates G and has very few edges? The answer of course depends on what we mean by approximates and very few; we show that if we consider the space of *weighted undirected graphs* the answer is always yes in a certain very strong sense.

We consider the spectral notion of approximation introduced by Spielman and Teng [6]. To define it, we recall that the *Laplacian matrix* of a weighted graph $G = (V, E, w)$ (where $w_{ij} \geq 0$ are edge weights) is given by

$$L_G = \sum_{ij \in E} w_{ij}(\delta_i - \delta_j)(\delta_i - \delta_j)^T = \sum_{ij \in E} w_{ij} b_{ij} b_{ij}^T, \tag{1}$$

where δ_i is the canonical basis vector with a 1 in position i and zeros elsewhere, and $b_{ij} = (\delta_i - \delta_j)$ to ease notation. We now say that H is a *spectral sparsifier* for

G if:

$$x^T L_G x \le x^T L_H x \le \kappa \cdot x^T L_H x, \qquad \forall x \in \mathbf{R}^n, \tag{2}$$

where κ is some (constant) approximation factor. The above guarantee may be written as

$$L_G \preceq L_H \preceq \kappa \cdot L_G,$$

where $A \preceq B$ means that $B - A$ is positive semidefinite. It is a natural and useful notion for the following reasons:

(1) The best κ for which (2) holds is actually the relative condition number of L_G and L_H, a well-established measure of distance in numerical linear algebra. It implies that one can solve systems of linear equations in L_G by solving systems of linear equations in L_H (this known as *preconditioning*). Spielman and Teng considered (2) precisely for this purpose; in particular, using it they devised a very fast algorithm which solves systems $L_G x = b$ in time nearly linear in its number of edges.

(2) Substituting $x \in \{0,1\}^V$, we find that (2) implies as a special case that for each cut $S \subset V$, the total weight of edges leaving S is approximately equal in G and in H. This weaker notion of cut approximation was introduced earlier by Benczur and Karger [1], who showed that for all G one can efficiently find H with $O(n \log n)$ edges satisfying (2) for $x \in \{0,1\}^V$.

(3) The Courant-Fischer theorem tells us that the quadratic form $x^T L_G x$ determines all of the eigenvalues of L_G, which in turn reflect many combinatorial properties of G and random walks on G by results in spectral graph theory.

Our main result [2] is that every G has a spectral sparsifier H with $N = O(n)$ edges.

Theorem 1. *Suppose $G = (V, E, w)$ is an undirected weighted graph on n vertices. Then for every $N > n$, there is a graph H with N edges which satisfies:*

$$L_G \preceq L_H \preceq \frac{(1 + \sqrt{n/N})^2}{(1 - \sqrt{n/N})^2} \cdot L_G.$$

Theorem 1 improves [1] and [5], which obtained $N = O(n \log n)$ for cut and spectral approximation respectively. The Marchenko-Pastur type dependence of the approximation quality on N is only a factor of two worse than that achieved by the celebrated *Ramanujan graphs* [3], which are the best possible approximations of the complete graph $G = K_n$.

The theorem is actually a special case of a more general result regarding sparse approximations of sums of outer products.

Theorem 2. *Suppose $v_1, \ldots v_m \in \mathbf{R}^n$ are vectors and $N \ge 1$. Then there are nonnegative weights $w_i \ge 0$, at most N of which are nonzero, for which*

$$\left(1 - \sqrt{n/N}\right)^2 \sum_i v_i v_i^T \preceq \sum_i w_i v_i v_i^T \preceq \left(1 + \sqrt{n/N}\right)^2 \sum_i v_i v_i^T.$$

Moreover, the numbers w_i can be computed in deterministic $O(mn^2 N)$ time.

The above theorem gives a constant factor approximation when $N = O(n)$; we note that a $N = O(n \log n)$ can be obtained by the well-known result of M. Rudelson [4].

The proof of Theorem 2 is iterative and involves choosing the weights w_i one by one. In particular, we construct a series of matrices $A_0, A_1, \dots A_N$ where $A_0 = 0$ and $A_k = A_{k-1} + t_k z_k z_k^T$ where z_k is some vector from our collection $\{v_i\}$ and $t_k \geq 0$ is some increment of its weight. The spectra of the A_k are controlled by evaluating the Stieltjes transform at pre-determined, steadily increasing *real* numbers $u_1, \dots u_N$ and $\ell_1, \dots \ell_N$ and choosing the t_k, z_k so that the following properties are maintained:

(1) $\ell_k I \prec A_k \prec u_k I$.
(2) $\operatorname{tr}(u_k I - A_k)^{-1} = \operatorname{tr}(u_k I - A_{k-1} - t_k z_k z_k^T)^{-1}$
$\leq \operatorname{tr}(u_{k-1} I - A_{k-1})^{-1}$.
(3) $\operatorname{tr}(A_k - \ell_k I)^{-1} = \operatorname{tr}(A_{k-1} + t_k z_k z_k^T - \ell_k I)^{-1} \leq \operatorname{tr}(A_{k-1} - \ell_{k-1} I)^{-1}$.

This is mathematically tractable because the updated Stieltjes transforms as in (2) and (3) can be explicitly computed using the Sherman-Morisson formula for rank-one updates of the inverse of a matrix. We show that a suitable t_k, z_k must always exist by an averaging argument which relies in (2) and (3) as an induction hypothesis. Setting parameters appropriately we can obtain ℓ_N and u_N so that (1) implies the bound in Theorem 2.

References

[1] András A. Benczúr and David R. Karger. Approximating s-t minimum cuts in O(n2) time. In *STOC '96*, pages 47–55, New York, NY, USA, 1996. ACM.

[2] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. In *STOC '09*, pages 255–262, New York, NY, USA, 2009. ACM.

[3] Lubotzky, A., Phillips, R., and Sarnak, P. Ramanujan graphs. *Combinatorica 8*, 3 (1988), 261–277.

[4] Mark Rudelson. Contact points of convex bodies. *Israel J. Math*, 101(1):92–124, 1997.

[5] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *STOC '08*, pages 563–568, New York, NY, USA, 2008. ACM.

[6] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC '04*, pages 81–90, New York, NY, USA, 2004. ACM.

Limit theorems for spectrum of products of large random matrices

Alexander Tikhomirov

(joint work with Friedrich Götze, Nikita Alexeev)

Let for some fixed $m \geq 1$ given $n = p_0 \leq p_1 \leq \dots \leq p_m$. Let $\mathbf{X}^{(\nu)} = \frac{1}{\sqrt{p_{\nu-1}}}(X_{jk}^{(\nu)})_{1 \leq j \leq p_{\nu-1},\, 1 \leq k \leq p_\nu}$ be random matrices with mutually independent random entries $X_{jk}^{(\nu)}$ and $\mathbb{E}\, X_{jk} = 0$ and $\mathbb{E}\,|X_{jk}|^2 = 1$. Consider matrix $\mathbf{W} = \prod_{\nu=1}^{m} \mathbf{X}^{(\nu)}$. Let $\mathbf{\Sigma} = \mathbf{W}\mathbf{W}^*$. Denote by $\mathcal{F}_n(x)$ empirical spectral distribution

of matrix $\mathbf{\Sigma}$ and $F_n(x) = \mathbb{E}\,\mathcal{F}_n(x)$ Assuming that $\lim_{n\to\infty} \frac{p_\nu}{n} = y_\nu$ is shown that $F_n(x)$ converge to the limit distribution $G(x)$ with Stieltjes transform $S(z)$ satisfying the equation

$$1 + zS(z) - S(z)\prod_{\nu=1}^{m}(1 - y_\nu - y_\nu zS(z)) = 0.$$

In the case $y_\nu = 1$, for $\nu = 1, \ldots, m$, the moments of distribution $G(x)$ are Fuss–Catalan numbers with parameter m, $M_m(p) = \binom{mp+p}{p}$.

Let $y_1 = \cdots = y_m = 1$. Denote by μ_n empirical spectral measure of matrix $\mathbf{W}$. It is shown that μ_n converge to the distribution on the unit disc on the complex plane with density $p(x, y) = \frac{1}{\pi m(x^2+y^2)^{\frac{m-1}{m}}}$ under assumptions that the entries $X_{jk}^{(\nu)}$ are i.i.d and have the finite second moment. Furthermore, we shall discuss the limit distributions of spectrum of products of rectangular random matrices. under minimal moment assumption is shown that the Stieltjes transform $S(z, \alpha)$ of spectral distribution of symmetrized shifted matrix $\begin{bmatrix} \mathbf{0} & \mathbf{W} - z\mathbf{I} \\ \mathbf{W}^* - \overline{z}\mathbf{I} & \mathbf{O} \end{bmatrix}$ satisfies the system of equations

$$(1) \qquad \begin{cases} 1 + wS(z,\alpha) + S^2(z,\alpha)\prod_{\nu=1}^{m-1}(1 - y_\nu - y_\nu wS(z,\alpha)) = 0 \\ (w-\alpha)^2 S(z,\alpha) + (w-\alpha) - |z|^2 S(z,\alpha) = 0 \end{cases}$$

In particular, equations (1) imply that for $m = 2$ the expectation of empirical distribution of eigenvalues of matrix $\mathbf{W}$ has a limit with density

$$p(u, v) = \frac{1}{\pi\sqrt{(1 - y_1)^2 + 4y_1(u^2 + v^2)}}\mathbb{I}\{u^2 + v^2\}$$

References

[1] N. Alexeev, F. Götze, A. N. Tikhomirov, *On the asymptotic distribution of singular values of power of random matrices*, Lithuanian mathematical journal, **50** (2010), 121–132.

[2] N. Alexeev, F. Götze, A. N. Tikhomirov, *On the singular spectrum of powers and products of random matrices*, Doklady mathematics, **82** (2010), 505–507.

[3] N. Alexeev, F. Götze, A. N. Tikhomirov, *On the asymptotic distribution of singular values of products of large rectangular random matrices*, Preprint. arXiv:1012.258.

[4] F. Götze, A. N. Tikhomirov, *On the Asymptotic Spectrum of Products of Independent Random Matrices*, Preprint. arXiv:1012.2710

Limiting Empirical Singular Value Distribution of DFT Submatrices

Brendan Farrell

This work addresses the singular values of a random submatrix of the discrete Fourier transform (DFT) matrix. The DFT matrix is unitary and in dimension n has entries

$$F_{jk} = \frac{1}{\sqrt{n}}e^{-2\pi i(j-1)(k-1)/n}.$$

For $T_n, \Omega_n \subset \{1, ..., n\}$ we define $F_{\Omega_n T_n}$ to be the matrix obtained from the $n \times n$ DFT matrix by keeping only rows with index in Ω_n and columns with index in T_n. We include each index in Ω_n independently with probability $(1-q)$ and, also independently, in T_n with probability $(1-p)$.

Three questions are: in dependence on p and q, how does the distribution of all the singular values behave, how does the largest singular value behave and how does the smallest singular value behave? We recently provided an answer to the first question, the limiting distribution [4]. Two natural conjectures follow from this distribution for the other two questions. We state this result and then give a short discussion.

Theorem For $i = 1, ..., n$ let i be contained in Ω_n independently with probability $(1-q)$ and, also independently, let i be included in T_n with probability $(1-p)$. Then the empirical distributions of the $\min(|T_n|, |\Omega_n|)$ largest eigenvalues of $F_{\Omega_n T_n} F^*_{\Omega_n T_n}$ converges almost surely to

$$f_{p,q}(x) = \frac{\sqrt{(1-\frac{r_-}{x})(\frac{r_+}{x}-1)}}{2\pi(1-x)(1-\max(p,q))} \cdot I_{(r_-,r_+)}(x) + \frac{\max(0, 1-(p+q))}{1-\max(p,q)} \cdot \delta(x-1) \tag{1}$$

where

$$r_- = (\sqrt{p(1-q)} - \sqrt{q(1-p)})^2$$

and

$$r_+ = (\sqrt{p(1-q)} + \sqrt{q(1-p)})^2.$$

Our proof relies on the Stieltjes transform; this general approach was developed by Marčenko and Pastur to determine the limiting distribution for Wishart matrices [5]. Using resolvent identities, many matrix expansions and basic large deviations results, we obtain an implicit equation in terms of the Stieltjes transform of the limiting distribution. The equation is, quite fortunately, easily solvable, thus allowing us to recover the Stieltjes transform and apply the inversion formula. Plots of the distribution for several parameter pairs are given in the figure.

The two conjectures corresponding to the two questions posed above are:

Conjecture 1: If $p + q > 1$, then for $\epsilon > 0$

$$\|F_{\Omega_n T_n}\| < \sqrt{p(1-q)} + \sqrt{q(1-p)} + \epsilon \tag{2}$$

with high probability.

Conjecture 2: If $p \neq q$, then for $\epsilon > 0$,

$$\|(F_{\Omega_n T_n})^{-1}\| < \left(|\sqrt{p(1-q)} - \sqrt{q(1-p)}| - \epsilon\right)^{-1}. \tag{3}$$

with high probability.

There have been many works investigating conditions on the cardinalities of Ω_n and T_n such that the largest and smallest singular values of $F_{\Omega_n T_n}$ are bounded away from 1 and 0. Most of these take a probabilistic approach [1, 2, 6, 8], though [3, 7], for example, present deterministic results. The threshold given by the condition $p + q > 1$ can be seen in the following theorem of Tao.

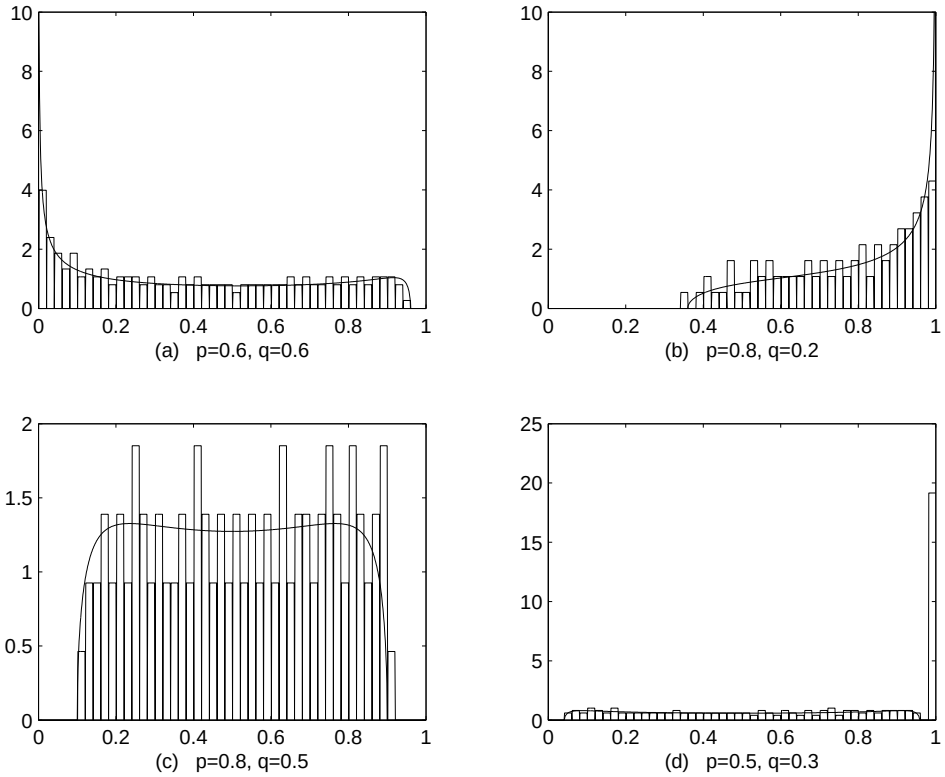

FIGURE 1. Empirical eigenvalue distribution for *one* realization plotted against continuous part of asymptotic distribution. In each case, the original DFT matrix had dimensions 500×500.

Theorem (Tao, [7]) If n is prime and $|\Omega_n| + |T_n| < n$, then $\|F_{\Omega_n T_n}\| < 1$.

Due to the almost sure convergence in our theorem and the existence of infinitely many primes, for any $\epsilon > 0$, there exists a prime n and subsets Ω_n and T_n satisfying $\frac{|\Omega_n|}{n} + \frac{|T_n|}{n} < 1$ and $\|F_{\Omega_n T_n}\| \geq 1 - \epsilon$. Conjecture 1 is needed to bound $\|F_{\Omega_n T_n}\|$ away from 1 for most sets when $\frac{|\Omega_n|}{n} + \frac{|T_n|}{n}$ is held away from 1. This conjecture would imply that the behavior seen in Tao's result holds for almost all subsets of appropriate cardinality for any integer n and also give a sharp bound on $\|F_{\Omega_n T_n}\|$ for most such sets. We view such a result as a general uncertainty principle, and it is the focus of our current work.

REFERENCES

[1] E.J. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math. **6** (2006), no. 2, 227–254.

[2] E.J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier measurements*, IEEE Trans. on Info. Theory **52** (Feb. 2006), no. 2, 489–509.

[3] D.L. Donoho and P.B. Stark, *Uncertainty principles and signal recovery*, SIAM J. of Appl. Math. **49** (June 1989), no. 3, 906–931.

[4] B. Farrell, *Limiting empirical singular value distribution of restrictions of discrete Fourier matrices*, Journal of Fourier Analysis and Applications, to appear.

[5] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536.
[6] M. Rudelson and R. Vershynin, *Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements*, Information Sciences and Systems, 2006 40th Annual Conference on, march 2006, pp. 207 –212.
[7] T. Tao, *An uncertainty principle for cyclic groups of prime order*, Math. Res. Lett. **12** (2005), no. 1, 121–127.
[8] J. A. Tropp, *On the linear independence of spikes and sines*, Journal of Fourier Analysis and Applications **14** (2008), no. 5, 838–858.

Log gases and Tracy-Widom laws

Brian Rider
(joint work with Manjunath Krishanpur, José A. Ramírez, Bálint Virág)

Consider n real particles $\lambda_1, \lambda_2, \ldots, \lambda_n$ distributed according to the law with density proportional to

$$\prod_{j<k} |\lambda_j - \lambda_k|^\beta \, e^{-\frac{\beta}{4}\sum_{k=1}^n \lambda_k^2}. \tag{1}$$

When $\beta = 1$ or $\beta = 2$, these particles may be realized as the eigenvalues of the classical Gaussian Orthogonal or Unitary Ensembles (G{O/U}E), random real symmetric or complex hermitian matrices comprised of independent Gaussians of mean zero and mean-square one. In both cases these are integrable ensembles: all correlation functions may be expressed in determinantal or pfaffian form, built from the standard Hermite polynomials. These explicit formulas pave the way for various detailed local limit theorems, such as the Tracy-Widom laws for the fluctuations of the largest eigenvalues described in terms of a special solution to Painlevé II ODE.

When $\beta > 0$ is not one or two (or four, which corresponds to the so-called Gaussian Symplectic Ensemble and is a pfaffian process) the laws (1) are no longer integrable in the same way, but remain interesting as they possess the interpretation of (one-dimensional caricatures) couloumb (or "log") gases. Here $\sqrt{\beta}$ is viewed as the common charge, and a special choice is made in terms of the quadratic potential.

In [1] the authors establish a limit theorem for the largest point under any of these "β-Hermite" ensembles. It is proved that $n^{1/6}(\lambda_{\max} - \sqrt{n})$ converges to a random variable TW_β defined by (the equality in law):

$$TW_\beta = \sup_{f \in L} \left\{ \frac{2}{\sqrt{\beta}} \int_0^\infty f^2(x) db(x) - \int_0^\infty \left[(f'(x))^2 + x f^2(x) \right] dx \right\},$$

in which $x \mapsto b(x)$ is a standard Brownian motion and L is the space of functions f which vanish at the origin and satisfy $\int_0^\infty f^2(x)dx = 1$, $\int_0^\infty [(f'(x))^2 + x f^2(x)]dx < \infty$. Said another way, $-TW_\beta$ is the ground state eigenvalue for the Stochastic Airy Operator, $\mathcal{H}_\beta = -\frac{d^2}{dx^2} + x + \frac{2}{\sqrt{\beta}} b'(x)$.

The proof of the above rests on a tridiagonal representation discovered for general beta by Dumitriu and Edelman. Let $g_1, g_2, \dots g_n$ be independent Gaussians with mean 0 and variance 2. Let also χ_β, $\chi_{2\beta}$, $\dots$, $\chi_{(n-1)\beta}$ be independent χ random variables of the indicated parameter. Then the eigenvalues of the random matrix

$$H_{n,\beta} = \frac{1}{\sqrt{\beta}} \begin{bmatrix} g_1 & \chi_{\beta(n-1)} & & & \\ \chi_{\beta(n-1)} & g_2 & \chi_{\beta(n-2)} & & \\ & \ddots & \ddots & \ddots & \\ & & \chi_{\beta 2} & g_{n-1} & \chi_\beta \\ & & & \chi_\beta & g_n \end{bmatrix}$$

have joint law (1). What is proved is that $n^{1/6}(\sqrt{n}I_n - H_{n,\beta})$ almost surely Γ-converges to $\mathcal{H}_\beta$. A result of this type had been conjectured by Edelman and Sutton (appropriate references are contained in the bibliography of [1]).

An allied result is available at the random matrix hard edge. Take a law of type (1), restricted to positive particles ($\lambda_k \in \mathbb{R}^+$) and with the Hermite weight $e^{-\beta\lambda^2/4}$ replaced by the Laguerre weight $\lambda^{\beta(a+1)/2-1}e^{-\beta\lambda/2}$, $a > -1$. When $\beta = 1, 2$ these are the eigenvalue laws for certain Gaussian sample covariance (Wishart) matrices. For all β, the results of [1] imply that the scaling limit of the largest eigenvalue is again TW_β . The smallest eigenvalue behaves differently. In [2] it is proved that $n^2\lambda_{\min} \to \Lambda(\beta, a)$ in which

$$\Lambda(\beta, a) = \inf_{f:f(0)=0, f\neq 0} \frac{\int_0^\infty (f'(x))^2 e^{-ax-\frac{2}{\sqrt{\beta}}b(x)}dx}{\int_0^\infty (f(x))^2 e^{-(a+1)x-\frac{2}{\sqrt{\beta}}b(x)}dx}.$$

The corresponding differential operator is (formally)

$$\mathfrak{G}_{\beta,a} = e^x \left[\frac{d^2}{dx^2} - (a + \frac{2}{\sqrt{\beta}}b'(x))\frac{d}{dx}\right],$$

a generator for a version of the Brox diffusion. What is important to know is that the hard edge and "soft" edge are knit together $a^{2/3} - a^{-4/3}\Lambda(\beta, 2a)$ converges in law to TW_β as $a \to \infty$ [2].

A new result (with M. Krisnapur and B. Virág) advertised here is the extension of the above to more general potentials. In particular, replace $e^{-\beta\lambda^2/4}$ in (1) with $e^{-\beta V(\lambda)}$ for a strictly convex, polynomial V. Then the theorem is that there are constants c_0, c_1 so that $n^{1/6}c_1(\lambda_{max} - \sqrt{n}c_0) \Rightarrow TW_\beta$.

This follows from yet another tridiagonal representation. Denote by $T = T(A, B)$ the symmetric tridiagonal matrix with $T_{i,i} = A_i$ for $i \leq n$ and $T_{i,i+1} = T_{i+1,i} = B_i$ for $i \leq n-1$. Then the nice fact is that, if (A, B) is sampled from the density

$$e^{-\,n\beta \mathrm{tr}\, \mathrm{V}(\mathrm{T}(\mathbf{A},\mathbf{B}))} \prod_{k=1}^{n-1} B_k^{\beta(n-k)-1}$$

on $(\mathbb{R}^+)^n \times \mathbb{R}^n$, the eigenvalues of $T = T(\mathbf{A}, \mathbf{B})$ have density

$$e^{-n\beta \sum_{k=1}^n V(\lambda_k)} \prod_{1 \le i < k \le n} |\lambda_i - \lambda_k|^\beta.$$

Note of course that $\operatorname{tr} T^2 = \sum (A_k^2 + 2B_k^2)$, giving back the independent Gaussians and χ's of the $H_{n,\beta}$ matrix corresponding to quadratic potentials. If V has degree greater than two, the entries of T are no longer independent, still the proof can be built along the lines of the main result of [1].

References

[1] J. Ramírez, B. Rider, B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, To appear, JAMS (2011).

[2] J. Ramírez, B. Rider, *Diffusion at the random matrix hard edge*, Comm. Math. Phys. **288** (2009), 1319–1343.

Some connections between random matrix theory and statistics

Noureddine El Karoui

Spectral properties of matrices play a central role in many methods of multivariate analysis and various related problems in applied mathematics. Properties of random matrices have therefore long been of interest to statisticians involved in multivariate analysis.

In recent years, practical problems have started involving datasets of larger and larger size. In particular, it is now common to be working with data for which the number of measurements per observation, p, is of the same order of magnitude as the number of observations, n. From an asymptotic stand point, this is radically different from the classical paradigm where p was assumed to be much smaller than n.

I have described some of the interactions between modern random matrix theory and statistics in this high-dimensional context.

Sums and Products of Free Random Variables and Applications

Friedrich Götze
(joint work with S. Bobkov, G. Chistyakov)

Let μ denote a probability measure on $\mathbb{R}$ and let $\mu^{n\boxplus}$ denote it's n-fold free additive convolution measure. Write $F_n(x) := \mu^{n\boxplus}((-\infty, \sqrt{n}x]))$ for the distribution function of the corresponding normalized sum $S_n := n^{-1/2}(X_1 + \ldots + X_n)$ of free random variables with distribution μ. Define $\beta_q(\mu) := \int |x|^q \mu(du)$ and assume that $m_1(\mu) = 0, \quad m_2(\mu) = 1$. Let $L_{q,n} := \frac{\beta_q}{n^{(q-2)/2}}$ denote the q-th Liapunov fraction. Let w denote Wigner's half-circle law with density $p_w(x) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}$.

Then the following analog of the classical Berry-Essen bound in the CLT holds.

Theorem 1. *(Chistyakov, Götze (2010))*
For a sequence $\lim_n \eta_q(n) = 0$ *we have*

$$\begin{aligned}\sup_x|\mu_n^{n\boxplus}(x) - w(x)| &\le c\Big(\eta_q(n)\frac{\beta_q}{n^{(q-2)/2}} + n^{-1}\Big) \ll L_{q,n}, \quad \text{if } \beta_q < \infty,\ 2 < q < 3,\\ &\le c\frac{\beta_3}{n^{1/2}} = cL_{3,n}, \qquad \text{if } \beta_3 < \infty.\end{aligned}$$

Let $U_m(x) = U_m(\cos\theta) := \frac{\sin(m+1)\theta}{\sin\theta}$, $m = 1, 2, \dots$ denote the Chebyshev polynomials of 2nd kind. As an analog to classical Edgeworth expansions we show

Theorem 2. *(Chistyakov-Götze (2008,2010))*
Assume that $\beta_q(\mu) < \infty$, $q \ge 3$. *Then there exists* $c > 0$ *and a sequence* $\eta_{q2}(n)$ *as above such that*

$$F_n(x) = \mu_w((-\infty, x]) - \frac{1}{3}a_nU_2(x/2)p_w(x) + \rho_{n1}(x), \quad a_n := \frac{m_3}{\sqrt{n}}$$

where

$$\begin{aligned}\sup_x|\rho_{n1}(x)| &\le c\Big(\eta_{q2}(n)L_{q,n} + L_{3,n}^2 + |a_n|^{3/2}\Big), \quad \text{if} \quad \beta_q < \infty, \quad 3 \le q < 4,\\ &\le c\left(L_{4,n} + |a_n|^{3/2}\right) = \mathcal{O}(n^{-1}), \qquad \text{if} \quad \beta_q < \infty, \quad q \ge 4.\end{aligned}$$

Writing $b_n := (m_4 - m_3^2 - 1)$, *the approximation of 3rd order is given by*

$$\begin{aligned}F_n(x + a_n) &= \mu_w((-\infty, x])\\ &+ \Big(-\frac{a_n^2}{2}U_1\big(\frac{x}{2}\big) + \frac{a_n}{3}(3 - U_2\big(\frac{x}{2}\big)) - \frac{b_n - a_n^2 - 1/n}{4}U_3\big(\frac{x}{2}\big)\Big)p_w(x)\\ &+ \rho_{n2}(x),\end{aligned}$$

where

$$|\rho_{n2}(x)| \le c\begin{cases}\eta_{q3}(n)L_{qn} + L_{4n}^{3/2} & \text{if} \quad \beta_q < \infty,\ 4 \le q < 5\\ L_{5n} & \text{if} \quad \beta_q < \infty,\ q \ge 5,\end{cases}$$

for all $x \in \mathbb{R}$, $n \ge m_4$.

For $m_3(\mu) = 0$ this approximation simplifies to

$$F_n(x) = \mu_w((-\infty, x]) - \frac{m_4(\mu) - 2}{4n}U_3\big(\frac{x}{2}\big)p_w(x) + \rho_{n2}(x).$$

It is known that for $n \ge n_0$ and μ nontrivial, $F_n'(x) =: p_n(x)$ exists and we may apply our results to get corresponding expansions for the densities p_n which converge to p_w. Using Voiculescu's free entropy

$$\chi(\nu) = \int\int_{\mathbb{R}\times\mathbb{R}} \log|x - y|\,\nu(dx)\nu(dy) + \chi_0, \quad \chi_0 = \frac{3}{4} + \frac{1}{2}\log 2\pi,$$

which is maximized by Wigner's w in the class of centered and normalized measures ν we may study the distance of $\mu_n = \mathcal{D}(S_n)$ to w via the (relative) entropic distance

$$D(\nu) := \chi(w) - \chi(\nu).$$

In particular we obtain the following quantitative results.

Theorem 3. *(Chistyakov-Götze (2011)) Assume* $\mathrm{supp}(\mu) \subset [-\mathrm{L}, \mathrm{L}]$ *and* $m_1(\mu) = 0$, $m_2(\mu) = 1$. *Then for all* $n \geq n_0$ *and some* $|\theta| \leq 1$,

$$\begin{aligned} \int\limits_{\mathbb{R}} |p_n(x) - p_w(x)|\, dx &= \frac{2|m_3|}{\pi\sqrt{n}} + \theta\frac{c(\mu)}{n}, \qquad \text{if} \quad m_3 \neq 0, \\ &= \tilde{c}_0 \frac{|m_4 - 2|}{n} + \theta\frac{c(\mu)}{n^{3/2}}, \quad \text{if} \quad m_3 = 0, \\ \text{Furthermore,} \quad D(\mu_n) &= \frac{m_3^2}{6n} + \theta\frac{c(\mu)}{n^{3/2}}. \end{aligned}$$

Let X denote a random variable with mean a, variance $\sigma^2 > 0$ and density p. Let $\varphi_{a,\sigma}(x)$ denote the corresponding normal density. Using the fact that the classical (continuous) entropy is minimal for the normal law in the class of all random variables X as above, the entropic (or Kullback-Leibler) distance to the class of normal distributions is defined by

$$D(X) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{\varphi_{a,\sigma}(x)}\, dx \ \in [0, +\infty].$$

Improving upon previous results, like [1], [2] and [3], requiring much stronger assumptions, we have

Theorem 4. *(Bobkov-Chistyakov-Götze (2011))*
Assume that X, X_j, $j \in \mathbb{N}$ *are i.i.d.* $\beta := \mathbf{E}X^s < \infty, s \geq 2$. *If* $D(S_{n_0}) < \infty$, $n_0 \geq 1$ *then for* $s_0 := [(s-2)/2)] \in \mathbb{N}$ *we have*

$$D(S_n) = c_1\, n^{-1} + c_2\, n^{-2} + \ldots + c_{s_0}\, n^{s_0} + o\big((n \log n)^{-(s-2)/2}\big),$$

where c_j *depend on the cumulants of* X. *For instance* $c_1 = \frac{1}{12}\,\gamma_3^2$.

The implied sequence $o(1)$ in the error term depends on the tail of the sth moment as well as n_0 and $D(S_{n_0})$ only. The result extends to multidimensional random vectors $X, X_j \in \mathbb{R}^d$ when s is an integer and to stable laws as well, see [4] and [5]. An explicit upper bound of order $O(n^{-1})$ of Berry-Esseen type has been shown as well, see [6].

References

[1] Artstein, S., Ball, K.M., Barthe, F., and Naor, A. On the rate of convergence in the entropic central limit theorem. Probab. Theory Related Fields, 129 (2004), no. 3, 381-390.

[2] Barron, A.R. Entropy and the central limit theorem. Ann. Probab. 14 (1986), no. 1, 336-342.

[3] Barron, A.R., and Johnson, O. Fisher information inequalities and the central limit theorem. Probab. Theory Related Fields, 129 (2004), no. 3, 391-409.

[4] Bobkov, S.G., Chistyakov, G.P and Götze, F. Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. arXiv:1104.3994v1 [math.PR], (2011).

[5] Bobkov, S.G., Chistyakov, G.P and Götze, F. Convergence to stable laws in relative entropy. arXiv: 1104.4360v1 [math.PR], (2011).

[6] Bobkov, S.G., Chistyakov, G.P and Götze, F. Berry-Esseen bounds in the entropic central limit theorem. arXiv: 1105.4119v1 [math.PR], (2011).

Geometry of log-concave vectors and log-concave Ensembles of random matrices

Rafał Latała
(joint work with Radosław Adamczak, Alexander E. Litvak, Alain Pajor, Nicole Tomczak-Jaegermann)

An N dimensional random vector is called log-concave if it has a log-concave distribution, i.e. for any compact nonempty sets $A, B \subset \mathbb{R}^N$ and $\lambda \in (0,1)$,

$$\Pr(X \in \lambda A + (1-\lambda)B) \geq \Pr(X \in A)^{\lambda} \Pr(X \in B)^{1-\lambda},$$

where $\lambda A + (1-\lambda)B = \{\lambda x + (1-\lambda)y \colon x \in A, y \in B\}$. By the result of Borell [4] a vector X with full dimensional support is log-concave if and only if it has a density of the form e^{-f}, where $f \colon \mathbb{R}^N \to (-\infty, \infty]$ is a convex function. Log-concave vectors are frequently studied in convex geometry, since by the Brunn-Minkowski inequality uniform distributions on convex sets as well as their lower dimensional marginals are log-concave.

An N-dimensional random vector $X = (X(1), \ldots, X(N))$ is isotropic if $\mathbb{E}X(i) = 0$ and $\mathrm{Cov}(X(i), X(j)) = \delta_{i,j}$ for all $i, j \leq N$. Equivalently, a random vector in $\mathbb{R}^N$ with mean zero is isotropic if $\mathbb{E}\langle t, X\rangle^2 = |t|^2$ for any $t \in \mathbb{R}^N$. For any nondegenerate log-concave vector X there exists an affine transformation T such that TX is isotropic.

In recent years there were derived numerous important properties of log-concave vectors. One of such results is the Paouris concentration of mass [6] that states that for any isotropic log-concave vector X in $\mathbb{R}^N$,

$$\mathbb{P}(|X| \geq Ct\sqrt{N}) \leq \exp(-t\sqrt{N}) \quad \text{for } t \geq 1, \tag{1}$$

where $|x| = (\sum_{i=1}^N |x_i|^2)^{1/2}$ denotes the Euclidean norm of x and C denotes an absolute constant.

It is natural to ask if one may extend the Paouris result to l_r norms, that is derive upper bounds for $\mathbb{P}(\|X\|_r \geq t)$, where $\|x\|_r = (\sum_{i=1}^N |x_i|^r)^{1/r}$. For $1 \leq r \leq N$ by the Hölder inequality $\|x\|_r \leq N^{1/2-1/r}|x|$, therefore (1) gives

$$\mathbb{P}(\|X\|_r \geq CtN^{1/r}) \leq \exp(-t\sqrt{N}) \quad \text{for } t \geq 1, r \in [1,2].$$

It is not hard to construct examples showing that this bound is optimal [5].

For $r > 2$ the situation is less trivial. If coordinates of X are independent and have symmetric exponential distribution with variance 1, then for $N \geq C^r$, $\mathrm{Med}(\|X\|_r) \geq rN^{1/r}/C$ and for such vectors

$$\mathbb{P}\Big(\|X\|_r \geq \frac{1}{C}trN^{1/r}\Big) \geq \exp(-trN^{1/r}) \quad \text{for } t \geq 1, r \geq 2. \tag{2}$$

To get the reverse inequality we write

$$\|X\|_r = \Big(\sum_{i=1}^N |X(i)|^r\Big)^{1/r} = \Big(\sum_{i=1}^N |X^*(i)|^r\Big)^{1/r} \leq \Big(\sum_{k=0}^{s-1} 2^k |X^*(2^k)|^r\Big)^{1/r},$$

where $s = \lfloor \log_2 n \rfloor$ and $X^*(1) \geq X^*(2) \geq \ldots \geq X^*(N)$ denote the nonincreasing rearrangement of $|X_1|, \ldots, |X_n|$ (*order statistics of vectors* X). So to derive concentration inequalities for l_r norms of X we may look at the tail inequalities for $X^*(l)$, $1 \leq l \leq n$.

Theorem 1. *Let X be an N-dimensional log-concave isotropic vector. Then*

$$\mathbb{P}(X^*(k) \geq t) \leq \exp\Big(-\frac{1}{C}\sqrt{k}t\Big) \quad \textit{for } t \geq C\log\Big(\frac{eN}{k}\Big).$$

Threshold $\log(eN/k)$ cannot be improved as shows the example of the isotropic vector X with the product symmetric exponential distribution.

Theorem 1 yields the following reverse to the inequality (2).

Theorem 2. *For any $\delta > 0$ there exists a constant $C_1(\delta) \leq C\delta^{-1/2}$ such that for any N-dimensional log-concave isotropic vector X,*

$$\mathbb{P}(\|X\|_r \geq C_1(\delta) r N^{1/r}) \leq \exp(-trN^{1/r}) \quad \textit{for } t \geq 1, r \geq 2+\delta.$$

We believe that the constant C_1 in Theorem 2 should not depend on δ.

Tail estimates for order statistics yield also the following uniform version of the Paouris result (1),

$$(3) \quad \forall_{t\geq 1}\ \mathbb{P}\Big(\sup_{|I|=m} |P_I X| \geq Ct\sqrt{m}\log\Big(\frac{eN}{m}\Big)\Big) \leq \exp\Big(-\frac{t\sqrt{m}}{\sqrt{\log(em)}}\log\Big(\frac{eN}{m}\Big)\Big),$$

where for $\emptyset \neq I \subset \{1, \ldots, N\}$ and $x \in \mathbb{R}^N$, $P_I x$ denotes the canonical projection of x onto $\{y \in \mathbb{R}^N\colon\ \mathrm{supp}(y) \subset I\}$.

Estimates presented so far work for all isotropic log-concave vectors and do not take into account specific geometric properties of vectors X. To formulate more precise bounds we introduce a weak L_p-norm $\sigma_X(p)$ of a random vector X in $\mathbb{R}^N$:

$$\sigma_X(p) := \sup_{t \in S^{N-1}} (\mathbb{E}|\langle t, X\rangle|^p)^{1/p} \quad p \geq 2.$$

For isotropic log-concave vectors X, $\sigma_X(p) \leq p/\sqrt{2}$ and $\sigma_X^{-1}(t) \geq \sqrt{2}t$.

Combining few results from [6] one gets a stronger version of (1) for isotropic log-concave vectors X in $\mathbb{R}^N$,

$$\mathbb{P}(|X| \geq Ct\sqrt{N}) \leq \exp(-\sigma_X^{-1}(t\sqrt{N})) \quad \text{for } t \geq 1.$$

One may also show improved tail estimates for order statistics.

Theorem 3. *For any N-dimensional log-concave isotropic vector X,*

$$\mathbb{P}(X^*(l) \geq t) \leq \exp\Big(-\sigma_X^{-1}\Big(\frac{1}{C}t\sqrt{l}\Big)\Big) \quad \textit{for } t \geq C\log\Big(\frac{eN}{l}\Big).$$

Theorem 3 yields the following improvement of (3) for isotropic log-concave vectors X in $\mathbb{R}^N$ and $t \geq 1$,

$$\mathbb{P}\left(\sup_{|I|=m} |P_I X| \geq Ct\sqrt{m}\log\Big(\frac{eN}{m}\Big)\right) \leq \exp\left(-\sigma_X^{-1}\Big(\frac{t\sqrt{m}\log\Big(\frac{eN}{m}\Big)}{\sqrt{\log(em/m_0)}}\Big)\right),$$

where

$$m_0 = m_0(X,t) = \sup\Big\{k \le m\colon\ k\log\Big(\frac{eN}{k}\Big) \le \sigma_X^{-1}\Big(t\sqrt{m}\log\Big(\frac{eN}{m}\Big)\Big)\Big\}.$$

In many applications there appear sums of independent log-concave vectors. Let $X_1,\dots,X_n$ be independent isotropic log-concave vectors and $Y = \sum_{i=1}^n x_i X_i$. Then

$$\sigma_Y(p) \le C(\sqrt{p}|x| + p\|x\|_\infty) \quad \text{for } p \ge 2$$

and Theorem 3 implies in this case

$$\mathbb{P}(Y^*(l) \ge t) \le \exp\Big(-\frac{1}{C}\min\Big\{\frac{t^2 l}{|x|^2}, \frac{t\sqrt{l}}{\|x\|_\infty}\Big\}\Big) \quad \text{for } t \ge |x|\log\Big(\frac{eN}{l}\Big).$$

If we assume that $|x| \le 1$ and $\|x\|_\infty \le b \in [1/\sqrt{m}, 1]$ then for any $t \ge 1$,

$$\mathbb{P}\Big(\sup_{|I|=m}|P_I Y| \ge Ct\sqrt{m}\log\Big(\frac{eN}{m}\Big)\Big) \le \exp\left(-\frac{t\sqrt{m}\log\left(\frac{eN}{m}\right)}{b\sqrt{\log(e^2b^2m)}}\right).$$

Above estimates may be applied to derive uniform bound for norms of $k \times m$ submatrices of $n \times N$ random matrices with independent log-concave rows.

Theorem 4. *Let A be an $n \times N$ random matrix with independent log-concave isotropic rows $X_1,\dots,X_n \in \mathbb{R}^N$. For $k \le n, m \le N$ and $t \ge 1$ we have*

$$\mathbb{P}\left(\sup_{|I|=k,|J|=m}\|A_{|I\times J}\|_{\ell_2^m \to \ell_2^k} \ge Ct\lambda_{mk}\right) \le \exp\Big(-\frac{t\lambda_{mk}}{\sqrt{\log(3m)}}\Big),$$

where

$$\lambda_{mk} = \sqrt{\log\log(3m)}\sqrt{m}\log\Big(\frac{e\max(N,n)}{m}\Big) + \sqrt{k}\log\Big(\frac{en}{k}\Big).$$

Case $m = N$ was investigated in [1]. Theorem 4 plays a crucial role in the study of Restricted Isoperimetry Properties of random matrices with independent log-concave rows [2, 3].

References

[1] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles*, Journal of AMS **234** (2010), 535–561.

[2] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Geometry of log-concave Ensembles of random matrices and approximate reconstruction*, preprint, http://arxiv.org/abs/1103.0201.

[3] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Tail estimates for norms of sums of log-concave random vectors*, in preparation.

[4] C. Borell, *Convex measures on locally convex spaces*, Ark. Math. **12** (1974), 239–252.

[5] R. Latała, *Order statistics and concentration of l_r norms for log-concave vectors*, J. Funct. Anal. **261** (2011), 681–696.

[6] G. Paouris, *Concentration of mass on convex bodies*, Geom. Funct. Anal. **16** (2006), 1021–1049.

Approximate reconstruction and sharp bounds for singular values for matrices in log-concave Ensemble

Nicole Tomczak-Jaegermann

(joint work with Radosław Adamczak, RafałLatała, Alexander Litvak, Alain Pajor)

1. Overview

The talk is based on a joint research by Radosław Adamczak, Rafał Latała, Alexander Litvak, Alain Pajor and Nicole Tomczak-Jaegermann.

We study the Restricted Isometry Property and approximate reconstruction problems for random matrices Γ with independent isotropic rows. This is shown to depend on the behaviour of parameter $\Gamma_{k,m}$ that controls uniformly the operator norm of sub-matrices of Γ with k rows and m columns. A similar argument combined with an approximation trick shows that a behaviour of a simpler parameter Γ_k, in the case of uniformly sub-exponential rows, is responsible for sharp bounds for the extremal (non-zero) singular values of $\Gamma^*\Gamma$, which are quantitative counterparts of Bai-Yin theorem known for random matrices with i.i.d. entries.

The behaviour of $\Gamma_{k,m}$ has been established in a very recent paper [3] by Adamczak, Latała, Litvak, Pajor and Tomczak-Jaegermann, while Γ_k was investigated in the earlier paper [1].

2. Main results

We present some recent results from [3] and [2]. Other related results are also discussed in the lectures of Rafał Latała and Alexander Litvak. We also send the reader to these reports as well as to above papers for all basic definitions and notations not explained in this report. Finally, we refer the reader to the above papers for many references to related results and their history.

General motivation for our recent results was coming from several sources:

- Convexity, reconstruction problems and the Restricted Isometry Property;
- point of view of Random Matrix Theory;
- Kannan-Lovász-Simonovits question.

Recall the notation connected with the Compressed Sensing and reconstruction.

Let $n, N \geq 1$ be arbitrary integers (no relation between them will be assumed). Let $T \subset \mathbb{R}^N$ and Γ be an $n \times N$ matrix.

One of problems considered in the theory of Compressed Sensing is to reconstruct any vector $x \in T$ from the data $\Gamma x \in \mathbb{R}^n$, with a fast algorithm. We clearly need some hypothesis on T and on Γ. The common hypothesis is that T consists of sparse vectors. Recall that if $m \leq N$, then $x \in \mathbb{R}^N$ is m-sparse if $|\mathrm{supp}\, x| \leq m$. We let

$$U_m = \{x \in S^{N-1} : x \text{ is } m - \text{sparse}\}.$$

For any $T \subset S^{N-1}$ we let

$$\delta_T(\Gamma) = \sup_{x \in T} \left| |\Gamma x|^2 - \mathbb{E}\,|\Gamma x|^2 \right|.$$

Denote the rows of Γ by $Y_1, \ldots, Y_n \in \mathbb{R}^N$, and assume that they are random vectors which are independent. (In reconstruction problems we look for vectors given by their n independent *measurements*, with $n \ll N$.)

Define the parameter $\Gamma_k(T)$ by

$$\Gamma_k(T)^2 = \sup_{y \in T} \sup_{\substack{I \subset \{1,\ldots,n\} \\ |I|=k}} \sum_{i \in I} |\,\langle Y_i, y\rangle\,|^2.$$

We have the following fundamental lemma:

Fundamental Lemma: *Let n, N. Let $Y_1, \ldots, Y_n \in \mathbb{R}^N$ be independent isotropic, $T \subset S^{N-1}$ finite. Let $0 < \theta < 1$ and $B \geq 1$. Then with probability at least $1 - |T| \exp\left(-3\theta^2 n/8B^2\right)$,*

$$\begin{aligned}\delta_T\left(\frac{\Gamma}{\sqrt{n}}\right) &= \sup_{y \in T} \left| \frac{1}{n} \sum_{i=1}^n (|\langle Y_i, y\rangle|^2 - \mathbb{E}\,|\langle Y_i, y\rangle|^2) \right| \\ &\leq \theta + \frac{1}{n}\left(\Gamma_k(T)^2 + \mathbb{E}\,\Gamma_k(T)^2\right).\end{aligned}$$

where $k \leq n$ is the largest integer satisfying $k \leq (\Gamma_k(T)/B)^2$.

For $m \leq N$ we let

$$\delta_m(\Gamma) = \delta_{U_m}(\Gamma).$$

This is so-called the RIP parameter first introduced by E.J. Candés and T. Tao in [5] and extensively studied in many papers that followed. (We also mention related results of D. Donoho.) The interest in this parameter stems from the fact that if $\delta_{2m}(\Gamma)$ is appropriately small then every m-sparse vector x can be reconstructed from Γx by the ℓ_1-minimization method.

For $k \leq n$ we write $\Gamma_{k,m} = \Gamma_k(U_m)$. In other words, $\Gamma_{k,m}$ is equal to the maximal operator norm in a Hilbert space of all sub-matrices of Γ with k rows and m columns.

We easily get the following Corollary for RIP:

Corollary: *Let Y_i, Γ, $0 < \theta < 1$ and $B \geq 1$, as before. Assume that $m \leq N$ satisfies*

$$m \log \frac{7N}{m} \leq \frac{3\theta^2 n}{16B^2}.$$

Then with probability at least $1 - \exp\left(-\frac{3\theta^2 n}{16B^2}\right)$ *one has*

$$\delta_m\left(\frac{\Gamma}{\sqrt{n}}\right) \leq 4.5\ \theta + \frac{4.5}{n}\left(\Gamma_{k,m}^2 + \mathbb{E}\Gamma_{k,m}^2\right),$$

where $k \leq n$ is the largest integer satisfying $k \leq (\Gamma_{km}/B)^2$.

One of the main results of [3] (see also the lecture by R. Latała) gives upper bounds for $\Gamma_{k,m}$ and $\mathbb{E}\,\Gamma_{k,m}$. Using these bounds we obtain an RIP Theorem for matrices with independent rows:

Theorem: *Let $n, N \geq 1$ and $0 < \theta < 1$. Let Γ be an $n \times N$ matrix, whose rows are independent isotropic log-concave random vectors Y_i, $i \leq n$.*
There exists $c = c(\theta)$ depending on θ only, such that if $m \leq N$ satisfies

$$m \ \log\log 3m \left(\log \frac{3\max\{N,n\}}{m}\right)^2 \leq c\left(\frac{\theta}{\log(3/\theta)}\right)^2 \ n$$

then

$$\delta_m(\Gamma/\sqrt{n}) \leq \theta$$

with overwhelming probability.

The theorem is optimal up to a $\log\log$ factor. For unconditional distributions these factors can be removed (see [4] and the lecture of A. Litvak).

A similar argument allows to study a simpler parameter $\Gamma_k = \Gamma_{k,N}$, which was introduced and discussed in [1], and improve (in [2]) on the main convergence result from [1]. In the case of matrices with uniformly sub-exponential rows it gives sharp bounds for the extremal (non-zero) singular values of $\Gamma^*\Gamma$, which are quantitative counterparts of Bai-Yin theorem known for random matrices with i.i.d. entries.

Theorem: *Let $n, N \geq 1$. Let $X_1, \ldots, X_N \in \mathbb{R}^n$ be isotropic independent random vectors satisfying*

$$\sup_{i \leq N} \sup_{y \in S^{n-1}} \| \, |\langle X_i, y\rangle| \, \|_{\psi_1} \leq \psi,$$

for some $\psi < \infty$. Assume that for some $K < \infty$,

$$\mathbb{P}\left(\max_{i \leq N} |X_i|/\sqrt{n} > K \max\{1, (N/n)^{1/4}\}\right) \leq \exp\left(-\sqrt{n}\right).$$

Then with probability at least $1 - 2\exp\left(-c\sqrt{n}\right)$ one has

$$\sup_{x \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(|\langle X_i, x\rangle|^2 - \mathbb{E}\,|\langle X_i, x\rangle|^2\right) \right| \leq C\,(\psi + K)^2 \ \sqrt{\frac{n}{N}}.$$

Moreover, let A be a random $n \times N$ matrix, whose columns are $X_1, \ldots, X_N$. Let $\lambda_{\min}$ and $\lambda_{\max}$ be the smallest and the largest eigenvalues of $A\,A^*$.

Then with probability at least $1 - 2\exp(-c\sqrt{n})$,

$$1 - C\,(\psi + K)^2 \sqrt{\frac{n}{N}} \leq \frac{\lambda_{\min}}{N} \leq \frac{\lambda_{\max}}{N} \leq 1 + C\,(\psi + K)^2 \sqrt{\frac{n}{N}}.$$

The proof is based on the Fundamental Lemma. Namely, we consider $\Gamma = A^*$ which is $N \times n$ matrix, and $T = S^{n-1}$.

This improves the estimates from [2] for log-concave isotropic vectors. There, N was proportional to n, which was sufficient to answer the Kannan-Lovász-Simonovits question, however, for bigger N, it was off by a logarithmic factor. Here the estimate is sharp, in particular of the same order as in the Gaussian case.

References

[1] R. Adamczak, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. **2**3 (2010), 535–561.

[2] R. Adamczak, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of the empirical covariance matrix*, C.R. Math. Acad. Sci. Paris, **3**49 (2011), 195–200

[3] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Geometry of log-concave Ensembles of random matrices and approximate reconstruction*, C.R. Math. Acad. Sci. Paris, to appear.

[4] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *A Chevet type inequality and norms of submatrices*. Preprint.

[5] E.J. Candés and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **5**1 (2005), 4203–4215.

The single ring theorem

Manjunath Krishnapur

(joint work with Alice Guionnet, Ofer Zeitouni)

If A_n is a random matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$, its ESD is the empirical spectral distribution $n^{-1}\sum_{k=1}^{n}\delta_{\lambda_k}$ and its LSD is the limit of L_n, if it exists. For hermitian random matrices, there are various approaches to finding the LSD, for example the method of moments which exploits the identity $\int x^p dL_n(x) = n^{-1}\mathrm{trace}(A_n^p)$ to express the moments of the ESD in terms of the entries of the matrix. For non-hermitian A_n (more precisely, if $A_n^*A_n \neq A_nA_n^*$) the moments $\int z^p\bar{z}^q dL_n(z)$ characterize the ESD but are not easy to express in terms of the entries of the matrix. Conversely, the asymptotics of quantities $n^{-1}\mathrm{trace}(P(A_n, A_n^*))$ for any non-commutative polynomial P is not difficult to analyse, but does not lead to any direct information about L_n.

Owing to this difficulty, there are only a few models of non-hermitian random matrices whose LSD has been proved rigorously to exist. The most well known examples are of i.i.d entries with zero mean and finite variance where the LSD is the uniform distribution on the unit disk. This statement, well-known as the circular law, is the result of a long list of papers by many authors, a few important ones being those of Girko [2], Bai [5], Götze and Tikhomirov [6], [7], Pan and Zhou [8], and finally that of Tao and Vu [9] which settled it under the weakest hypothesis of second moments. Even LSD of matrices with i.i.d heavy tailed entries have been obtained by Bordenave, Chafai and Caputo [10] and given a spectacular description in terms of an operator on the PWIT.

Details of the last result may be found in abstract by Bordenave. Two other abstracts in this conference that deal with LSD of non-hermitian matrices are that of Götze and Tikhomirov on products of random matrices and that of Adamczak on matrices whose columns have log-concave distribution.

In this lecture (which is a summary of the paper [1]) we consider the random matrix A_n with density proportional to $\exp\{-n\mathrm{tr}V(A_n^*A_n)\}$ on the space of $n \times n$

complex matrices, where V is a polynomial of even degree and positive leading coefficient. The main result is that LSD μ of A_n exists, and is supported on a single connected annulus in the complex plane. One can also obtain an expression for the density of μ, but we do not present it here.

A few remarks on the result. When $V(x) = x$, the matrix has i.i.d standard complex Gaussian entries with mean zero and variance $1/n$, and hence μ is the uniform measure on the unit disk. For general V the exact density of eigenvalues is not known. The circular symmetry of μ follows obviously from the unitary invariance of A_n, that is $UA_nV \stackrel{d}{=} A_n$ for any fixed unitary matrices U and V.

We first describe the single ring phenomenon, predicted by Feinberg and Zee [4] and proved in the infinite dimensional setting of free probability by Haagerup and Larsen [3]. For any matrix X, let H_X denote the block matrix $\begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}$. A complex number z is an eigenvalue of A_n if and only if 0 is an eigenvalue of $(zI - A_n)^*(zI - A_n)$ or equivalently, if 0 is an eigenvalue of $H_{zI-A_n} = H_{zI} - H_{A_n}$. Heuristically assume that this relationship persists in the limit. Let μ denote the LSD (assuming they exist) of A_n and let ν^z denote the LSD of H_{zI-A_n}. Thus z is in the support of μ is and only if 0 is in the support of ν^z. The crucial point is that for a large class of random matrices, H_{A_n} and H_{zI} are freely independent, and hence $\nu^z = \nu^0 \boxplus \lambda^z$ where $\lambda^z = \frac{1}{2}\delta_{\pm|z|}$ is the ESD and LSD of H_{zI}. Haagerup and Larsen [3] have shown that for a symmetric, compactly supported probability measure on $\mathbb{R}$ such as ν^0, zero is in the support of $\nu^0 \boxplus \lambda^t$ if and only if $\left(\int u^{-2} d\nu^0(u)\right)^{-1} \leq t \leq \int u^2 d\nu^0(u)$. Evidently, this proves that the support of μ is the annulus with in-radius and out-radius given by these two bounds.

The actual determination of the limit also goes through the hermitian matrices H_{zI-A_n}, following the approach of Girko. Modulo some technicalities, it leads to the result

$$\int \phi(z) dL_n(z) \to \int \phi(z) \frac{1}{2\pi} \Delta h(z) dm(z), \qquad \phi \in C_c^{(2)}(\mathbb{R}^2)$$

where $h(z) = \int \log|x| \nu^z(dx)$. The heart of the matter is in proving the technicalities referred to above.

This consists in proving that $\int \log|x| dL_n(x) \to \int \log|x| \nu^z(dx)$ which does not automatically follows from the weak convergence of ESD of H_n^z to ν^z. One source of difficulty is the possibility that H_n^z has a very small singular value, for example of order e^{-n}. In the model at hand, we show that the smallest singular value of H_n^z is at least n^{-C}. The second difficulty is that there could be a crowding of δn singular values below $n^{-C/2}$. To overcome this problem, we show that the Stieltjes' transform G_n^z of H_n^z is uniformly bounded above by a constant on the set $\Im z \geq n^{-C'}$. From here, it is simple to control the number of singular values that fall in an interval around zero.

The heart of the proof consists in showing the uniform boundedness of G_n^z on the set $\Im z \geq n^{-C'}$. The main techniques used here are the Schwinger-Dyson

equations for Haar unitary matrices and the concentration of measure for Haar unitary matrices. Here is a very short summary of the idea. Our random matrix A_n has the property that UAV has the same distribution as A, for any unitary matrices U and V. This invariance can be expressed as a collection of differential equations which are what are known as Schwinger-Dyson equations (see chapter 5 of [11] for details). From this, one can write an equation for the ESD of H_n^z that converges to the free convolution equation in the limit. The finite equation can be thought of as an approximation of the limiting equation and the error terms can be bounded using concentration results for Haar measure on the unitary group.

References

[1] A. Guionnet, M. Krishnapur and O. Zeitouni, *The single ring theorem*, Ann. Math. *To appear.*

[2] V.L Girko, *The circular law.*, Teor. Veroyatnost. i Primenen. **29** (1984), 669-679.

[3] U. Haagerup and F. Larsen, *Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras*, J. Funct. Anal. **176** (2000), 331–367.

[4] Feinberg, J. and Zee, A., *Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism*, Nuclear Phys. B, **501** (1997), 643–669.

[5] Z. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.

[6] F. Götze and A. Tikhomirov, *On the Circular Law.*, unpublished, *arXiv:math.PR/0702386.*

[7] F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Prob. **38** (2010), 1444–1491.

[8] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101** (2010), 645-656

[9] T Tao and V. Vu, with an appendix by M. Krishnapur, *Random matrices: universality of ESDs and the circular law.*, Ann. Probab. **38** (2010), 2023-2065.

[10] C. Bordenave, P. Caputo and D. Chafai, *Spectrum of non-Hermitian heavy tailed random matrices*, Commun. Math. Phys. *to appear.*

[11] G. Anderson, A. Guionnet and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010.

Some recent works on the random matrices whose entries are almost independent

Jun Yin

(joint work with L. Erdös, A. Knowles, B. Schlein, H.-T. Yau)

The well known GOE (Gaussian orthogonal ensemble) has two important properties. First, for any fixed orthogonal matrix O, the ensemble is invariant under every transformation. Second, the variant elements h_{jk}'s ($j \leq k$) are statistically independent random variables. On the other hand, if a random matrix ensemble is orthogonally invariant and statistically independent, then it must be GOE. The orthogonally invariant random matrices have been well studied in 1990's. In this talk, we will introduce some recent works [1, 2, 3, 4, 5, 6, 7] on the random matrices whose entries are almost independent.

This type of random matrices including Wigner matrix, H_{ij}:

$$H = (h_{kj})_{1 \leq k,j \leq N}, \quad h_{ji} = h_{ij} \quad \text{i. i. d.}$$

$$\mathbb{E}h_{ij} = 0,\ \mathbb{E}|h_{ij}|^2 = \frac{1}{N} + \delta_{ij}\frac{1}{N}.$$

Generalized Wigner matrix which is similar to Wigner matrix, but variant elements h_{jk} can have different distribution, and the variance of h_{ij} only need to statisfy

$$\mathbb{E}|h_{ij}|^2 = O(1/N), \quad \sum_j \mathbb{E}|h_{ij}|^2 = 1$$

Covariance matrix, Non-Hermitian (i.i.d.) Matrix, sparse random matrix whose most entries equal to zero and band Wigner matrix whose non-zero entries are located near the diagonal line. For some of these matrices, we proved some properties of their local statistics. For example, it is well known that correlation function for the k eigenvalues (in the bulk) of GUE has sine kernel. We proved that [2, 3, 4, 6] the bulk universality holds for generalized Wigner ensembles if

$$\mathbb{E}|h_{ij}|^{4+\epsilon} \le C,$$

i.e., for $-2 < E < 2$, $b = N^{-1+\epsilon'}, \epsilon' > 0$

$$\lim_{N\to\infty} \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \Big(p^{(k)}_{H,N} - p^{(k)}_{GUE,N}\Big)\Big(E' + \frac{b_1}{N}, \ldots, E' + \frac{b_k}{N}\Big) = 0 \quad \text{weakly}$$

where $p^{(k)}_{H,N}$ and $p^{(k)}_{GUE,N}$ are the correlation function for the k eigenvalues of H and GUE. For the eigenvalues at edge, it is well known that they satisfy the TW law. We also proved edge universality for generalized Wigner matrix as follows[4, 6]. Suppose $H^{\mathbf{v}} = (h^{\mathbf{v}}_{ij})$ and $H^{\mathbf{w}} = (h^{\mathbf{w}}_{ij})$ are generalized Wigner matrices. Assume

$$\mathbb{E}^{\mathbf{v}}(x^{\mathbf{v}}_{ij})^2 \ = \ \mathbb{E}^{\mathbf{w}}(x^{\mathbf{w}}_{ij})^2$$

and for both ensembles we have

$$\mathbb{E}|x_{ij}|^{12} < C$$

Then for any $s \in \mathbb{R}$ we have

$$\mathbb{P}^{\mathbf{v}}\big(N^{2/3}(\lambda_N - 2) \le s\big) - \mathbb{P}^{\mathbf{w}}\big(N^{2/3}(\lambda_N - 2) \le s\big) \to 0$$

We will also introduce some similar results on covariance matrices [1] and sparse matrices [5, 6].

Besides eigenvalues, we also study the eigenvectors of these matrix ensembles. Let $\mathbf{v}_\alpha$, $1 \le \alpha \le N$ be the normalized eigenvectors. For sparse [5], generalized Wigner matrix [4] we proved that

$$\max_\alpha \|\mathbf{v}_\alpha\|_\infty \ \le \ \frac{(\log N)^C}{\sqrt{N}}$$

For Band Wigner matrix [2],

$$\max_\alpha \|\mathbf{v}_\alpha\|_\infty \ \le \ \frac{(\log N)^C}{\sqrt{M}}$$

where M is the bandwidth of the matrix. Furthermore, in recent work [7], we studied the distribution of eigenvectors of Wigner matrix.

In the end of the talk, we introduce our main ideas on the proof.

References

[1] Erdős, L., Schlein, B., Yau, H.-T., Yin, J.: The local relaxation flow approach to universality of the local statistics for random matrices. To appear in *Ann. Inst. H. Poincaré Probab. Statist.* Preprint arXiv:0911.3687.

[2] Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Preprint arXiv:1001.3453.

[3] Erdős, L., Yau, H.-T., Yin, J.: Universality for generalized Wigner matrices with Bernoulli distribution. To appear in *J. Combinatorics*. Preprint arXiv:1003.3813.

[4] Erdős, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized Wigner matrices. Preprint arXiv:1007.4652.

[5] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of sparse random matrices I: local semicircle law. Preprint arXiv:1103.1919.

[6] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral Statistics of Erdos-Renyi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues. Preprint arXiv:1103.3869

[7] Knowles, A., Yin, J.: Eigenvector distribution of Wigner matrices. Preprint arXiv:1102.0057.

Structure from Local Optima: Learning Subspace Juntas via Higher Order PCA

Ying Xiao

(joint work with Santosh Vempala)

A fundamental computational problem is that of learning a *k-junta*, a function of k relevant variables in an n-dimensional space. In this problem, introduced by A. Blum, one is given examples in $\mathbb{R}^n$ according to some fixed distribution. These examples are labeled by a boolean function that depends only on k of the n coordinates, and the goal is to learn the relevant k coordinates, and the labeling function. Naive enumeration of k subsets of the coordinates leads to an algorithm of complexity roughly n^k; Mossel et al gave an algorithm of complexity roughly $O(n^{0.7k})$ assuming labeled examples are drawn from the uniform distribution on the boolean hypercube. They refer to understanding the complexity of this restricted variant of the problem as "the single most important open question in uniform distribution learning".

We consider the problem of learning a labeling function $\ell(x)$ which is dependent only on some k-dimensional subspace V, the *relevant* subspace. We call such a function a *k-subspace junta* or *k-sjunta*. More precisely, let π_V denote the projection to the subspace V, then we are interested in labeling functions which have the form:

$$\ell(x) = \ell(\pi_V(x))$$

The problem is to learn the unknown concept ℓ, i.e., to find a function that agrees with ℓ on most of the distribution from which examples are drawn.

We make a natural assumption on the input distribution on examples: let F be a distribution over $\mathbb{R}^n$ and ℓ be a hypothesis with a relevant subspace V. We assume that F can be factored as a product of independent marginal distributions F_V and F_W on the subspaces V and $W = V^\perp$, i.e., $F = F_V F_W$ in which case we call F *factorizable*. Thus, a random point in F is generated by first picking

its coordinates in V according to F_V and then independently picking coordinates in W according to F_W. One may view the distribution on W as irrelevant noisy attributes introduced by an irrelevant subspace, and so we call this the *irrelevant noisy attributes* model. The full statement of our learning problem in this model is as follows:

Problem 1 (Learning a k-subspace junta)**.** *Given examples drawn from a factorizable distribution $F = F_V F_W$, and labeled by ℓ (a k-subspace junta with relevant subspace V), learn the concept ℓ.*

Specific instances of the learning problem have been studied under the assumption that the full distribution F is a Gaussian. Our primary algorithmic result is an efficient randomized algorithm when the distribution F_W on the irrelevant attributes is a Gaussian while the distribution on the relevant subspace needs only mild assumptions – in particular the relevant attributes can have dependencies or correlations. For example, a sample point x in the relevant subspace might have some independent coordinates and others being functions of these; moreover, we might only see an affine transformation of x. The complexity of our algorithm is polynomial in n times a function of $k, 1/\varepsilon$ and parameters of the k-dimensional concept class. This learning result under Gaussian noise is a substantial generalization of recent results on special cases of this problem, specifically for learning intersections of halfspaces and convex concepts assuming the full distribution is Gaussian. We also consider the case when the distribution over the noisy attributes is completely arbitrary – our algorithms can factorize and learn in this setting in a stronger computational model.

We examine higher moments of the distributions involved, and from these, we compute a basis for the relevant subspace. This approach is inspired by viewing PCA as optimising the bilinear form defined by the covariance matrix over the unit sphere. For example, the top eigenvector is the solution to a matrix optimisation problem:

$$\max_{\|v\|=1} v^T A v = \sum_{i_1,i_2} A_{i_1,i_2} v_{i_1} v_{i_2}$$

where A is the covariance matrix. Our higher moment method is a essentially a tensor method which optimises the *multi*linear form defined by higher moments:

$$\max_{\|v\|=1} A(v,\ldots,v) = \sum_{i_1,\ldots,i_r} A_{i_1,\ldots,i_r} v_{i_1} \ldots v_{i_r}.$$

While computing global maxima and minima to a bilinear form defined by a matrix is well-understood and can be done efficiently to arbitrary accuracy, similarly optimizing a multilinear form defined by a tensor is NP-hard. Indeed, for $\alpha > 16/17$, it is NP-hard to approximate the optimum to better than factor $\alpha^{\lfloor r/4 \rfloor}$, and the best known approximation factor is roughly $n^{r/2}$. There is actually a substantial literature regarding the computation of tensor optima, including several well known algorithms. Almost nothing, however, is known about convergence rates: even the strongest results prove convergence only in the limit.

Our primary contribution is to sidestep this NP-hardness: instead of trying to compute *global* optima to the multilinear form, we use *local* optima, and recover the relevant subspace from these. The use of local optima appears to be a novel and useful technique, and can be viewed as an effective realization of higher-order PCA; previous tensor methods have required the global optimum – for example, the planted clique algorithms of Brubaker and Vempala. Using local optima, we are able to give efficient algorithms which run in polynomial time, which has so far not been possible in the global optima setting.

The main algorithmic idea is to find local optima of the m^{th} moment tensor $f_m(u) = \mathbb{E}\,(x^T u)^m$. We prove that such a vector must lie entirely in V or its complement W unless its moments are identical to that of a Gaussian. This key property is inspired by the work of Frieze, Jerrum and Kannan, who under further assumptions use fourth moments to propose an algorithm for learning a vector of n fully independent random variables. Our algorithms have two main ingredients that might be of independent interest: (a) a rigorous second-order gradient method for computing approximate local optima and (b) a robust version of the Schwartz-Zippel lemma for testing approximate polynomial identities.

To state our results precisely, we require a few definitions: for a random vector $x \in \mathbb{R}^n$ with distribution F, the m^{th} moment tensor is a tensor of order m with n^m entries given by:

$$M^m_{i_1,\ldots,i_m} = \mathbb{E}\, x_{i_1} \ldots x_{i_m}.$$

The m^{th}*-moment distance* of two distributions F, G over $\mathbb{R}^n$ is the metric defined as

$$\begin{aligned} d_m(F,G) &= \max_{\|u\|=1} \left|\mathbb{E}_F\left((x^{\mathrm{T}}u)^m\right) - \mathbb{E}_G\left((x^{\mathrm{T}}u)^m\right)\right| \\ &= \max_{\|u\|=1} \left|(M^m_F - M^m_G)\cdot u \otimes \cdots \otimes u)\right| = \left\|M^m_F - M^m_G\right\|_2. \end{aligned}$$

Let Γ^n be the standard Gaussian distribution over $\mathbb{R}^n$ and γ_m denote the m^{th} moment of a standard Gaussian random variable:

$$\gamma_m = \begin{cases} 0 & \text{if } m \text{ is odd} \\ (m-1)!! & \text{if } m \text{ is even} \end{cases}$$

We say that a distribution G over $\mathbb{R}^k$ is (m,η)*-moment-distinguishable* along a direction $u \in \mathbb{R}^k$, $\|u\| = 1$, if either there exists $j \le m$:

$$\left|\mathbb{E}\,(x^T u)^j - \gamma_j\right| \ge \eta$$

or there exist $\{v_1, \ldots, v_t\} \subset u^\perp$ where $t \le m$ such that

$$\left|\mathbb{E}\,(x^T u)^{m-t} \Pi_{i=1}^t (x^T v_i) - \mathbb{E}\,(x^T u)^{m-t}\,\mathbb{E}\,\Pi_{i=1}^t (x^T v_i)\right| \ge \eta.$$

This definition simply states that G doesn't look too Gaussian in the direction of u, or that there are correlations of u with some vectors orthogonal to it.

To extend to learning, we need the following property for concept classes and distributions. For a distribution F and a k-dimensional concept class $\mathcal{H}$, we say that the triple $(k, F, \mathcal{H})$ is (m,η)*-moment learnable* if:

(1) $F = F_V F_W$ is a factorizable distribution where $dim(V) = k$.
(2) $\mathcal{H}$ is a set of k-subspace juntas whose relevant subspaces are contained in V.
(3) For $\ell \in \mathcal{H}$ with minimal (with respect to dimension) relevant subspace $P \subseteq V$, for each $u \in P$ with $\|u\| = 1$, either F_V of F_V^+ is (m, η)-moment distinguishable along u.

To state our learning guarantee, we need one more definition:

A triple $(k, F, \mathcal{H})$ is *robust* if for any subspace U of dimension at most k with orthonormal basis $\{u_i\}$ such that $|u_i^T \pi_V(u_i)| \geq 1-\varepsilon$, then $\ell(\pi_U(x))$ labels correctly $1 - g(\varepsilon)$ fraction of $\mathbb{R}^n$ under F, where g is a monotone function which goes to 0 at a polynomial rate as $\varepsilon \to 0$. The definition requires the distribution F and labeling function ℓ to be robust under small perturbations of the relevant subspace. We can now state our main learning result:

Theorem 2 (Learning, Gaussian noise)**.** *Let $\varepsilon, \delta > 0$, let $\ell \in \mathcal{H}$ where $(k, F, \mathcal{H})$ is (m, η)-moment learnable and robust, and let $F_W = \Gamma^{n-k}$ be Gaussian. Suppose that we are given labeled examples from F, then Algorithm* ***LearnUnderGaussian*** *identifies a subspace U and a hypothesis h such that h correctly classifies $1 - \varepsilon$ of F with probability at least $1 - \delta$. The time and sample complexity of the algorithm are bounded by $C_F(m, \varepsilon)T(k, \varepsilon)(kn)^m \mathrm{poly}(n, \eta, k, 1/\varepsilon, \log(1/\delta), M)$ where T is the complexity of learning the k-dimensional concept class $\mathcal{H}$.*

Another RIP-property and non-Euclidean embeddings

Omer Friedland
(joint work with Olivier Guédon)

1. Sparsity and Kašin-type embeddings

Let $0 < r \leq p < \infty$. We denote by $\mathrm{sparse}(m) = \{x \in \mathbb{R}^n : |\mathrm{supp}(x)| \leq m\}$ the set of vectors in $\mathbb{R}^n$ of cardinality of the support smaller than m.

We say that an operator $A : \ell_p^n \to \ell_r^k$ satisfies property $\mathcal{P}_1(m)$ iff

$$\forall x \in \mathrm{sparse}(m) \quad \alpha|x|_p \leq |Ax|_r \leq \beta|x|_p$$

We call this property *the restricted isomorphism property*, which is a generalization of the RIP-property that was introduced in [6]. There are other known modification of this RIP-property cf. [2, 10, 5]. In the sequel, we shall assume that $m, k \in O(n)$ and α, β do not depend on the dimension n.

Now, we present our main theorem:

Theorem 1. *Let $0 < r \leq p < \infty$ and $r \leq 1$. Let $A : \ell_p^n \to \ell_r^k$ be an operator that satisfies property $\mathcal{P}_1(m)$. Denote $U = \frac{1}{\beta} m^{1/q} A$. Then for any $x \in \ell_p^n$*

$$c_r \left(\frac{\alpha}{\beta}\right) \left(\frac{m}{n}\right)^{1/q} |x|_p \leq \frac{1}{n^{1/q}} \left(|Ux|_r + |x|_r\right) \leq 2|x|_p$$

where $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$*, and* c_r *is a positive constant depending only on* r*.*

This is what we call a Kašin-type embedding, for operators that satisfy property $\mathcal{P}_1(m)$ we have

$$\frac{1}{n^{1/q}}\begin{pmatrix}\mathrm{Id}_n\\ U\end{pmatrix}:\ell_p^n\to\ell_r^{n+k}$$

This is a general framework, for constructing Kašin-type embeddings from ℓ_p^n into ℓ_r^{n+k}, which is a deterministic statement. Note that for the Euclidean case, where $p=2$, there are a lot of results of this spirit, cf. [12, 20, 21, 19, 14, 15, 1, 18], which also follow from theorem 1.

2. Applications

We start by giving an example of an operator that satisfies property $\mathcal{P}_1$, based on this operator, we can later present our applications.

Let $0<r<p<2$ and $r\leq 1$, let $\eta>0$ and let $N\geq n$ be natural numbers such that $N=(1+\eta)n$. Let $(e_i)_{1\leq i\leq N}$ be the canonical basis of $\mathbb{R}^N$, and Y denotes a random vector taking the values $\{\pm e_{n+1},\dots,\pm e_N\}$ with probability $\frac{1}{2\eta n}$. Let $(Y_{i,j})$ be a sequence of independent copies of Y, where $1\leq i\leq n$, $j\in\mathbb{N}$.

We define the operator

$$T:\ell_p^n\to\ell_r^{\eta n}$$

$$x=(x_1,\dots,x_n)\mapsto\frac{\sigma_{p,r}}{(\eta n)^{1/q}}\sum_{i=1}^n x_i\sum_{j\geq 1}\frac{1}{j^{1/p}}Y_{i,j}\tag{1}$$

where $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$ and $\sigma_{p,r}$ is a normalization constant depending only on p,r.

This operator, T, approximates a random p-stable operator cf. [17, 13]. The rows of T are dependent. In other words, it is not an operator with independent entries, which makes the situation hard to deal with. In [11] the same type of operator was used, the difference is that Y is a random vector taking the values $\{\pm e_1,\dots,\pm e_N\}$ with probability $\frac{1}{2N}$.

The next proposition tells us that the operator T satisfies property $\mathcal{P}_1$:

Proposition 2. *Let* $\delta\leq\eta/\log(1/\eta)$*. Then with probability greater than* $1-2\exp(-c_{p,r}\eta n)$ *the operator* T *satisfies property* $\mathcal{P}_1(\delta n)$*. More precisely,*

$$\forall x\in\mathrm{sparse}(\delta n)\quad\frac{1}{2}|x|_p^r\leq|Tx|_r^r\leq\frac{3}{2}|x|_p^r$$

where $c_{p,r}$ *is a positive constant depending only on* p,r*.*

For the proof of this proposition, we need two known results, which are analogous to the main lemmas in [17]. The first one is a consequence of well-known results about $p-$stable random variables. The second one follows from results about scalar martingale difference (see also [8, Proposition 2]).

2.1. **Non-Euclidean embeddings.** Random embeddings in the non-Euclidean case attracts a lot of attention. The first major break through was done by Johnson and Schechtman [8], who proved that for any $\varepsilon > 0$ $\ell_p^n \overset{1+\varepsilon}{\hookrightarrow} \ell_r^N$, where $N = C(p, r, \varepsilon)n$. Later, Pisier [17] gave a different proof and extended their result to the case of general finite dimensional normed space E, proving that for any $\varepsilon > 0$ $\ell_p^n \overset{1+\varepsilon}{\hookrightarrow} E$, where n depends only on ε and on the stable-type constant of E. For $r = 1$, Naor and Zvavitch [16] proved that for any $\eta > 0$ $\ell_p^n \overset{C}{\hookrightarrow} \ell_1^{(1+\eta)n}$, where $C = (c \log n)^{(1-\frac{1}{p})(1+\frac{1}{\eta})}$. It is important to note that they provide an explicit definition of a random operator, which satisfies the desired property. Slightly after, Johnson and Schechtman [9] proved that for any $1 \leq r < p < 2$, there exists an operator $T : \ell_p^n \to \ell_r^{(1+\eta)n}$, such that $\|T\|\|T^{-1}_{|\mathrm{Im}T}\| \leq C(\eta)$. However, their proof depends heavily on a result of Bourgain, Kalton and Tzafriri [3], which in turn is based on a theorem of Elton [7]. Moreover, it doesn't give any explicit construction of the random operator T. The result of Naor and Zvavitch has been also extended by Bernués and López-Valdes [4] for $r \leq 1$, who proved that $\ell_p^n \overset{C(\log n, \eta, r)}{\hookrightarrow} \ell_r^{(1+\eta)n}$. The question, whether one can give an explicit construction of a random operator that embeds ℓ_p^n into ℓ_r^N, and the isomorphism constant depends only on p, r, recently answered by Friedland and Guédon [11], the isomorphism constant is $c_{p,r}^{1/\eta}$. Now, as a corollary of the main theorem, we improve the result of [11]. First, we manage to get an optimal isomorphism constant, i.e. polynomial in $1/\eta$, and second we present a full variety of operators that satisfy it.

Let us denote $S = \left(\frac{2}{3}\right)^{1/r} (\delta n)^{1/q} T$, where T is the operator defined in (1). Now, that we know that property $\mathcal{P}_1(m)$ holds for T, we conclude

Theorem 3. *Let $0 < \frac{2p}{p+2} < r < p < 2$ with $r \leq 1$. For any $\eta > 0$ and any natural number n, with probability greater than $1 - 2\exp(-c_{p,r}\eta n)$, for any $x \in \ell_p^n$*

$$\frac{c\eta^{1/q}}{(\log(1/\eta))^{1/q}}|x|_p \leq \frac{1}{n^{1/q}}\left(|x|_r + |Sx|_r\right) \leq 2|x|_p$$

where c is a positive constant depending only on r.

Theorem 3 gives us a new operator that embeds ℓ_p^n into ℓ_r^N with optimal bounds, where $N = (1 + \eta)n$.

2.2. **Gelfand width.** Let K be a bounded subset of a normed space E. Let $k \geq 0$ be an integer. Its k^{th} Gelfand width is defined as

$$d^k(K, E) := \inf_S \sup_{x \in S \cap T} \|x\|_E$$

where $\|\,.\,\|_E$ denotes the norm of E and where the infimum is taken over all linear subspaces S of codimension $\leq k$.

We have

$$d^k(B_r^N, \ell_p^N) = \inf_{\mathrm{codim}\, S \leq k} \mathrm{diam}_p(S \cap B_r^N)$$

In the next proposition we bound the p-diameter of a kernel of a given operator A with respect to the best α and β as defined in property $\mathcal{P}_1(m)$.

Proposition 4. *Let $A \in \mathbb{R}^{n\times N}$ be a matrix. Then*

$$\operatorname{diam}_p(\ker A \cap B_r^N) \leq \frac{(1+(\beta/\alpha)^p)^{1/p}}{m^{1/q}}$$

Proposition 2 combined with the above, give us the optimal known bounds for Gelfand width of B_r^n.

Theorem 5. *Let $\eta > 0$, denote $k = \eta n$. Then*

$$d^k(B_r^n, \ell_p^n) \leq c_{p,r}\left(\frac{\log(1/\eta)}{\eta n}\right)^{1/q}$$

where $c_{p,r}$ is a positive constant depending only on p, r.

2.3. **Compressed sensing.** Let $0 < r \leq p < \infty$ and let A be an $n \times N$ matrix. Now, let x be an unknown vector, and $y = Ax$ is the given data. The ℓ_r-minimization problem is the following:

$$\min_{At=Ax} \left\{|t|_r : t \in \mathbb{R}^N\right\} \tag{2}$$

Note that the above property is not specific to the matrix A but rather a property of its null space.

Proposition 6. *The following are equivalent*

(i) $\forall x \in \operatorname{sparse}(m)$ *problem (2) has a unique solution equal to x*
(ii) $\forall h \in \ker A$, $h \neq 0$ *and* $\forall I \subset [N]$, $|I| \leq m$: $|h_I|_r < |h_{I^c}|_r$

As a corollary of propositions 6 and 4 we get that the operator T, as defined in (1), is a good operator in the sense of compressed sensing:

Theorem 7. *Let A be a $n \times N$ matrix and let $1 \leq m \leq N$. If*

$$\operatorname{diam}_p(\ker A \cap B_r^N) < \frac{1}{2m^{1/q}}$$

then A satisfies ℓ_r-minimization problem for any $x \in \operatorname{sparse}(m)$.

References

[1] Artstein-Avidan, Shiri; Friedland, Omer; Milman, Vitali; Sodin, Sasha *Polynomial bounds for large Bernoulli sections of ℓ_1^N*. Israel J. Math. 156 (2006), 141-155.

[2] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, M. J. Strauss *Combining geometry and combinatorics: A unified approach to sparse signal recovery.* 2008.

[3] Bourgain, J. ; Kalton, N. J. ; Tzafriri, L. *Geometry of finite-dimensional subspaces and quotients of L_p*. Geometric aspects of functional analysis (1987–88), 138-175, Lecture Notes in Math., 1376, Springer, Berlin, 1989.

[4] Bernués, J. ; López-Valdes, M. . *Tail estimates and a random embedding of l_p^n into $l_r^{(1+\epsilon)n}, 0 < r < p < 2$*. Publ. Math. Debrecen 70 (2007), no. 1-2, 9-18.

[5] D. Chafaï, O. Guédon, G. Lecué, A. Pajor *Interactions between compressed sensing random matrices and high dimensional geometry.* to appear.

[6] Candes, E. J.; Tao, T. *Decoding by linear programming.* IEEE Trans. Inform. Theory 51 (2005), no. 12, 42034215.
[7] Elton, J. *Sign-embeddings of l_1^n.* Trans. Amer. Math. Soc. 279 (1983), no. 1, 113-124.
[8] Johnson, W. B.; Schechtman, G. *Embedding l_p^m into l_1^n.* Acta Math. 149 (1982), no. 1-2, 71-85.
[9] Johnson, W. B.; Schechtman, G. *Very tight embeddings of subspaces of L_p, $1 \leq p < 2$, into l_p^n.* Geom. Funct. Anal. 13 (2003), no. 4, 845-851.
[10] Foucart, S.; Lai, M.-J. *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$.* Appl. Comput. Harmon. Anal. 26 (2009), no. 3, 395407.
[11] Friedland, O.; Guédon, O. *Random embedding of ℓ_p^n into ℓ_1^N*, Mathematische Annalen.
[12] Kašin, B. S. *The widths of certain finite-dimensional sets and classes of smooth functions.* (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 41 (1977), no. 2, 334-351, 478.
[13] LePage, R.; Woodroofe, M.; Zinn, J. *Convergence to a stable distribution via order statistics.* Ann. Probab. 9 (1981), no. 4, 624-632.
[14] Litvak A. E.; Pajor A.; Rudelson M.; Tomczak-Jaegermann N. *Smallest singular value of random matrices and geometry of random polytopes.* Adv. Math. 195 (2005), 491-523.
[15] Litvak, A. E.; Pajor, A.; Rudelson, M.; Tomczak-Jaegermann, N.; Vershynin, R. *Euclidean embeddings in spaces of finite volume ratio via random matrices.* J. Reine Angew. Math. 589 (2005), 1-19.
[16] Naor, A.; Zvavitch, A. *Isomorphic embedding of l_p^n, $1 < p < 2$, into $l_1^{(1+\epsilon)n}$.* Israel J. Math. 122 (2001), 371-380.
[17] Pisier, G. *On the dimension of the l_p^n-subspaces of Banach spaces, for $1 \leq p < 2$.* Trans. Amer. Math. Soc. 276 (1983), no. 1, 201-211.
[18] Rudelson, M. *Invertibility of random matrices: norm of the inverse.* Ann. of Math., 168 (2008), 575-600.
[19] Schechtman, G. *Special orthogonal splittings of L_1^{2k}.* Israel J. Math. 139 (2004), 337-347.
[20] Szarek, S. *On Kashin's almost Euclidean orthogonal decomposition of l_n^1.* Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. 26 (1978), no. 8, 691-694.
[21] Szarek, S.; Tomczak-Jaegermann, N. *On nearly Euclidean decomposition for some classes of Banach spaces.* Compositio Math. 40 (1980), no. 3, 367-385.

Polynomial Littlewood-Offord Problems

Kevin P. Costello

Consider a standard random walk on the integers, where at each step the walker independently moves left one step or right one step with probability 1/2. Let X_n be the location of the walker at time n. Intuitively, we expect X_n to become more dispersed as n increases. There are several ways in which we can quantify this dispersion:

- **Variance:** At time n, we have $\sqrt{\mathbf{E}(X^2)} = \sqrt{n}$.
- **Concentration:** If you take an interval I of size much larger than $\sqrt{n}$ centered at 0, the walker will with very high probability lie in I
- **Anti-Concentration:** If you take an interval I of size much **smaller** than $\sqrt{n}$, the walker will with very high probability *not* lie in I. In particular, the likelihood that X_i takes on any particular value is $O(n^{-1/2})$.

Note that the anti-concentration bound is the inverse of the variance bound; this is to be expected since the limiting normal distribution of the walk is smooth.

Now suppose we remove the requirement that all step sizes be equal, but still fix a set of (non-zero) step sizes in advance. Is it possible, by clever choices of step sizes, to increase the concentration on one value to something larger than $O(n^{-1/2})$? This question was first address by Littlewood and Offord in their study of the roots of random polynomials [6] and was answered in a strong sense by Erdős [3]: No matter what the step sizes, the concentration on one value is never larger than in the case when all step sizes are equal.

Now suppose we exclude this example by requiring all step sizes distinct. Assuming the step sizes are still integers (intuitively, the concentration on one value is maximized when all the step sizes are on a single lattice), this increases the standard deviation to at least $n^{3/2}$. Sárközy and Szémeredi [9] showed the corresponding anticoncentration result: The concentration on one value is at most $n^{-3/2}$. A similar phenomenon happens when we forbid solutions to $a_1+a_2 = a_3+a_4$ among the step sizes: The standard deviation is upped to $n^{5/2}$, and a matching anti-concentration bound of $n^{-5/2}$ was provided by Halász [4]. More recently, Tao and Vu [10] and Rudelson and Vershynin [8] have considered results in the reverse direction: If the concentration on one value is large, then the step sizes must be very highly structured (lying in a generalized arithmetic progression).

These results have had many applications in the study of random matrices with independent entries, since many of the key linear algebraic quantities of these matrices (e.g. the determinant, or the distance from one row to the span of the remaining rows) are linear in the entries of each row or column. These applications have included bounds on the singularity probability (first shown to be $o(1)$ by Komlós [5]) and on the smallest singular value ([10, 8]). All of these results, however, depend heavily on the independence of the entries.

In a random symmetric matrix (with entries above the main diagonal independent, and below the main diagonal forced by symmetry), we lose this independence. Furthermore, the determinant of the matrix becomes a *quadratic* form in the entries of a particular row or column. This difficulty motivated the author (with Tao and Vu) [2] to consider the question of developing a quadratic analogue of the Littlewood-Offord problem: If we take a quadratic polynomial in n variables with many nonzero coefficients and choose each input randomly, how quickly must the concentration on one value decrease as n increases? Note that if "quadratic" is replaced by "linear" here we recover the original Littlewood-Offord problem.

One new wrinkle which arises here is that the variance heuristic from the linear case no longer holds. A quadratic polynomial in n variables with non-zero integer coefficients is the sum of n^2 terms, and so has standard deviation at least n. Nevertheless the form $(x_1 + \cdots + x_n)^2$ has concentration of $n^{-1/2}$ on 0. In other words, degeneracy (or near-degeneracy) can cause spikes in the distribution where the concentration on one value is much larger than on surrounding values. For Bilinear forms, I have shown [1] that this degeneracy is essentially the only reason the heuristic fails by a large amount:

Theorem 1. *Let $f(x_1, \ldots x_n, y_1, \ldots y_n) = x^T B y$ be a bilinear form whose matrix B has at least k nonzero entries in every row, and each x_i and y_j are independently*

set to 1 *or* -1 *with equal probability. If* f *takes on a single value with probability* $k^{-1+\epsilon}$*, then* B *contains a rank one submatrix of size* $n(1-o_\epsilon(1))$*.*

The -1 in the exponent is tight here; if B is a random ± 1 matrix, then $x^T By$ concentrates on one value with probability $\Omega(n^{-1})$ but with high probability B contains no rank one submatrix of size larger than $O(\log n)$. I conjecture the same holds for quadratic forms, but cannot prove it at this point. However, the analogue of the original Littlewood-Offord-Erdős result does (nearly) hold [1]

Theorem 2. *Let* $f(x_1, \dots x_n) = x^T Ax$ *be a quadratic form whose matrix* A *has at least* mn *nonzero entries. Then for any* $\epsilon > 0$*, the probability* f *takes on a given value is at most* $O\left(m^{-1/2+\epsilon}\right)$*.*

This result improves on [2], which had the 1/2 replaced by a 1/8. Again the 1/2 is tight here, as is shown by the form $(x_1 + \cdots + x_m)(x_1 + \cdots + x_n)$. A consequence of this result is a bound of $O(n^{-1/2+\epsilon})$ on the singularity probability of a random symmetric matrix. Recent work of Nguyen [7] and Vershynin [11] has improved this bound still further by developing further inverse results for the Quadratic Littlewood-Offord problem.

References

[1] K. Costello. Bilinear and quadratic Variants on the Littlewood-Offord problem. Preprint at arXiv:0902.1538

[2] K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely non-singular, *Duke Math J.* **135** (2006) 395-413.

[3] P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898-902.

[4] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, *Period. Math. Hungar.* **8** (1977), no. 3-4, 197-211.

[5] J. Komlós, On the determinant of $(0,1)$ matrices, *Studia Sci. Math. Hungar.* **2** (1967) 7-22.

[6] J.E. Littlewood and A.C. Offord, On the real roots of a random algebraic equation III., *Rec. Math. [Mat. Sbornik] N.S.* **12** (1943), 277-286

[7] H. Nguyen, Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. Preprint at arXiv:1101.3074

[8] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices, *Adv. Math.* **218** (2008) 600-633

[9] A. Sárközy and E. Szemerédi, Über ein Problem von Erdös und Moser, *Acta Arith.* **11** (1965), 205-208

[10] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, *Annal. Math* **169** (2009), 595-632

[11] R. Vershynin, Invertibility of symmetric random matrices, preprint at arXiv:1102.0300.

A Chevet type inequality and norms of submatrices of a random matrix in the unconditional case.

Alexander E. Litvak
(joint work with Radosław Adamczak, Rafał Latała, Alain Pajor, Nicole Tomczak-Jaegermann)

Let n, N be positive integers and let $X_1, \dots, X_n$ be independent isotropic log-concave unconditional vectors in $\mathbb{R}^N$. Below we consider an $n \times N$ random matrix A, whose rows are X_i's. The canonical basis on $\mathbb{R}^d$ is denoted by $\{e_i\}_i$.

We prove a Chevet type inequality for the matrix A. Namely, let $K \subset \mathbb{R}^N$ and $L \subset \mathbb{R}^n$ be origin symmetric convex bodies, $\|\cdot\|_K$ and $\|\cdot\|_L$ be the corresponding norms on $\mathbb{R}^N$ and $\mathbb{R}^n$, and let E_i's be i.i.d. symmetric exponential random variables. Then

$$\mathbb{E}\|A : (\mathbb{R}^N, \|\cdot\|_K) \to (\mathbb{R}^n, \|\cdot\|_L)\| \le C\|Id : (\mathbb{R}^N, \|\cdot\|_K) \to \ell_2^N\| \cdot \mathbb{E}\left\|\sum_{i=1}^n E_i e_i\right\|_L + C\|Id : \ell_2^n \to (\mathbb{R}^n, \|\cdot\|_L)\| \cdot \mathbb{E}\left\|\sum_{i=1}^N E_i e_i\right\|_{K^0},$$

where $\|\Gamma : X \to Y\|$ denotes the operator norm of Γ, considered as a linear operator from X to Y, Id stays for the formal identity operator, and C is an absolute constant. We provide the corresponding probability estimates. For the case of Gaussian random matrices with E_i's substituted by the standard Gaussian random variables the inequality is due to Chevet (see [7] and references therein). It plays an important role in Asymptotic Geometric Analysis. Our proof consist of two steps: first, using the corresponding Latała Theorem ([5]), we reduce the case of general matrix A to the case of the exponential matrix, i.e. the matrix whose entries are i.i.d. symmetric exponential random variables. Then, using Talagrand's result ([6]) on relations between some random processes and so-called γ_p functionals and computing those functionals, we obtain the result.

We apply our Chevet inequality to obtain sharp uniform bounds on norms of submatrices of A. More precisely, for any subsets $J \subset \{1, \dots, n\}$ and $I \subset \{1, \dots, N\}$ denote the submatrix of A consisting of the rows indexed by elements from J and the columns indexed by elements from I by $A(J, I)$. Given $k \le n$ and $m \le N$ define the parameter $A_{k,m}$ by

$$A_{k,m} = \sup \|A(J, I) \ : \ \ell_2^m \to \ell_2^k\|,$$

where the supremum is taken over all subsets $J \subset \{1, \dots, n\}$ and $I \subset \{1, \dots, N\}$ with cardinalities $|J| = k$, $|I| = m$. That is, $A_{k,m}$ is the maximal operator norm of a submatrix of A with k rows and m columns. We prove that

$$A_{km} \le \sqrt{m} \log\left(\frac{3N}{m}\right) + \sqrt{k} \log\left(\frac{3n}{k}\right),$$

with high probability. This estimate is sharp. We conjecture that the condition of unconditionality of vectors is not needed for this bound, however we can prove it only with an additional factor $\ln\ln(3m)$ and the proof is much more involved.

Furthermore, we provide applications of this result to the restricted isometry property of A. More precisely, define RIP parameter (of order m) of A as the smallest number $\delta = \delta_m(A)$ such that

$$(1-\delta)|x|^2 \leq \frac{1}{n}|Ax|^2 \leq (1+\delta)|x|^2$$

for every $x \in \mathbb{R}^N$ with at most m non-zero coordinates. We show that given $\theta \in (0,1)$ with overwhelming probability $\delta_m(A) \leq \theta$ provided that either
(i) $N \leq n$ and

$$m \approx \min\left\{N,\ \frac{\theta^2 n}{\ln^3(3/\theta)}\right\}$$

or
(ii) $N \geq n$ and

$$m \leq c\ \frac{\theta n}{\ln(3N/(\theta n))}\ \min\left\{\frac{1}{\ln(3N/(\theta n))}, \frac{\theta}{\ln^2(3/\theta)}\right\},$$

where c is an absolute constant. The proof is similar to the one presented in [4].

Finally, we construct examples showing that the condition of unconditionality of vectors is important in the Latała Theorem and in our Chevet inequality. Thus this scheme cannot be used to provide estimates for $A_{k,m}$ in the general case.

The results appear in [1] and were partially announced in [2]. For the related estimates in the non-unconditional case, see [2] and [3].

References

[1] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *A Chevet type inequality and norms of submatrices*, preprint.

[2] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Geometry of log-concave Ensembles of random matrices and approximate reconstruction*, preprint.

[3] R. Adamczak, R. Latała, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Tail estimates for norms of sums of log-concave random vectors*, preprint.

[4] R. Adamczak, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of empirical covariance matrix*, C.R. Math. Acad. Sci. Paris **349** (2011), 195–200.

[5] R. Latała, *On weak tail domination of random vectors*, Bull. Polish Acad. Sci. Math. **57** (2009), 75–80.

[6] M. Talagrand, *The generic chaining. Upper and lower bounds of stochastic processes*, Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.

[7] N. Tomczak-Jaegermann, *Banach-Mazur distances and finite-dimensional operator ideals*, Pitman Monographs and Surveys in Pure and Applied Mathematics, 38. Longman Scientific & Technical, Harlow; John Wiley & Sons, Inc., New York, 1989.

The circular law for random matrices with independent logarithmically-concave rows

Radosław Adamczak

For an $n \times n$ matrix M define its spectral measure μ_M as

$$\mu_M = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(M)},$$

where $\lambda_i(M)$ are (complex) eigenvalues of M (counted with multiplicities) and δ_x stands for the Dirac mass at x.

It is well known (and goes back to the work of Ginibre and Mehta) that if A_n is a random $n \times n$ matrix with independent standard complex Gaussian entries then the spectral measure of $n^{-1/2}A_n$ converges as $n \to \infty$ to the uniform measure on the unit disc. It was conjectured that this type of limit behaviour was in fact universal and held for all random matrices with independent mean zero variance one entries. The first general approach towards the proof of this conjecture was proposed by Girko [8, 9] in the eighties and later developed by many authors (e.g. in [4, 10, 12]) resulting in substantial weakening of the conditions on the distribution of the random matrix under which the limiting behaviour of the spectral measure was the same as in the complex Gaussian case.

In a recent article [13], Tao and Vu solved the conjecture by proving the following

Theorem 1 (The circular law)**.** *Let $(X_{ij})_{i,j<\infty}$ be an infinite array of i.i.d. mean zero, variance one complex random variables. Let $A_n = (A_{ij})_{i,j\leq n}$. Then the spectral measure of $n^{-1/2}A_n$ converges almost surely as $n \to \infty$ to the uniform measure on the unit disc.*

The integrability assumptions of the above theorem are known to be the weakest possible as the limiting behaviour of the spectral distribution of matrices with i.i.d. infinite variance is known to be qualitatively different from the circular case (see [6] or the abstract of the talk by Charles Bordenave). In this context a natural question arises, whether the independence assumption can be relaxed and if so, to what extent.

Our work may be considered the first step in this direction and is an attempt to extend the circular law to matrices with independent rows. It is easy to believe that a corresponding result should hold for random matrices with independent rows distributed on properly normalized ℓ_p^n balls (thanks to [5] one can compare such matrices to matrices with independent entries), therefore it is natural to have a look at a generalization of such distributions, i.e. at random matrices with independent, isotropic, log-concave rows.

Recall that a random vector in $\mathbb{R}^n$ is called isotropic if it is centered and its covariance matrix is equal to identity.

An n-dimensional random vector not supported on a proper affine subspace of $\mathbb{R}^n$ is called log-concave if it has density of the form $\exp(-V(x))$, where $V\colon \mathbf{R}^n \to$

$(-\infty, \infty]$ is convex. Examples include in particular Gaussian measures and uniform distributions on convex bodies. Log-concave isotropic vectors have many nice geometric properties which allow to prove good bounds on the largest and smallest singular values of random matrices with independent isotropic log-concave rows (see e.g. [2], [3] or abstracts of talks by Rafał Latała, Alexander Litvak and Nicole Tomczak-Jaegermann). Combining this properties with the Tao and Vu approach one can obtain

Theorem 2. *Let A_n be a sequence of $n \times n$ random matrices with independent rows $X_1^{(n)}, \dots, X_n^{(n)}$ (defined on the same probability space). Assume that for each n and $i \leq n$, $X_i^{(n)}$ has a log-concave isotropic distribution. Then, with probability one, the spectral measure $\mu_{\frac{1}{\sqrt{n}}A_n}$ converges weakly to the uniform distribution on the unit disc.*

The above theorem under an additional assumption that $X_i^{(n)}$ are unconditional (i.e. their distribution is invariant under reflections with respect to coordinate hyperplanes) was proved in 2010 and published in [1]. The general case is a recent result (yet unpublished). The basic idea behind the proof is using the replacement principle by Tao and Vu [13] and reducing the problem to the analysis of

$$\frac{1}{n} \log |\det(\frac{1}{\sqrt{n}} A_n - z\mathrm{id})|$$

for $z \in \mathbb{C}$, which can be further reduced to the analysis of extreme singular values of the matrix $\frac{1}{\sqrt{n}}A_n - z\mathrm{id}$ and the spectral measure of $M_n = (\frac{1}{\sqrt{n}}A_n - z\mathrm{id})(\frac{1}{\sqrt{n}}A_n - z\mathrm{id})^*$.

The difference between the original proof in the unconditional case and the general one consists in the approach to the limit theorem for μ_{M_n}. In [1] it was obtained via moment method by exploiting the sign invariance of the matrix, in the general log-concave case it relies on the Stieltjes transform techniques developed in [7]. In both cases the crucial property of isotropic log-concave vectors that steps in for independence is Klartag's thin shell inequality [11], which asserts that the Euclidean norm of an isotropic random vector in dimension n is with high probability very close to $\sqrt{n}$ (see Olivier Guédon's talk for more information).

Let us point out that the general approach suggested by Tao and Vu seems very robust and that the circular law is likely to hold for matrices with independent rows under much milder regularity assumptions than log-concavity. Proving a more general theorem along the same lines would require analysis of extreme singular values for random matrices with independent rows, which is a problem of independent interest, with potential applications in statistics or numerical analysis.

References

[1] R. Adamczak. *The Marchenko-Pastur and circular laws for some classses of random matrices with dependent entries.* Electronic Journal of Probability Vol. 16 (2011), 1065–1095.

[2] R. Adamczak, O. Guédon, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann. *Condition number of a square matrix with i.i.d. columns drawn from a convex body.* To appear in Proc. Amer. Math. Soc.

[3] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann. *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles.* J. Amer. Math. Soc. 23 (2010), no. 2, 535–561.
[4] Z.D. Bai. *Circular law.* Ann. Probab. 25 (1997), no. 1, 494–529.
[5] F. Barthe, O. Guédon, S. Mendelson, A. Naor. *A probabilistic approach to the geometry of the l_p^n-ball.* Ann. Probab. 33 (2005), no. 2, 480513.
[6] C. Bordenave, P. Caputo, D. Chafaï. *Spectrum of non-hermitian heavy tailed random matrices.* To appear in Comm. Math. Phys.
[7] R.B. Dozier, J.W. Silverstein. *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices.* J. Multivariate Anal. 98 (2007), no. 4, 678–694.
[8] V.L. Girko. *The circle law.* Teor. Veroyatnost. i Mat. Statist. No. 28, (1983), 15–21.
[9] V.L. Girko *The strong circular law. Twenty years later. I.* Random Oper. Stochastic Equations 12 (2004), no. 1, 49–104.
[10] F. Goetze, A. Tikhomirov. *The circular law for random matrices.* Ann. Probab. 38 (2010), no. 4, 14441491.
[11] B. Klartag. *Power-law estimates for the central limit theorem for convex sets.* J. Funct. Anal. 245 (2007), no. 1, 284–310.
[12] G. Pan, W. Zhou. *Circular law, extreme singular values and potential theory.* J. Multivariate Anal. 101 (2010), no. 3, 645656.
[13] R. Tao, V. Vu. *Random matrices: Universality of ESDs and the circular law. With an Appendix by M. Krishnapur.* Ann. Probab 38 (2010), no. 5, 2023-2065.

Estimating Population Eigenvalues From Large Dimensional Sample Covariance Matrices

Jack W. Silverstein

This talk introduces a way to estimate the population eigenvalues from those of a sample covariance matrix, with an application to wireless communications. Let $B_n = (1/N)T_n^{1/2}X_nX_n^*T_n^{1/2}$ where $X_n = (X_{ij})$ is $n \times N$ with i.i.d. complex standardized entries, and $T_n^{1/2}$ is a Hermitian square root of the nonnegative definite Hermitian matrix T_n. This matrix can be viewed as the sample covariance matrix of N i.i.d. samples of the n dimensional random vector $T_n^{1/2}(X_n)_{\cdot 1}$, the latter having T_n for its population covariance matrix. Quite a bit is known about the behavior of the eigenvalues of B_n when n and N are large but on the same order of magnitude. These results are relevant in situations in multivariate analysis where the vector dimension is large, but the number of samples needed to adequately approximate the population matrix (as prescribed in standard statistical procedures) cannot be attained. Before presenting the results we introduce some concepts. For matrix A ($p \times p$) with real eigenvalues, define F^A, the empirical distribution function (d.f.) of the eigenvalues of A, to be

$$F^A(x) \equiv (1/p) \cdot (\text{number of eigenvalues of } A \leq x).$$

For any p.d.f. G the Stieltjes transform of G is defined as

$$m_G(z) \equiv \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}.$$

We have the well-known inversion formula

$$G\{[a,b]\} = (1/\pi) \lim_{\eta \to 0^+} \int_a^b \Im\, m_G(\xi + i\eta) d\xi$$

(a, b continuity points of G).

The first result is on the limiting behavior of F^{B_n}. It is shown in Silverstein (1995) that if T_n and X_n are independent, and $N = N(n)$ with $n/N \to c > 0$ as $n \to \infty$ and F^{T_n} converges almost surely in distribution to a p.d.f. H on $[0, \infty)$ as $n \to \infty$, then, almost surely, F^{B_n} converges in distribution, as $n \to \infty$, to a (nonrandom) p.d.f. F, whose Stieltjes transform $m(z)$ ($z \in \mathbb{C}^+$) satisfies

$$m = \int \frac{1}{t(1 - c - czm) - z} dH(t),$$

in the sense that, for each $z \in \mathbb{C}^+$, $m = m(z)$ is the unique solution to the equation in $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}^+\}$.

We have

$$\begin{aligned} F^{(1/N)X^*TX} &= (1 - \frac{n}{N}) I_{[0,\infty)} + \frac{n}{N} F^{(1/N)XX^*T} \\ &\xrightarrow{a.s.} (1-c) I_{[0,\infty)} + cF \equiv \underline{F}. \end{aligned}$$

Notice m_F and $m_{\underline{F}}$ satisfy

$$\frac{1-c}{cz} + \frac{1}{c} m_{\underline{F}}(z) = m_F(z) = \int \frac{1}{-z m_{\underline{F}} t - z} dH(t).$$

Therefore, $\underline{m} = m_{\underline{F}}$ solves

$$z = -\frac{1}{\underline{m}} + c \int \frac{t}{1 + t\underline{m}} dH(t).$$

Of course this result only reveals the proper proportion of eigenvalues inside the support of F. Work was achieved on exact separation of eigenvalues in Bai and Silverstein (1998),(1999). For any $d > 0$ and d.f. G, let $F^{d,G}$ denote the limiting spectral d.f. of $(1/N)X_n^* T_n X_n$ corresponding to limiting ratio d and limiting F^{T_n} G. Assume the interval $[a, b]$ with $a > 0$ lies in an open interval outside the support of F^{c_n, H_n} for all large n. Then, under certain assumptions on the entries of X_n, it is proven in Bai and Silverstein (1998) that

$$P(\text{ no eigenvalue of } B_n \text{ appears in } [a,b] \text{ for all large } n\ ) = 1.$$

In Bai and Silverstein (1999) it is shown that

(1) If $c[1 - H(0)] > 1$, then x_0, the smallest value in the support of $F^{c,H}$, is positive, and with probability one $\lambda_N^{B_n} \to x_0$ as $n \to \infty$. The number x_0 is the maximum value of the function

$$z(m) = -\frac{1}{m} + c \int \frac{t}{1 + tm} dH(t)$$

for $m \in \mathbb{R}^+$.

(2) If $c[1-H(0)] \le 1$, or $c[1-H(0)] > 1$ but $[a,b]$ is not contained in $[0,x_0]$ then $m_{F^{c,H}}(b) < 0$. Let for large n integer $i_n \ge 0$ be such that

$$\lambda_{i_n}^{T_n} > -1/m_{F^{c,H}}(b) \qquad \text{and} \qquad \lambda_{i_n+1}^{T_n} < -1/m_{F^{c,H}}(a)$$

(eigenvalues arranged in non-increasing order). Then

$$P(\lambda_{i_n}^{B_n} > b \quad \textit{and} \quad \lambda_{i_n+1}^{B_n} < a \quad \text{for all large } n\ ) = 1.$$

From this work Mestre (2008) developed a scheme for estimating eigenvalues of T_n from those of B_n. For fixed n, N, and $H_n = F^{T_n}$, let $\underline{m} = \underline{m}(z) = m_{F^{c_n,H_n}}(z)$. Suppose T_n has positive eigenvalue t_1 with multiplicity n_1, and that exact separation occurs for the eigenvalues of B_n for all n large, associated with t_1, that is, with probability one, for all n large there will be n_1 eigenvalues of B_n associated with t_1, clustered together, and separated from the remaining eigenvalues of B_n. Then, it is shown in Mestre (2008) that

$$t_1 = -\frac{N}{n_1}\frac{1}{2\pi i}\oint \frac{z\underline{m}'(z)}{\underline{m}(z)}dz,$$

the contour, C, only containing the support of F^{c_n,H_n} associated with t_1. Let $\underline{m}_n = m_{F^{(1/N)X_n^* T_n X_n}}$. We have, with probability 1,

$$\sup_{z\in \mathrm{C}} \max\left|\underline{m}(z) - \underline{m}_n(z)|, |\underline{m}'(z) - \underline{m}_n'(z)\right| \to 0,$$

as $n \to \infty$. Thus

$$-\frac{N}{n_1}\frac{1}{2\pi i}\oint \frac{z\underline{m}_n'(z)}{\underline{m}_n(z)}dz$$

can be taken as an estimate of t_1. This quantity equals

$$\frac{N}{n_1}\left(\sum_{\lambda_j\in[x_a,x_b]}\lambda_j - \sum_{\mu_j\in[x_a,x_b]}\mu_j\right),$$

where λ_j's are the eigenvalues of B_n, μ_j's are the zeros of $\underline{m}_n(z)$. We have

$$\underline{m}_n(z) = \frac{1}{N}\sum_{j=1}^{n}\frac{1}{\lambda_j - z} + \frac{N-n}{N}\frac{1}{-z} = 0$$

$$\Longleftrightarrow \frac{1}{N}\sum_{j=1}^{n}\frac{\lambda_j}{\lambda_j - z} = 1.$$

The solutions are the eigenvalues of the matrix

$$\mathrm{Diag}(\lambda_1,\ldots,\lambda_n) - N^{-1}ss^*,$$

where $s = (\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n})^*$.

This result is extended to power estimation of multiple signal sources in multi-antenna fading channels (Couillet, Silverstein, Bai, and Debbah):

Consider K entities transmitting data. Transmitter $k \in \{1,\ldots,K\}$ has (unknown) transmission power P_k with n_k antennas. They transmit data to N sensing devices (receiver). The multiple antenna channel matrix between transmitter

k and the receiver is denoted by $H_k \in \mathbb{C}^{N\times n_k}$, where the entries of $\sqrt{N}H_k$ are i.i.d. standardized.

At time instant $m \in \{1,\dots,M\}$, transmitter k emits signal $x_k^{(m)} \in \mathbb{C}^{n_k}$, entries independent and standardized, independent for different m's. At the same time the receive signal is impaired by additive noise $\sigma w^{(m)} \in \mathbb{C}^N$ ($\sigma > 0$), the entries of $w^{(m)}$ are i.i.d. standardized (independent across m). Therefore at time m the receiver senses the signal

$$y^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} H_k x_k^{(m)} + \sigma w^{(m)}.$$

Therefore, with $Y = [y^{(1)},\dots,y^{(M)}] \in \mathbb{C}^{N\times M}$, $X_k = [x_k^{(1)},\dots,x_k^{(M)}] \in \mathbb{C}^{n_k\times M}$, and $W = [w^{(1)},\dots,w^{(M)}] \in \mathbb{C}^{N\times M}$ we have

$$Y = \sum_{k=1}^{K} \sqrt{P_k} H_k X_k + \sigma W \quad = \quad HP^{1/2}X + \sigma W,$$

where, with $n = n_1 + \cdots + n_K$, $H = [H_1,\dots,H_K]$,

$$X = [X_1^T,\cdots,X_K^T]^T \in \mathbb{C}^{n\times M},$$

and $P^{1/2}$ is the positive square root of the $n \times n$ diagonal matrix P having first n_1 diagonal entries equal to P_1, next n_2 diagonal matrices equal to P_2, etc.

The goal is to estimate the P_k's. Notice Y is the first N rows of

$$\begin{pmatrix} HP^{1/2} & \sigma I_N \\ 0_1 & 0_2 \end{pmatrix} \begin{pmatrix} X \\ W \end{pmatrix},$$

(I_N $N\times N$ identity matrix, 0_1, $n\times n$, 0_2 $n\times N$ zero matrices) so with the weakening of assumptions on the entries of X_n in Silverstein (1995) and Bai and Silverstein (1998),(1999), the previous results apply. In Couillet, Silverstein, Bai, and Debbah the following are proven:

Theorem. Assume σ and K are fixed, $M/N \to c > 0$, and each $N/n_k \to c_k > 0$, as $N \to \infty$. Let $B_N = (1/M)YY^*$. Then, almost surely, F^{B_N} converges in distribution, as $N \to \infty$, to a (nonrandom) p.d.f., whose Stieltjes transform, $m_F(z)$ ($z \in \mathbb{C}^+$) satisfies

$$m_F(z) = cm_{\underline{F}}(z) + (c-1)\frac{1}{z},$$

where $m_{\underline{F}}$ is the unique solution with positive imaginary part to the equation

$$\frac{1}{m_{\underline{F}}} = -\sigma^2 + \frac{1}{f} - \sum_{k=1}^{K} \frac{1}{c_k} \frac{P_k}{1+P_k f}$$

with

$$f = (1-c)m_{\underline{F}} - czm_{\underline{F}}^2.$$

Theorem. Assuming $M > N$, $n < N$, $P_1 < P_2 < \cdots < P_K$, and certain assumptions on the size of c, and the c_k's, exact separation occurs. Let λ_i denote the

i-th smallest eigenvalue of B_N and $s = (\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})^T$. Then with probability 1 $\hat{P}_k \to P_k$ as $N \to \infty$ where

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i \in \mathrm{N}_k} (\eta_i - \mu_i),$$

where $\mathrm{N}_k = \{N - \sum_{i=k}^{K} n_i + 1, \ldots, N - \sum_{i=k+1}^{K} n_i\}$, the η_i's are the ordered eigenvalues of $\mathrm{diag}(\lambda_1, \ldots, \lambda_N) - (1/N)ss^*$, and the μ_i's are the ordered eigenvalues of $\mathrm{diag}(\lambda_1, \ldots, \lambda_N) - (1/M)ss^*$.

References

[1] Z.D. Bai and J.W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices*, Annals of Probability **26**(1) (1998), 316–345.

[2] Z.D. Bai and J.W. Silverstein, *Exact separation of eigenvalues of large dimensional sample covariance matrices*, Annals of Probability **27**(3) (1999), 1536–1555.

[3] R. Couillet, J.W. Silverstein, Z.D. Bai, and M. Debbah, *Eigen-inference for Energy Estimation of Multiple Sources*, to appear in IEEE Transactions on Information Theory.

[4] X. Mestre, *Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates*, IEEE Transactions on Information Theory **54**(11) (2008), 5113–5219.

[5] J.W. Silverstein, *Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices*, Journal of Multivariate Analysis **55**(2) (1995), 331–339.

Random matrices with prescribed row and column sums

Alexander Barvinok

Let us fix positive integer vectors $R = (r_1, \ldots, r_m)$ and $C = (c_1, \ldots, c_n)$, called *margins*, such that

$$r_1 + \ldots + r_m = c_1 + \ldots + c_n = N. \tag{1}$$

We consider two sets: the set $A_0(R, C)$ of $m \times n$ matrices with 0-1 entries, row sums R and column sums C and the set $A_+(R, C)$ of $m \times n$ matrices with non-negative integer entries, row sums R and column sums C. A well-known theorem of Gale and Ryser states the necessary and sufficient conditions for $A_0(R, C)$ to be non-empty. The set $A_+(R, C)$ is non-empty as long as the balance condition (1) holds. Assuming that $A_0(R, C)$ and $A_+(R, C)$ are non-empty, we consider the sets as finite probability spaces with the uniform measure. We are interested in what a random matrix $D \in A_0(R, C)$ and a random matrix $D \in A_+(R, C)$ are likely to look like.

We recall that a random variable x is *Bernoulli* if

$$\mathbf{P}(x = 1) = p \quad \text{and} \quad \mathbf{P}(x = 0) = q$$

for some $p, q \geq 0$ such that $p + q = 1$.

Let

$$h(y) = y \ln \frac{1}{y} + (1 - y) \ln \frac{1}{1 - y} \quad \text{for} \quad 0 \leq y \leq 1$$

be the entropy of the Bernoulli random variable with expectation y. As is known, $h(y)$ is strictly concave. We extend function h to $m \times n$ matrices $Y = (y_{ij})$ such that $0 \le y_{ij} \le 1$ for all i and j by

$$H(Y) = \sum_{i,j} h\left(y_{ij}\right).$$

Given margins R and C, we consider the polytope $P_0(R, C)$ of all $m \times n$ matrices $Y = (y_{ij})$ with row sums R, column sums C and such that $0 \le y_{ij} \le 1$ for all i and j. Then $P_0(R, C) \neq \emptyset$ if and only if $A_0(R, C) \neq \emptyset$. Since H is strictly concave, it attains its maximum on $P_0(R, C)$ at a unique matrix $Z_0 = Z_0(R, C)$.

In many respects, a random matrix $D \in A_0(R, C)$ looks like the matrix of independent Bernoulli random variables with expectation Z_0.

Theorem 1. *Let us assume that $A_0(R, C) \neq \emptyset$ and let $Z_0 = Z_0(R, C)$ be the matrix constructed as above. Let $X = (x_{ij})$ be the $m \times n$ matrix of independent Bernoulli random variables such that $\mathbf{E}(X) = Z_0$. Then*

(1) *For every $D \in A_0(R, C)$ we have*

$$\mathbf{P}(X = D) = e^{-H(Z_0)}.$$

In particular, the distribution of X conditioned on $A_0(R, C)$ is uniform.

(2) *We have*

$$\mathbf{P}\left(X \in A_0(R, C)\right) \;\ge\; (mn)^{-\gamma(m+n)}$$

for some absolute constant $\gamma > 0$.

Theorem 1 implies that any event which is sufficiently rare for the matrix X of independent Bernoulli random variables with $\mathbf{E}(X) = Z_0$ is also rare for a random $D \in A_0(R, C)$. In particular, it follows that asymptotically, as m and n grow, we have $\mathbf{E}(D) \approx Z_0$. Similarly, the sum of a set of, say, 1% of the entries of a random $D \in A_0(R, C)$ is likely to be very close to the sum of the corresponding entries of Z_0 as m and n grow.

The distribution of X is the maximum entropy distribution among those supported on the set of $m \times n$ matrices with 0-1 entries and the expectation in the affine subspace of the $m \times n$ real matrices with row sums R and column sums C.

We recall that a random variable x is *geometric* if

$$\mathbf{P}(x = k) = pq^k \quad \text{for} \quad k = 0, 1, 2, \ldots$$

and some $p, q \ge 0$ such that $p + q = 1$.

Let

$$g(y) = (y + 1)\ln(1 + y) - y \ln y \quad \text{for} \quad y \ge 0$$

be the entropy of the geometric random variable with expectation y. Then $g(y)$ is strictly concave. We extend function g to $m \times n$ non-negative matrices $Y = (y_{ij})$ by

$$G(Y) = \sum_{i,j} g\left(y_{ij}\right).$$

Given margins R and C, we consider the polytope $P_+(R,C)$ of all $m \times n$ non-negative matrices $Y = (y_{ij})$ with row sums R and column sums C. Since G is strictly concave, it attains its maximum on $P_+(R,C)$ at a unique matrix $Z_+ = Z_+(R,C)$.

In many respects, a random $D \in A_+(R,C)$ looks like the matrix of independent geometric random variables with expectation Z_+.

Theorem 2. *Let $Z_+ = Z_+(R,C)$ be the matrix constructed as above. Let $X = (x_{ij})$ be the $m \times n$ matrix of independent geometric random variables such that $\mathbf{E}(X) = Z_+$. Then*

(1) *For every $D \in A_+(R,C)$ we have*

$$\mathbf{P}(X = D) = e^{-G(Z_+)}.$$

In particular, the distribution of X conditioned on $A_+(R,C)$ is uniform.

(2) *We have*

$$\mathbf{P}\left(X \in A_+(R,C)\right) \;\geq\; N^{-\gamma(m+n)}$$

for some absolute constant $\gamma > 0$, where N is the total sum of the margins from the balance condition (1).

Theorem 2 implies that any event which is sufficiently rare for the matrix X of independent geometric random variables with $\mathbf{E}(X) = Z_+$ is also rare for a random $D \in A_+(R,C)$. In particular, it follows that asymptotically, as m and n grow, we have $\mathbf{E}(D) \approx Z_+$. Similarly, the sum of a set of, say, 1% of the entries of a random $D \in A_+(R,C)$ is likely to be very close to the sum of the corresponding entries of Z_+.

The distribution of X is the maximum entropy distribution among those supported on the set of $m \times n$ matrices with non-negative integer entries and the expectation in the affine subspace of $m \times n$ real matrices with row sums R and column sums C.

A survey, containing sketches of proofs of Theorems 1 and 2 as well as references to detailed proofs can be found in [1]. Connections to the maximum entropy principle are discussed in [2].

One interesting question is whether the distribution of an *individual entry* of a random $D \in A_+(R,C)$ is asymptotically geometric as m and n grow. In a related development, Chatterjee, Diaconis and Sly showed that the properly scaled entry of a random $n \times n$ doubly stochastic matrix (a non-negative matrix with all row and column sums equal to 1) is asymptotically exponential as n grows [3]. The question seems to be somewhat harder in the absence of symmetry.

Here is one curious phenomenon. Suppose that $m = n$ and that $R = C = (3n, n, \ldots, n)$. One can show that the $(1,1)$ entry of the matrix Z_+ of Theorem 2 grows linearly in n (it remains bounded by a constant if $3n$ is replaced by $2n$). Is it true that the expectation of the corresponding entry of a random $D \in A_+(R,C)$ also grows linearly in n? Our methods seem to be too crude to answer this question.

References

[1] A. Barvinok, *Matrices with prescribed row and column sums*, preprint `arXiv:10105706`, 2010, to appear in Linear Algebra and its Applications.

[2] A. Barvinok and J.A. Hartigan, *Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes*, Advances in Applied Mathematics **45** (2010), 252–289.

[3] S. Chatterjee, P. Diaconis and A. Sly, *Properties of uniform doubly stochastic matrices*, preprint `arXiv:1010.6136`, 2010.

Stein's method and multivariate normal approximation for random matrices from the compact classical groups

Michael Stolz

(joint work with Christian Döbler)

One aspect of random matrix theory concerns random elements of compact Lie groups. A classical result, published by Persi Diaconis and Mehrdad Shahshahani [2] in 1994, is as follows: Let M_n be an element of $\mathrm{U}_n, \mathrm{O}_n$, or USp_{2n}, distributed according to Haar measure. Then, as $n \to \infty$, the vector

$$(\mathrm{Tr}(M_n), \mathrm{Tr}(M_n^2), \ldots, \mathrm{Tr}(M_n^d))$$

converges weakly to a vector of independent, (real or complex) Gaussian random variables. The proof deduced this from exact moment formulae, valid for n sufficiently large (see also [13]). A different approach to these moment computations, also taking care of SO_n, has been proposed in [10].

Subsequently, the speed of convergence in the univariate version of this result was studied by Charles Stein [12] (in the orthogonal case), who proved that the error decreases faster than any power of the dimension, and Kurt Johansson [5], who obtained exponential convergence. While Johansson's approach had its roots in Szegö's limit theorem for Toeplitz determinants, Stein used a set of techniques that he had been developing since 1972 (see [11]) and that nowadays is referred to as "Stein's method". It basically exploits the observation that the equality

$$\mathbb{E}(f'(X)) = \mathbb{E}(Xf(X))$$

for a suitable class of test functions characterizes a random variable X as standard normal. In the "exchangeable pairs" version of this method, which Stein used in his work on traces of powers, a random variable W_n, that is to be proven close to normal in the large n limit, is considered together with a random variable W_n' such that (W_n, W_n') has the same distribution as (W_n', W_n).

Recently, Jason Fulman [4] has proposed a different construction of an exchangeable pair in the traces of powers setting, based on the reversibility of Brownian motion on compact Lie groups with respect to Haar measure. Combined with moment formulae from the proof of the original Diaconis-Shahshahani result and a formula on the action of the Laplace-Baltrami operator on power sum symmetric polynomials that is based on Schur-Weyl duality (see [8, 6]), this construction leads to a transparent proof for all classical groups that seems to lend itself to

various generalizations. On the other hand, the rate is only $O(j/n)$, where the power j in question may depend on the dimension n.

In joint work with Christian Döbler [3], Fulman's approach has been extended to the multivariate case, yielding, to the best of our knowledge, the first speed of convergence results in the multivariate central limit theorem for traces of powers. Multivariate versions of the exchangeable pair method have become available only recently in work of Sourav Chatterjee and Elizabeth Meckes [1] on the one hand, and Gesine Reinert and Adrian Röllin [9] on the other. There are substantial differences between the univariate and multivariate versions of Stein's method, since a first order Stein type characterization of the multivariate normal distribution seems not to be available. A further refinement provided by Meckes ([7]) (and a complex extension thereof) proves suitable for the traces of powers problem. In the end, one obtains for $\mathrm{U}_n, \mathrm{SO}_n$ and USp_{2n} bounds of order $O(d^{7/2}/n)$ in Wasserstein distance between the vector $(\mathrm{Tr}(M_n), \mathrm{Tr}(M_n^2), \ldots, \mathrm{Tr}(M_n^d))$ and a (real or complex) d-dimensional normal distribution with covariance matrix $\mathrm{diag}(1, 2, \ldots, d)$. Here d may vary with n. In the orthogonal and symplectic cases, the traces of even powers have to be shifted by -1 and $+1$, respectively.

References

[1] Sourav Chatterjee and Elizabeth Meckes. *Multivariate normal approximation using exchangeable pairs,* ALEA Lat. Am. J. Probab. Math. Stat. **4** (2008), 257–283.

[2] Persi Diaconis and Mehrdad Shahshahani, *On the eigenvalues of random matrices*, J. Appl. Probab. **31A** (1994), 49–62.

[3] Christian Döbler and Michael Stolz, *Stein's method and the multivariate CLT for traces of powers on the classical compact groups*, 2010, available on arXiv.org:1012.3730.

[4] Jason Fulman, *Stein's method, heat kernel, and traces of powers of elements of compact Lie groups*, 2010, available on arXiv.org:1005.1306.

[5] Kurt Johansson. *On random matrices from the compact classical groups*, Ann. of Math. (2) **145** (1997), 519–545.

[6] Thierry Lévy, *Schur-Weyl duality and the heat kernel measure on the unitary group*, Adv. Math. **218** (2008), 537–575.

[7] Elizabeth Meckes. *On Stein's method for multivariate normal approximation*, In: High dimensional probability **V**: The Luminy volume, pages 153–158. IMS, 2009.

[8] E. M. Rains, *Combinatorial properties of Brownian motion on the compact classical groups*, J. Theoret. Probab. **10** (1997), 659–679

[9] Gesine Reinert and Adrian Röllin, *Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition.* Ann. Probab. **37** (2009), 2150–2173.

[10] L. Pastur and V. Vasilchuk, *On the moments of traces of matrices of classical groups*, Comm. Math. Phys. **252** (2004), 149–166.

[11] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, 583–602.

[12] Charles Stein, *The accuracy of the normal approximation to the distribution of the traces of powers of random orthogonal matrices*, Department of Statistics, Stanford University, Technical Report No. **470**, 1995

[13] Michael Stolz, *On the Diaconis-Shahshahani method in random matrix theory*, J. Algebraic Combin. **22** (2005), 471–491.

Reporter: Radosław Adamczak

Participants

Prof. Dr. Radoslaw Adamczak
Instytut Matematyki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
POLAND

Prof. Dr. Greg W. Anderson
School of Mathematics
University of Minnesota
127 Vincent Hall
206 Church Street S. E.
Minneapolis MN 55455-0436
USA

Prof. Dr. Shiri Artstein-Avidan
School of Mathematical Sciences
Tel Aviv University
Ramat Aviv
Tel Aviv 69978
ISRAEL

Prof. Dr. Guillaume Aubrun
Institut Camille Jordan
Universite Claude Bernard Lyon 1
43 blvd. du 11 novembre 1918
F-69622 Villeurbanne Cedex

Prof. Dr. Alexander Barvinok
Department of Mathematics
The University of Michigan
2074 East Hall
525 E. University Ave.
Ann Arbor MI 48109-1043
USA

Dr. Charles Bordenave
Institut de Mathematiques de Toulouse
Universite Paul Sabatier
118, route de Narbonne
F-31062 Toulouse Cedex 9

Prof. Dr. Djalil Chafai
Marne-la-Vallee University
Faculte de Mathematiques
5, Blvd. Descartes
Champs-sur-Marne
F-77454 Marne-la-Vallee Cedex 2

Dr. Kevin P. Costello
School of Mathematics
Georgia Institute of Technology
686 Cherry Street
Atlanta , GA 30332-0160
USA

Sandrine Dallaporta
Laboratoire de Statistique et
Probabilites
Universite Paul Sabatier
118, route de Narbonne
F-31062 Toulouse Cedex 4

Dr. Hanna Döring
Fachbereich Mathematik
Technische Universität Berlin
Straße des 17. Juni 135
10623 Berlin

Prof. Dr. Peter Eichelsbacher
Fakultät für Mathematik
Ruhr-Universität Bochum
NA 3/66
44780 Bochum

Prof. Dr. Noureddine El Karoui
Department of Statistics
University of California, Berkeley
367 Evans Hall
Berkeley CA 94720-3860
USA

Dr. Brendan Farrell
Lehrstuhl für Theoretische Informationstechnik
Technische Universität München
Karlstr. 45-47
80333 München

Dr. Omer Friedland
Institut de Mathematiques
Projet Analyse Fonctionnelle
Boite 186, Universite Paris 6
4, place Jussieu
F-75005 Paris

Prof. Dr. Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld

Prof. Dr. Yehoram Gordon
Department of Mathematics
Technion - Israel Institute of Technology
Haifa 32000
ISRAEL

Dr. Olivier Guedon
Laboratoire d'Analyse et Math. Appliq.
Universite de Paris-Est Marne-la-Vallee
Cite Descartes, 5 BD Descartes
Champs-sur-Marne
F-77454 Marne-La-Vallee Cedex

Prof. Dr. William B. Johnson
Department of Mathematics
Texas A & M University
College Station , TX 77843-3368
USA

Dr. Subhash A. Khot
Courant Institute of Mathematical Sciences
New York University
251, Mercer Street
New York , NY 10012-1110
USA

Prof. Dr. Bo'az Klartag
Department of Mathematics
School of Mathematical Sciences
Tel Aviv University
Ramat-Aviv, Tel Aviv 69978
ISRAEL

Dr. Holger Kösters
Fakultät für Mathematik
Universität Bielefeld
Universitätsstr. 25
33615 Bielefeld

Prof. Dr. Manjunath Krishnapur
Department of Mathematics
Indian Institute of Science
Bangalore 560 012
INDIA

Prof. Dr. Rafal Latala
Instytut Matematyki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
POLAND

Prof. Dr. Michel Ledoux
Institut de Mathematiques de Toulouse
Universite de Toulouse
F-31062 Toulouse

Prof. Dr. Alexander Litvak
Dept. of Mathematical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1
CANADA

Prof. Dr. Wenbo V. Li
Department of Mathematical Sciences
University of Delaware
501 Ewing Hall
Newark , DE 19716-2553
USA

Prof. Dr. Mark W. Meckes
Department of Mathematics
Case Western Reserve University
10900 Euclid Avenue
Cleveland OH 44106-7058
USA

Prof. Dr. Krzysztof Oleszkiewicz
Instytut Matematyki
Uniwersytet Warszawski
ul. Banacha 2
02-097 Warszawa
POLAND

Joscha Prochno
Mathematisches Seminar
Christian-Albrechts-Universität Kiel
Ludewig-Meyn-Str. 4
24098 Kiel

Prof. Dr. Brian Rider
Department of Mathematics
University of Colorado
Boulder , CO 80309-0395
USA

Stiene Riemer
Mathematisches Seminar
Christian-Albrechts-Universität Kiel
Ludewig-Meyn-Str. 4
24098 Kiel

Prof. Dr. Mark Rudelson
Department of Mathematics
University of Michigan
530 Church Street
Ann Arbor , MI 48109-1043
USA

Prof. Dr. Gideon Schechtman
Department of Mathematics
The Weizmann Institute of Science
P. O. Box 26
Rehovot 76 100
ISRAEL

Prof. Dr. Carsten Schütt
Mathematisches Seminar
Christian-Albrechts-Universität Kiel
Ludewig-Meyn-Str. 4
24098 Kiel

Prof. Dr. Jack W. Silverstein
Department of Mathematics
North Carolina State University
Campus Box 8205
Raleigh , NC 27695-8205
USA

Prof. Dr. Miklos Simonovits
Alfred Renyi Institute of Mathematics
Hungarian Academy of Sciences
P.O.Box 127
H-1364 Budapest

Dr. Nikhil Srivastava
School of Mathematics
Institute for Advanced Study
1 Einstein Drive
Princeton , NJ 08540
USA

Dr. Michael Stolz
Fakultät für Mathematik
Ruhr-Universität Bochum
NA 3/66
44780 Bochum

Prof. Dr. Stanislaw Jerzy Szarek
Department of Mathematics
Case Western Reserve University
10900 Euclid Avenue
Cleveland OH 44106-7058
USA

Prof. Dr. Alexander Tikhomirov
Mathematical Department
KOMI Science Center
Ural Div.,Russian Acad. of Science
Chernova St. 3a
Syktyvkar 167982
RUSSIA

Prof. Dr. Nicole Tomczak-Jaegermann
Dept. of Mathematical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1
CANADA

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

Martin Venker
Fakultät für Mathematik
Universität Bielefeld
Universitätsstr. 25
33615 Bielefeld

Prof. Dr. Roman Vershynin
Department of Mathematics
University of Michigan
East Hall, 525 E. University
Ann Arbor , MI 48109-1109
USA

Prof. Dr. Van H. Vu
Department of Mathematics
Rutgers University
Hill Center, Busch Campus
110 Frelinghuysen Road
Piscataway , NJ 08854-8019
USA

Ying Xiao
School of Electrical and
Computer Engineering
Georgia Institute of Technology
Atlanta , GA 30332-0250
USA

Dr. Jun Yin
Department of Mathematics
Harvard University
Science Center
One Oxford Street
Cambridge MA 02138-2901
USA

Prof. Dr. Shuheng Zhou
Department of Statistics
University of Michigan
439 West Hall
1085 South University
Ann Arbor MI 48109-1107
USA