

# RESEARCH STATEMENT

ROMAN VERSHYNIN

My research interests are in the areas of high-dimensional probability and mathematical data science. They span a broad range of problems extending from fundamental, theoretical questions in probability to applications in various branches of data science. My research focuses on random geometric structures that appear in random matrix theory, geometric functional analysis, convex and discrete geometry, geometric combinatorics, high dimensional statistics, information theory, learning theory, signal processing, numerical analysis, and network science.

My work has created new connections between high-dimensional probability and various other branches of mathematics and data science, such as computational geometry [26, 53, 28], high-dimensional statistics [25, 27, 13], sparse signal recovery [34, 20, 21, 24] and computer science [2, 8, 32], and it led to applications in biomedical research (see e.g. [3, 6]).

I will now briefly mention some of the major directions of my research in an approximately chronological order. I will skip all the history and partial results on these problems, which you will find in the actual publications cited here.

## 1. RANDOM PROCESSES AND THEIR COMBINATORICS

In a series of papers [19, 31, 46], we developed a combinatorial theory of stochastic processes for general classes of functions. The classical results on VC dimension form a cornerstone of statistical learning theory. They were developed in 1970-80's for classes of *Boolean*, or 0/1 valued, functions. Due to the combinatorial flavor of the VC theory, attempts to extend it for general *real-valued* functions had met serious technical challenges. Our work was the first to achieve a lossless extension of the VC theory to general functions. It helped to bridge combinatorial, analytical and probabilistic approaches to stochastic processes.

In particular, [19] established an optimal connection between the metric entropy and combinatorial dimension. In [31] we solved a couple of basic conjectures on empirical processes, showing that the central limit theorem holds uniformly over a given class of functions  $\mathcal{F}$  if the square root of the combinatorial dimension of  $\mathcal{F}$  is integrable. Our results and methods led to a range of applications to geometric functional analysis [19], probability, convex geometry and operator theory [31, 46] and geometry of numbers [46].

## 2. RANDOM MATRICES

During the last decade, I was heavily involved in the development of a theory of random matrices in the non-asymptotic regime, whose results would be valid for matrices or fixed dimensions (as opposed to in the limit), and which has a tremendous potential for applications. My key achievements in this area include the optimal results on the invertibility of random matrices [35, 37, 50, 39], sampling in random matrices [32, 51, 12], and delocalization of eigenvectors [41, 40, 39]. My ICM survey [38] and the tutorial [52] describe some of these developments in an educational way. Below I will touch upon two of these areas – invertibility and delocalization.

The fundamental *invertibility problem* asks whether random matrices are well conditioned. The goal here is to bound the condition number of a random matrix, which is known to be a guarantee in randomized algorithms of numerical linear algebra. The condition number is the ratio of the largest to the smallest singular values. The largest singular value is the easier one to bound, but this can still be met with considerable challenges; see e.g. my work [49, 29]. The smallest singular value serves as a quantitative indicator of invertibility, and it is even more challenging to analyze. I addressed the invertibility problem in [18, 35, 37, 36, 49, 38, 50, 39, 4].

We found optimal bounds on the smallest singular values of square random matrices with independent entries [35] and extended this result for rectangular random matrices in [37, 49]. Along the way, we developed an arithmetic structural theory for random vectors and subspaces, which can be viewed as an extension of Littlewood-Offord theory for random sums. For random matrices with dependent entries, the invertibility problem is even more arduous. We established quantitative invertibility for symmetric random matrices in [50], and for random unitary and orthogonal perturbations of fixed matrices, in [39].

Eigenvectors of random matrices are harder to analyze than eigenvalues, and less is known about them. In [41] and [40] we established *delocalization* of eigenvectors of general random matrices. Intuitively, delocalization means that the energy ( $L^2$  norm) is spread evenly among the coefficients of all eigenvectors. Proving delocalization is challenging even for symmetric random matrices, and it had been unknown for matrices with all independent entries. In [41] we developed a completely new, geometric approach to delocalization, and showed for the first time that eigenvectors of a random matrix with all independent entries are completely delocalized. We found a different geometric, non-spectral approach in [40], which leads to delocalization for many ensembles of random matrices at once, simultaneously treating symmetric, skew-symmetric, and all-independent random matrices.

## 3. COMPRESSED SENSING

I have contributed to the development of the young area of *compressed sensing* since its inception [2, 30, 34, 33, 8, 7, 21, 20, 22, 24, 25, 26, 27, 28, 9, 1]. The field of compressed sensing lies at the crossroads of signal processing, learning theory, information theory

and numerical linear algebra, and it draws insights from probability theory, statistics, theoretical computer science, and geometric functional analysis.

We gave a benchmark proof of the restricted isometry property for *random Fourier matrices* in [34]. Our method based on stochastic processes has become a starting point for many later developments in the area; it led to various extensions and applications for sparse signal recovery from structured measurements. The book [5] describes some of these consequences; our paper [34] has 574 citations according to Google Scholar.

We developed the first *iterative algorithm for sparse recovery* with almost optimal (linear) number of measurements [20, 21]. This work reconciled the two main algorithmic approaches to compressed sensing – optimization-based and iterative methods. Our work inspired numerous improvements, ramifications, and extensions in the areas of compressed sensing and signal recovery. Together with the follow-up paper [23] by Joel Tropp and my former student Deanna Needell, these papers attracted over 4,000 citations.

In [24, 25, 1], we developed an algorithmic theory of *1-bit compressed sensing* for extremely quantized, 0/1 valued measurements. We found that, surprisingly, such severe quantization has almost no effect on the required number of measurements, and the recovery can be achieved by linear programming. We established a connection of 1-bit compressed sensing with a wonderful problem in discrete geometry, which asks how many hyperplane cuts are needed to break a given set  $K$  in  $\mathbb{R}^n$  into small pieces. We found an almost optimal bound on the number of cuts in terms of the Gaussian width of  $K$  [26], and related it to problems of coding and dimension reduction.

#### 4. HIGH-DIMENSIONAL INFERENCE

Our results on 1-bit compressed sensing extend beyond the area of signal processing. They led to the first rigorous and algorithmic analysis of high-dimensional *logistic regression* in [25] and *generalized linear models* in [28]. A further step was made in [27] where we showed that the classical methods for structured linear regression such as Lasso work provably and without change even for *non-linear data*. The non-linearity can be very general: discontinuous, not one-to-one, and even unknown. In spite of this, we found that Lasso stays unharmed even in presence of such nonlinearities. This dramatically extended the range of statistical models for which data analysis can be done rigorously. Our work led to cross-fertilization of several branches of information theory, high-dimensional statistics [27], sparse signal recovery, quantization [10], statistical learning [16], computational geometry [26], and high-dimensional probability [15]. The survey [53] puts these developments in the context of high-dimensional geometry.

Very recently, we found in [15] a particularly general and pretty *matrix deviation inequality*, which quantifies how much a random matrix distorts vectors from a given geometric set  $T$  in  $\mathbb{R}^n$ . One can use this inequality to recover many previously known results and prove new ones in dimension reduction (far-reaching generalizations of Johnson-Lindenstrauss lemma), compressed sensing (restricted isometry property), high-dimensional statistics (covariance estimation, model selection for constrained linear models), random

matrix theory (bounds on extreme singular values), and high-dimensional geometry (general escape theorems and  $M^*$  bounds). I feel that the matrix deviation inequality will become a standard and useful tool in high-dimensional probability and mathematical data sciences.

*Covariance estimation* of high-dimensional distributions is a basic problem in statistics. For a general distribution in  $\mathbb{R}^n$ , one needs  $O(n \log n)$  samples for covariance estimation, and the common belief is that one can do away with logarithmic oversampling for “most” distributions. In [51, 45] we showed that this is indeed the case if the distribution has finite  $2 + \varepsilon$  moments. While dimension reasons prohibit covariance estimation with fewer than  $n$  samples, we showed in [12] how one can still have sub-linear number of samples to estimate *sparse covariance matrices*.

## 5. NETWORKS

In the area of network analysis, my recent work contributed to the development of some of the first algorithmic methods for *finding structure in extremely sparse networks* [14, 11, 13].

Classical methods, such as spectral algorithms, are known to fail for sparse networks due to a wild variation of the degrees of the nodes. For example, the vertices of abnormally high degrees (“hubs”) cause the failure of concentration in the adjacency matrix, while the vertices of abnormally low degrees (e.g. isolated vertices, leaves of trees) prevent the Laplacian matrix to concentrate. We found in [14] a very general and simple regularization recipe to circumvent this failure. If one lowers the weights of the edges connected to the hubs, the adjacency matrix will provably concentrate. Alternatively, one can boost the weights of the edges slightly, which amounts to allowing random hops – just like in *PageRank algorithm*. We proved that this forces the Laplacian matrix to concentrate. These results rigorously justify the empirical predictions of network scientists and enable some of the most basic spectral methods (like spectral clustering) for data mining in sparse networks.

Our paper [11] made it possible for the first time to analyze *semidefinite programming* methods for structure discovery in sparse networks. Our argument featured the first application of the fundamental Grothendieck’s inequality to give an arbitrarily fine approximation to the hidden structure of a network. Both of our approaches for structure mining in sparse networks are applicable for a far wider class of networks than the benchmark class of stochastic block models in network science.

## 6. OTHER WORK ON ALGORITHMS

In [32] we studied *sampling from matrices* in numerical linear algebra. We demonstrated how to find a good approximation to a matrix  $A$  based on a random submatrix of  $A$ . Our work improved upon known algorithms for computing low rank approximations, singular value decomposition, and approximations to MAX-CSP problems.

My work [48, 47] simplified and improved the *smoothed analysis of the simplex algorithm* in linear programming. Smoothed analysis is concerned with the expected running time of an algorithm under slight random perturbations of arbitrary inputs. We showed that the length of walk in the simplex method is *poly-logarithmic* in the number of constraints  $n$  in the linear program. The previously known best bound, due to D. Spielman and S.-H. Teng, was polynomial in  $n$ . From a geometric perspective, this work demonstrated how random perturbations or polytopes create very short paths between vertices, thus beating the generic bound on the diameter of polytopes given by Hirsch conjecture.

In [42, 43, 44] we gave the first analysis of a randomized version of the classical *Kaczmarz algorithm* for solving system of linear equations. This simple algorithm constructs a sequence of approximations to the solution by iteratively projecting a point onto the solution spaces of each equation. The previously known guarantees for Kaczmarz algorithm had been pessimistic and complicated. We demonstrated how to exponentially speed-up Kaczmarz algorithm and give very simple guarantees in terms of the condition number. All one has to do is to *randomly choose* which equation to project onto at each iteration step. In addition to the speed-up, the randomization also considerably simplifies the analysis of the algorithm.

## 7. SOME BROADER IMPACTS OF MY RESEARCH

I have written a number of *educational texts* in order to make modern mathematical methods a standard toolkit for applied scientists. My tutorial on non-asymptotic methods in random matrix theory [52] is a standard text for beginning graduate students in various quantitative areas. My recent survey [53] promotes building bridges between high-dimensional probability and high-dimensional inference.

I am writing the first *textbook on high-dimensional probability with applications in data science*; you can download a draft at [here](#):

<http://www-personal.umich.edu/~romanv/papers/HDP-book/HDP-book.html>

Graduate students in various areas of data science are experiencing a significant gap in their preparation for modern research. The basic probabilistic methods that underlie mathematically rigorous work in data sciences are not covered systematically by any existing textbook. In the time when data sciences are moving forward very fast, the gap is widening between the standard university probability curriculum and the mathematical sophistication that is expected from a beginning researcher. A textbook written for a sufficiently general audience and on a basic level would help to close this gap immensely. The book in preparation would be the first to do it. It would emphasize the analysis of high-dimensional random structures, such as high dimensional distributions, random vectors, and random matrices – the structures form the mathematical skeleton of modern data sciences. The book will feature applications in network theory, computer science, high dimensional statistics, signal processing and statistical learning theory. The textbook will be intended for graduate students in mathematics (pure and applied) and all of the data sciences, including statistics, electrical engineering, and theoretical computer science.

Together with my five distinguished colleagues – Luc Devroye, Gabor Lugosi, Elchanan Mossel, Mike Steele, and Alexandre Tsybakov – we are starting up a new journal *Mathematical Statistics and Learning*. The journal will be published by the European Mathematical Society, and its announcement will be made soon. The focus of the new journal will be on new and important mathematical ideas and techniques that are inspired by applications. Surprisingly as it may sound, this will be the first journal in the area of mathematical statistics. It will be devoted to research articles of the highest quality in all aspects of mathematical statistics and learning, including those studied in theoretical computer science and signal processing.

An example of my contributions to *training a new generation of researchers* is my former Ph.D. student *Deanna Needell*. Her research career started by our joint discovery of almost optimal iterative methods for compressed sensing [20, 21, 22]. After that, Deanna went on to become one of the leaders in *compressed sensing* with her independent research program. She received dozens of awards for her work including Sloan Research Fellowship, NSF CAREER Award and 2016 IMA Prize. With over 5,600 citations according to Google Scholar, she is one of the most cited young mathematicians, and possibly the most cited female mathematician of her generation. Deanna has been a great role model as a professor at Claremont McKenna College and UCLA.

I am affiliated with *Michigan Institute for Data Science*, an interdisciplinary enterprise that fosters connections among mathematicians, statisticians, computer scientists, researchers in biomedical fields, and it facilitates transition of fundamental research ideas to address important problems in science, engineering and business.

#### REFERENCES

- [1] A. Ai, A. Lapanowski, Y. Plan, R. Vershynin, *One-bit compressed sensing with non-Gaussian measurements*, Linear Algebra and Applications 441 (2014), 222–239.
- [2] E. Candes, M. Rudelson, T. Tao, R. Vershynin, *Error correction via Linear Programming*, FOCS 2005 (46th Annual Symposium on Foundations of Computer Science), 668–681.
- [3] S. Chretien, C. Guyeux, M. Boyer-Guittaut, R. Delage-Mouroux, F. Descotes, *Investigating gene expression array with outliers and missing data in bladder cancer*, preprint.
- [4] B. Farrell, R. Vershynin, *Smoothed analysis of symmetric random matrices with continuous distributions*, Proceedings of the AMS 144 (2016), 2257–2261.
- [5] S. Foucart, H. Rauhut, *A mathematical introduction to compressive sensing*. Birkhäuser/Springer, New York, 2013.
- [6] M. Genzel, G. Kutyniok, *A mathematical framework for feature selection from real-world data with non-linear observations*, preprint.
- [7] A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, *Algorithmic linear dimension reduction in the  $L^1$  norm for sparse vectors*, Allerton 2006 (44th Annual Allerton Conference on Communication, Control, and Computing), invited paper.

- [8] A. Gilbert, M. Strauss, J. Tropp, R. Vershynin, *One sketch for all: fast algorithms for compressed sensing*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 237–246, ACM, New York, 2007.
- [9] R. Giryes, Y. Plan, R. Vershynin, *On the effective measure of dimension in analysis cosparse model*, IEEE Transactions on Information Theory 61 (2015), 5745–5753.
- [10] X. Gu, S. Tu, H.-J. Shi, M. Case, D. Needell, Y. Plan, *Optimizing quantization for Lasso recovery*, preprint (2016).
- [11] O. Guedon, R. Vershynin, *Community detection in sparse networks via Grothendieck's inequality*, Probability Theory and Related Fields 165 (2016), 1025–1049.
- [12] E. Levina, R. Vershynin, *Partial estimation of covariance matrices*, Probability Theory and Related Fields 153 (2012), 405–419.
- [13] C. Le, E. Levina, R. Vershynin, *Optimization via low-rank approximation, with applications to community detection in networks*, Annals of Statistics 44 (2016), 373–400.
- [14] C. Le, E. Levina, R. Vershynin, *Concentration and regularization of random graphs*, Random Structures and Algorithms, to appear.
- [15] C. Liaw, A. Mehrabian, Y. Plan, R. Vershynin, *A simple tool for bounding the deviation of random matrices on geometric sets*, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, Springer, Berlin, to appear (2016).
- [16] J. Liu, J. Jia, *Vanilla Lasso for sparse classification under single index models*, preprint (2016).
- [17] Yu. Lyubarskii, R. Vershynin, *Uncertainty principles and vector quantization*, IEEE Transactions on Information Theory 56 (2010), 3491–3501.
- [18] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, *Euclidean embeddings in spaces of finite volume ratio via random matrices*, Journal für die Reine und Angewandte Mathematik 589 (2005), 1–19.
- [19] S. Mendelson, R. Vershynin, *Entropy and the combinatorial dimension*, Inventiones Mathematicae 152 (2003), 37–55.
- [20] D. Needell, R. Vershynin, *Uniform Uncertainty Principle and signal recovery via Regularized Orthogonal Matching Pursuit*, Foundations of Computational Mathematics 9 (2009), 317–334.
- [21] D. Needell, R. Vershynin, *Signal recovery from incomplete and inaccurate measurements via Regularized Orthogonal Matching Pursuit*, IEEE Journal of Selected Topics in Signal Processing 4 (2010), 310–316.
- [22] D. Needell, R. Vershynin, *Greedy signal recovery and uncertainty principles*, Computational Imaging VI, IS&T/SPIE 19th Annual Symposium, 2008. Proc. SPIE Vol. 6814, 68140J.
- [23] D. Needell, J. Tropp, *CoSaMP: iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal. 26 (2009), 301–321.
- [24] Y. Plan, R. Vershynin, *One-bit compressed sensing by linear programming*, Communications on Pure and Applied Mathematics 66 (2013), 1275–1297.
- [25] Y. Plan, R. Vershynin, *Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach*, IEEE Transactions on Information Theory 59 (2013), 482–494.

- [26] Y. Plan, R. Vershynin, *Dimension reduction by random hyperplane tessellations*, Discrete and Computational Geometry 51 (2014), 438–461.
- [27] Y. Plan, R. Vershynin, *The generalized Lasso with non-linear observations*, IEEE Transactions on Information Theory, to appear.
- [28] Y. Plan, R. Vershynin, E. Yudovina, *High-dimensional estimation with geometric constraints*, Information and Inference 0 (2016), 1–40.
- [29] E. Rebrova, R. Vershynin, *Norms of random matrices: local and global problems*, *Advances of Mathematics*, submitted (2016).
- [30] M. Rudelson, R. Vershynin, *Geometric approach to error correcting codes and reconstruction of signals*, International Mathematical Research Notices 64 (2005), 4019–4041.
- [31] M. Rudelson, R. Vershynin, *Combinatorics of random processes and sections of convex bodies*, Annals of Mathematics 164 (2006), 603–648.
- [32] M. Rudelson, R. Vershynin, *Sampling from large matrices: an approach through geometric functional analysis*, Journal of the ACM (2007), Art. 21, 19 pp.
- [33] M. Rudelson, R. Vershynin, *Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements*, CISS’06 (40th Annual Conference on Information Sciences and Systems), 2006, 207–212.
- [34] M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Communications on Pure and Applied Mathematics 61 (2008), 1025–1045.
- [35] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600–633.
- [36] M. Rudelson, R. Vershynin, *The least singular value of a random square matrix is  $O(n^{-1/2})$* , Comptes rendus de l’Académie des sciences - Mathématique 346 (2008), 893–896.
- [37] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.
- [38] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians. Volume III, 1576–1602, Hindustan Book Agency, New Delhi, 2010.
- [39] M. Rudelson, R. Vershynin, *Invertibility of random matrices: unitary and orthogonal perturbations*, Journal of the AMS 27 (2014), 293–338.
- [40] M. Rudelson, R. Vershynin, *No-gaps delocalization for general random matrices*, Geometric and Functional Analysis 26 (2016), 1716–1776.
- [41] M. Rudelson, R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Mathematical Journal 164 (2015), 2507–2538.
- [42] T. Strohmer, R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, Journal of Fourier Analysis and Applications 15 (2009), 262–278.
- [43] T. Strohmer, R. Vershynin, *A randomized solver for linear systems with exponential convergence*, Approximation, randomization and combinatorial optimization, 499–507, Lecture Notes in Comput. Sci., 4110, Springer, Berlin, 2006.
- [44] T. Strohmer, R. Vershynin, *Comments on the randomized Kaczmarz method*, Journal of Fourier Analysis and Applications 15 (2009), 437–440.

- [45] N. Srivastava, R. Vershynin, *Covariance estimation for distributions with  $2 + \varepsilon$  moments*, *Annals of Probability* 41 (2013), 3081–3111.
- [46] R. Vershynin, *Integer cells in convex sets*, *Advances in Mathematics* 197 (2005), 248–273.
- [47] R. Vershynin, *Beyond Hirsch Conjecture: walks on random polytopes and smoothed complexity of the simplex method*, *SIAM Journal on Computing* 39 (2009), 646–678 (invited paper).
- [48] R. Vershynin, *Beyond Hirsch Conjecture: walks on random polytopes and smoothed complexity of the simplex method*, *FOCS 2006 (47th Annual Symposium on Foundations of Computer Science)*, 133–142.
- [49] R. Vershynin, *Spectral norm of products of random and deterministic matrices*, *Probability Theory and Related Fields* 150 (2011), 471–509.
- [50] R. Vershynin, *Invertibility of symmetric random matrices*, *Random Structures and Algorithms* 44 (2014), 135–182.
- [51] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?* *Journal of Theoretical Probability* 25 (2012), 655–686.
- [52] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*. Compressed sensing, 210–268, Cambridge Univ. Press, Cambridge, 2012.
- [53] R. Vershynin, *Estimation in high dimensions: a geometric perspective*. Sampling Theory, a Renaissance, 3–66, Birkhauser Basel, 2015.