UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

On the Singular Values of Gaussian Random Matrices

Jianhong Shen

August 2000

CAM Report 00-30

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

On the Singular Values of Gaussian Random Matrices

Jianhong Shen
Department of Mathematics
University of California
Los Angeles, CA 90095
jhshen@math.ucla.edu
AND
School of Mathematics
University of Minnesota
Minneapolis, MN 55455
jhshen@math.umn.edu

In the memory of my dear mentor and friend — Gian-Carlo Rota.

Abstract

This short note is about the singular value distribution of Gaussian random matrices (i.e. Gaussian Ensemble or GE) of size N. We present a new approach for deriving the p.d.f. of the singular values directly from the SVD form (singular value decomposition), which also takes advantage of the rotational invariance of GE and the Lie algebra of the orthogonal group. Our method is more direct and general than the conventional approach that relies on the Wishart Ensemble and the combination of QR and Cholesky decomposition. Directly based on this p.d.f., and its interpretation by statistical mechanics, we give the physics proof that in the thermodynamic limit ($N \to \infty$), the singular value distribution satisfies the *quadrant law*, similar to the celebrated *semi-circle law* established by Wigner more than forty years ago for the spectral distribution of Gaussian Orthogonal (or Unitary) Ensembles. This quadrant law was also proved earlier and mathematically more rigorously by some authors based on the probabilistic estimations and the moment method, but not directly from the p.d.f. formula.

AMS classification 2000: 15A52, 65F35, 31A99, 82B05.

Keywords: Random matrices, singular values, Gaussian Ensemble, Wishart Ensemble, thermodynamic limit, pseudo-Coulomb gas, circle law, quadrant law.

To appear in *Linear Algebra and Its Applications*, 2000.

1 Introduction

Scientists in several different fields all study the subject of random matrices. Therefore, it is worthwhile to point out to our readers in the very beginning that the standpoint of this note is *numerical linear algebra.*

Random matrix theory is currently an attractive area because of its rich content of physics, statistics, and mathematics. Its motivations and applications can be found in several important areas: condensed matter physics, statistical mechanics and chaotic systems [14, 26, 27], multivariate statistics [8, 10, 13, 15, 19, 28], the Riemann hypothesis [16, 17], 2-D potential theory and orthogonal polynomials [1, 2, 18], and numerical linear algebra [3, 4, 5, 6, 23]). From the physics point of view, the dominance of spectral analysis for random matrices is mostly due to the significant physics meaning of eigenvalues, i.e., the correspondence between eigenvalues and nuclear energy levels, and between eigenvalues and Coulomb particles [14, 18].

To numerical analysts, on the other hand, eigenvalue study is often restricted to the symmetric or Hermitian systems of linear algebraic equations (which can be further traced back to symmetric or Hermitian differential systems in the continuous world such as the Sturm-Liouville problems, the Laplacian operator, and Schrödinger equations. See Strang [21] and Golub and Ortega [9], for examples). For general systems of linear equations or the currently highly active effort in digital databank analysis (like eigen-face analysis from human face databank and internet text search engines), singular values become more crucial. It is also a household advice among numerical analysts that on facing a new general linear system, the first right question to ask is "What is the condition number κ?" The condition number essentially charactizes the relative dynamic range of the singular value spectrum since (in the Euclidean world)

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = \frac{\sigma_{\max}}{\sigma_{\min}},$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the two ends. It was mostly such awareness of the importance of singular values in analyzing linear systems (also see Smale [20] and Demmel [3]) that had motivated the remarkable thesis of the contemporary numerical analyst Alan Edelman [5]. To the best knowledge of the author, about ten years later, this thesis still remains the only work in the theory of Gaussian random matrices that has been solely and deeply devoted to the understanding of numerical linear algebra. In the same spirit, the current paper tries to improve or complement some aspects of Edelman [5] in the study of singular values of random matrices.

Perhaps spoiled by the spectral analysis in the random matrix theory, most of the existing works (including Edelman's thesis) transformed singular value analysis to eigenvalue analysis through the Wishart Ensemble $W(N)$ (and more generally, $W(N, M)$ from Gaussian Ensemble $G(N, M)$, see Edelman [5] and Muirhead [15]), namely the ensemble of N by N random positive matrices $M = AA^T$, with $A \in G(N) = G(N, N)$. The major advantage of such a approach is that one can immediately benefit from many works on Wishart Ensembles (see [8, 10, 13, 12, 15, 19], for examples) in the literature of multivariate statistics. The pities are, if singular values (σ_k's) could indeed speak for themselves, the merciless defiance of their independent "civil rights" in the kingdom of linear algebra and linear transforms. In the singular value decomposition (SVD),

$$A = USV^T = \sum_{k=1}^{N} \sigma_k \ \mathbf{u}_k \cdot \mathbf{v}_k^T,$$

all the three ingredients — the left singular vectors $\mathbf{u}_k$, their conjugate vectors $\mathbf{v}_k$, and the singular values σ_k — have their own intrinsic meaning in the geometric picture of linear transforms in Euclidean spaces (see for example, Strang [22] and Trefethen and Bau [23]). The popular transition from a singular value problem to an eigenvalue problem (as in most numerical algorithms) only explains the deficiency of human beings, not the singular values or vectors.

To taste the pity, let us first check out two examples, through which we intend to argue that the singular value variable σ of A is more natural and pleasing to work with than its square $\lambda = \sigma^2$, or the eigenvalue of AA^T. The first example is the eigenvalue density for Wishart Ensemble [5, 15]:

$$P(d\lambda) = \frac{1}{Z_N} \exp(-\frac{1}{2} \sum \lambda_k) \ \prod_k \lambda_k^{-\frac{1}{2}} \ \prod_{i<j} |\lambda_j - \lambda_i| d\lambda, \tag{1}$$

where Z_N is a normalization constant or the *partition function* in the context of statistical mechanics. *Throughout the paper, we shall not elaborate on the exact forms of the Z's, since they can be found in the standard literature [5, 14, 15, for examples].* The first exponential term and the third term of differences are familiar objects in the well studied Gaussian Orthogonal Ensembles (GOE). Is there any significant statistical meaning of the term with a $-1/2$ power? The answer is no. Its existence is purely caused by the squares:

$$d\sigma = \prod_k d\sigma_k = 2^{-N} \prod_k \lambda_k^{-\frac{1}{2}} d\lambda_k = 2^{-N} \prod_k \lambda_k^{-\frac{1}{2}} d\lambda.$$

The other example is one of the major contributions of Edelman's thesis [5, Theorem 5.1], which states that the p.d.f of $N\lambda_{\min}$ of Wishart Ensemble $W(N)$ converges to

$$f(x) = \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})}.$$

This "ugly" formula, though representing a critical asymptotic result for numerical linear algebra, looks much simpler and more pleasing under the singular value variable:

$$\text{Prob}(\hat{\sigma}_{\min} \geq y) = e^{-(\frac{y^2}{2}+y)}, \tag{2}$$

where $\hat{\sigma}_{\min}$ is the (weak) limit of the normalized smallest singular value $\sqrt{N}\sigma_{\min}$. Though simplicity and convenience are often quite psychological, these two examples do at least show that looking at singular values directly is not a bad idea.

Therefore, in this paper, our first main result is to derive the distribution density Eq. (1) without turning to eigenvalues or Wishart Ensembles (Theorem 1 in Section 2). This approach starts right from the SVD form, and is made simpler by utilizing the geometric and algebraic properties of GE and the orthogonal groups. In our opinion, it is more intrinsic and direct than the approach in Edelman's thesis [5, Chapter 3] and those in multivariate statistics [15]. The latter were based on the combination of LQ factorization and Cholesky factorization $L^T L$, and Wishart Ensembles, which we shall agree from our proof are extra and unnecessary structures for studying singular values. [1]

Based on Theorem 1, in Section 3, we establish via the statistical mechanics approach the second result about the thermodynamic limit (as $N \to \infty$) of the singular value distributions (Theorem 2, Section 3). This is very similar to Wigner's celebrated semi-circle law for eigenvalues of GOE, and we call it *the quadrant law* since singular values are nonnegative. Simple and heuristic applications are outlined at the end of the section. We shall also mention briefly in the beginning of the section some earlier and more precise mathematical proofs by other authors. These were proofs mostly based on the probabilistic estimations of certain numerical transforms of the random matrices, and/or the moment method. But to the best knowledge of the author, none of them started directly from the p.d.f. information of the singular values.

2 The Distribution of Singular Values of $G(N)$

We shall mainly consider N by N real square random matrices. All the argument can be modified easily for more general non-square and/or complex ensembles, and such modifications will be briefly mentioned.

2.1 The Gaussian Ensemble $G(N)$

Let $\mathsf{gl}(N)$ denote the general linear algebra of all N by N real matrices. An element or an individual matrix in $\mathsf{gl}(N)$ is denoted by $M = (a_{ij})$. Equip $\mathsf{gl}(N)$ with the Euclidean structure by defining the Frobenius inner product:

$$\langle M,\ L\rangle = \mathrm{trace}(M\ L^{\mathrm{T}}), \qquad M, L \in \mathsf{gl}(N).$$

Let $d_v M$ denote the infinitesimal volume element of the Lebesgue measure of $(\mathsf{gl}(N), \langle\cdot,\cdot\rangle)$. We thereby reserve the notation dM for the ordinary differential 1-form. Then the Gaussian Ensemble (GE) $G(N)$ is a randomization of $\mathsf{gl}(N)$ under the probability measure

$$P(d_v M) = \frac{1}{Z_N} e^{-\frac{\beta}{2}\mathrm{trace}(MM^{\mathrm{T}})} d_v M, \tag{3}$$

[1] Upon submission, it came to the author's attention that the same idea and philosophy also appeared in Edelman's lecture note in Berkely [7], where one can learn the much broader context of the subject from the numerical linear algebra point of view.

The constant β denotes the inverse variance in statistics, and the inverse temperature $1/kT$ in statistical physics (k is the Boltzmann constant).

It can be shown by the similar argument for GOE as in Mehta's classical book [14] that Gaussian Ensemble is the unique measure $\mu(d_v M)$ on $\mathfrak{gl}(N)$ that meets the following two requirements:

(i) (**Two-side Rotational Invariance**) For any N by N orthogonal matrix Q,

$$\mu(d_v(Q\ M)) = \mu(d_v(M\ Q)) = \mu(d_v M).$$

(ii) (**Independence of Entries**) For any $1 \leq i, j \leq N$, define the (ij)-entry random variable X_{ij} on $(\mathfrak{gl}(N),\ \mu)$ by $X_{ij}(M) = a_{ij}$. Then the N^2 random variables $(X_{ij}\ |1 \leq i, j \leq N)$ are independent.

(Note: This assertion is not true for GOE. As Mehta [14] showed, requirements (i) and (ii) still allow another degree of freedom for the measure μ — the mass center, though it must be a scalar matrix.)

From the independence condition, we immediately see from (3) that X_{ij} must be a normal random variable $N(0, 1/\beta)$. In fact, in numerical linear algebra, the GE is always generated element-wise, instead of by the above axiomatic approach.

2.2 Distribution of Singular Values of $G(N)$

As promised in the beginning, the main goal of this section is to derive the p.d.f of the singular values directly from SVD.

Theorem 1 (p.d.f of singular values for GE) *The probability density function for the singular values of $G(N)$ is*

$$P_N(d\sigma) = \frac{1}{Z'_N} e^{-\frac{\beta}{2}\sum_k \sigma_k^2} \prod_{j>i} |(\sigma_j^2 - \sigma_i^2)| d\sigma, \tag{4}$$

where $\sigma = (\sigma_1, \cdots, \sigma_N)$ is the singular value vector (unnecessarily in the conventional descending order). (The partition function Z'_N normalizes the integral on $[0,\ \infty)^N$.)

Proof. Any given N by N matrix M_0 allows a singular value decomposition:

$$M_0 = U_0 S_0 V_0^{\mathrm{T}}, \tag{5}$$

where U_0 and V_0 are orthogonal matrices and $S_0 = \mathrm{diag}(\sigma_1^{(0)}, \cdots, \sigma_N^{(0)})$. The singular values are almost surely distinct in the Lebesgue measure. Therefore, a neighborhood of M_0 in $\mathfrak{gl}(N)$ allows the *unique* SVD coordinate system:

$$M = USV^{\mathrm{T}}, \qquad \text{for any } M \text{ in the neighborhood}, \tag{6}$$

such that we have N explicit singular value coordinates from $S = \mathrm{diag}(\sigma_1, \cdots, \sigma_N)$, and a pair of $\binom{N}{2}$ implicit coordinates from U and V on the orthogonal (Lie) group

$O(N)$. Here we certainly assume that $\|U - U_0\|, \|V - V_0\|$, and $\|S - S_0\|$ are all small.

Thanks to the two-side rotational invariance of Gaussian Ensembles, we can assume that $U_0 = V_0 = I_N$. Then U and V are both near the unit element I_N in the Lie group $O(N)$. Recall that the Lie algebra $\mathsf{A}(N)$ of $O(N)$ is the linear subspace of $\mathsf{gl}(N)$ consisting of all anti-symmetric matrices, and the exponential mapping

$$A \to Q = \exp(-A)$$

provides an isomorphism between the neighbors of 0 in $\mathsf{A}(N)$ and I_N in $O(N)$. Hence, U and V allow the explicit coordinates A and B in $\mathsf{A}(N)$ such that

$$U = e^{-A} \quad \text{and} \quad V = e^{-B},$$

and at $A = B = 0$, we have $dU = -dA$ and $dV = -dB$.

A total differentiation of (6) gives

$$dM = dU\ SV^{\mathrm{T}} + US\ dV^{\mathrm{T}} + U\ dS\ V^{\mathrm{T}},$$

and at $U_0 = V_0 = I_N$ (i.e. $A = B = 0$) it simplifies to

$$dM = dS + dU\ S + S\ dV^{\mathrm{T}} = dS + S\ dB - dA\ S. \tag{7}$$

Define $E_{ij} = \mathbf{e}_i \otimes \mathbf{e}_j = \mathbf{e}_i \mathbf{e}_j^{\mathrm{T}}$ with $\mathbf{e}_i = (0, \cdots, 0, \overset{i}{1}, 0, \cdots, 0)^{\mathrm{T}}$. Then (7) becomes

$$dM = \sum_k d\sigma_k E_{kk} + \sum_{i \neq j} (-\sigma_j da_{ij} + \sigma_i db_{ij}) E_{ij}. \tag{8}$$

Since $(E_{ij} \mid 1 \leq i, j \leq N)$ is an orthonormal basis in $(\mathsf{gl}(N), \langle \cdot, \cdot \rangle)$, we thus obtain the formula for the volume element $d_v M$:

$$d_v M = \pm d\sigma_1 \wedge \cdots \wedge d\sigma_N \wedge \prod_{i \neq j} (-\sigma_j da_{ij} + \sigma_i db_{ij}). \tag{9}$$

These are exterior products, and the correction sign $\pm$ is because Lebesgue measure $d_v M$ does not distinguish orientations. Noticing that for any $i \neq j$,

$$\begin{aligned} &(-\sigma_j da_{ij} + \sigma_i db_{ij}) \wedge (-\sigma_i da_{ji} + \sigma_j db_{ji}) \\ =\ &(-\sigma_j da_{ij} + \sigma_i db_{ij}) \wedge (\sigma_i da_{ij} - \sigma_j db_{ij}) \\ =\ &(\sigma_j^2 - \sigma_i^2) da_{ij} \wedge db_{ij}, \end{aligned}$$

we have

$$d_v M = 2^{-\binom{N}{2}} \prod_{j > i} |(\sigma_j^2 - \sigma_i^2)|\ d\sigma\ d_v A\ d_v B,$$

where $d_v A = \sqrt{2}^{\binom{N}{2}} \prod_{j>i} da_{ij}$ is the volume element of $\mathsf{A}(N)$ (as a submanifold of $\mathsf{gl}(N)$). Since the exponential mapping is an isomorphism, we have $d_v U = d_v A$ and $d_v V = d_v B$. Thus

$$d_v M = d(USV^{\mathrm{T}}) = 2^{-\binom{N}{2}} \prod_{j>i} |(\sigma_j^2 - \sigma_i^2)| \; d\sigma \; d_v U \; d_v V. \tag{10}$$

Here $d_v U$ and $d_v V$ are understood as the intrinsic volume element of the orthogonal group.

From the SVD

$$M = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}} + \cdots + \sigma_N \mathbf{u}_N \mathbf{v}_N^T,$$

we see that almost surely (i.e. when σ_k's are distinct) each conjugate singular vector pair $(\mathbf{u}_k, \mathbf{v}_k)$ is unique up to a reflection. This introduces a multiplicity factor 2^N as we integrate. Thus from the GE distribution formula (3), we conclude: at a given $\sigma = (\sigma_1, \cdots, \sigma_N)$,

$$\begin{aligned} P(d\sigma) &= \frac{1}{Z_N} e^{-\frac{\beta}{2}\sum_k \sigma_k^2} \;\cdot\; 2^{-\binom{N}{2}} \prod_{j>i} |(\sigma_j^2 - \sigma_i^2)| \; d\sigma \; \frac{1}{2^N} \int_{O(N)\times O(N)} d_v U \; d_v V \\ &= \frac{1}{Z_N'} e^{-\frac{\beta}{2}\sum_k \sigma_k^2} \prod_{j>i} |(\sigma_j^2 - \sigma_i^2)| \; d\sigma, \end{aligned}$$

where the new partition function Z_N' is given by

$$\frac{1}{Z_N'} = \frac{2^{-\binom{N}{2}}}{2^N Z_N} \mathrm{vol}(O(N))^2 = \frac{2^{-\binom{N+1}{2}}}{Z_N} \mathrm{vol}(O(N))^2.$$

The volume formula for the orthogonal group or more general Stiefel manifolds can be found in [5, 14, 15]. This completes the proof. □

The approach here we have taken is very general. It contains more geometric and algebraic structures (based on rotational invariance and the Lie algebra of $O(N)$) compared with the conventional more analytic one based on Jacobian evaluation [5, 14, 15]. [2] Moreover, our approach allows one to "zoom" into the Jacobian and see clearly the algebraic meaning of the mysterious factors $|\lambda_j - \lambda_i|^k$ that frequently appear in the theory of random matrices. Here come our more detailed comments along this line.

Remark 1. (*Spectral distribution for GOE.*) Our approach also applies to Gaussian Orthogonal Ensembles. M is symmetric and thus S is replaced by its eigenvalue

[2]It is interesting to point out that all matrix factorizations familiar to numerical analysts have nice Jacobians [5, 7, 15], — "a stroke of luck?" in Trefethen's language [23].

matrix diag$(\lambda_1, \cdots, \lambda_N)$ and $V = U$ is the eigenvector matrix. Therefore, in Eq.(8), $a_{ij} = b_{ij}$, and

$$dM = \sum_k d\lambda_k E_{kk} + \sum_{j>i} (\lambda_i - \lambda_j)\ da_{ij}\ (E_{ij} + E_{ji}).$$

Since $(E_{kk}, E_{ij} + E_{ji} \mid k; j > i)$ is an orthogonal basis in GOE, we have

$$d_v M = d\lambda \prod_{j>i} |(\lambda_j - \lambda_i)|\ d_v A.$$

This easily gives the spectral distribution density of GOE.

Furthermore, each difference factor $\lambda_j - \lambda_i$ also sees its clear algebraic meaning from our approach. By Eq.(7) and (8), if we define

$$\Lambda_{ij} = E_{ij} - E_{ji} \in \mathsf{A}(N)$$

for each pair $i < j$, then the "finite difference" $\lambda_j - \lambda_i$ of the spectra exactly comes from the Lie bracket (which is an algebraic generalization of taking differentiation!):

$$[\Lambda_{ij}, S] = \Lambda_{ij}\ S - S\ \Lambda_{ij} = (\sigma_j - \sigma_i)(E_{ij} + E_{ji}). \tag{11}$$

Remark 2. (*Real and complex ensembles: an algebraic approach.*) Our approach also easily offers a unified viewpoint on real ensembles and complex ensembles (and even quaternion ensembles). It is well known that the only difference in the spectral density functions between GOE and GUE is the factor of 2: in GUE, each spectral difference is squared —$(\lambda_j - \lambda_i)^2$. The above approach offers a very general algebraic explanation. For GUE, a complete orthogonal basis for the anti-Hermitian algebra consists of the elements Λ_{ij} as defined above, as well as

$$\Lambda'_{ij} = \sqrt{-1}(E_{ij} + E_{ji}), \qquad \text{for each pair} \quad i < j.$$

Thus we have another bracket besides (11):

$$[\Lambda'_{ij}, S] = (\sigma_j - \sigma_i)\sqrt{-1}\Lambda_{ij}. \tag{12}$$

Since $(E_{ij} + E_{ij})$ and $\sqrt{-1}\Lambda_{ij}$ are orthogonal directions in GUE, the measure product along the directions in Eq. (11) and (12) gives the squared factor. The same discussion applies to quaternion ensembles.

Remark 3. (*Non-square Gaussian Ensemble G(N, n).*)

Our approach applies easily to any non-square Gaussian ensembles $G(N, n)$ with, say, $N > n$. Then in Eq. (5), U_0, S_0, and V_0 have sizes $N \times n$, $n \times n$, and $n \times n$. Complete U_0 to a square orthogonal matrix $\hat{U}_0$. Then the rotational invariance allows us (by considering $\hat{U}_0^{\mathrm{T}} \cdot M_0 \cdot V_0$) to consider only the much simpler case of

$U_0 = (I_n, 0)^{\mathrm{T}}$ and $V_0 = I_n$. Near such U_0 and V_0, we have explicit coordinates: $A \in \mathsf{A}(n)$ and W of $N-n$ by n for U, and $B \in \mathsf{A}(n)$ for V (of course valid only in a small neighborhood), such that $U = (e^{-A}, W)^{\mathrm{T}}$ and $V = e^{-B}$. Now Eq.(7) becomes

$$dM = \begin{bmatrix} dS + S\ dB - dA\ S \\ dW\ S \end{bmatrix}.$$

(Note: W is a "free" variable (to the first order) near $W_0 = 0_{N-n,n}$.) The only new interaction comes from S and dW, which, after taking exterior products, contributes a factor of $(\sigma_1\sigma_2\cdots\sigma_n)^{N-n}$ to the volume element d_vW, since each σ_k is multiplied to a column of dW of length $N-n$. The remaining analysis is the same as in the proof. Readers can find that our approach clearly explains the meaning of the different factors in the p.d.f, first obtained by Fisher, Hsu and Roy in three different papers in 1939 (see [5, 15]).

3 The Quadrant Law of the Thermodynamic Limit

In this section, we give the *physics proof* of the quadrant law for the thermodynamic limit of the singular value distributions. We shall follow the standard statistical mechanics approach for GOE and GUE [14]. We called it "the physics proof" upon the consideration of two factors: (1) first it is a proof since the correspondence between the distribution of the singular values and Coulomb or quasi-Coulomb many-body interactions is exact, and there is no approximation; (2) but it is the "physics" proof, since the statistical mechanics approach only predicts the equilibrium state of the thermodynamic limit of the many-body problem, and it does not offer the accurate information regarding how such state is achieved, or mathematically speaking, in the sense of weak convergence, in measure, or almost surely? The very recent mathematical paper of Kiessling and Spohn [11] attempted to address this problem in a more accurate way.

There might exist earlier works, but from what the author has learned, the first paper that gave the rigorous mathematical proof of the quadrant law appeared in Wegmann [25, 1976], in which the author studied more general random matrices from the non-commutative algebra $\mathbb{C}[A, A^*]$, where A is a random complex matrix. The major tool is the moment method. Another interesting paper that explicitly mentioned the "quarter-circle law" was by Trotter [24, 1984], a concise summary of which can also be found in Edelman's lecture note [7]. The main techniques of [24] include numerical transforms, the direct probabilistic estimations, and the Jacobi tri-diagonal matrix associated with the three-term relation of the Hermitian orthogonal polynomials. None of these works utilized the information of the p.d.f. of the singular values in Theorem 1, however. The author would be glad to receive any other information concerning the literature of the quadrant law.

3.1 γ-Coulomb gases

We have observed the major difference between the spectral distribution of Gaussian Orthogonal Ensembles and the singular value distribution of Gaussian Ensembles.

From the statistical mechanics point of view, this is the difference between Coulomb and non-Coulomb gases. But the non-Coulomb gas corresponding to the singular values are not too far away from the Coulomb gas as we shall explain below.

For any $\gamma \geq 1$, define a "*γ-Coulomb gas*" by the Hamiltonian

$$H_\gamma(\sigma) = H_\gamma(\sigma_1, \cdots, \sigma_N) = \sum_k V(\sigma_k) - \sum_{j>i} \ln |\sigma_j^\gamma - \sigma_i^\gamma|, \tag{13}$$

where $V(x)$ is an external potential field acting on singletons. The admissible domain (for the gas particles) is $\sigma_k \geq 0, k = 1, \cdots, N$. (This seemingly artificial constraint can be better explained by the energy barrier at the origin for all even integers.)

If $\gamma = 1$, this represents the classical Coulomb gas (in an external field). The spectral analysis of GOE and GUE falls into this category. An extensive study on such a gas is also motivated from the approximation theory and can be found in the excellent monograph by Saff and Totik [18]. In the case of singular values of GE, however, we have

$$V(x) = \frac{\beta}{2} x^2, \qquad \text{and} \quad \gamma = 2,$$

and in terms of the Hamiltonian, the p.d.f becomes

$$P(d\sigma) = \frac{1}{Z} e^{-H_2(\sigma)}\, d\sigma.$$

Now we follow the standard practice to compute the thermodynamic limit. In the limit, suppose the ratio of the total number of particles on $[x,\ x+dx)$ is $\phi(x)\, dx$. Then the total energy $\langle \phi\ |\frac{1}{2} H_\gamma(x, y)|\ \phi\rangle$ (using the standard quantum mechanics notation) should be minimized. In other words, $\phi(x)$ solves the following constraint quadrature:

$$\min \langle \phi\ |\frac{1}{2} H_\gamma(x, y)|\ \phi\rangle \qquad \text{under the constraint} \quad \int_0^\infty \phi(x) dx = 1. \tag{14}$$

Notice that this is NOT the familiar Rayleigh quotient problem and thus NOT an eigenvalue problem. In fact, the stationary equation for (14) is

$$V(x) - \int_0^\infty \ln |x^\gamma - y^\gamma| \phi(y)\, dy = C,$$

where C is the Lagrange multiplier. Taking differentiation leads to the

$$(\text{P.V.}) \int_0^\infty \frac{\gamma x^{\gamma-1} \phi(y)}{x^\gamma - y^\gamma}\, dy = V'(x).$$

Set $F(x) = V'(x)$ — the negative external force, physically speaking. Then

$$(\text{P.V.}) \int_0^\infty \frac{\gamma \phi(y)}{x^\gamma - y^\gamma}\, dy = x^{1-\gamma} F(x).$$

The beautiful property about this last equation is its scaling law. Define

$$\hat{\phi}(x) = 1/\gamma\ x^{-1+1/\gamma}\phi(x^{1/\gamma}) \qquad \text{and} \qquad \hat{F}(x) = 1/\gamma\ x^{-1+1/\gamma}F(x^{1/\gamma}). \tag{15}$$

Then the last equation becomes

$$(\text{P.V.}) \int_0^\infty \frac{\hat{\phi}(y)}{x-y}\, dy = \hat{F}(x). \tag{16}$$

Notice that the scaling transform (15) preserves the total integral of ϕ. This conforms to the constraint that $\phi(x)$ is a probability density function. Eq. (16) is just like the equilibrium equation for Coulomb gases (only that here the domain is half of the real line)! As Mehta [14] pointed out, we only require the last equation to be valid on the support of the distribution $\hat{\phi}(x)$ (due to the non-negativity of a density).

3.2 The quadrant law for singular values

Let us apply the above result to the singular value distribution of Gaussian Ensembles with inverse temperature β. The equilibrium distribution depends on β, and thus is denoted by ϕ_β. We have $\gamma = 2$ and $F(x) = V'(x) = \beta x$ and

$$\hat{F}(x) = 1/2\ x^{-1+1/2}F(x^{1/2}) = \beta/2.$$

Thus we only need solve

$$(\text{P.V.}) \int_0^\infty \frac{\hat{\phi}_\beta(y)}{x-y}\, dy = \frac{\beta}{2}.$$

A further change of variable $x \to x/\beta$ leads to

$$(\text{P.V.}) \int_0^\infty \frac{\hat{\phi}_1(y)}{x-y}\, dy = \frac{1}{2}, \tag{17}$$

such that

$$\hat{\phi}_\beta(x) = \beta\hat{\phi}_1(\beta x). \tag{18}$$

This last equation clearly shows the re-scaling role of the inverse temperature β!

The Cauchy integral equation (17) now is quite standard. In approximation theory, Eq. (18) corresponds to the Laguerre weights $w(x) = e^{-x/2}$ on $[0, \infty)$ [18]. Therefore according to [18, Theorem 1.11 and Example 5.4], the support of $\hat{\phi}_1$ is $[0, 4]$, and

$$\hat{\phi}_1(x) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}, \tag{19}$$

which is a Beta distribution $B(3/2, 1/2)$ if linearly scaled to the unit interval $[0, 1]$. The combination of the two scaling transforms (15) and (17) eventually gives

$$\phi_\beta(x) = 2\beta^2 x \hat{\phi}_1(\beta^2 x^2) = \beta\phi_1(\beta x), \tag{20}$$

where

$$\phi_1(x) = 2x\hat{\phi}_1(x^2) = 2x \cdot \frac{1}{2\pi}\sqrt{\frac{4-x^2}{x^2}} = \frac{1}{\pi}\sqrt{4-x^2},$$

for $x \in (0, 2)$. Thus as in the classical spectral analysis for GOE, we have proved the quadrant law for singular value distribution of GE in the thermodynamic limit.

Theorem 2 (Quadrant law for singular values) *For Gaussian Ensembles with inverse temperature $\beta = 1$, the thermodynamic equilibrium distribution $\phi_1(\sigma)$ for the singular values obeys the quadrant law:*

$$\phi_1(\sigma) = \frac{1}{\pi}\sqrt{4-\sigma^2}, \qquad \sigma \in [0, 2]. \tag{21}$$

For general β, the distribution is $\phi_\beta(\sigma) = \beta\phi_1(\beta\sigma)$, and is thus supported on $[0, 2/\beta]$. Therefore, asymptotically for large N, the histogram of the singular values can be approximated by

$$h_N(\sigma) = \sqrt{N}\phi_\beta(\frac{\sigma}{\sqrt{N}})$$

(since we need $\int h_N(\sigma)\, d\sigma = N$).

Before ending this paper, we would like to make some profits from the quadrant law. As in the case of GOE or GUE, the thermodynamic limit can offer very important results without complicated algebra, though very often, the rigorous proofs are surprisingly tedious. Physicists usually employ the thermodynamic limit to guess or heuristically obtain many important results, whose rigorous proofs might be missing. Here we follow such practice:

(a) Given β, almost surely for any sequence of Gaussian matrices $(M_N \in G(N) \mid N = 1, 2, \cdots)$, we have

$$\lim_{N\to\infty} \frac{\|M_N\|_2}{\sqrt{N}} = 2/\beta.$$

This result is easy to guess from the quadrant law, but its rigorous proof (including non-Gaussian ensembles) is highly non-trivial and now has become a classical result in the random matrix theory owing to Geman [8].

(b) Denote the median of a collection of real numbers $x = \{x_1, x_2, \cdots, x_N\}$ by $\mathrm{med}(x)$. Then almost surely for any sequence of Gaussian matrices $(M_N \in G(N) \mid N = 1, 2, \cdots)$, we have

$$\lim_{N\to\infty} \frac{\mathrm{med}(\sigma(M_N))}{\sqrt{N}} = (2/\beta)\ \cos\theta_m,$$

where $\theta_m \in (0, \pi/2)$ is the unique solution to $2\theta - \sin 2\theta = \pi/2$, so that the vertical line $x = \cos\theta_m$ divides the quarter-disk

$$x^2 + y^2 = 1, \qquad x, y \geq 0$$

into two equal-area parts. The construction of a rigorous proof shall be another good statistical problem.

(c) Finally comes a more interesting result that gives the leading order information of one of Edelman's major contributions [5] in his thesis — the thermodynamic limit of the p.d.f of $\sigma_{\min}$, whose exact form has been re-formulated in Eq. (2) in the introduction section. Again our argument is heuristically based on the quadrant law. In the thermodynamic limit $N \to \infty$, the singular value sequence of a sample matrix from the Gaussian Ensemble can be seen as independent samples from the quadrant law. Assume $\beta = 2$ to make computation clear. For any $\epsilon \in (0, 1)$, define

$$p_\epsilon = \frac{4}{\pi} \int_0^\epsilon \sqrt{1 - x^2} dx = \frac{4}{\pi}\epsilon + O(\epsilon^3).$$

Then,

$$\text{Prob}(\frac{\sigma_{\min}}{\sqrt{N}} \geq \epsilon) = \text{Prob}(\text{all } \sigma_k \geq \epsilon) = \prod_{k=1}^{N} \text{Prob}(\sigma_k \geq \epsilon) = (1 - p_\epsilon)^N.$$

Noticing that $p_\epsilon = O(\epsilon)$, in the limit $N \to \infty$, we have:

$$\text{Prob}(\sqrt{N}\sigma_{\min} \geq \epsilon) = \text{Prob}(\frac{\sigma_{\min}}{\sqrt{N}} \geq \frac{\epsilon}{N}) = (1 - \frac{(4/\pi)\epsilon}{N})^N = e^{-(4/\pi)\epsilon}.$$

The last two equalities are understood in the sense of the leading order and the limit as $N \to \infty$. Compared to Edelman's exact formula (2), this heuristic argument does catch the right scaling for the smallest singular value $\sigma_{\min} = O(1/\sqrt{N})$, which is far from being obvious according to the joint p.d.f, and the leading order $e^{-O(\epsilon)}$ of the density function, though the heuristic coefficient $4/\pi = 1.2732...$ is not right (the correct coefficient is 1 according to (2)). In numerical linear algebra, however, such heuristic estimation is already quite useful since ϵ is always small and its order is the most important.

Acknowledgments and A Note

The author thanks the Division of Applied Mathematics at Brown University for her generous support of a visiting position. The author would also like to thank Professors Strang and Edelman of MIT, and Maier of Arizona for their help.

Exactly one year after Gian-Carlo's permanent departure from this planet, the author would like to dedicate this paper to him, with the evergreen memory of a good friend, to whom age gaps never mattered.

The author may have missed some important references. Thus he is willing to accept criticism, or more happily, the kind help from our readers to complete the references.

References

[1] H. Bateman. *Higher Transcendental Functions,* Edited by A. Erdelyi, chapter 10.10 and 10.13. McGraw Hill, New York, 1953.

[2] P. Deift. *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach.* Number 3 in Courant Lecture Notes. Courant Institute, New York, 1999.

[3] J. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988.

[4] A. Edelman. Eigenvalues and condition number of random matrices. *SIAM J. Matrix Anal. Appl.*, 9:543–560, 1988.

[5] A. Edelman. *Eigenvalues and Condition Numbers of Random Matrices.* PhD thesis, MIT, 1989.

[6] A. Edelman. On the distribution of a scaled condition number. *Math. Comp.*, 58:185–190, 1992.

[7] A. Edelman. Eigenvalues of random matrices. Course notes for Math 273, UC Berkeley. Available at: *http://www.math.berkeley.edu/˜ edelman* , 1993, 1999.

[8] S. Geman. A limit theorem for the norm of random matrices. *Ann. Prob.*, 8:252–261, 1980.

[9] G. H. Golub and J. M. Ortega. *Scientific Computing and Differential Equations.* Academic Press, 1992.

[10] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12:1–38, 1982.

[11] M. K.-H. Kiessling and H. Spohn. A note on the eigenvalue density of random matrices. *Commun. Math. Phys.*, 199:683–695, 1999.

[12] P. R. Krishnaiah and T. C. Cheng. On the exact distributions of the extreme roots of Wishart and MANOVA matrices. *J. Multivariate Anal.*, 1:108–116, 1971.

[13] P. R. Krishnaiah and T. C. Cheng. On the exact distributions of the smallest roots of the Wishart matrix using zonal polynomials. *Ann. I. Math. Stat.*, 23:293–295, 1971.

[14] M. L. Mehta. *Random Matrices.* Academic Press, New York, 1991.

[15] R. J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, 1982.

[16] A. M. Odlyzko. On the distribution of spacings between zeros of the zeta function. *Math. Comp.*, 48(177):273–308, 1987.

[17] A. M. Odlyzko. Fast algorithms for multiple evaluations of the Riemann zeta function. *Trans. Amer. Math. Soc.*, 309(2):797–809, 1988.

[18] E. B. Saff and V. Totik. *Logarithmic Potentials with External Fields.* Grundl. d. Math. Wiss. **316**. Springer, 1997.

[19] J. W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Prob.*, 13:1364–1368, 1985.

[20] S. Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc.*, 13:87–121, 1985.

[21] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Massachusetts, 1986.

[22] G. Strang. *Introduction to Linear Algebra.* Wellesley-Cambridge Press, Massachusetts, 1993.

[23] L. N. Trefethen and D. Bau. *Numerical Linear Algebra.* SIAM, 1997.

[24] H. F. Trotter. Eigenvalue distribution of large Hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock, and Szegö. *Adv. Math.*, 54:67–82, 1984.

[25] R. Wegmann. The asymptotic eigenvalue-distribution for a certain class of random matrices. *J. Math. Anal. Appl.*, 56:113–132, 1976.

[26] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.*, 62:548–564, 1955.

[27] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. Math.*, 67:325–327, 1958.

[28] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–43, 1928.