IMPROVED BOUNDS ON RESTRICTED ISOMETRY CONSTANTS FOR GAUSSIAN MATRICES[*]

BUBACARR BAH[†] AND JARED TANNER[†]

Abstract. The restricted isometry constant (RIC) of a matrix A measures how close to an isometry is the action of A on vectors with few nonzero entries, measured in the ℓ^2 norm. Specifically, the upper and lower RICs of a matrix A of size $n \times N$ are the maximum and the minimum deviation from unity (one) of the largest and smallest, respectively, square of singular values of all $\binom{N}{k}$ matrices formed by taking k columns from A. Calculation of the RIC is intractable for most matrices due to its combinatorial nature; however, many random matrices typically have bounded RIC in some range of problem sizes (k, n, N). We provide the best known bound on the RIC for Gaussian matrices, which is also the smallest known bound on the RIC for any large rectangular matrix. Our results are built on the prior bounds of Blanchard, Cartis, and Tanner [*SIAM Rev.*, to appear], with improvements achieved by grouping submatrices that share a substantial number of columns.

Key words. Wishart matrices, compressed sensing, sparse approximation, restricted isometry constant, phase transitions, Gaussian matrices, singular values of random matrices

AMS subject classifications. Primary, 15B52, 60F10, 94A20; Secondary, 94A12, 90C25

DOI. 10.1137/100788884

1. Introduction. Interest in parsimonious solutions to underdetermined systems of equations has seen a spike with the introduction of compressed sensing [10, 7, 6]. Much of the analysis in this new topic has relied upon a new matrix quantity, the restricted isometry constant (RIC), also referred to as the restricted isometry property (RIP) constant. Let A be a matrix of size $n \times N$ and define the set of N-vectors with at most k nonzero entries as

$$(1.1) \qquad \chi^N(k) := \{x \in \mathbb{R}^N : \|x\|_0 \leq k\}.$$

Upper and lower RICs of A, $U(k, n, N; A)$ and $L(k, n, N; A)$, respectively, are defined as [5, 2]

$$(1.2) \qquad U(k, n, N; A) := \min_{c \geq 0} c \quad \text{subject to} \quad (1+c)\|x\|_2^2 \geq \|Ax\|_2^2 \quad \forall x \in \chi^N(k),$$

$$(1.3) \qquad L(k, n, N; A) := \min_{c \geq 0} c \quad \text{subject to} \quad (1-c)\|x\|_2^2 \leq \|Ax\|_2^2 \quad \forall x \in \chi^N(k).$$

RICs differ from standard singular values squared in their combinatorial nature. $U(k, n, N; A)$ and $L(k, n, N; A)$ measure the maximum and the minimum deviation from unity (one) of the largest and smallest, respectively, square of the singular values of all $\binom{N}{k}$ submatrices of A of size $n \times k$ constructed by taking k columns from A. The RICs can be equivalently defined as

$$(1.4) \qquad U(k, n, N; A) := \max_{K \subset \Omega, |K|=k} \lambda^{\max}\left(A_K^* A_K\right) - 1$$

and

$$(1.5) \qquad L(k, n, N; A) := 1 - \min_{K \subset \Omega, |K|=k} \lambda^{\min} (A_K^* A_K),$$

where $\Omega := \{1, 2, \ldots, N\}$, A_K is the restriction of the columns of A to a support set $K \subset \Omega$ with cardinality k ($|K| = k$), and $\lambda^{\max} (B)$ and $\lambda^{\min} (B)$ are the smallest and largest eigenvalues of B, respectively.

The standard notion of general position is $L(n, n, N; A) < 1$, and Kruskal rank [19] is the largest k such that $L(k, n, N; A) < 1$.

Many of the theorems in compressed sensing rely upon a "sensing matrix" having suitable bounds on its RIC. Unfortunately, computing the RICs of a matrix A is in general NP-hard [22]. Efforts are underway to design algorithms that compute accurate bounds on the RICs of a matrix [9, 17], but to date these algorithms have had limited success, with the bounds effective only for $k \sim n^{1/2}$. Lacking the ability to efficiently calculate the RICs of a given matrix, the research community is actively computing probabilistic bounds for various random matrix ensembles. These efforts have followed three research programs:

- Determination of the largest ensemble of matrices such that as the problem sizes (k, n, N) grow, the RICs $U(k, n, N; A)$ remain bounded and the $L(k, n, N; A)$ is bounded away from 1 [21].
- Computing bounds as accurate as possible for particular ensembles, such as the Gaussian ensemble [5, 2], where the entries of A are drawn independently and identically distributed (i.i.d.) from the standard Gaussian normal $\mathcal{N}(0, 1/n)$. (In part as a model for i.i.d. mean zero ensembles.)
- Computing bounds as accurate as possible for the partial Fourier ensemble [24], where A is formed from random rows, j, or samples, t_l, of a Fourier matrix with entries $F_{j,l} = e^{2\pi i j t_l}$. (In part as a model for matrices possessing a fast matrix-vector product.)

This manuscript focuses on the second of these research programs, accurate bounds for the Gaussian/Wishart ensemble. Candès and Tao derived the first set of RIC bounds for the Gaussian ensemble using a union bound over all $\binom{N}{k}$ submatrices and bounding the singular values of each submatrix using concentration of measure bounds [7]. Blanchard, Cartis, and Tanner derived the second set of RIC bounds for the Gaussian ensemble, similarly using a union bound over all $\binom{N}{k}$ submatrices, but achieved substantial improvements by using more accurate bounds on the probability density function of Wishart matrices [2]. These bounds are presented here in Theorems 2.8 and 2.10, respectively. This paper presents yet further improved bounds for the Gaussian ensemble (see Theorem 2.3 and Figure 2.1) by grouping submatrices with overlapping support sets (say, A_K and $A_{K'}$ with $|K \cap K'| \gg 1$) for which we expect the singular values to be highly correlated. These are the first RIC bounds that exploit this structure. In addition to asymptotic bounds for large problem sizes, we present bounds valid for finite values of (k, n, N).

This paper is organized as follows: Our improved asymptotic bounds are stated in section 2.1, and their derivation is described in section 2.2. Prior bounds are presented in section 2.3 and are compared with those in Theorem 2.3. Bounds valid for finite values are presented in section 2.4, and the implications of the RIC bounds for compressed sensing are given in section 2.5. Proof of technical lemmas used or assumed in our discussion come in the appendix.

2. RIC bounds. We focus our attention on bounding the RIC for the Gaussian ensemble in the setting of *proportional-growth asymptotics*.

DEFINITION 2.1 (proportional-growth asymptotics). *A sequence of problem sizes* (k, n, N) *is said to follow proportional-growth asymptotics if*

$$(2.1) \quad \frac{k}{n} = \rho_n \to \rho \quad and \quad \frac{n}{N} = \delta_n \to \delta \quad for \quad (\delta, \rho) \in (0, 1)^2 \quad as \quad (k, n, N) \to \infty.$$

In this asymptotic we provide quantitative values, above which it is exponentially unlikely that the RIC will exceed. In section 2.4 we show how our derivation of these bounds can also supply probabilities for specified bounds and finite values of (k, n, N).

2.1. Improved RIP bounds. The probability density functions (pdf) of the RIC for the Gaussian ensemble are currently unknown, but asymptotic probabilistic bounds have been proven. Our bounds, and earlier ones, for the RIC of the Gaussian ensemble are built upon the bounds of the pdf's of the extreme eigenvalues of Gaussian (Wishart) matrices due to Edelman [13, 12]. All earlier bounds on the RIC have been derived using union bounds that consider each of the $\binom{N}{k}$ submatrices of size $n \times k$ individually [2, 7]. We consider groups of submatrices where the columns of the submatrices in a group are from at most $m \geq k$ distinct columns of A. We present our improved bounds in Theorem 2.3, preceded by the definition whose terms are defined in Definition 2.2. Plots of these bounds are displayed in Figure 2.1.

DEFINITION 2.2. *Let* $(\delta, \rho) \in (0, 1)^2$, $\gamma \in [\rho, \delta^{-1}]$, *and denote the Shannon entropy with base e logarithms as* $H(p) := p \ln(1/p) + (1 - p) \ln(1/(1 - p))$. *Let*

$$(2.2) \qquad \psi_{\min}(\lambda, \gamma) := H(\gamma) + \frac{1}{2} \Big[(1 - \gamma) \ln \lambda + \gamma \ln \gamma + 1 - \gamma - \lambda \Big],$$

$$(2.3) \qquad \psi_{\max}(\lambda, \gamma) := \frac{1}{2} \Big[(1 + \gamma) \ln \lambda - \gamma \ln \gamma + 1 + \gamma - \lambda \Big].$$

Define $\lambda^{\min}(\delta, \rho; \gamma)$ *and* $\lambda^{\max}(\delta, \rho; \gamma)$ *as the solution to* (2.4) *and* (2.5), *respectively:*

$$(2.4) \quad \delta \psi_{\min}\left(\lambda^{\min}(\delta, \rho; \gamma), \gamma\right) + H(\rho\delta) - \delta\gamma H(\rho/\gamma) = 0 \quad for \ \lambda^{\min}(\delta, \rho; \gamma) \leq 1 - \gamma,$$

$$(2.5) \quad \delta \psi_{\max}\left(\lambda^{\max}(\delta, \rho; \gamma), \gamma\right) + H(\rho\delta) - \delta\gamma H(\rho/\gamma) = 0 \quad for \ \lambda^{\max}(\delta, \rho; \gamma) \geq 1 + \gamma.$$

Let $\lambda^{\min}(\delta, \rho) := \max_\gamma \lambda^{\min}(\delta, \rho; \gamma)$ *and* $\lambda^{\max}(\delta, \rho) := \min_\gamma \lambda^{\max}(\delta, \rho; \gamma)$, *and define*

$$(2.6) \qquad \mathcal{L}^{BT}(\delta, \rho) := 1 - \lambda^{\min}(\delta, \rho) \quad and \quad \mathcal{U}^{BT}(\delta, \rho) := \lambda^{\max}(\delta, \rho) - 1.$$

That for each $(\delta, \rho; \gamma)$, (2.4) and (2.5) have a unique solution $\lambda^{\min}(\delta, \rho; \gamma)$ and $\lambda^{\max}(\delta, \rho; \gamma)$, respectively, was proven in [2]. That $\lambda^{\min}(\delta, \rho; \gamma)$ and $\lambda^{\max}(\delta, \rho; \gamma)$ have unique maxima and minima, respectively, over $\gamma \in [\rho, \delta^{-1}]$ is established in Lemma 2.6.

THEOREM 2.3. *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Let δ and ρ be defined as in (2.1), and let $\mathcal{L}^{BT}(\delta, \rho)$ and $\mathcal{U}^{BT}(\delta, \rho)$ be defined as in Definition 2.2. For any fixed $\epsilon > 0$, in the proportional-growth asymptotic,*

$$\mathbf{P}(L(k, n, N) < \mathcal{L}^{BT}(\delta, \rho) + \epsilon) \to 1 \quad and \quad \mathbf{P}(U(k, n, N) < \mathcal{U}^{BT}(\delta, \rho) + \epsilon) \to 1$$

exponentially in n.

In the spirit of reproducible research, software and web forms that evaluate $\mathcal{L}^{BT}(\delta, \rho)$ and $\mathcal{U}^{BT}(\delta, \rho)$ are publicly available at [14].
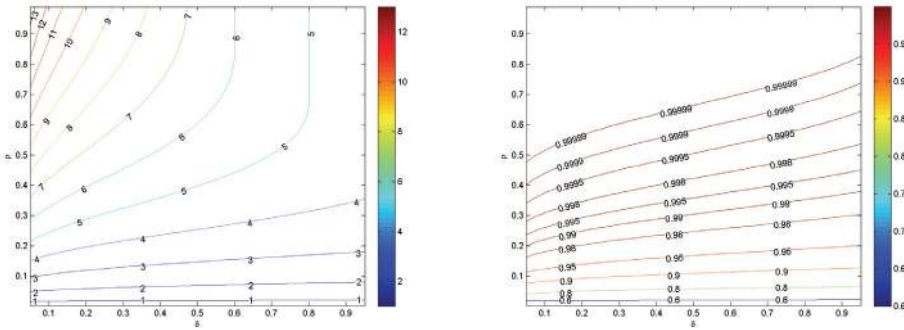
FIG. 2.1. $\mathcal{U}^{BT}(\delta, \rho)$ (left panel) and $\mathcal{L}^{BT}(\delta, \rho)$ (right panel) from Definition 2.2 for $(\delta, \rho) \in (0, 1)^2$.
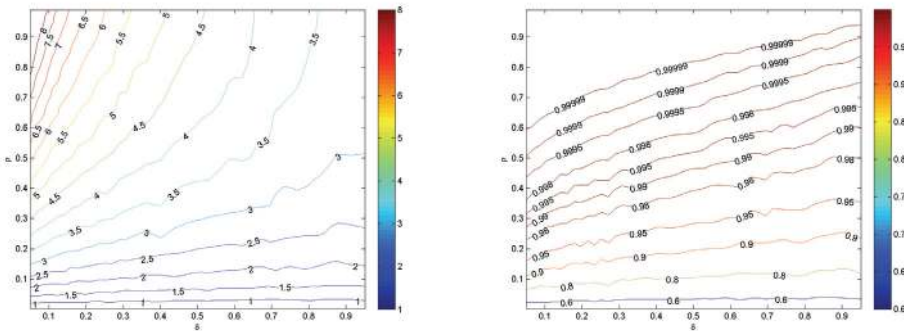

FIG. 2.2. Empirically observed lower bounds on RIC for A Gaussian. Observed lower bounds of $U(k, n, N; A)$ (left panel) and $L(k, n, N; A)$ (right panel). Although there is no computationally tractable method for calculating the RICs of a matrix, there are efficient algorithms that perform local searches for extremal eigenvalues of submatrices, allowing for observable lower bounds on the RICs. Algorithms for observing $L(k, n, N)$ [11] and $U(k, n, N)$ [16] were applied to hundreds of A drawn i.i.d. $\mathcal{N}(0, 1/n)$ with $n = 400$ and N increasing from 420 to 8000.

Sharpness of the bounds can be probed by comparison with empirically observed lower bounds on the RIC for finite-dimensional draws from the Gaussian ensemble. There exist efficient algorithms for calculating lower bounds of RICs [11, 16]. These algorithms perform local searches for submatrices with extremal eigenvalues. The new bounds in Theorem 2.3 (see Figure 2.1) can be compared with empirical data displayed in Figure 2.2.

To further demonstrate the sharpness of our bounds, we compute the maximum and minimum "sharpness ratios" of the bounds in Theorem 2.3 to empirically observed lower bounds; for each ρ, the maximum and minimum of the ratio are taken over all $\delta \in [0.05, 0.9524]$. These are the same δ values used in Figure 2.2. These ratios are shown in the left panel of Figure 2.3 and are below 1.57 of the empirically observed lower bounds on $L(k, n, N)$ and $U(k, n, N)$ observed with $n = 400$.

2.2. Discussion on the construction of improved RIC bounds. The bounds in Theorem 2.3 improve upon the earlier results of [2] by grouping matrices A_K and
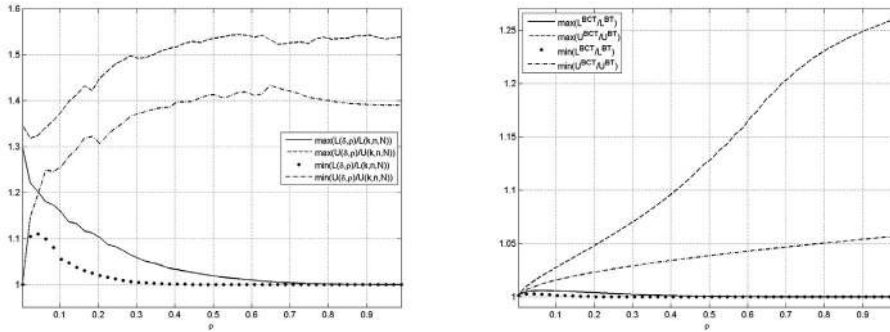
FIG. 2.3. *Left panel: The maximum and minimum (over δ) sharpness ratios, $\frac{\mathcal{U}^{BT}(\delta,\rho)}{U(k,n,N;A)}$ and $\frac{\mathcal{L}^{BT}(\delta,\rho)}{L(k,n,N;A)}$, as functions of ρ, with the maximum and minimum taken over all δ ∈ [0.05, 0.9524], the same δ values used in Figure 2.2. Right panel: The maximum and minimum (over δ) improvement ratios over the previous best known bounds, $\frac{\mathcal{U}^{BCT}(\delta,\rho)}{\mathcal{U}^{BT}(\delta,\rho)}$ and $\frac{\mathcal{L}^{BCT}(\delta,\rho)}{\mathcal{L}^{BT}(\delta,\rho)}$, as a function of ρ, with the maximum and minimum also taken over δ ∈ [0.05, 0.9524].*

$A_{K'}$ that share a significant number of columns from A. This is manifest in Definition 2.2 through the introduction of the free parameter γ associated with the number of groups considered. In this section we first discuss the way in which we construct these groups and the sense in which the bounds in Theorem 2.3 are optimal for this construction. Equipped with a suitable construction of groups, we discuss the way in which this grouping is employed to improve the RIC bounds from [2].

2.2.1. Construction of groups. We construct our groups of A_K by selecting a subset M_i from $1, 2, \ldots, N$ of cardinality $|M_i| = m \geq k$ and setting $\mathcal{G}_i := \{K\}_{K \subset M_i, |K|=k}$, the set of all sets $K \subset M$ of cardinality k. The group $\mathcal{G}_i$ has $\binom{m}{k}$ members, with any two members sharing at least $2k - m$ elements. Hence, the quantity $\gamma = m/n$ in Definition 2.2 is associated with the cardinality of the groups $\mathcal{G}_i$. In order to calculate bounds on the RIC of a matrix, we need a collection of groups whose union includes all $\binom{N}{k}$ sets of cardinality k from $\Omega := \{1, 2, \ldots, N\}$; that is, we need $\{\mathcal{G}_i\}_{i=1}^{u}$ such that $G := \bigcup_{i=1}^{u} \mathcal{G}_i$ with $|G| = \binom{N}{k}$. From simple counting, the minimum number of groups $\mathcal{G}_i$ needed for this covering is at least $r := \binom{N}{k}\binom{m}{k}^{-1}$. Although the construction of a minimal covering is an open question [18], even a simple random construction of the $\mathcal{G}_i$'s typically requires only a polynomial multiple of r groups, hence achieving the optimal large deviation rate.

LEMMA 2.4 (see [18]). *Set $r = \binom{N}{k}\binom{m}{k}^{-1}$, and draw $u := rN$ sets M_i each of cardinality m, drawn uniformly at random from the $\binom{N}{m}$ possible sets of cardinality m. With G defined as above,*

$$(2.7) \qquad \mathbf{P}\left(|G| < \binom{N}{k}\right) < C(k/N)N^{-1/2}e^{-N(1-\ln 2)},$$

where $C(p) \leq \frac{5}{4}(2\pi p(1-p))^{-\frac{1}{2}}$.

Proof. Select one set $K \subset \Omega$ of cardinality $|K| = k$ prior to the draw of the sets M_i. The probability that it is contained in one set M_i is $1/r$, and with each M_i drawn independently, the probability that it is not contained in any of the u sets M_i

is $(1 - r^{-1})^u \le e^{-u/r}$. Applying a union bound over all $\binom{N}{k}$ sets K yields

$$\mathbf{P}\left(|G| < \binom{N}{k}\right) < \binom{N}{k} e^{-u/r}.$$

Noting from Stirling's inequality that

(2.8) $$\frac{16}{25}(2\pi p(1-p)N)^{-\frac{1}{2}} e^{NH(p)} \le \binom{N}{pN} \le \frac{5}{4}(2\pi p(1-p)N)^{-\frac{1}{2}} e^{NH(p)},$$

with $H(p) \le \ln 2$ for $p \in [0, 1]$, and substituting the selected value of u completes the proof. Note that an exponentially small probability can be obtained with u just larger than $rNH(\delta\rho)$, but the smaller polynomial factor is negligible for our purposes. $\square$

COROLLARY 2.5. *Given Lemma* 2.4, *as* $n \to \infty$ *in the proportional-growth asymptotics, the probability that all the* $\binom{N}{k}$ *k-subsets of* $1, 2, \dots, N$ *are covered by* G *converges to one exponentially in* n.

2.2.2. Decreasing the combinatorial term. We illustrate the way the groups $\mathcal{G}_i$ are used to improve the RIC bound on the upper RIC bound $U(k, n, N; A)$; the bounds for $L(k, n, N; A)$ follow by a suitable replacement of maximizations/minimizations and sign changes. All previous bounds on the RIC for the Gaussian ensemble have overcome the combinatorial maximization/minimization by use of a union bound over all $\binom{N}{k}$ sets $K \subset \Omega$ and then using a tail bound on the pdf of the extreme eigenvalues of $A_K^* A_K$; for some $\lambda_{\max}^* > 0$,

$$\mathbf{P}\left(\max_{K \subset \Omega, |K|=k} \lambda^{\max}(A_K^* A_K) > \lambda_{\max}^*\right) \le \binom{N}{k} \mathbf{P}\left(\lambda^{\max}(A_K^* A_K) > \lambda_{\max}^*\right).$$

The fact that the random variables $\lambda^{\max}(A_K^* A_K)$ are treated as independent is the principal deficiency of this bound. To exploit dependencies of this variable for K and K' with significant overlap, we exploit the groupings $\mathcal{G}_i$, which, at least for m moderately larger than k, contain sets with significant overlap. For the moment we assume the groups $\{\mathcal{G}_i\}_{i=1}^u$ cover all $K \subset \Omega$, and we replace the above maximization over K with a double maximization

$$\mathbf{P}\left(\max_{K \subset \Omega, |K|=k} \lambda^{\max}(A_K^* A_K) > \lambda_{\max}^*\right) = \mathbf{P}\left(\max_{i=1,\dots,u} \max_{K \subset \mathcal{G}_i, |K|=k} \lambda^{\max}(A_K^* A_K) > \lambda_{\max}^*\right).$$

The outer maximization can be bounded over all u sets $\mathcal{G}_i$, again, using a simple union bound; however, with a smaller combinatorial term.

The expected dependencies between $\lambda^{\max}(A_K^* A_K)$ for $K \subset \mathcal{G}_i$ can, at times, be better controlled by replacing the maximization over $K \subset \mathcal{G}_i$ by $\lambda^{\max}(A_{M_i}^* A_{M_i})$, where M_i is the subset of cardinality m containing all $K \subset \mathcal{G}_i$:

(2.9) $$\mathbf{P}\left(\max_{i=1,\dots,u} \max_{K \subset \mathcal{G}_i, |K|=k} \lambda^{\max}(A_K^* A_K) > \lambda_{\max}^*\right) \le u\mathbf{P}\left(\lambda^{\max}(A_M^* A_M) > \lambda_{\max}^*\right).$$

Selecting $m = k$ recovers the usual union bound with u equal to $\binom{N}{k}$. Larger values of m decrease the combinatorial term at the cost of increasing $\lambda^{\max}(A_M^* A_M)$. The efficacy of this approach depends on the interplay between these two competing factors. In the proportional-growth asymptotic, this interplay is observed through the optimization over $\frac{m}{n} = \gamma \in [\rho, \delta^{-1}]$. Definition 2.2 uses the tail bounds
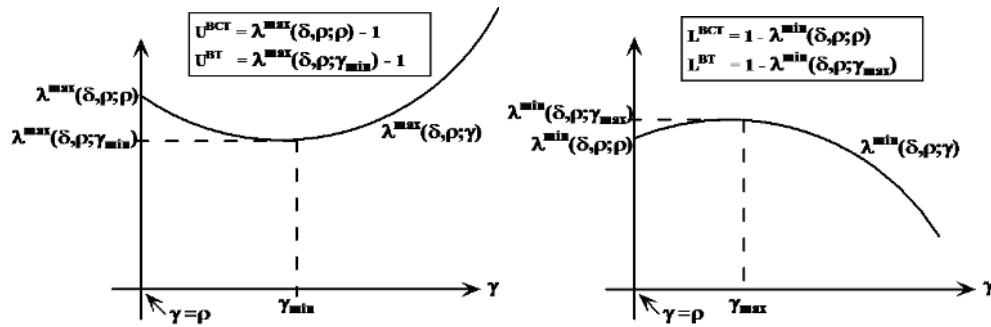
FIG. 2.4. *Left panel: The relationship between the new bound* $\mathcal{U}^{BT}(\delta, \rho)$, *Theorem 2.3, and the previous smallest bound* $\mathcal{U}^{BCT}(\delta, \rho)$, *Theorem 2.10, where in the two bounds* $\lambda^{\max}(\delta, \rho; \gamma)$ *is evaluated at* $\gamma = \gamma_{\min}$ *and* $\gamma = \rho$, *respectively. Right panel: The relationship between the new bound* $\mathcal{L}^{BT}(\delta, \rho)$, *Theorem 2.3, and the previous smallest bound* $\mathcal{L}^{BCT}(\delta, \rho)$, *Theorem 2.10, where in the two bounds* $\lambda^{\min}(\delta, \rho; \gamma)$ *is evaluated at* $\gamma = \gamma_{\max}$ *and* $\gamma = \rho$, *respectively.*
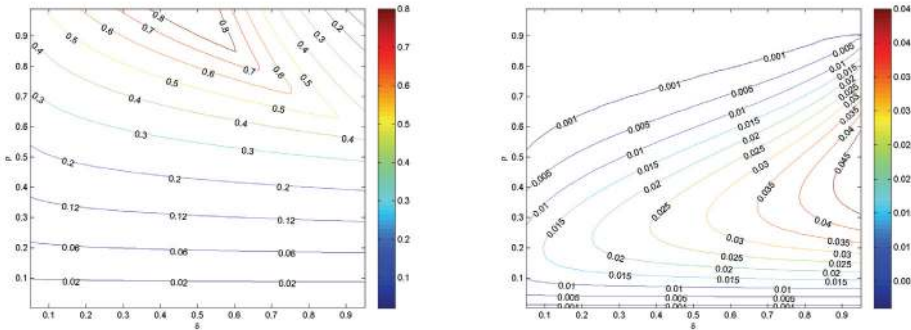

FIG. 2.5. *Optimal choice of* $\gamma - \rho$ *for* $\mathcal{U}^{BT}(\delta, \rho)$ *(left panel) and* $\mathcal{L}^{BT}(\delta, \rho)$ *(right panel).*

on the extreme eigenvalues of Wishart matrices derived by Edelman [12] to bound $\mathbf{P}\left(\lambda^{\max}(A_M^* A_M) > \lambda_{\max}^*\right)$. The previously best known bound on the RIC for the Gaussian ensemble is recovered by selecting $\gamma = \rho$ in Definition 2.2 [2]. This is illustrated in Figure 2.4. The innovation of the bounds in Theorem 2.3 follows from there always being a unique $\gamma > \rho$ such that $\lambda^{\max}(\delta, \rho; \gamma)$ is less than $\lambda^{\max}(\delta, \rho; \rho)$.

LEMMA 2.6. *Suppose* $\lambda^{\min}(\delta, \rho; \gamma)$ *and* $\lambda^{\max}(\delta, \rho; \gamma)$ *are solutions to* (2.4) *and* (2.5), *respectively. For any fixed* (δ, ρ) *there exist a unique* $\gamma_{\min} \in [\rho, \delta^{-1}]$ *that minimizes* $\lambda^{\max}(\delta, \rho; \gamma)$ *and a unique* $\gamma_{\max} \in [\rho, \delta^{-1}]$ *that maximizes* $\lambda^{\min}(\delta, \rho; \gamma)$. *Furthermore,* $\gamma_{\min}$ *and* $\gamma_{\max}$ *are strictly larger than* ρ.

The optimal choices of $\gamma - \rho$ for $\mathcal{U}^{BT}(\rho, \delta)$ and $\mathcal{L}^{BT}(\rho, \delta)$ in $(\rho, \delta) \in (0, 1)^2$ are displayed in Figure 2.5. The proof of Lemma 2.6 is presented in Appendix A.1.

2.3. Prior RIP bounds. There have been two previous quantitative bounds for the RIC of the Gaussian ensemble in the proportional-growth asymptotics. The first bounds on the RIC of the Gaussian ensemble were supplied in [7] by Candès and Tao using union bounds and concentration of measure bounds on the extreme eigenvalues of Wishart matrices from [20]. These bounds are stated in Theorem 2.8 with Definition 2.7 defining some of the terms used in the theorem, and plots of these bounds are displayed in Figure 2.6.
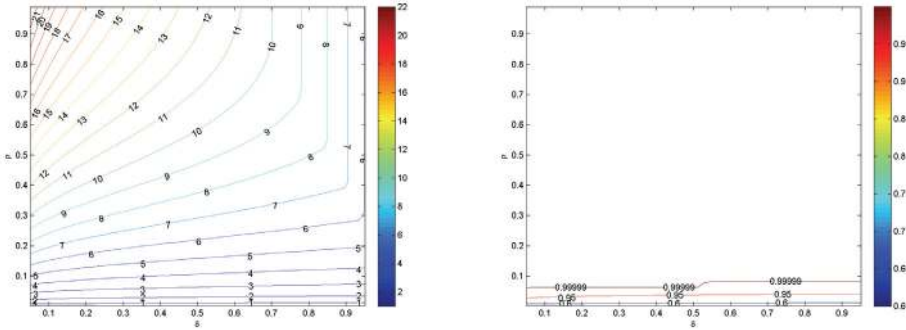
FIG. 2.6. $\mathcal{U}^{CT}(\delta, \rho)$ (left panel) and $\mathcal{L}^{CT}(\delta, \rho)$ (right panel) from Definition 2.7 for $(\delta, \rho) \in (0,1)^2$.

DEFINITION 2.7. *Let* $(\delta, \rho) \in (0,1)^2$, *and define*

$$\mathcal{U}^{CT}(\delta, \rho) := \left[1 + \sqrt{\rho} + (2\delta^{-1}H(\delta\rho))^{1/2}\right]^2 - 1$$

and

$$\mathcal{L}^{CT}(\delta, \rho) := 1 - \max\left\{0, \left[1 - \sqrt{\rho} - (2\delta^{-1}H(\delta\rho))^{1/2}\right]^2\right\}.$$

THEOREM 2.8 (Candès and Tao [7]). *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Let δ and ρ be defined as in (2.1), and let $\mathcal{L}^{CT}(\delta, \rho)$ and $\mathcal{U}^{CT}(\delta, \rho)$ be defined as in Definition 2.7. For any fixed $\epsilon > 0$, in the proportional-growth asymptotic,*

$$\mathbf{P}(L(k,n,N) < \mathcal{L}^{CT}(\delta, \rho) + \epsilon) \to 1 \quad and \quad \mathbf{P}(U(k,n,N) < \mathcal{U}^{CT}(\delta, \rho) + \epsilon) \to 1$$

exponentially in n.

The bounds in Theorem 2.3 follow the construction of the second bounds on the RIC for the Gaussian ensemble, presented in [2]. Removing the optimization of $\gamma \in [\rho, \delta^{-1}]$ in Definition 2.2 and fixing $\gamma = \rho$ recovers the bounds on $L(k, n, N; A)$ and the first of two bounds on $U(k, n, N; A)$ presented in [2]. The first bound on $U(k, n, N; A)$ in [2] suffers from excessive overestimation when $\delta\rho \approx 1/2$ because of the combinatorial term. In fact, this overestimation is so severe that for some (δ, ρ) with $\delta\rho \approx 1/2$, smaller bounds are obtained at $(\delta, 1)$. This overestimation is somewhat ameliorated by our noting the monotonicity of $U(k, n, N; A)$ in k, obtaining the improved bound; see (2.12). These bounds are stated in Theorem 2.10 with Definition 2.9 defining some of the terms used in the theorem, and plots of these bounds are displayed in Figure 2.7.

DEFINITION 2.9. *Let $(\delta, \rho) \in (0,1)^2$, and denote the Shannon entropy with base e logarithms as $H(p) := p\ln(1/p) + (1-p)\ln(1/(1-p))$. Let $\psi_{\min}(\lambda, \rho)$ and $\psi_{\max}(\lambda, \rho)$ be defined as in (2.2) and (2.3), respectively. Define $\lambda_{BCT}^{\min}(\delta, \rho)$ and $\lambda_{BCT}^{\max}(\delta, \rho)$ as the solution to (2.10) and (2.11), respectively:*

$$(2.10) \qquad \delta\psi_{\min}(\lambda_{BCT}^{\min}(\delta, \rho), \rho) + H(\rho\delta) = 0 \quad for \quad \lambda_{BCT}^{\min}(\delta, \rho) \leq 1 - \rho,$$

$$(2.11) \qquad \delta\psi_{\max}(\lambda_{BCT}^{\max}(\delta, \rho), \rho) + H(\rho\delta) = 0 \quad for \quad \lambda_{BCT}^{\max}(\delta, \rho) \geq 1 + \rho.$$
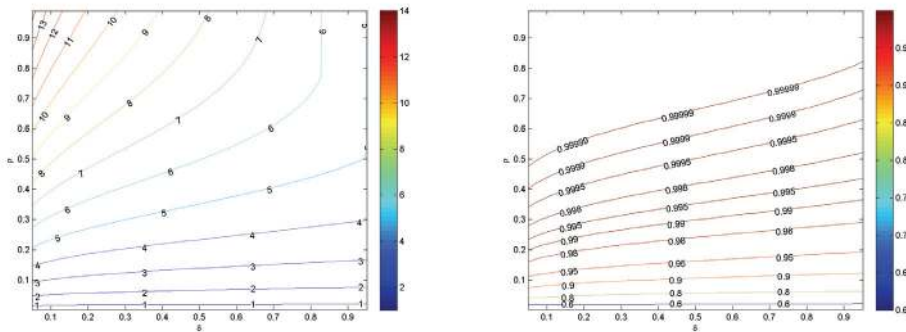
FIG. 2.7. $\mathcal{U}^{BCT}(\delta,\rho)$ (left panel) and $\mathcal{L}^{BCT}(\delta,\rho)$ (right panel) from Definition 2.9 for $(\delta,\rho) \in (0,1)^2$.

Define

$$(2.12) \quad \mathcal{L}^{BCT}(\delta,\rho) := 1 - \lambda_{BCT}^{\min}(\delta,\rho) \quad and \quad \mathcal{U}^{BCT}(\delta,\rho) := \min_{\nu\in[\rho,1]} \lambda_{BCT}^{\max}(\delta,\nu) - 1.$$

THEOREM 2.10 (Blanchard, Cartis, and Tanner [2]). *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/n)$. Let δ and ρ be defined as in (2.1), and let $\mathcal{L}^{BCT}(\delta,\rho)$ and $\mathcal{U}^{BCT}(\delta,\rho)$ be defined as in Definition 2.9. For any fixed $\epsilon > 0$, in the proportional-growth asymptotic,*

$$\mathbf{P}(L(k,n,N;A) < \mathcal{L}^{BCT}(\delta,\rho)+\epsilon) \to 1 \quad and \quad \mathbf{P}(U(k,n,N;A) < \mathcal{U}^{BCT}(\delta,\rho)+\epsilon) \to 1$$

exponentially in n.

Figures 2.6 and 2.7 show that the bounds in Theorem 2.10 are a substantial improvement on those in Theorem 2.8. The bounds presented here in Definition 2.2 and Theorem 2.3 are a further improvement over those in [2], as implied by Lemma 2.6.

COROLLARY 2.11. *Let $\mathcal{L}^{BT}(\delta,\rho)$ and $\mathcal{U}^{BT}(\delta,\rho)$ be defined as in Definition 2.2, and let $\mathcal{L}^{BCT}(\delta,\rho)$ and $\mathcal{U}^{BCT}(\delta,\rho)$ be defined as in Definition 2.9. For any fixed $(\delta,\rho) \in (0,1)^2$,*

$$\mathcal{L}^{BT}(\delta,\rho) < \mathcal{L}^{BCT}(\delta,\rho) \quad and \quad \mathcal{U}^{BT}(\delta,\rho) < \mathcal{U}^{BCT}(\delta,\rho).$$

The right panel of Figure 2.3 shows the ratio of the previously best known bounds, Theorem 2.10, to the new bounds, Theorem 2.3; for each ρ, the ratio is maximized over $\delta \in [0.05, 0.9524]$.

2.4. Finite N interpretations. The method of proof we have used to obtain the proportional-growth asymptotic bounds in Theorem 2.3 also provides, albeit less elegantly, bounds valid for finite values of (k,n,N) and specified probabilities of the bound being satisfied. For a specified problem instance (k,n,N) and ϵ, bounds on the probabilities $\mathbf{P}\left(U(k,n,N) > \mathcal{U}^{BT}(\delta_n,\rho_n) + \epsilon\right)$ and $\mathbf{P}\left(L(k,n,N) > \mathcal{L}^{BT}(\delta_n,\rho_n) + \epsilon\right)$ for discrete values of δ and ρ are given in Propositions 2.12 and 2.13, respectively.

PROPOSITION 2.12. *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/n)$. Define $\mathcal{U}^{BT}(\delta_n,\rho_n)$ as in Definition 2.2 for discrete values of δ*

and ρ. Then for any $\epsilon > 0$,

$$\mathbf{P}\Big(U(k,n,N) > \mathcal{U}^{BT}(\delta_n, \rho_n) + \epsilon\Big)$$

$$\leq p'_{\max}\Big(n, \lambda^{\max}(\delta_n, \rho_n) + \epsilon\Big) \exp\left(n\epsilon \cdot \frac{d}{d\lambda}\psi_U\Big(\lambda^{\max}(\delta_n, \rho_n)\Big)\right)$$

(2.13) $$+ \frac{5}{4}(2\pi k(1 - k/N))^{-1/2} \exp(-N(1 - \ln 2)),$$

where

(2.14) $$p'_{\max}(n, \lambda) := \left(\frac{8}{\pi}\right)^{1/2} \frac{2n^{-7/2}}{\sqrt{\gamma\lambda}} \left(\frac{5}{4}\right)^3 \left(\frac{nN(\gamma - \rho)}{\gamma\delta(1 - \rho\delta)}\right)^{1/2}$$

and

$$\psi_U(\lambda, \gamma) := \delta^{-1}\left[H(\rho\delta) - \delta\gamma H\left(\frac{\rho}{\gamma}\right) + \delta\psi_{\max}(\lambda, \gamma)\right]$$

for $\psi_{\max}(\lambda, \gamma)$ defined in (2.3).

PROPOSITION 2.13. *Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Define $\mathcal{L}^{BT}(\delta_n, \rho_n)$ as in Definition 2.2 for discrete values of δ and ρ. Then for any $\epsilon > 0$,*

$$\mathbf{P}\Big(L(k,n,N) > \mathcal{L}^{BT}(\delta_n, \rho_n) + \epsilon\Big)$$

$$\leq p'_{\min}\Big(n, \lambda^{\min}(\delta_n, \rho_n) + \epsilon\Big) \exp\left(n\epsilon \cdot \frac{d}{d\lambda}\psi_L\Big(\lambda^{\min}(\delta_n, \rho_n)\Big)\right)$$

(2.15) $$+ \frac{5}{4}(2\pi k(1 - k/N))^{-1/2} \exp(-N(1 - \ln 2)),$$

where

(2.16) $$p'_{\min}(n, \lambda) := \left(\frac{5}{4}\right)^3 \frac{e\sqrt{\lambda}}{\pi\sqrt{2}} \left(\frac{nN(\gamma - \rho)}{\gamma\delta(1 - \rho\delta)}\right)^{1/2}$$

and

$$\psi_L(\lambda, \gamma) := \delta^{-1}\left[H(\rho\delta) - \delta\gamma H\left(\frac{\rho}{\gamma}\right) + \delta\psi_{\min}(\lambda, \gamma)\right]$$

for $\psi_{\min}(\lambda, \gamma)$ defined in (2.2).

The proofs of Propositions 2.12 and 2.13 are presented in Appendix A.2 and also serve as the proof of Theorem 2.3, which follows by taking the appropriate limits. From Propositions 2.12 and 2.13 we calculated bounds for a few example values of (k, n, N) and ϵ. Table 2.1 shows bounds on $\mathbf{P}\left(U(k,n,N) > \mathcal{U}^{BT}(\delta_n, \rho_n) + \epsilon\right)$ for a few values of (k, n, N) with two different choices of ϵ. It is remarkable that these probabilities are already close to zero for these small values of (k, n, N) and even for $\epsilon \ll 1$. Table 2.2 shows bounds on $\mathbf{P}\left(L(k,n,N) > \mathcal{L}^{BT}(\delta_n, \rho_n) + \epsilon\right)$ for the same values of (k, n, N) as in Table 2.1, but with even smaller values for ϵ. Again, it is remarkable that these probabilities are extremely small, even for relatively small values of (k, n, N) and ϵ.

TABLE 2.1
Prob is an upper bound on $\mathbf{P}\left(U(k,n,N) > \mathcal{U}^{BT}(\delta_n, \rho_n) + \epsilon\right)$ for the specified (k,n,N) and ϵ.

k	n	N	ϵ	Prob
100	200	2000	10^{-3}	2.9×10^{-2}
200	400	4000	10^{-3}	9.5×10^{-3}
400	800	8000	10^{-3}	2.9×10^{-3}
100	200	2000	10^{-10}	3.2×10^{-2}
200	400	4000	10^{-10}	1.1×10^{-2}
400	800	8000	10^{-10}	4.0×10^{-3}

TABLE 2.2
Prob is an upper bound on $\mathbf{P}\left(L(k,n,N) > \mathcal{L}^{BT}(\delta_n, \rho_n) + \epsilon\right)$ for the specified (k,n,N) and ϵ.

k	n	N	ϵ	Prob
100	200	2000	10^{-5}	2.8×10^{-18}
200	400	4000	10^{-5}	9.1×10^{-32}
400	800	8000	10^{-5}	2.8×10^{-58}

2.5. Implications for sparse approximation and compressed sensing.
The RIC was introduced by Candès and Tao [7] as a technique to prove that in certain
conditions the sparsest solution of an underdetermined system of equations $Ax = b$ (A
of size $n \times N$ with $n < N$) can be found using linear programming. The RIC is now a
widely used technique in the study of sparse approximation algorithms, allowing the
analysis of sparse approximation algorithms without specifying the measurement ma-
trix A. For instance, in [5] it was proven that if $\max(L(2k,n,N;A), U(2k,n,N;A)) <
\sqrt{2} - 1$, then if $Ax = b$ has a unique k-sparse solution as its sparsest solution, then
$\arg\min \|z\|_1$ subject to $Az = b$ will be this k-sparse solution. A host of other RIC
based conditions have been derived for this and other sparsifying algorithms. How-
ever, the values of (k,n,N) when these conditions on the RIC are satisfied can be
determined only once the measurement matrix A has been specified [1].

The RIC bounds for the Gaussian ensemble discussed here allow one to state
values of (k,n,N) when sparse approximation recovery conditions are satisfied, and,
from these, guarantee the recovery of k-sparse vectors from (A,b). Unfortunately, all
existing sparse approximation bounds on the RICs are so sufficiently small that they
are satisfied only for $\rho \ll 1$, typically on the order of 10^{-3}. Although the bounds pre-
sented here are a strict improvement over the previously best known bounds, and for
some (δ, ρ) achieve as much as a 20% decrease (see Figure 2.3), the improvements for
$\rho \ll 1$ are meager, approximately $0.5 - 1\%$. This limited improvement for compressed
sensing algorithms is in large part due to the previous bounds being within 30% of
empirically observed lower bounds on RIC for $n = 400$ when $\rho < 10^{-2}$ [2].

In [3], using RIC bounds from [2], lower bounds on the phase transitions for
exact recovery of k-sparse signals for three greedy algorithms and l_1-minimization
were presented. These curves are functions $\rho_S^{sp}(\delta)$ for subspace pursuit (SP) [8],
$\rho_S^{csp}(\delta)$ for compressive sampling matching pursuit (CoSaMP) [23], $\rho_S^{iht}(\delta)$ for iter-
ative hard thresholding (IHT) [4], and $\rho_S^{l_1}(\delta)$ for l_1-minimization [15]. Figure 1 in
[3] shows a plot of these phase transition curves, which are raised by approximately
$0.5 - 1\%$ when the bounds from [2] are replaced with the improved bounds from
Theorem 2.3.

A. Appendix. Here we present the proofs of the key theorems and lemmas stated in this paper. For theorems and lemmas stated or used without proof, see the appendix of [2].

A.1. Proof of Lemma 2.6. We start by showing that $\lambda^{\max}(\delta, \rho; \gamma)$ has a unique minimum for each fixed δ, ρ, and $\gamma \in [\rho, \delta^{-1}]$. Equation (2.5) gives the implicit relation between $\lambda^{\max}$ and γ as

$$\delta \psi_{\max}(\lambda^{\max}, \gamma) + H(\rho \delta) - \delta \gamma H(\rho/\gamma) = 0 \quad \text{for} \quad \lambda^{\max} \geq 1 + \gamma,$$

where

$$\psi_{\max}(\lambda^{\max}, \gamma) = \frac{1}{2}\left[(1 + \gamma) \ln(\lambda^{\max}) - \gamma \ln \gamma + 1 + \gamma - \lambda^{\max}\right].$$

Therefore,

$$\frac{d}{d\gamma}(\lambda^{\max}) = \frac{\lambda^{\max}}{\lambda^{\max} - (1 + \gamma)} \ln\left[\frac{\lambda^{\max} \cdot (\gamma - \rho)^2}{\gamma^3}\right]$$

is equal to zero when

(A.1) $$\lambda^{\max} \cdot (\gamma - \rho)^2 = \gamma^3.$$

Let $\gamma_{\min}$ satisfy (A.1). Since $\lambda^{\max} \geq 1 + \gamma > 0$, $\frac{d}{d\gamma}(\lambda^{\max})$ is negative for $\gamma \in [\rho, \gamma_{\min})$, is zero at $\gamma_{\min}$, and is positive for $\gamma \in (\gamma_{\min}, \delta^{-1}]$, equation (2.5) has a unique minimum over $\gamma \in [\rho, \delta^{-1}]$, and the γ that obtains the minimum is strictly greater than ρ.

Similarly, we show that $\lambda^{\min}(\delta, \rho; \gamma)$ has a unique maximum for each fixed δ, ρ and $\gamma \in [\rho, \delta^{-1}]$. Equation (2.4) gives the implicit relation between $\lambda^{\min}$ and γ as

$$\delta \psi_{\min}(\lambda^{\min}, \gamma) + H(\rho \delta) - \delta \gamma H(\rho/\gamma) = 0 \quad \text{for} \quad \lambda^{\min} \leq 1 - \gamma,$$

where

$$\psi_{\min}(\lambda^{\min}, \gamma) := H(\gamma) + \frac{1}{2}\left[(1 - \gamma) \ln(\lambda^{\min}) + \gamma \ln \gamma + 1 - \gamma - \lambda^{\min}\right].$$

Therefore,

$$\frac{d}{d\gamma}(\lambda^{\min}) = \frac{\lambda^{\min}}{(1 - \gamma) - \lambda^{\min}} \ln\left[\frac{\gamma^3 \cdot \lambda^{\min}}{(1 - \gamma)^2 \cdot (\gamma - \rho)^2}\right]$$

is equal to zero when

(A.2) $$\gamma^3 \cdot \lambda^{\min} = (1 - \gamma)^2 (\gamma - \rho)^2.$$

Let $\gamma_{\max}$ satisfy (A.2). Since $0 < \lambda^{\min} \leq 1 + \gamma$, $\frac{d}{d\gamma}(\lambda^{\min})$ is positive for $\gamma \in [\rho, \gamma_{\max})$, is zero at $\gamma_{\max}$, and is negative for $\gamma \in (\gamma_{\max}, \delta^{-1}]$, equation (2.4) has a unique maximum over $\gamma \in [\rho, \delta^{-1}]$, and the γ that obtains the maximum is strictly greater than ρ. □

A.2. Proof of main results, Theorem 2.3 and Propositions 2.12 and 2.13. Here we give a proof similar to that in [2], but we take great care with the non-exponential terms necessary for the calculation of bounds on probabilities for finite values of (k, n, N) in section 2.4. We present the proof for $\mathcal{U}^{BT}(\delta, \rho)$ in detail and sketch the proof of $\mathcal{L}^{BT}(\delta, \rho)$, which follows similarly.

The following lemma regarding the bound on the probability distribution function of the maximum eigenvalue of a Wishart matrix due to Edelman [2, 3, 12] is central to our proof.

LEMMA A.1 (see [12], presented in this form in [2]). *Let A_M be a matrix of size $n \times m$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Let $f_{\max}(m, n; \lambda)$ denote the distribution function for the largest eigenvalue of the derived Wishart matrix $A_M^* A_M$ of size $m \times m$. Then $f_{\max}(m, n; \lambda)$ satisfies*

(A.3)

$$f_{\max}(m, n; \lambda) \le \left[(2\pi)^{\frac{1}{2}} (n\lambda)^{-\frac{3}{2}} \left(\frac{n\lambda}{2} \right)^{\frac{n+m}{2}} \frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \right] e^{-\frac{n\lambda}{2}} =: g_{\max}(m, n; \lambda).$$

It is helpful at this stage to rewrite Lemma A.1, separating the exponential and polynomial parts (with respect to n) of $g_{\max}(m, n; \lambda)$ as follows.

LEMMA A.2. *Let $\gamma_n = m/n \in [\rho_n, \delta_n^{-1}]$, and define*

(A.4)
$$\psi_{\max}(\lambda, \gamma) := \frac{1}{2}\left[(1 + \gamma) \ln \lambda - \gamma \ln \gamma + 1 + \gamma - \lambda \right].$$

Then

(A.5)
$$f_{\max}(m, n; \lambda) \le g_{\max}(m, n; \lambda) \le p_{\max}(n, \lambda; \gamma) \exp\left(n \cdot \psi_{\max}(\lambda, \gamma) \right),$$

where $p_{\max}(n, \lambda; \gamma)$ is a polynomial in n, λ, and γ, given by

(A.6)
$$p_{\max}(n, \lambda; \gamma) = \left(\frac{8}{\pi} \right)^{1/2} \gamma^{-1} n^{-7/2} \lambda^{-3/2}.$$

Proof. Let $\gamma_n = \frac{m}{n}$ and $\frac{1}{n} \ln[g_{\max}(m, n; \lambda)] = \Phi_1(m, n; \lambda) + \Phi_2(m, n; \lambda) + \Phi_3(m, n; \lambda)$, where

$$\Phi_1(m, n; \lambda) = \frac{1}{2n} \ln(2\pi) - \frac{3}{2n} \ln(n\lambda), \quad \Phi_2(m, n; \lambda) = \frac{1}{2}\left[(1 + \gamma) \ln\left(\frac{n\lambda}{2}\right) - \lambda \right],$$

and

$$\Phi_3(m, n; \lambda) = -\frac{1}{n} \ln\left(\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) \right).$$

We simplify $\Phi_3(m, n; \lambda)$ by using the second Binet's log gamma formula [25]

(A.7)
$$\ln\left(\Gamma(z)\right) \ge (z - 1/2) \ln z - z + \ln\sqrt{2\pi}.$$

Thus we have

$$\Phi_2(m, n; \lambda) + \Phi_3(m, n; \lambda) \le \frac{1}{2}\left[(1 + \gamma_n) \ln \lambda - \gamma_n \ln \gamma_n + 1 + \gamma_n - \lambda \right] - n^{-1} \ln\left(\pi \gamma_n n^2 / 2 \right).$$

Incorporating $\Phi_1(m, n; \lambda)$ and $-n^{-1} \ln\left(\pi \gamma_n n^2 / 2 \right)$ into $p_{\max}(n, \lambda; \gamma)$ and defining the exponent (A.4) as

$$\psi_{\max}(\lambda, \gamma) := \lim_{n\to\infty} \frac{1}{n} \ln\left(g_{\max}(m, n; \lambda) \right) = \frac{1}{2}\left[(1 + \gamma) \ln \lambda - \gamma \ln \gamma + \gamma + 1 - \lambda \right]$$

completes the proof. ☐

The upper RIC bound $\mathcal{U}^{BT}(\delta, \rho)$ is obtained by constructing the groups $\mathcal{G}_i$ according to Lemma 2.4, taking a union bound over all $u = rN$ groups, and bounding the extreme eigenvalues within a group by the extreme eigenvalues of the Wishart matrices $A_{M_i}^* A_{M_i}$; see (2.9). In preparation for bounding the right-hand side of (2.9), we compute a bound on $rN g_{\max}(m, n; \lambda)$.

From Lemma A.2 and (2.8) we have

$$(A.8) \qquad 2\lambda N \binom{N}{k}\binom{m}{k}^{-1} g_{\max}(m, n; \lambda) \le p'_{\max}(n, \lambda) e^{n\psi_U(\lambda, \gamma)},$$

where

$$\psi_U(\lambda, \gamma) := \delta^{-1}\left[H(\rho\delta) - \delta\gamma H\left(\frac{\rho}{\gamma}\right) + \delta\psi_{\max}(\lambda, \gamma)\right]$$

and

$$(A.9) \qquad p'_{\max}(n, \lambda) := 2\lambda \left(\frac{5}{4}\right)^3 \left(\frac{nN(\gamma - \rho)}{\gamma\delta(1 - \rho\delta)}\right)^{1/2} p_{\max}(n, \lambda; \gamma).$$

The proof of Proposition 2.12 then follows.

Proof of Proposition 2.12. For $\epsilon > 0$ with $\lambda^{\max}(\delta, \rho) = \min_\gamma \lambda^{\max}(\delta, \rho; \gamma)$ being the optimal solution to (2.5),

$$\mathbf{P}\Big(U(k, n, N) > U(\delta_n, \rho_n) + \epsilon\Big) = \mathbf{P}\Big(U(k, n, N) > \lambda^{\max}(\delta_n, \rho_n) - 1 + \epsilon\Big)$$

$$= \mathbf{P}\Big(1 + U(k, n, N) > \lambda^{\max}(\delta_n, \rho_n) + \epsilon\Big)$$

$$= N\binom{N}{k}\binom{m}{k}^{-1} \int_{\lambda^{\max}(\delta_n, \rho_n)+\epsilon}^{\infty} f_{\max}(m, n; \lambda)d\lambda$$

$$(A.10) \qquad \le N\binom{N}{k}\binom{m}{k}^{-1} \int_{\lambda^{\max}(\delta_n, \rho_n)+\epsilon}^{\infty} g_{\max}(m, n; \lambda)d\lambda.$$

To bound the final integral in (A.10), we write $g_{\max}(m, n; \lambda)$ as a product of two separate functions, one of λ and the other of n and γ_n, as $g_{\max}(m, n; \lambda) = \varphi(n, \gamma_n)\lambda^{-\frac{3}{2}}\lambda^{\frac{n}{2}(1+\gamma_n)}e^{-\frac{n}{2}\lambda}$, where

$$\varphi(n, \gamma_n) = (2\pi)^{\frac{1}{2}}(n)^{-\frac{3}{2}}\left(\frac{n}{2}\right)^{\frac{n}{2}(1+\gamma_n)}\frac{1}{\Gamma\left(\frac{n}{2}\gamma_n\right)\Gamma\left(\frac{n}{2}\right)}.$$

With this and using the fact that $\lambda^{\max}(\delta_n, \rho_n) > 1 + \gamma_n$ and that $\lambda^{\frac{n}{2}(1+\gamma_n)}e^{-\frac{n}{2}\lambda}$ is strictly decreasing in λ on $[\lambda^{\max}(\delta_n, \rho_n), \infty)$, we can bound the integral in (A.10) as follows:

$$\int_{\lambda^{\max}(\delta_n, \rho_n)+\epsilon}^{\infty} g_{\max}(m, n; \lambda)d\lambda$$

$$\le \varphi(n, \gamma_n)[\lambda^{\max}(\delta_n, \rho_n) + \epsilon]^{\frac{n}{2}(1+\gamma_n)}e^{-\frac{n}{2}(\lambda^{\max}(\delta_n, \rho_n)+\epsilon)} \int_{\lambda^{\max}(\delta_n, \rho_n)+\epsilon}^{\infty} \lambda^{-\frac{3}{2}}d\lambda$$

$$= [\lambda^{\max}(\delta_n, \rho_n) + \epsilon]^{\frac{3}{2}}g_{\max}\big[m, n; \lambda^{\max}(\delta_n, \rho_n) + \epsilon\big] \int_{\lambda^{\max}(\delta_n, \rho_n)+\epsilon}^{\infty} \lambda^{-\frac{3}{2}}d\lambda$$

$$(A.11) \qquad = 2[\lambda^{\max}(\delta_n, \rho_n) + \epsilon]g_{\max}\big[m, n; \lambda^{\max}(\delta_n, \rho_n) + \epsilon\big].$$

Thus (A.10) and (A.11) together give

$$\mathbf{P}\Big(U(k,n,N) > U(\delta_n, \rho_n) + \epsilon\Big)$$
$$\leq 2[\lambda^{\max}(\delta_n, \rho_n) + \epsilon] r N g_{\max}\big[m, n; \lambda^{\max}(\delta_n, \rho_n) + \epsilon\big],$$
$$\leq p'_{\max}\Big(n, \lambda^{\max}(\delta_n, \rho_n) + \epsilon\Big) \exp\big[n \cdot \psi_U\big(\lambda^{\max}(\delta_n, \rho_n) + \epsilon\big)\big]$$

(A.12)
$$\leq p'_{\max}\Big(n, \lambda^{\max}(\delta_n, \rho_n) + \epsilon\Big) \exp\Big[n\epsilon \cdot \frac{d}{d\lambda}\psi_U\Big(\lambda^{\max}(\delta_n, \rho_n)\Big)\Big],$$

where $r = \binom{N}{k}\binom{m}{k}^{-1}$, and the last inequality is due to $\psi_U(\lambda)$ being strictly concave. $\square$

The following is a corollary to Proposition 2.12.

COROLLARY A.3. *Let $(\delta, \rho) \in (0,1)^2$, and let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Define $\mathcal{U}^{BT}(\delta, \rho) = \lambda^{\max}(\delta, \rho) - 1$, where $\lambda^{\max}(\delta, \rho; \gamma)$ is the solution of (2.5) for each $\gamma \in [\rho, \delta^{-1}]$ and $\lambda^{\max}(\delta, \rho) := \min_\gamma \lambda^{\max}(\delta, \rho; \gamma)$. Then for any $\epsilon > 0$, in the proportional-growth asymptotics*

$$\mathbf{P}\left(U(k,n,N) > \mathcal{U}^{BT}(\delta, \rho) + \epsilon\right) \to 0$$

exponentially in n.

Proof. From (A.12), since $\frac{d}{d\lambda}\psi_U\left(\lambda^{\max}(\delta_n, \rho_n)\right) < 0$ is strictly bounded away from zero and all the limits of (δ_n, ρ_n) are smoothly varying functions, we conclude for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(U(k,n,N) > \mathcal{U}^{BT}(\delta, \rho) + \epsilon\right) \to 0. \quad \square$$

Thus we finish the proof for $\mathcal{U}^{BT}(\delta, \rho)$. We sketch the similar proof for Proposition 2.13 and $\mathcal{L}^{BT}(\delta, \rho)$. Bounds on the probability distribution function of the minimum eigenvalue of a Wishart matrix are given in the following lemma.

LEMMA A.4 (see [12], presented in this form in [2]). *Let A_M be a matrix of size $n \times m$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. Let $f_{\min}(m, n; \lambda)$ denote the distribution function for the smallest eigenvalue of the derived Wishart matrix $A_M^* A_M$, of size $m \times m$. Then $f_{\min}(m, n; \lambda)$ satisfies*

$$f_{\min}(m, n; \lambda) \leq \left(\frac{\pi}{2n\lambda}\right)^{\frac{1}{2}} \left(\frac{n\lambda}{2}\right)^{\frac{n-m}{2}} \left[\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m+1}{2}\right)\Gamma\left(\frac{n-m+2}{2}\right)}\right] e^{-\frac{n\lambda}{2}}$$

(A.13)
$$=: g_{\min}(m, n; \lambda).$$

Again an explicit expression of $f_{\min}(m, n; \lambda)$ in terms of exponential and polynomial parts leads to the following lemma.

LEMMA A.5. *Let $\gamma_n = m/n$, and define*

(A.14)
$$\psi_{\min}(\lambda, \gamma) := H(\gamma) + \frac{1}{2}\Big[(1-\gamma)\ln \lambda + \gamma \ln \gamma + 1 - \gamma - \lambda\Big].$$

Then

(A.15)
$$f_{\min}(m, n; \lambda) \leq g_{\min}(m, n; \lambda) \leq p_{\min}(n, \lambda) \exp\Big(n \cdot \psi_{\min}(\lambda, \rho, \delta; \gamma)\Big),$$

where $p_{\min}(n, \lambda; \gamma)$ is a polynomial in n, λ, and γ, given by

(A.16)
$$p_{\min}(n, \lambda; \gamma) = \frac{e}{2\pi\sqrt{2\lambda}}.$$

The proof of Lemma A.5 follows that of Lemma A.2 and is omitted for brevity. Equipped with Lemma A.5, a large deviation analysis yields

$$(A.17) \qquad 2\lambda N \binom{N}{k} \binom{m}{k}^{-1} g_{\max}(m,n;\lambda) \leq p'_{\min}(n,\lambda) e^{n\psi_L(\lambda,\gamma)},$$

where

$$\psi_L(\lambda,\gamma) := \delta^{-1} \left[H(\rho\delta) - \delta\gamma H\left(\frac{\rho}{\gamma}\right) + \delta\psi_{\min}(\lambda,\gamma) \right]$$

and

$$(A.18) \qquad p'_{\min}(n,\lambda) := 2\lambda \left(\frac{5}{4}\right)^3 \left(\frac{nN(\gamma-\rho)}{\gamma\delta(1-\rho\delta)}\right)^{1/2} p_{\min}(n,\lambda).$$

With Lemma A.5 and (A.17), Proposition 2.13 follows similarly to the proof of Proposition 2.12 stated earlier in this section. The bound $\mathcal{L}^{BT}(\delta,\rho)$ is a corollary of Proposition 2.13 (Corollary A.6), the proof of which is the same as the proof of Corollary A.3.

COROLLARY A.6. *Let $(\delta,\rho) \in (0,1)^2$, and let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0,1/n)$. Define $\mathcal{L}^{BT}(\delta,\rho) := 1 - \lambda_{\min}(\delta,\rho)$, where $\lambda_{\min}(\delta,\rho;\gamma)$ is the solution of (2.4) for each $\gamma \in [\rho,\delta^{-1}]$ and $\lambda^{\min}(\delta,\rho) := \min_\gamma \lambda^{\min}(\delta,\rho;\gamma)$. Then for any $\epsilon > 0$, in the proportional-growth asymptotic*

$$\mathbf{P}\left(L(k,n,N) > \mathcal{L}^{BT}(\delta,\rho) + \epsilon\right) \to 0$$

exponentially in n.

REFERENCES

[1] J. D. BLANCHARD, C. CARTIS, AND J. TANNER, *Decay properties for restricted isometry constants*, IEEE Signal Process. Lett., 16 (2009), pp. 572–575.

[2] J. D. BLANCHARD, C. CARTIS, AND J. TANNER, *Compressed sensing: How sharp is the RIP?*, SIAM Rev., accepted.

[3] J. D. BLANCHARD, C. CARTIS, J. TANNER, AND A. THOMPSON, *Phase transitions for greedy sparse approximation algorithms*, Appl. Comput. Harmon. Anal., to appear.

[4] T. BLUMENSATH AND M. E. DAVIES, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 265–274.

[5] E. J. CANDÈS, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 589–592.

[6] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[7] E. J. CANDÈS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[8] W. DAI AND O. MILENKOVIC, *Subspace pursuit for compressive sensing signal reconstruction*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2230–2249.

[9] A. D'ASPREMONT, F. BACH, AND L. EL GHAOUI, *Optimal solutions for sparse principal component analysis*, J. Mach. Learn. Res., 9 (2008), pp. 1269–1294.

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

[11] C. DOSSAL, G. PEYRÉ, AND J. FADILI, *A numerical exploration of compressed sampling recovery*, Linear Algebra Appl., 432 (2010), pp. 1663–1679.

[12] A. EDELMAN, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 543–560.

[13] A. EDELMAN AND N. R. RAO, *Random matrix theory*, Acta Numer., 14 (2005), pp. 233–297.

[14] EDINBURGH COMPRESSED SENSING RESEARCH GROUP, http://ecos.maths.ed.ac.uk.

[15] S. FOUCART AND M.-J. LAI, *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 395–407.

[16] M. JOURNÉE, Y. NESTEROV, P. RICHTÁRIK, AND R. SEPULCHRE, *Generalized power method for sparse principal component analysis*, J. Mach. Learn. Res., 11 (2010), pp. 517–553.

[17] A. JUDITSKY AND A. NEMIROVSKI, *On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization*, Math. Program. Ser. B., submitted.

[18] D. KLEITMAN, *private communication*, 2009.

[19] J. B. KRUSKAL, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and Appl., 18 (1977), pp. 95–138.

[20] M. LEDOUX, *The Concentration of Measure Phenomenon*, Math. Surveys Monogr. 89, American Mathematical Society, Providence, RI, 2001.

[21] S. MENDELSON, A. PAJOR, AND N. TOMCZAK-JAEGERMANN, *Reconstruction and subgaussian operators in asymptotic geometric analysis*, Geom. Funct. Anal., 17 (2007), pp. 1248–1282.

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

[23] D. NEEDELL AND J. TROPP, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–321.

[24] H. RAUHUT, *Compressive sensing and structured random matrices*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, Radon Series Comp. Appl. Math. 9, Walter de Gruyter, Berlin, 2010, pp. 1–92.

[25] E. T. WHITTAKER AND G. N. WATSON, *A Course of Modern Analysis*, 4th ed., Cambridge University Press, London, 1946.