Phase Transitions for Restricted Isometry Properties

Jeffrey D. Blanchard
Department of Mathematics
University of Utah
Salt Lake City, UT 84112-0090
jeff@math.utah.edu

Coralia Cartis
School of Mathematics
University of Edinburgh
Edinburgh EH9 3JZ
coralia.cartis@ed.ac.uk

Jared Tanner
School of Mathematics
University of Edinburgh
Edinburgh EH9 3JZ
jared.tanner@ed.ac.uk

***Abstract*—Currently there is no framework for the transparent comparison of sparse approximation recoverability results derived using different methods of analysis. We cast some of the most recent recoverability results for ℓ_1-regularization in terms of the phase transition framework advocated by Donoho. To allow for quantitative comparisons across different methods of analysis a particular random matrix ensemble must be selected; here we focus on Gaussian random matrices. Methods of analysis considered include the Restricted Isometry Property of Candès and Tao, geometric covering arguments of Rudelson and Vershynin, and convex polytopes formulations of Donoho.**

I. Introduction

There is no widely agreed upon framework for the quantitative comparison of theoretical sparse approximation results derived from different methods of analysis. Lacking such a framework it is becoming increasingly unclear if new results are improvements over existing results, or which methods of analysis are achieving the best results. In an effort to alleviate this shortcoming we cast some of the most cited results of sparse approximation in a simple format which allow for transparent comparison. We focus on ℓ_1-regularization (2) and Gaussian random matrices, for which there is a well established literature.

Let A be an $n \times N$ matrix with $n < N$ and $x \in \mathbb{R}^N$ be a real N-dimensional vector with $k < n$ nonzero entries. Let the set of *k-sparse* vectors be denoted $\chi^N(k) = \{x \in \mathbb{R}^N : \|x\|_0 \le k\}$ where $\|\cdot\|_0$ counts the number of nonzero entries. Let $b = Ax$ and from (b, A), we seek the sparsest vector x such that $b = Ax$. Now standard in sparse approximation, we seek the sparsest solution,

$$\min \|x\|_0 \quad \text{subject to } b = Ax. \tag{1}$$

Rather than solve (1) directly through a combinatorial search, we relax the problem [1] to solving

$$\min \|x\|_1 \quad \text{subject to } b = Ax. \tag{2}$$

When the solution to (2) is identical to the solution of (1), x is called a point of *ℓ_1/ℓ_0-equivalence*. Our goal is to determine the largest sparsity level $k < n$, for a given n and N, such that every vector $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. We follow the convention advocated by Donoho in [2], [3]; with $\rho = \frac{k}{n}$ and $\delta = \frac{n}{N}$, we define regions of (δ, ρ) in which there is a high probability on the draw of a Gaussian matrix A that for large problem sizes, $(k, n, N) \to \infty$, all $x \in \chi^N(k)$ are points of ℓ_1/ℓ_0-equivalence. This region where ℓ_1/ℓ_0-equivalence typically occurs for all $x \in \chi^N(k)$ is given by (δ, ρ) for $\rho \le (1-\epsilon)\rho_S(\delta)$ for any $\epsilon > 0$; lower bounds on $\rho_S(\delta)$ provide an easily interpreted and comparable quantitative condition. We compare the best known lower bounds on the strong (ℓ_1/ℓ_0-equivalence for *all* $x \in \chi^N(k)$) phase transition curves associated with the following techniques of analysis: convex polytopes [2], geometric functional analysis [4], and the restricted isometry property [5], [6].

II. The Phase Transition Framework

In this section, we recast known results in sparse approximation in terms of lower bounds on the strong phase transition curve, $\rho_S(\delta)$, for Gaussian random matrices. The function $\rho_S(\delta)$ defines a curve below which there is exponentially high probability on the draw of a matrix A with Gaussian i.i.d. entries that every k-sparse vector is a point of ℓ_1/ℓ_0-equivalence. That is to say for any problem instance with parameters (k, n, N), if $\frac{k}{n} = \rho < (1-\epsilon)\rho_S(\delta)$ for any $\epsilon > 0$, then with high probability on the draw of a matrix A with entries drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$, every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. The lower bounds define the region in which the associated sufficient condition is satisfied with high probability on the draw of a Gaussian matrix.

Donoho [2] provided a necessary and sufficient condition on any matrix A of size $n \times N$ such that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence; namely the projection of the unit ℓ_1 ball, C^N, under A preserves all k-faces. This is the notion of k *central-neighborliness*. Using k central-neighborliness, Donoho developed a lower bound on the phase transition, $\rho_S(\delta, C^N)$, for Gaussian matrices which satisfy the k central-neighborliness condition [3] and therefore exactly recover every $x \in \chi^N(k)$ via (2). In the limit as $\delta \to 0$, the lower bound of Donoho, $\rho_S(\delta, C^N)$, approaches the true phase transition $\rho_S(\delta)$, [7].

Theorem 1 (Donoho [2])**.** *For any $\epsilon > 0$, as $(k, n, N) \to \infty$, there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho < (1-\epsilon)\rho_S(\delta, C^N)$. $\rho_S(\delta, C^N)$ is displayed as the black curve in Figure 3.*

Using covering/net techniques of geometric functional analysis, Rudelson and Vershynin [4] provided a sufficient condition under which Gaussian matrices will recover all $x \in$

$\chi^N(k)$. Here we reformulate their result in terms of a lower bound on the phase transition, $\rho_S^{RV}(\delta)$, for Gaussian matrices.

Theorem 2 (Rudelson and Vershynin [4]). *For any $\epsilon > 0$, as $(k, n, N) \to \infty$, there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho < (1-\epsilon)\rho_S^{RV}(\delta)$. $\rho_S^{RV}(\delta)$ is defined as the solution of (3).*

$$\rho = \frac{1}{12 + 8\log(1/\rho\delta)\cdot\alpha^2(\rho\delta)} \quad (3)$$

$$\textit{with} \quad \alpha(\rho\delta) := \exp\left(\frac{\log(1+2\log(e/\rho\delta))}{4\log(e/\rho\delta)}\right).$$

$\rho_S^{RV}(\delta)$ is displayed as the red curve in Figure 3.

With $\rho_S^{RV}(\delta)$, we can directly compare Theorem 2 to the central-neighborliness phase transition $\rho_S(\delta, C^N)$. The red curve in Figure 4 shows the fraction $\rho_S^{RV}(\delta)/\rho_S(\delta, C^N)$.

III. Restricted Isometry Properties

Candès and Tao [8] introduced the notion of the *Restricted Isometry Property* (RIP) and went on to prove various sufficient conditions on the matrix A such that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. These sufficient conditions impose restrictions on the RIP constants. A matrix A is said to have the *RIP constant* $R(k,n,N) = \delta_k$ when $\forall x \in \chi^N(k)$

$$R(k,n,N) := \arg\min_{c\ge 0}\, (1-c)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+c)\|x\|_2^2. \quad (4)$$

We break from the standard notation in the literature and refer to the restricted isometry constant as $R(k,n,N)$ in order to make explicit the dependence on all three problem parameters (k,n,N) and to avoid conflict with another standard notation, namely $\delta = \frac{n}{N}$.

The most recent sufficient RIP condition for ℓ_1/ℓ_0-equivalence derived by Candès is Theorem 3.

Theorem 3 (Candès [5]). *If $R(2k,n,N) < \sqrt{2}-1$ then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

The constant $R(k,n,N)$ measures the maximum deviation from unity of the smallest and largest singular values of all submatrices of A with size $n \times k$. By its symmetric definition, $R(k,n,N)$ must satisfy both inequalities in (4), taking on the value of the largest deviation from unity of the smallest and largest singular values of the $n \times k$ submatrices of A. Let $\Lambda \subset \{1,\dots,N\}, |\Lambda| = k$, be an index set selecting the k columns of A for the submatrix A_Λ. One of the central triumphs of random matrix theory is the characterization of the distribution of the eigenvalues of random Wishart matrices $A_\Lambda^* A_\Lambda$; as $(k,n,N) \to \infty$ with $\rho = \frac{k}{n}$, the expected value of the largest and smallest eigenvalues of the Wishart matrix $A_\Lambda^* A_\Lambda$ tend to $(1+\sqrt{\rho})^2$ [9] and $(1-\sqrt{\rho})^2$ [10], respectively. The asymmetric deviation from unity of the expected value of the largest and smallest eigenvalues suggests that the largest eigenvalue of the Wishart matrix $A_\Lambda^* A_\Lambda$ dominates the RIP constant $R(k,n,N)$. See Figure 1.

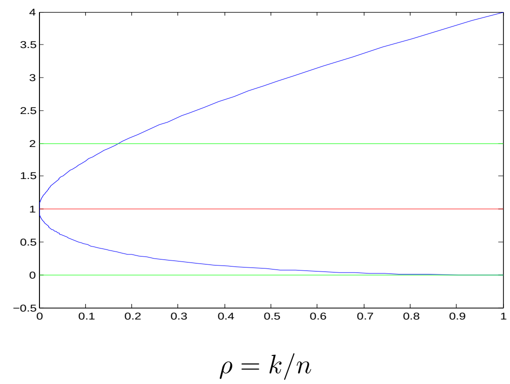

Fig. 1. The expected value of the largest and smallest eigenvalues of the random Wishart matrix $A_\Lambda^* A_\Lambda$ with A_Λ of size $n \times k$ with entries drawn i.i.d. from $\mathcal{N}(0, 1/\sqrt{n})$.

To increase the family of matrices satisfying a sufficient RIP condition, we remove the unnecessary restriction on $R(k,n,N)$ imposed by the symmetry inherent in (4). The matrix A is said to have the *asymmetric RIP constants* $L(k,n,N)$ and $U(k,n,N)$ when $\forall x \in \chi^N(k)$

$$L(k,n,N) := \arg\min_{c\ge 0}\, (1-c)\|x\|_2^2 \le \|Ax\|_2^2, \quad (5)$$

$$U(k,n,N) := \arg\min_{c\ge 0}\, (1+c)\|x\|_2^2 \ge \|Ax\|_2^2. \quad (6)$$

It is straightforward to reformulate the known RIP statements in terms of the asymmetric restricted isometry constants $L(k,n,N)$ and $U(k,n,N)$. For example, the generalization of Theorem 3 becomes:

Theorem 4. *If $(1+\sqrt{2})L(2k,n,N) + U(2k,n,N) < \sqrt{2}$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

Clearly, if $L(2k,n,N) = U(2k,n,N)$ in Theorem 4, we recover Theorem 3.

IV. Phase Transitions for RIP

Using bounds developed by Edelman [11] on the probability distribution functions for the largest and smallest eigenvalues of the $k \times k$ Wishart matrix $A_\Lambda^* A_\Lambda$, we perform a large deviation analysis on the most extreme eigenvalues of the $\binom{N}{k}$ matrices $A_\Lambda^* A_\Lambda$ derived from A. From this analysis, we derive upper bounds, $\mathcal{L}(\delta,\rho)$ and $\mathcal{U}(\delta,\rho)$, on the constants $L(k,n,N)$ and $U(k,n,N)$, satisfied with high probability on the draw of A as $(k,n,N) \to \infty$ with ρ, δ fixed [6]. These bounds permit the formulation of lower bounds on the phase transition $\rho_S(\delta)$ which hold with high probability on the draw of a matrix A with Gaussian i.i.d. entries. The lower bounds on the phase transition clearly identify the regions of (δ,ρ) such that for large problem instances (k,n,N), $\rho = \frac{k}{n}$, $\delta = \frac{n}{N}$, a Gaussian matrix A satisfies the associated sufficient RIP condition. Therefore, with high probability on the draw of A, every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. These regions allow comparison of RIP based results with those

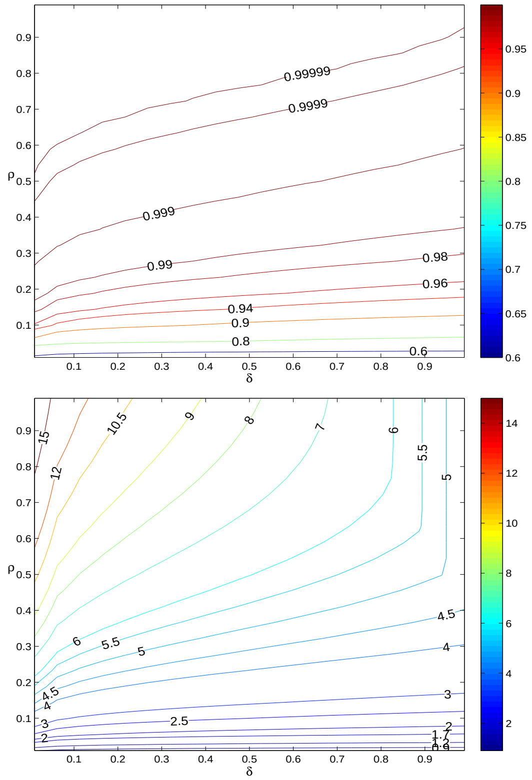

Fig. 2. Bounds, $L(\delta,\rho)$ and $U(\delta,\rho)$ (top and bottom respectively), for which it is exponentially unlikely that the RIP constants $L(k,n,N)$ and $U(k,n,N)$ will exceed; A drawn i.i.d. $N(0,1/\sqrt{n})$ and in the limit as $n\to\infty$ with $\frac{k}{n}\to\rho$ and $\frac{n}{N}\to\delta$; see Theorem 5.

obtained from other methods of analysis, such as Theorems 1 and 2.

Theorem 5 (Blanchard, Cartis, and Tanner [6])**.** *Let A be a matrix of size $n\times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/\sqrt{n})$ and let $n\to\infty$ with $\frac{k}{n}\to\rho$ and $\frac{n}{N}\to\delta$. For any $\epsilon>0$, as $n\to\infty$,*

$$\textit{Prob}\left(L(k,n,N)<L\left(\delta,\rho\right)+\epsilon\right)\to 1 \quad \textit{and}$$
$$\textit{Prob}\left(U(k,n,N)<U\left(\delta,\rho\right)+\epsilon\right)\to 1,$$

with $L(\delta,\rho)$ and $U(\delta,\rho)$ displayed in Figure 2; formulae for their calculation are presented in [6].

The details of the proof of Theorem 5 appear in [6], but a sketch is as follows. The bounds are derived using a simple union bound over all $\binom{N}{k}$ of the $k\times k$ Wishart matrices $A_\Lambda^*A_\Lambda$ formed from columns of A. Bounds on the tail behavior of the probability distribution function for the largest and smallest eigenvalues of $A_\Lambda^*A_\Lambda$ can be expressed in the form $p(n,\lambda)\exp(n\psi(\lambda,\rho))$ with p a polynomial in n,λ and ψ defined in [6]. Following standard practices in large deviation analysis, the tails of the probability distribution functions are balanced against the exponentially large number of Wishart matrices to define upper and lower bounds on the most extreme eigenvalues of all $\binom{N}{k}$ Wishart matrices, with bounds $\lambda_{min}(\delta,\rho)$ and $\lambda_{max}(\delta,\rho)$, respectively. Overestimation of the union bound over the combinatorial number of $\binom{N}{k}$ Wishart matrices cause these bounds to be pessimistic; however they appear to be the best available bounds at the time of writing. The asymptotic bounds of the asymmetric RIP constants, $L(\delta,\rho)$ and $U(\delta,\rho)$, follow directly.

The symmetry inherent in (4) implies that $R(k,n,N)=\max\{L(k,n,N),U(k,n,N)\}$. Therefore, Theorem 5 also provides an upper bound for the symmetric RIP constant $R(k,n,N)$; clearly $R(\delta,\rho)$ is the maximum of the bounds $L(\delta,\rho)$ and $U(\delta,\rho)$.

Corollary 1. *Let A be a matrix of size $n\times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/\sqrt{n})$, let $n\to\infty$ with $\frac{k}{n}\to\rho$ and $\frac{n}{N}\to\delta$, and let $L(\delta,\rho)$ and $U(\delta,\rho)$ be defined as in Theorem 5. Define $R(\delta,\rho):=\max\{L(\delta,\rho),U(\delta,\rho)\}$. For any $\epsilon>0$, as $n\to\infty$,*

$$\textit{Prob}\left(R(k,n,N)<R\left(\delta,\rho\right)+\epsilon\right)\to 1.$$

For all $\rho=\frac{k}{n}$ and $\delta=\frac{n}{N}$, $U(\delta,\rho)>L(\delta,\rho)$ and therefore $R(\delta,\rho)=U(\delta,\rho)$; this is consistent with the conjecture that the more rapid deviation from unity of the largest eigenvalue suggests that the symmetric RIP (4) is controlled by $U(k,n,N)$. Using the bound $R(\delta,2\rho)$ from Corollary 1 we establish a lower bound on the phase transition, $\rho_S^C(\delta)$, for Theorem 3 when A is an $n\times N$ Gaussian matrix.

Theorem 6. *For any $\epsilon>0$, as $(k,n,N)\to\infty$, there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that every $x\in\chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho<(1-\epsilon)\rho_S^C(\delta)$. $\rho_S^C(\delta)$ is defined as the solution of $R(\delta,2\rho)=\sqrt{2}-1$ and is displayed as the green curve in Figures 3 and 5.*

We observe that $\rho_S^C(\delta)$ is substantially lower than $\rho_S^{RV}(\delta)$ and $\rho_S(\delta,C^N)$ (see the green curve of Figure 3) capturing less than 2% of the region defined by $\rho_S(\delta,C^N)$ (see the green curve of Figure 4). Since $U(\delta,\rho)>L(\delta,\rho)$, the heavier weighting of $L(2k,n,N)$ allows a larger portion of Gaussian matrices to satisfy Theorem 4 than Theorem 3. Using $L(\delta,2\rho)$ and $U(\delta,2\rho)$ from Theorem 5 we establish a lower bound on the phase transition, $\rho_S^{BCT}(\delta)$, for Theorem 4. See the blue curves of Figures 3-5.

Theorem 7. *For any $\epsilon>0$, as $(k,n,N)\to\infty$, there is an exponentially high probability on the draw of A with Gaussian i.i.d. entries that every $x\in\chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if $\rho<(1-\epsilon)\rho_S^{BCT}(\delta)$. $\rho_S^{BCT}(\delta)$ is defined as the solution of $(1+\sqrt{2})L(\delta,2\rho)+U(\delta,2\rho)=\sqrt{2}$ and is displayed as the blue curve in Figures 3 and 5.*

From these phase transitions we see a nontrivial performance improvement from Theorem 3 to Theorem 4 via an improved lower bound on the phase transition for ℓ_1/ℓ_0-equivalence from Theorem 6 to Theorem 7 (see Figure 5). However, even with this improvement, $\rho_S^{BCT}(\delta)$ still fails to capture the vast majority of the Gaussian matrices which satisfy Theorem 1 (see the blue curve of Figure 4).

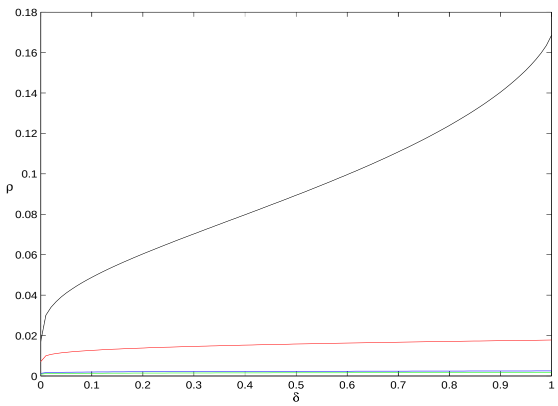

Fig. 3. Lower bounds on the ℓ_1/ℓ_0-equivalence phase transition for Gaussian random matrices; $\rho_S(\delta, C^N)$ (black), $\rho_S^{RV}(\delta)$ (red), $\rho_S^{BCT}(\delta)$ (blue), and $\rho_S^{C}(\delta)$ (green).

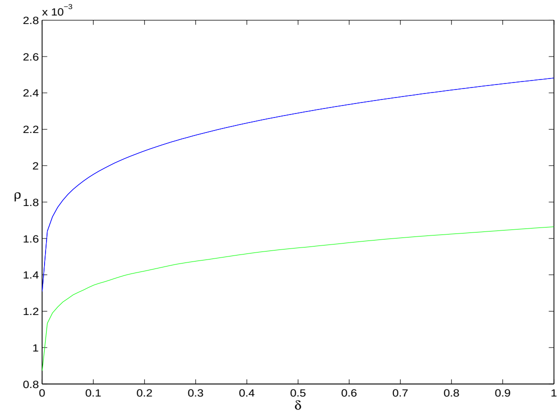

Fig. 5. Lower bounds on the ℓ_1/ℓ_0-equivalence phase transition for Gaussian random matrices; $\rho_S^{BCT}(\delta)$ (blue) and $\rho_S^{C}(\delta)$ (green).

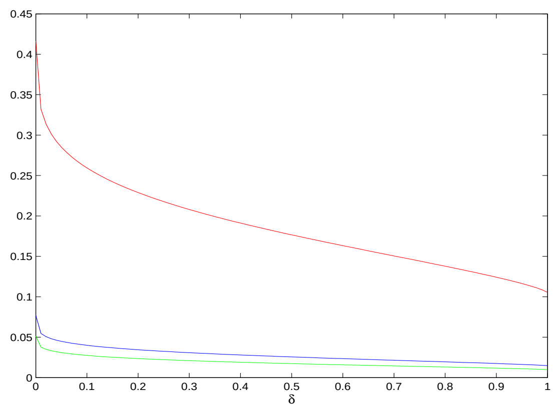

Fig. 4. The fractions $\rho_S^{RV}(\delta)/\rho_S(\delta, C^N)$ (red), $\rho_S^{BCT}(\delta)/\rho_S(\delta, C^N)$ (blue), and $\rho_S^{C}(\delta)/\rho_S(\delta, C^N)$ (green).

V. Discussion

The lower bounds on the phase transitions permit straightforward comparisons of Theorems 1, 2, 6, and 7. Although the ordering of $\rho_S^C(\delta) < \rho_S^{BCT}(\delta) < \rho_S^{RV}(\delta) < \rho_S(\delta, C^N)$ makes clear the advantage of using the inherent geometry of ℓ_1-regularization in its analysis, other advantages and disadvantages of the techniques of analysis exist. For instance, the conditions based upon the RIP also have related results, for yet smaller ρ, which ensure stability when x is not strictly k-sparse. Lower bounds on the ℓ_1/ℓ_0-equivalence phase transition have been developed for specified stability constants [6]. This analysis, including stability, has also been extended to the setting of ℓ_q-regularization for $q \in (0, 1]$, (where ℓ_q replaces ℓ_1 in (2)) [6].

No such stability analysis exists for either of the geometric techniques of analysis presented here. However, the geometric approach of Donoho can be extended to derive weak phase transitions, $\rho_W(\delta, C^N)$ characterizing where ℓ_1/ℓ_0-equivalence is satisfied for *most* $x \in \chi^N(k)$, or to take into account other models of data sparsity [12], [13], [14]. The superior performance of Theorems 1 and 2 in the setting of exact sparsity suggests that stability analysis using these techniques of geometric analysis are likely to result in substantial improvements over results derived using the RIP.

Acknowledgements

The authors thank E. Candès for providing an early draft of [5]. JDB acknowledges support from NSF DMS (VIGRE) grant 0602219. JT acknowledges support from the Alfred P. Sloan Foundation and thanks John E. and Marva M. Warnock for their generous support in the form of an endowed chair.

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159 (electronic), 2001, reprinted from SIAM J. Sci. Comput. **2**0 (1998), no. 1, 33–61.

[2] D. L. Donoho, “Neighborly polytopes and sparse solution of underdetermined linear equations,” 2004, technical Report, Department of Statistics, Stanford University.

[3] ——, “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” *Discrete Comput. Geom.*, vol. 35, no. 4, pp. 617–652, 2006.

[4] M. Rudelson and R. Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements,” *Comm. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.

[5] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *C. R. Math. Acad. Sci. Paris*, vol. 346, no. 9-10, pp. 589–592, 2008.

[6] J. D. Blanchard, C. Cartis, and J. Tanner, “The restricted isometry property and ℓ^q-regularization: phase transitions for sparse approximation,” 2009, submitted.

[7] D. L. Donoho and J. Tanner, “Counting faces of randomly-projected polytopes when the projection radically lowers dimension,” *J. of the AMS*, vol. 22, no. 1, pp. 1–53, 2009.

[8] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[9] S. Geman, “A limit theorem for the norm of random matrices,” *Ann. Probab.*, vol. 8, no. 2, pp. 252–261, 1980.

[10] J. W. Silverstein, “The smallest eigenvalue of a large-dimensional Wishart matrix,” *Ann. Probab.*, vol. 13, no. 4, pp. 1364–1368, 1985.

[11] A. Edelman, “Eigenvalues and condition numbers of random matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, 1988.

[12] D. L. Donoho and J. Tanner, “Neighborliness of randomly-projected simplices in high dimensions,” *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9452–9457, 2005.

[13] ——, “Sparse nonnegative solutions of underdetermined linear equations by linear programming,” *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, 2005.

[14] ——, “Counting the faces of randomly-projected hypercubes and orthants, with applications,” 2008, submitted.