Condition Numbers of Random Matrices*

STANISLAW J. SZAREK

Department of Mathematics, Case Western Reserve University, Cleveland, Ohio, and
Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France

Received October 2, 1990

1. INTRODUCTION AND THE MAIN RESULTS

Let $G = G_n(\omega)$ be an $n \times n$ random matrix with independent Gaussian
entries $g_{ij}$ (real or complex), defined on some probability space $(\Omega, \mathcal{P})$ and
distributed according to the $N(0,1)$ law. In the theory of computational
complexity it is of interest to consider the "random condition number"
$\|G_n^{-1}\| \cdot \|G_n\|$; in particular a question about the exact order of

$$\mathbb{E} \log(\|G_n^{-1}\| \cdot \|G_n\|),$$

where $\|\cdot\|$ denotes the operator norm on the Euclidean space and $\mathbb{E}$ the
expected value, was asked in (Smale, 1985) (this quantity may be interpreted as the average "loss of precision" when solving large systems of
linear equations).

It is very well known that $n^{-1/2}\mathbb{E}\|G_n\| \to 2$ as $n \to \infty$; moreover, $\mathbb{E}\|G_n\| < 2n^{1/2}$, $\mathbb{E} \left| n^{-1/2}\|G_n\| - 2 \right| \to 0$, $\mathbb{P}(\|G_n\| > \beta n^{1/2}) \leq C \exp(-c\beta^2 n)$, $\mathbb{P}(\|G_n\| < \alpha n^{1/2}) \leq (Ca)^n$, etc. Consequently, as far as the condition numbers are
concerned, essentially the only unknown is the behavior of $G_n^{-1}$. Above
and in what follows, $C, c, \text{ etc.}$, denote universal (effectively computable)
umerical constants, most notably independent of $n$; however, identical
symbols may represent different numbers in different places.

In this paper we deal with a more general setup, which covers, e.g., the
case when $\mathbb{R}^n$ is endowed with the $l_p^n$ norm $\|\cdot\|_p$ (for some $p \in [1, \infty]$); we
denote the corresponding operator norm by $\|\cdot\|_{p \to p}$ (in fact our methods
allow us to handle arbitrary norms, on both the domain and the range of
$G$; see Remark 4.1). We then have the following

* Research partially supported by a grant from the National Science Foundation. A Pre-
    liminary version of this paper was circulated as a preprint IHES/89/36, May 1989.
Theorem 1.1. There exist universal constants $C, c > 0$ such that, for any positive integer $n$ and any $p \in [1, \infty]$, 

$$c l(n, p) \leq \exp(E \log(\|G_n^{-1}\|_{p \to p} \cdot \|G_n\|_{p \to p})) \equiv Cl(n, p),$$

where $l(n, p) = n^{1 + |1/2 - 1/p|} (\min\{q^*, 1 + \log n\})^{1/2}$ and $q^* = \max\{p, p/(p - 1)\}$. For $p = 2$, one can take as the upper estimate $(2e)^{3/2} n$. 

Remark. One has similar estimates for $(E(\|G_n\|_{p \to p} \cdot \|G_n\|_{p \to p})^{1/p})$ when $p < 1$, with $C$ depending on $p$ as $p \to 1$ (resp. $p < 2$, $p \to 2$ in the complex case). This comment also applies to Theorem 1.3.

We point out that, for $p$ bounded away from 1 and $\infty$, $l(n, p)$ reduces effectively to $n^{1 + |1/2 - 1/p|}$. If $p = 2$, one gets an answer to the question of Smale mentioned above. In that case, however, the estimates involved were obtained independently by several researchers over the last year or so; see the comments following Theorem 1.2 below. If $p = \infty$ (or 1; these cases arise naturally), we get the order $n^{3/2} (1 + \log n)^{1/2}$.

For an $n \times n$ matrix $A$, let $(s_j(A))_{1 \leq j \leq n}$ be its $s$-numbers (i.e., the eigenvalues of $|A|$, multiplicities counted), which we choose to arrange in the nondecreasing order. Since, e.g., $\|A^{-1}\| = s_1(A)^{-1}$, the study of condition numbers quickly leads to the study of $s$-numbers and eigenvalues of random matrices. This is a subject that has been studied extensively: in nuclear physics, beginning with the work of E. Wigner (see also Porter, 1965; Mehta, 1967; Carmeli, 1983), and especially in the multivariate statistics (see Silverstein, 1986, and references therein). Here, we derive Theorem 1.1 from the following result, which is of independent interest.

Theorem 1.2. There exist universal constants $C_1, c_1, C_2, c_2, \beta_0 > 0$ such that, for any positive integer $n$,

$$\exp(-C_1 \beta^2 j^2) \leq \mathbb{P}(s_j(G_n) > \beta j/n^{1/2}) \leq \exp(-c_1 \beta^2 j^2)$$

(1.1)

for $\beta \geq \beta_0$ and $j \leq n/2$;

$$(c_2 \alpha)^2 \leq \mathbb{P}(s_j(G_n) < \alpha j/n^{1/2}) \leq (C_2 \alpha)^2$$

(1.2)

for $\alpha \geq 0$ and $j \leq n$ (e.g., with $C_2 = (2e)^{1/2}$). In the complex case we need to replace $j^2$ by $2j^2$.

Inequalities (1.1) and (1.2) improve Theorem 1.3 in (Szarek, 1990) (where upper estimates of the form $C \exp(-cj^2)$ were obtained for some fixed $\alpha, \beta, \gamma, c > 0$, and used in some constructions in the local theory of Banach spaces) and are proved by appropriately modifying the arguments from that paper.
Let us observe that the case $p = 2$ of Theorem 1.1 (i.e., the answer to Smale's question) follows immediately from Theorem 1.2. Indeed, since $s_1(A)^{-1} = \|A^{-1}\|$, (1.2) applied with $j = 1$ shows that

$$P(\|G_n^{-1}\| > tn^{1/2}) \leq C_2/t;$$

hence, $E \log \|G_n^{-1}\| \leq \frac{1}{2} \log(C_2en)$ while, by the previous estimates, $E \log \|G_n\| < \log(2n^{1/2})$, and so

$$E \log(\|G_n^{-1}\| \cdot \|G_n\|) < \log(2C_2en).$$

The corresponding lower estimate is obtained similarly from the inequalities

$$P(s_n(G_n) = \|G_n\| < \alpha n^{1/2}) \leq (C_2\alpha)^2,$$

$$P(s_1(G_n)^{-1} = \|G_n^{-1}\| < \beta^{-1}n^{1/2}) \leq \exp(-c_1\beta^2),$$

which are just special cases of (1.2) and (1.1). Alternatively, as was observed by S. Heinrich (preprint), it may be simply derived even from Theorem 1.3 in (Szarek, 1990).

After a preliminary version of this paper was written, we learned that A. Edelman (1988) did show the upper estimate from Theorem 1.1 in the case $p = 2$; it is mentioned in that paper that E. Kostlan also proved the corresponding lower estimate; the paper being in preparation. One should also mention that somewhat weaker estimates in that direction were obtained earlier by A. Ocneanu (to appear) and E. Kostlan (1985) and that related problems were also considered in the meantime in (Blum and Shub, 1986; Demmel, 1988; Weiss et al., 1986). It appears, however, that none of the above-mentioned papers yields, for "nonextreme" s-numbers, the precise "distributional" information given by our Theorem 1.2.\footnote{A definitive treatment of the case of the extreme s-numbers and a very elegant exposition of some related topics can be found in the Ph.D. thesis of A. Edelman, MIT, May 1989.}

Since our estimates imply that $s_j(G_n)$ are virtually deterministic, one can perform all kinds of calculations. Roughly speaking, for any "reasonable" norm $\|\|$ on the set of matrices, one can (with some additional work and using perhaps the results of D. Slepian (1962), S. Chevet (1977–1978), Y. Gordon (1985), and others) in effect determine the distribution of $\|G_n\|$, $\|G_n^{-1}\|$, or $\|G_n^{-1}\| \cdot \|G_n\|$. For example, if $\|\|$ = $\|\|_{c_p}$ is the Schatten norm (i.e., $\|A\|_{c_p} = (\text{tr}|A|^p)^{1/p}$), one gets

**Theorem 1.3.** There exist universal constants $C$, $c$, $C'$, $c'$ > 0 such that, for any positive integer $n$,

$$cn^{1/2+1/p} \leq \exp(E \log \|G_n\|_{c_p}) \leq E \|G_n\|_{c_p} \leq Cn^{1/2+1/p}.$$
where $\beta(n, p) = n^{1/2}(\min\{q, 1 + \log n\})^{1/p}$ and $q = p/(p - 1)$.

The case $p = 2$ of Theorem 1.3 is clearly relevant to the "loss of precision" (see the comment at the beginning of this paper and cf. Smale (1985) and Kostlan (1985)), averaged also over "inputs" of linear systems.

Since existing software typically detects and rejects "nearly singular" matrices, it is of interest to analyze the quantity $(G^{-1/2}A(G)G^{-1/2}A(G))$ restricted to the set of "not-so-singular" matrices. We present here the following sample result (the real case only).

**Theorem 1.4.** For $K > 1$, let $E_K = \{G: G - an n \times n$ matrix with $(\|G^{-1/2}\|_2 + 2) \cdot (\|G\|_2 + 2) \leq K\}$. Then

$$c\delta(n, K) \leq \mathbb{P}(E_K)^{-1} \int_{E_K} \|G^{-1}\|_2 \cdot \|G\|_2 d\mathbb{P} \leq C\delta(n, K),$$

where $\delta(n, K) = K$ if $K \leq n$, $\delta(n, K) = n(1 + \log(K/n))$ if $K \geq n$, and $C, c > 0$ are universal constants.

It would be of some interest to consider analogous problems for Gaussian matrices, in which variances of entries depend on their positions, in particular for $k$-diagonal matrices (say, with $(i, j)$th entry distributed according to $N(0, 1)$ if $|i - j| < k/2$ and equal to 0 otherwise). Another interesting question would be to show similar estimates for non-Gaussian distributions, e.g., with $g_{ij}$'s distributed uniformly on, say, $[-\frac{1}{2}, \frac{1}{2}]$. We note here that Wigner's Semicircle Law, which was the motivation for Theorem 1.3 from (Szarek, 1988), is true just with mild moment assumptions (see Silverstein, 1986, and references therein).

The paper is organized as follows. Section 2 contains known and preliminary facts. In Section 3 we prove the main technical result of this paper, Theorem 1.2. In Section 4 we derive the remaining theorems from Theorem 1.2.

In our arguments and statements of results we concentrate on the real case; there are always similar (typically somewhat better) estimates in the complex case. We indicate the differences between the two cases where necessary.

2. KNOWN AND PRELIMINARY RESULTS

We begin by recalling "Chevet's inequality" (with an improvement due to Gordon and additional refinements). Let $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $Y = (\mathbb{R}^n, \|\cdot\|_Y)$ denote the space $\mathbb{R}^n$ endowed with two norms, and $\|\cdot\|_{X \to Y}$ the corresponding operator norm on $n \times n$ matrices. We then have
Lemma 2.1 (Chevet, 1977–1978; Gordon, 1985). Let $G = G_n$ be an $n \times n$ Gaussian matrix with independent $N(0, 1)$ entries, and let $g$ be the standard $\mathbb{R}^n$-valued Gaussian vector (i.e., $g = \sum_{1 \leq j \leq n} \gamma_j e_j$ with $\gamma_j$'s independent $N(0, 1)$ variables). Then

$$\max\{\mathbb{E}A, \mathbb{E}B\} \leq \mathbb{E}\|G\|_{\mathcal{X} \to \mathcal{Y}} \leq \mathbb{E}(A + B),$$

where $A = \|I\|_{\ell_2^2 \to \mathcal{Y}} \cdot \|g\|_{\mathcal{X}^*}$, $B = \|I\|_{\ell_2^2 \to \mathcal{X}} \cdot \|g\|_{\mathcal{Y}} (\|\cdot\|_{\mathcal{X}^*}$ is the norm dual to $\|\cdot\|_\mathcal{X}$ with respect to the standard scalar product). The same holds if we replace (everywhere) $\mathbb{E}\|\cdot\|_\mathcal{X}$ by $(\mathbb{E}\|\cdot\|_\rho)^{1/\rho}$ for some $\rho \in (0, \infty)$. Additionally, for any $t \geq 0$,

$$\mathbb{P}(\|G\|_{\mathcal{X} \to \mathcal{Y}} \leq t \max\{\mathbb{E}A, \mathbb{E}B\}) \leq Ct$$

and consequently

$$c\mathbb{E}(A + B) \leq \exp(\mathbb{E}\log\|G\|_{\mathcal{X} \to \mathcal{Y}}) \leq \mathbb{E}(A + B),$$

where $C, c > 0$ are universal constants.

We could not find the last two statements of the lemma in the literature. However, they follow easily from the preceding statements and Lemma 2.6, which we state and prove at the end of this section; see also the comments at the beginning of Section 4.

As the quantities $\mathbb{E}A$ and $\mathbb{E}B$ are usually easily computable, the lemma above gives nearly complete information about the distribution of $\|G\|_{\mathcal{X} \to \mathcal{Y}}$. We note in passing that the quantity $\mathbb{E}\|G\|_\mathcal{Y}$ is essentially the so-called "$l$-norm" of the formal identity $I = I_{1, \mathcal{Y}}; \ell_2^2 \to \mathcal{Y}$ and hence is closely related to the so-called "Levy mean" of $Y$ (cf. Milman and Schechtman, 1986). Note that if $X = Y = l_2^2$, then $\mathbb{E}A = \mathbb{E}B = \sqrt{n}$ (when $\rho = 2$). However, in that case we have slightly more precise information: the norm of $G$ is actually very close to $2\sqrt{n}$ on a set of nearly full measure if $n$ is large; we have, e.g.,

Lemma 2.2. Given $\varepsilon > 0$, there exists $N = N(\varepsilon)$ such that, for $n \geq N$, we have, in the notation of the previous lemma (with $X = Y = l_2^2$),

$$\mathbb{P}(2 - \varepsilon < n^{-1/2}\|G_n\|_{2 \to 2} < 2 + \varepsilon > 1 - \exp(-cn^2)),$$

where $c > 0$ is a universal constant. Also, one can taken $N \leq (c\varepsilon)^{-1} \log (1/\varepsilon)$.

Lemma 2.2 is most likely not optimal for small values of $\varepsilon$ (cf. Lemma 3.1). It is (easily) proved using, e.g., the method Silverstein (1986) or Szarek (1990); see also Geman (1980) for a similar statement in a more general (i.e., non-Gaussian) setting. We state explicitly the following well-known consequences of Lemmas 2.1 and 2.2.
Corollary 2.3. For the norm $\|G_n\| = \|G_n\|_{2 \to 2}$ we have

(a) $\mathbb{E}(n^{-1/2} \|G_n\|) < 2$ for all $n$,
(b) $\lim_{n \to \infty} \mathbb{E}(n^{-1/2} \|G_n\|) = 2$,
(c) $n^{-1/2} \|G_n\| \to 2$ almost surely as $n \to \infty$ (regardless of how $G_n$'s are related for different $n$).

The next lemma is needed to analyze the distribution of $G^{-1}$ and is quite well known (one of the inequalities is contained in (Marcus and Pisier, 1981, Sect. 5.1); the other one was observed later by Davis and Garling). We include the proof for completeness.

Lemma 2.4. In the notation of Lemma 2.1, let $U = U_n$ be $O(n)$-valued random matrix distributed uniformly on $O(n)$ (i.e., according to the Haar measure).

(i) Let $\|\cdot\|$ be a seminorm on the space of $n \times n$ matrices (e.g., $\|V\| = \|TV\|_{x \to y}$, where $T$ is a fixed matrix). Then

$$c \mathbb{E} \|U\| \leq \mathbb{E} \| n^{-1/2} G\| \leq C \mathbb{E} \|U\|,$$

where $C, c > 0$ are universal constants. One can replace in the above $\mathbb{E} \|\cdot\|$ by $(\mathbb{E} \|\rho\|^p)^{1/p}$ for any $\rho \in [1, n]$ (or for any $\rho \in [1, \infty)$ if we replace $C$ by $C\sqrt{1 + \rho/n}$).

(ii) Moreover, if $\|\cdot\|$ is an operator norm (i.e., $\|\cdot\| = \|\cdot\|_{x \to y}$), one additionally has

$$c \exp(\mathbb{E} \log \|U\|_{x \to y}) \leq \exp(\mathbb{E} \log \| n^{-1/2} G\|_{x \to y}) \leq C \exp(\mathbb{E} \log \|U\|_{x \to y}).$$

Remark 2.5. Part (ii) of the lemma fails, at least in some cases, for general seminorms: let $n = 2$ and $\|(a_{ij})\| = |a_{11} - a_{22}|$; then $\mathbb{P}(\|U\| = 0) = \frac{1}{2}$, while $\|2^{-1/2} G\|$ is distributed as the absolute value of an $N(0, 1)$ variable. We do not know what is the “right” most general statement in that direction.

Proof of Lemma 2.4. (i) We rely on the following observation: $n^{-1/2} G$ has the same distribution as $U' \Delta U''$, where $U'$ and $U''$ are independent copies of $U$ and $\Delta = (s_j \delta_{ij})$ is a (random) diagonal matrix with $s_j = s_j(n^{-1/2} G)$ (the singular number). Now, for fixed $\Delta$, $\Delta = \Sigma \lambda_j U_j$ with $\lambda_j \geq 0$, $\Sigma \lambda_j \leq s_n = \|\Delta\|_{2 \to 2} = n^{-1/2} \|G_n\|_{2 \to 2}$ and $U_j$’s-orthogonal diagonal matrices (depending on $\Delta$). Integrating, for fixed $\Delta$, over $U'$, $U''$ we get the second inequality with $C = \mathbb{E}(n^{-1/2} \|G_n\|_{2 \to 2}) < 2$; one just uses the fact that $U'U_jU''$ has the same distribution as $U$ (and the triangle inequality). The argument for general $\rho$ requires only minor modifications.

To show the lower estimate, we need, roughly speaking, to average $\Delta$ over permutations. For a permutation $\pi \in S(n)$, let $S_\pi = O(n)$ be defined by $S_\pi e_j = e_{\pi(j)}$. Then, clearly, the distribution of $G$ is the same as that of $U'S_\pi^{-1} \Delta S_\pi U''$, where $U'$, $U''$, and $\Delta$ are as before and $\pi$ varies (indepen-
dently of $U', U'', \Delta$) over $S(n)$ endowed with the normalized counting measure. Since $\frac{1}{n}S_n^{-1}A S_n = 1/n(S_{1<j<n} s_j)I$ and $\frac{1}{2} \leq \mathbb{E}(1/n \sum s_j^2) \leq 1$ (as follows easily from Lemma 2.1 and the fact that $\mathbb{E}(1/n \sum s_j^2) = 1$; in fact $\mathbb{E}(1/n \sum s_j) \to 8/(3\pi)$ as $n \to \infty$), we get the first inequality from Lemma 2.4 (i) with $c = \frac{1}{2}$.

(ii) The first inequality follows from the part (i) and the last statement of Lemma 2.1. Similarly, in order to prove the second inequality it suffices to show, e.g.,

$$\exp(\mathbb{E} \log \|U\|_{x \to y}) \geq c' \|I\|_{\ell^2 \to x^*} \mathbb{E} \|n^{-1/2}g\|_y.$$

Let $z \in \mathbb{R}^n$ be such that $\|z\|_2 = 1$, $\|z\|_2 = 1/\|I\|_{\ell^2 \to x^*} (= 1/\|I\|_{\ell^2 \to x^*})$. Then $\|U\|_{x \to y} \geq \|I\|_{\ell^2 \to x^*} \|Uz\|_y$ and $Uz$ is uniformly distributed on $S^{'-1}$. It remains to observe that

$$\mathbb{E} \log \|Uz\|_y + C' \geq \mathbb{E} \log \|n^{-1/2}g\|_y \geq \log (\mathbb{E} \|Uz\|_y) - C'',$

where the second inequality is a consequence of Lemma 2.6 (as indicated at the beginning of Section 4; cf. (4.3)), and the first one of the identity

$$\mathbb{E} (\log \|n^{-1/2}g\|_y) = \mathbb{E} (\log \|Uz\|_y) + \mathbb{E} (\log \|n^{-1/2}g\|_z),$$

the second term being handled again via (4.3).

The next lemma is the source of the last two statements from Lemma 2.1; it is also used in the discussion at the beginning of Section 4.

**Lemma 2.6.** Let $\mu = \mu_n$ be the standard Gaussian measure on $\mathbb{R}^n$. Let $b \in (0, 1)$ and let $K \subset \mathbb{R}^n$ be a symmetric convex body with $\mu(K) \leq b$. Then, for any $t \in [0, 1]$,

$$\mu(tK) \leq \kappa t \mu(K),$$

where $\kappa > 0$ depends only on $b$.

**Remark 2.7.** It seems plausible that the following stronger "isoperimetric" statement actually holds.

If $\mu = \mu_n$ and $K$ are as in the lemma, and if $a \geq 0$ is such that $\mu([-a, a]) = \mu(K)$, then, for any $t \in [0, 1]$, $\mu(tK) \leq \mu_1([-ta, ta])$. This would constitute a symmetric version of the theorem of Landau and Shepp (see Landau and Shepp, 1971, or Badrikian and Chevet, 1974, Cor. VIII.1.2). We were not able to find a proof or a reference in the literature.2

2 In the special case when $n = 2$, or when $K$ is symmetric with respect to the coordinate planes, this was shown recently by M. Sawa (to appear).
Proof of Lemma 2.6. Assume, for simplicity, that \( b = \mu(K) = \frac{1}{4} \). Let \( \lambda \) be the radius of the largest Euclidean ball (centered at 0) contained in \( K \). It follows that \( a \leq \lambda \leq n^{\frac{1}{2}} \), where \( a > 0 \) is such that \( \mu_i([-a, a]) = \frac{1}{2} (= \mu(K)) \); in general \( a \) depends on \( \mu(K) \)). Without loss of generality we can assume that \( K \subset \{ x = (x_j) : |x_1| \leq \lambda \} \); in particular, for any \( t \geq 0 \),

\[
\mu(tK) \leq \mu_i([-t\lambda, +\lambda]) \leq (2/\pi)^{1/2} t \lambda.
\]

This proves the lemma if \( \lambda \) is “not too large.” For “large” \( \lambda \), we argue differently: by Borell’s (1975) Gaussian isoperimetric inequality,

\[
\mu(tK) \leq \mu_i((1 - t)\lambda, \infty))
\]

for \( t \in [0, 1] \). This is an estimate much better than the one required by the lemma provided \( t \) is not “too small”; it certainly works for \( t \in [\lambda^{-1}, 1] \), in particular we have

\[
\mu(\lambda^{-1}K) \leq C_0 \lambda^{-1}.
\]

Note that if \( U = \lambda^{-1}K \), then \( U \subset \{ x : |x_1| \leq 1 \} \). We claim that, for such \( U \),

\[
\mu(tU) \leq C_1 t \mu(U) \tag{2.1}
\]

for \( t \in [0, 1] \). A combination of the facts above then clearly yields the lemma. To prove (2.1), let, for \( v \in \mathbb{R}^+ \), \( S_v \in L(\mathbb{R}^n) \) be defined by \( S_v(x_1, x_2, \ldots, x_n) = (v x_1, x_2, \ldots, x_n) \). It is well known that, for any (convex symmetric) \( U \), \( \mu(S_vU) \leq \mu(U) \), hence also

\[
\mu(vU) \leq \mu(S_vU) \tag{2.2}
\]

for \( v \in [0, 1] \) (see, e.g., Badrikian and Chevet, 1974, Lemme VI.2.2). On the other hand,

\[
\mu(S_vU) \leq C_1 v \mu(U) \tag{2.3}
\]

for any measurable \( U \subset \{ x : |x_1| \leq 1 \} \) (regardless of convexity, etc.). Indeed, denoting by \( \rho \) the density of \( \mu \) with respect to the Lebesgue measure, we have

\[
\mu(S_vU) = \int_{S_vU} \rho(x)dx = v \int_U \rho(S_vy)dy = v \int_U \frac{\rho(S_vy)}{\rho(y)} \rho(y)dy \\
\leq vC_1 \int_U \rho(y)dy = C_1 v \mu(U).
\]
The inequality follows just from the fact that, if $\rho_1(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ is the density of $\mu_1$, then $\rho_1(vx)/\rho_1(x) = e^{v^2/2} / \rho_1(1)$ if $v \in [0, 1]$ and $|x| \leq 1$. This shows (2.3); combining (2.3) with (2.2) we get (2.1), concluding the proof of the lemma.

3. The Technical Result

In this section we prove Theorem 1.2. We shall concentrate on the real case, which is somewhat more difficult to handle; see Remark 3.2(c) for comments regarding the complex version.

It will be more convenient to change the normalization. Let $l_1 \leq l_2 \leq \cdots \leq l_n$ be the eigenvalues of $\|G\|^2$; then (1.1) and (1.2) become

$$\langle c, \alpha \rangle k^2 \leq \mathbb{P}(l_k \leq \alpha^2 k^2) \leq \langle C, \alpha \rangle k^2$$

$$\exp(-C_1 k^2 \beta^2) \leq \mathbb{P}(l_k \geq \beta^2 k^2) \leq \exp(-c_2 k^2 \beta^2).$$

We follow the argument from (Szarek, 1990, Sect. 6) with a few additional subtleties needed to accommodate arbitrary $\alpha, \beta$. We use the well-known formula for the joint density of the singular numbers of a Gaussian matrix. If

$$L \equiv \{l = (l_j) : 0 \leq l_1 \leq \cdots \leq l_n \} \subset \mathbb{R}^n,$$

then the density is given by (see Carmeli; 1983, Krishnaiah and Chang, 1971, or Wilks, 1963)

$$\rho(l) = c(n) \exp \left( -\frac{1}{2n} \sum_{j=1}^n l_j \right) \prod_{1 \leq i < j \leq n} |l_i - l_j| \left( \prod_{1 \leq j \leq n} l_j \right)^{-1/2}$$

for $l \in L$ and $\rho = 0$ on $\mathbb{R}^n \setminus L$; the constant $c(n)$ is such that $\int \rho d\lambda = 1$ ($\lambda$ is the Lebesgue measure). We handle the upper estimates in (3.1), (3.2) first and then indicate the changes needed to obtain the lower ones (the latter are only marginally used in our applications). The main trick is as follows: given $E \subset L$ and a (piecewise smooth one-to-one) function $\Phi: E \rightarrow L$, one has

$$\mathbb{P}(E) = \int_E \rho(l) dl = \int_{\phi(E)} \rho(\Phi^{-1}(l')) |J_{\Phi^{-1}}(l')| dl'$$

$$= \int_{\phi(E)} \left[ \frac{\rho(\Phi^{-1}(l'))}{\rho(l')} |J_{\Phi^{-1}}(l')| \right] \rho(l') dl' \leq \sup_{l \in E} \frac{\rho(l)}{\rho(\Phi(l))} |J_{\Phi}(l)|^{-1}. \quad (3.4)$$
Proof of (3.1). The Upper Estimate: Fix $\alpha \in [0, 2^{-1/2}]$ and $k \leq n$, and denote

$$E_k = \{ l \in \mathcal{L} : l_k \leq \alpha^2 k^2 \}.$$  

Now define $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ to satisfy:

(a) $\varphi(0) = 0$, $\varphi(2\alpha^2 k^2) = k^2$,
(b) $\varphi$ is affine in $[0, 2\alpha^2 k^2]$ and $[2\alpha^2 k^2, \infty]$ with $\varphi(x) = x + b$ in the second interval (note that $b = k^2 - 2\alpha^2 k^2 < k^2$; see Fig. 1).

Let $\Phi : \mathcal{L} \to \mathcal{L}$ be given by $\Phi((l_j)_{1 \leq j \leq n}) = (\varphi(l_j))_{1 \leq j \leq n})$. By (3.4) and (3.3), we need to estimate, for $l \in E_k$,

$$\exp \left( \frac{1}{2n} \sum_{j=1}^{n} (\varphi(l_j) - l_j) \right) \prod_{1 \leq j \leq n} \left[ \left( \prod_{i < j} \frac{l_j - l_i}{\varphi(l_j) - \varphi(l_i)} \right) \cdot (\varphi(l_j))^{1/2} \cdot (\varphi'(l_j))^{-1} \right]$$

$$= P_0 \prod_{1 \leq j \leq n} P_j. \quad (3.5)$$

Since $\varphi(x) - x \leq b < k^2$ for all $x$, $P_0 \leq \exp(k^2/2)$. If $l_j \leq 2\alpha^2 k^2$, then $P_j = (\alpha^{-2}/2)^{-j + 1/2} \leq 1$. Since $l \in E_k$, this happens at least for $j = 1, 2, \ldots, k$ and so the contribution of such factors is $\leq (2^{1/2}\alpha)^k$. On the other hand, if $x > 2\alpha^2 k^2$ and $x' \leq \alpha^2 k^2$, then

$$\frac{x - x'}{\varphi(x) - \varphi(x')} \left( \frac{\varphi(x)}{x} \right)^{1/2} \leq 1.$$
Using this, one easily shows that if $l_j > 2\alpha^2 k^2$ (hence $j > k$, $\varphi'(l_j) = 1$), then $P_j \leq 1$. Putting these inequalities together, we get the upper estimate from (3.1) with $C_1 = (2e)^{1/2}$.

For the proof of (3.2) we need the following fact.

**Lemma 3.1.** There exist positive constants $\lambda_0$, $c$, $C$ (e.g., $\lambda_0 = 8$, $c = 2^{-4}$) such that, in the notation of this section, we have, for $\lambda \geq \lambda_0$, and $d \leq n$,

$$\exp(-C\lambda^2 n(n - d + 1)) \leq \mathbb{P}(l_d \geq \lambda^2 n^2) \leq \exp(-c\lambda^2 n(n - d + 1)).$$

The upper estimate is essentially Lemma 2.9 from (Szarek, 1990), the lower follows by a similar, but easier, argument (which we omit).

**Proof of (3.2).** First observe that Lemma 3.1 takes care of large $\beta$ ($\beta \geq \lambda_0 n/k$ or just $\beta \geq c_0 n/k$) and, consequently, of all $\beta$ when $k$ is comparable with $n$ (but, say, $\leq n/2$). Thus we may assume that $k < n/4$, $\beta < \lambda_0 n/k$. Fix such $k$, $\beta$ and set

$$F_k = \{l \in \mathcal{L} : l_k \geq \beta^2 k^2\}.$$

We define a piecewise affine function $\varphi: \mathbb{R}^+ \to \mathbb{R}^+$ to satisfy

(a) $\varphi(0) = 0$, $\varphi(\beta^2 k^2/2) = k^2$, $\varphi(\lambda_0^2 n^2) = \lambda_0^2 n^2$,

(b) $\varphi$ is affine in each of the intervals in between with $\varphi(x) = x$ for $x \in [\lambda_0^2 n^2, \infty)$ (see Fig. 2).
Note that, in particular, \( \phi'(x) = m \) for \( x \in [\beta^2k^2/2, \lambda_0n^2] \), with

\[
m^{-1} < 1 - (\beta^2/2 - 1)k^2/(\lambda_0^2n^2) \leq 1 - c'\beta^2k^2/n^2
\]

(with \( c' = 2^{-8} \) if we take \( \lambda_0 = 8 \) and \( \beta \geq 2 \)). Additionally set

\[
F = \{ l \in \mathcal{L} : L_{[5n]} \supseteq \lambda_0^2n^2 \}, \quad \tilde{E}_j = \{ l \in \mathcal{L} : l_j \leq c_0^2j^2 \},
\]

where \( c_0 = (2e^3)^{-1/2} \) and

\[
F_k' = F_k \setminus \bigcup_{j \geq \beta k/c_0} \tilde{E}_j. \tag{3.6}
\]

In view of the part of (4.1) already proved and Lemma 3.1, it is enough to estimate \( P(F_k') \). Consider the expression (3.5) for \( l \in F_k' \) and for the present choice of \( \phi \). Since \( \phi(x) \leq x \) for all \( x \), \( P_0 \leq 1 \). Also, \( P_j \leq 1 \) if \( l_j > \lambda_0^2n^2 \). If \( l_j < \beta^2k^2/2 \), then \( P_j = (2^{j-(1/2)} - 1) \); this also works as an upper bound for all \( P_j \)'s. Since there are at most \( k \) \( l_j \)'s, which are \( \leq \beta^2k^2 \), the contribution of the corresponding product of \( P_j \)'s is \( \leq (2^{k-1/2})k^2 \). If \( l_j \in [\beta^2k^2, \lambda_0^2n^2] \) (there are at least \( 3n/4 - k > n/2 \) such \( j \)'s), we write \( P_j = P_j'P_j'' \), where

\[
P_j' = \prod_{l_i < \beta k/c_0} \frac{l_j - l_i}{\phi(l_j) - \phi(l_i)}.
\]

Note that each factor in the above product is \( < 2 \) and that there are at most \( k \) of them, hence \( P_j' < 2^k \); we use this estimate for \( j \leq \beta k/c_0 \). On the other hand, if \( j > \beta k/c_0 \), then (as \( l \notin \tilde{E}_j \)) \( l_j > c_0^2j^2 \) and so, for \( x < \beta^3k^2/2 \),

\[
(l_j - x)/(\phi(l_j) - \phi(x)) \leq l_j/\phi(l_j) < l_j/(l_j - \beta^2k^2/2) < (1 - [\beta k/2c_0 j]^2)^{-1}.
\]

Consequently, for such \( j \),

\[
P_j' < (1 - [\beta k/(2c_0 j)]^2)^{-k} < \exp(\beta^2k^3/2c_0^2j^2).
\]

where we used the fact that \( \beta k/(2c_0 j) < 1 \) and the inequality \( (1 - s)^{-1} < \exp(2s) \), valid for \( s \in (0, 1) \). Concerning \( P_j'' \)'s, one easily sees that

\[
\prod P_j'' < m^{-\omega/18} < (1 - c'\beta^2k^2/n^2)^{\omega/18} < \exp(-c''\beta^2k^3).
\]

Combining all the inequalities, we get
\[ P(F'_l) \leq (2^{-l^2}\beta)^k \cdot (2^k)^{\beta k/\epsilon_0} \cdot \prod_{j>\beta k/\epsilon_0} \exp(\beta^2 k^3/2c_0^2 j^2) \cdot \exp(-c''\beta^2 k^2). \]

From this the required estimate follows by direct computation if \( \beta \) is large enough.

**The Lower Estimates:** We note that, in the notation of (3.4), we have

\[ P(E) \geq P(\Phi(E)) \inf_{l \in E} \frac{\rho(l)}{\rho(\Phi(l))} |J_0(l)|^{-1}. \quad (3.7) \]

To prove the lower estimate in (3.1) for some \( k \leq n \) we observe first that if we define \( \gamma \in (0, \infty) \) by the equality

\[ P(l < \gamma^2 k^2 < \alpha k^2) = \frac{1}{2}, \quad (3.8) \]

then it follows from the upper estimates that \( \alpha_0 \leq \gamma \leq \beta_0 \), where \( \alpha_0, \beta_0 \) are universal constants (even though \( \gamma \) depends on \( k \) and \( n \)).

Now fix \( \alpha > 0 \) and set \( E = \{ l \in \mathcal{E} : l < \alpha^2 k^2 \text{ and } l_{k+1} > \gamma^2 k^2 \} \). Let \( \varphi \) be a piecewise affine function with \( \varphi(0) = 0, \varphi(\alpha^2 k^2) = \gamma^2 k^2 = \varphi(\gamma^2 k^2); \varphi(x) = x \) for \( x \approx \gamma^2 k^2 \), and define \( \Phi \) as before (\( \Phi \) is one-to-one on \( E \)). It is then easily checked that the infimum from the right-hand side of (3.7) equals \( \Phi(E) = \text{exactly the set of measure } \frac{1}{2} \) described by (3.8). This shows the lower estimate from (3.1). (3.2) is treated very similarly: we use (3.4) with \( P(E) \) introduced into the right-hand side, \( E \) defined by (3.8) (with modifications analogous to (but simpler than in) (3.6), and \( \Phi(x) = (\beta/\gamma)^2 \) for \( x \in [0, \gamma^2 k^2], \varphi(x) = x + b \) for \( x \approx \gamma^2 k^2 \), to estimate from below

\[ P(l_k < \beta^2 k^2 < l_{k+1}). \]

**Remark 3.2.** (a) Note that above we do not use the hypothesis \( k \leq n/2 \). If \( k > n/2 \), however, the lower estimate from (3.2) may not be precise. The correct magnitude is given by Lemma 3.1.

(b) When proving the lower estimate from (3.1), we could as well add other conditions on \( l \) besides the one forced by the inequalities from (3.8). For example, since

\[ P(l_k < \gamma^2 k^2 < l_{k+1} \text{ and } l_{(n/2)} > \alpha_1 n^2) > \frac{1}{4} \]

for some absolute constant \( \alpha_1 \), it follows that

\[ P(l_k < \alpha^2 k^2 \text{ and } l_{(n/2)} > \alpha_1 n^2) \geq (c_1 \alpha)^2. \]

Similar comment applies to (3.2).
(c) The complex version of Theorem 1.2 is actually simpler. Again (see Carmeli, 1983) we have an explicit formula for density $\rho_c(l)$, which is, roughly, obtained from (3.3) by replacing $|l_j - l_i|$ with $|l_j - l_i|^2$ and removing the last factor. One argues then as in the real case, with some of the technical details just disappearing.

(d) Our proof of Theorem 1.2 is somewhat "heavy handed." It seems quite likely that with more natural choices of function $\varphi$ (or $\Phi$), one could streamline the argument, obtaining perhaps better constants. Also note that the statement (and the argument) has an "isoperimetric" favor (cf. Milman and Schechtman, 1986, Appendix I) and that there are some connections to classical orthogonal polynomials (cf. Szegö, 6.22.8, 6.72).

4. PROOFS OF THE THEOREMS

In this section we derive from Theorem 1.2 the remaining results stated in the Introduction. Before passing to the arguments, we make a few general comments.

For any Gaussian vector $g$ and any $p, q \in (0, \infty)$, the $L_p$- and $L_q$-norm of $g$ can differ at most by a numerical factor depending only on $p, q$; this is just the Kahane-Khinchine inequality (Kahane, 1985, or Lindenstrauss and Tzafriri, 1979, I.e.13). For example, if $\mathbb{E} \|g\| = M$, then (see Tomaszewski, 1982)

$$M \leq (\mathbb{E} \|g\|^2)^{1/2} \leq 3^{1/2}M.$$ 

It now follows that, for some universal constants $\alpha, \beta, \gamma > 0$,

$$\mathbb{P}(\alpha M \leq \|g\| \leq \beta M) > \gamma. \quad (4.1)$$

In particular, for any $t \in [0, 1]$,

$$\mathbb{P}(\|g\| < tM) \leq Ct \quad (4.2)$$

by Lemma 2.6 and consequently

$$\exp(\mathbb{E} \log \|g\|) \sim M \quad (4.3)$$

(here and in what follows, $f \sim h$ means "there exist universal constants $C, c > 0$ such that $ch \leq f \leq Ch$"). (4.3) is an extension of the Kahane-Khinchine inequality analogous to (Ulrich, 1988); most likely this was known earlier, but we could not find a suitable reference.

Of course all that we have said above applies equally well to the (say
n × n) Gaussian matrix G and to any seminorm on the space of such matrices. Less evidently, similar phenomena occur when dealing with \( G^{-1} \); this is because we are able to relate \( G^{-1} \) to an object involving "usual" Gaussian matrices. This will become clear in the sequel; let us just point out that since \( \mathbb{E} \|G^{-1}\| = \infty \) (resp. \( \mathbb{E} \|G^{-1}\|^2 = \infty \) in the complex case), \( \exp(\mathbb{E} \log \|G^{-1}\|) \) is comparable with \( (\mathbb{E} \|G^{-1}\|^p/p)^{1/p} \) only for \( p < 1 \) (resp. \( p < 2 \)), with the constant involved depending on \( p \) as \( p \to 1 \) (resp. \( p \to 2 \)).

**Proof of Theorem 1.1.** As was observed in Section 1, in the case \( p = 2 \) the conclusion follows immediately from Theorem 1.2. Consider now the case of an arbitrary \( p \in [1, \infty] \). Again, to estimate \( \mathbb{E} \|G\| \) (or \( \mathbb{E} \log \|G\| \)) one just needs to apply Chevet's inequality (Lemma 2.1). It is clear that, in the notation of Lemma 2.1,

\[
\|I\|_{2-v} = \begin{cases} 
    n^{1/v-1/2} & \text{if } v \leq 2 \\
    1 & \text{if } v \geq 2 
\end{cases} \quad (4.4)
\]

while

\[
\mathbb{E} \|G\|_v \sim \gamma(n, v), \quad (4.5)
\]

where \( \gamma(n, v) = s^{1/2}n^{1/v} \) with \( s = \min \{v, 1 + \log n\} \); in fact one has identical estimates for \( (\mathbb{E} \|G\|^p/p)^{1/p} \), \( p \in [0, r] \), with constants independent of \( r, \rho \).

From (4.4) and (4.5) we immediately conclude that

\[
\mathbb{E} \|G\|_{p\to p} \sim n^{1/p^*}, \quad (4.6)
\]

where \( p^* = \min \{p, p/(p - 1)\} \) (i.e., \( 1/p^* = 1/2 + |1/2 - 1/p| \)), and

\[
\exp[\mathbb{E} (\log \|G\|_{p\to p})] \sim n^{1/p^*}. \quad (4.7)
\]

The last estimate can also be deduced by letting \( p \to 0 \) in Lemma 2.1 (cf. the observation following (4.5)).

We now analyze the distribution of \( \|G^{-1}\|_{p\to p} \). We claim that

\[
\exp[\mathbb{E} (\log \|G^{-1}\|_{p\to p})] \sim n^{1/2} \cdot (\min(q^*, 1 + \log n))^{1/2}, \quad (4.8)
\]

where \( q^* = \max\{p, p/(p - 1)\} \); denote also \( q = p/(p - 1) \). Clearly, Theorem 1.1 follows by combining (4.7) and (4.8).

To prove (4.8), we observe first that, by duality, it is enough to consider the case \( p \in [1, 2] \); then \( q^* = q \). Then we again use the observation from the proof of Lemma 2.4, namely that \( G^{-1} \) has the same distribution as...
STANISLAW J. SZAREK

$U'^{-1}U''$, where $U'$, $U''$, $\Delta$ have the same meaning as in that proof. Denote further $D = \Delta^{-1}/n$; $D = (\sigma_j \delta_{ij})$ is a random diagonal matrix with $\sigma_j = [ns_j(G)]^{-1}$. For a fixed $(\sigma_j)$ we are in a position to apply Lemma 2.4(i) to obtain

$$\mathbb{E}'\|G^{-1}\|_{p \to p} \sim \mathbb{E}'\|G'DG''\|_{p \to p},$$

where $G'$, $G''$ are independent (also of $D$) copies of $G$ and $\mathbb{E}'$ is a shorthand for the conditional expectation $\mathbb{E}(\cdot | (\sigma_j))$. Now, for a fixed $(\sigma_j)$ (hence $D$), we repeatedly apply Lemma 2.1 and get

$$\mathbb{E}'\|G'DG''\|_{p \to p} \sim \mathbb{E}'\|G'D\|_{2 \to p} \cdot \mathbb{E}\|g\|_q + \mathbb{E}'\|G'Dg\|_p \cdot \|g\|_{2 \to q}
\sim \|\sigma\|_2 \cdot \|f\|_{2 \to q} \cdot \mathbb{E}\|g\|_q + \|\sigma\|_\infty \cdot \mathbb{E}\|g\|_p \cdot \|g\|_{2 \to q} + \mathbb{E}\|g\|_p \cdot \|f\|_{2 \to q}.$$

(4.9)

Note that, for the first application of Lemma 2.1, $\|u\|_{l_1} = \|G'Du\|_p$, etc. Substituting the values from (4.4), (4.5), one gets

$$\mathbb{E}'\|G^{-1}\|_{p \to p} \sim \|\sigma\|_\infty \cdot n \left(\min\{q, 1 + \log n\}\right)^{1/q} + \|\sigma\|_2 n^{1/p}. \quad (4.10)$$

By Theorem 1.2, both $\|\sigma\|_2$ and $\|\sigma\|\infty$ are typically of order $n^{-1/2}$; in particular we may conclude (say, from Theorem 1.3, whose proof uses only the case $p = 2$ of Theorem 1.1) that

$$-\frac{1}{2} \log(Cn) \leq \mathbb{E} \left(\log \|\sigma\|_\infty\right) \leq \mathbb{E} \left(\log \|\sigma\|_2\right) \leq -\frac{1}{2} \log(cn). \quad (4.11)$$

The upper estimate from (4.8) then follows immediately just replace $\|\sigma\|\infty$ by $\|\sigma\|_2$ (in (4.10); then the first term is dominating).

The lower estimate is obtained similarly using (4.9), (4.10), (4.11) and the fact that

$$\exp(\mathbb{E}' \log \|G'DG''\|_{p \to p}) \sim \exp(\mathbb{E}' \log \|U'^{-1}U''\|_{p \to p}),$$

which in turn is a consequence of Lemma 2.4(ii).

Remark 4.1  (i) As was mentioned in Section 1, our proof allows one to find the exact order of

$$\mathbb{E}(X, Y) \equiv \mathbb{E} \left[\log(\|G\|_{X \to Y} \|G^{-1}\|_{Y \to X})\right],$$

where $\|\cdot\|_X$, $\|\cdot\|_Y$ are any two norms on $\mathbb{R}^n$. Indeed, the arguments leading to (4.6), (4.7), and (4.9) were completely general. Since one always has
CONDITION NUMBERS OF RANDOM MATRICES

$\|I : l_2^2 \to Z\| \leq (\pi/2)^{1/2} \mathbb{E} \|g\|_Z$, it follows that

$$c \|\sigma\|_\infty \mathbb{E} \|g\|_\gamma \mathbb{E} \|g\|_\lambda \leq C \|G^{-1}\|_{X \to Y} \leq C \|\sigma\|_2 \mathbb{E} \|g\|_\gamma \mathbb{E} \|g\|_\lambda.$$ 

As $\|\sigma\|_\infty$ and $\|\sigma\|_2$ have “essentially” the same distribution according to Theorem 1.2 (and 1.3), one gets that

$$\mathcal{C}(X, Y) \sim n^{-1/2}(\|I\|_{l_2^2 \to X} \mathbb{E} \|g\|_\gamma + \|I\|_{l_2^2 \to Y} \mathbb{E} \|g\|_\lambda) \mathbb{E} \|g\|_\gamma + \mathbb{E} \|g\|_\lambda$$

(the lower estimate being handled as in the special setting of Theorem 1.1). The are, as usual, variants for $\rho$th moments with $\rho < 1$ (resp. $\rho < 2$ in the complex case). Let us note that, for any $X, Y, \mathcal{C}(X, Y)$ is at least of order $n$, and that $\mathcal{C}(X, X) \sim n$ implies that, for some $\alpha > 0$, $\|\cdot\|_\gamma \sim \alpha \|\cdot\|_2$. However, we have, e.g., $\mathcal{C}(l_p^2, l_q^2) \sim n$ when $1 < s \leq 2 \leq r < \infty$ (with uniformly bounded constants as long as $r, s$ remain bounded away from 1, $\infty$).

(ii) One may consider $\exp(\mathbb{E} \log \|G^{-1}\|_{X \to Y}) \sim n^{-1/2} \mathbb{E} \|g\|_\gamma \mathbb{E} \|g\|_\lambda$ to be a version of “Chevet’s inequality” for the inverse Gaussian matrix. It says that, in principle, $G^{-1}$ behaves as $n^{-1/2} g \otimes g'$ (a “generic” rank one operator whose norm, when acting on $l_2^2$, is $n^{1/2}$).

Theorems 1.3 and 1.4 are essentially immediate consequences of Theorem 1.2. We just indicate the main points and make a few additional comments.

Proof of Theorem 1.3. The inequalities involving $\|G\|_{C_p}$ follow directly from Corollary 2.3 and Theorem 1.2. Indeed, the former one implies

$$\mathbb{E} \|G\|_{C_p} \leq n^{1/p} \mathbb{E} \|G\|_{l_2 \to 2} \leq n^{1/2 + 1/p}, \quad (4.12)$$

while the latter one, applied with $j = \lceil n/2 \rceil + 1$, shows that, for $\alpha > 0$,

$$\mathbb{P}(\|G\|_{C_p} < \alpha/4 \cdot n^{1/2 + 1/p}) \leq (C\alpha)^{2/3}. \quad (4.13)$$

Consequently $\mathbb{E} \|G\|_{C_p} \sim n^{1/2 + 1/p}$ as required. The analogous estimate on $\exp(\mathbb{E} \|G\|_{C_p}^{2})$ follows from this and the comments from the beginning of this section, or directly from (4.12) and (4.13). We emphasize that Theorem 1.2, Lemma 2.2, and Corollary 2.3 yield much stronger facts about the distribution of $\|G\|_{C_p}$ than those contained in the assertion of Theorem 1.3.

Concerning $\mathbb{E} \|G^{-1}\|_{C_p}$, we have to use the full force of Theorem 1.2. The expression $b = b(p, n) = n^{1/2} (\min(q, 1 + \log n))^{1/q}$ appears because

$$b(p, n) \sim \left(\sum_{j=1}^{n} (n^{1/2}/j)^p\right)^{1/p},$$
the values $n^{1/2}/j$ being "approximately" the "most likely" ones for $[s_{j}(G)]^{-1} = s_{n-j+1}(G^{-1})$. More precisely, one has, by Theorem 1.2,
\[
P(\|G^{-1}\|_{C^p} < \alpha b(p, n)) \leq C_1 \alpha \quad \text{for } \alpha \geq 0
\]
\[
P(\|G^{-1}\|_{C^p} > \beta b(p, n)) \leq \exp(\epsilon \beta^2) \quad \text{for } \beta \geq \beta_1,
\]
where $C_1, \alpha, \beta_1$ are universal positive constants. From these, the required estimate on $\exp(\mathbb{E} \log \|G^{-1}\|_{C^p})$ follows immediately. In fact, we again have similar estimates on $(\mathbb{E}\|G^{-1}\|_{C^p})^{1/p}$ if $\rho < 1$ (resp. $\rho < 2$ in the complex case). We also point out that some additional information—implying, in particular, that $\mathbb{E} \|G\|_{C^p} \|G^{-1}\|_{C^{p'}} = \infty$ (in the real case) for any $\rho, \rho'$ is contained in Remark 3.2(b).

**Proof of Theorem 1.4.** Let $K \geq n$. We must be slightly careful since even though, for $t \geq n$, $P(\|G^{-1}\|_{L^2} > t \sqrt{n}) \sim n/t$ and $P(\|G\|_{L^2} > \sqrt{n})$ is "nearly" $1$, these two facts do not yield formally that $P(\|G\|_{L^2} \|G^{-1}\|_{L^2} > t) \sim n/t$. However, this is implied by Remark 3.2(b); the assertion then follows immediately.

If $K \leq n$, we again use Remark 3.2(b) to show that $P(E_K) \leq \exp(-C(n/K)^2)$. On the other hand, $P(E_K) \leq \exp(-c(n/K)^2)$ just by Theorem 1.2. This shows that $P(E_K)$ decreases fast as $K$ decreases, whence the required assertion follows.

**ACKNOWLEDGMENTS**

The author thanks S. Heinrich for bringing the subject to his attention, for sending a preprint (Heinrich, 1990) and for keeping him au courant with respect to the developments in the field, and H. Wozniakowski for several helpful suggestions.

**REFERENCES**


CONDITION NUMBERS OF RANDOM MATRICES


