Random Matrix Theory and Applications

alex olshevsky

October 11, 2004

Abstract

This summary will briefly describe some recent results in random matrix theory and their applications.

1 Motivation

1.1 Multiple Antenna Gaussian Channels

1.1.1 The deterministic case

Consider a gaussian channel with t transmitting and r receiving antennas. The received vector $y \in C^r$ will depend on the transmitted vector $x \in C^t$ by

$$y = Hx + n$$

where H is an rxt complex matrix of gains and n is a vector of zero-mean Gaussian noise with independent, equal variance real and imaginary parts, and x is a power-constrained input vector.

To compute the capacity of this channel we use the singular value decomposition for H: we write H as

$$H = U\Sigma V^T$$

where $U \in C^{rxr}, V \in C^{txt}$ are unitary while Σ is non-negative and diagonal. We now make a change of variables, $\tilde{x} = V^T x$, $\tilde{y} = U^T y$, $\tilde{n} = U^T n$. Note that $\tilde{n}$ has the same distribution as n, and because V is unitary out initial power constraint is unchanged. Thus, the original channel is equivalent to

$$\tilde{y} = \Sigma\tilde{x} + \tilde{n}$$

which can be written in terms of the singular values as

$$\tilde{y}_i = \lambda_i^{1/2}\tilde{x}_i + \tilde{n}_i$$

for $1 \leq i \leq min\{r,t\}$

The capacity of the channel can now be found via water-filling (see Section 10.4 of [3]): set

$$E[Re(\tilde{x}_i)^2] + E[Im(\tilde{x}_i)^2] = (\mu - \lambda_i^{-1})^+$$

where μ is chosen to meet the power constraint at $x^+ = max(x,0)$ This is the distribution that maximizes capacity; the corresponding formula for capacity is

$$C(u) = \sum_i (\ln \mu\lambda_i)^+$$

Example 1 Let $H = I_n$; in this case we simply have n parallel Gaussian channels. Then, the formula for capacity reduces to $C = nlog(1 + P/n)$ □

Note that capacity depends only on the singular values of H;two matrices that have the same singular values wil lead to channels with the same capacity.

1.1.2 The random case

A more realistic model is a random model for H. The matrix H will depend on the environment and so it is more prudent to characterize it in a statistical sense. We therefore consider the model

$$y = Hx + n$$

where the entries of H are generated randomly according to some distribution. We assume that each use of the channel results in an independent realization of H. We further assume that the transmitter has no knowledge of H but the receiver does.

We now have the problem of computing the capacity of this channel. The mutual information is given by

$$\begin{aligned} I(x;(y,H)) &= I(x;H) + I(x;y|H) \\ &= I(x;y|H) \\ &= E_H[I(x;y|H = h)] \end{aligned}$$

For a given realization of H, capacity may be obtained via the water-£lling method above. For a random x, the capacity is obtained at the distribution that maximizes the expectation over H. The distribution that maximizes $I(x;y|H)$ can be computed explicitly and the capacity can be shown to be (we omit the derivation) [10]

$$C = E[\sum_{i=1}^{m} \log(1 + \frac{P}{t}\lambda_i)] \tag{1}$$

where λ_i are the eigenvalues of $H^T H$. The question of computing capacity is now reduced to a question about the distribution of eigenvalues of a randomly generated matrix.

1.2 Power Law Graphs and the Internet

The internet has been observed to follow a power law distribution. Considering web pages as nodes and links as edges between nodes, the number of nodes with degree j is proportional to $j^{-\beta}$ for some exponent β [2].

This can be shown to be a consequence of the following fact: new web pages are more likely to link to web pages with large in-degree than to webpages with small in-degree. Indeed, webpage owners are more likely to have heard of the more popular pages and therefore more likely to link to them. This tends to reinforce the popularity of already-popular pages.

A similar statements holds for the graph of routers in internet providers. A number of similar models have been proposed to explain this effect [6].

The eigenvalues of the Internet-graph have been experimentally observed to follow a distribution that decays like the power law [9] Figuring out the distribution of the eigenvalues of a power law graph is thus an important problem in many internet-related applications.

For example, many techniques for clustering and £nding hidden patterns in graphs depend on the eigenvalues of the graph [1]. One such technique is the hubs-and-authorities algorithm used for internet search [7].

We now state the problem formally. Our problem statement follows [2].

For a given degree sequences $w = (w_1, \ldots, w_n)$ generate a random graph G_w on n vertices as follows: put an edge between i and j with probability $p_{ij} = w_i w_j \rho$ where $\rho = \frac{1}{\sum_{k=1}^{n} w_k}$. It is easy to check that the expected degree of i is w_i.

For a given graph G, the adjacency matrix A_G is de£ned as follows: $A_{ij} = 1$ if there is an edge between (i, j) and $A_{ij} = 0$ otherwise. The problem therefore assumes the following form: let w be a sequence with $w_i = ci^{-\frac{1}{\beta-1}}$ [1]. What is the distribution of the eigenvalues of A_{G_w}?

2 Some Theoretical Results

2.1 Wigner's Semicircle Law

The most fundamental result about matrix eigenvalues is Wigner's semicircle law [11], which says that in the limit as the matrix dimensions approach in£nity, the distributions of the eigenvalues approaches a semicircle.

That is, let A be a real symmetric matrix where entries are generated independently according to some distribution with $E[A_{ij}] = 0$ and $E[A_{ij}^2] = \sigma^2$ for all $i \neq j$; further, assume that the moments of $|A_{ij}|$ are £nite for £nite N. Then the density of eigenvalues of $\frac{A}{\sqrt{N}}$ approaches

$$\rho(\lambda) = (2\pi\sigma^2)^{-1}\sqrt{4\sigma^2 - \lambda^2} \quad |\lambda| < 2\sigma \tag{2}$$

and $\rho(\lambda) = 0$ if $\lambda \geq 2\sigma$.

A number of papers have been devoted to extending this result to different ensembles of random matrices. In this section, we will state the result for the Gaussian Orthogonal Ensemble and give a brief sketch of the method of proof. There summary here follows [8].

De£nition 1 *Consider an NxN matrix with entries $A_N(x,y) = a(x,y)$ that is real and symmetric, and let the collection $\{a(x,y)\}_{1 \leq x \leq y \leq N}$ be a family of random variables whose joint distributioin is one of Gaussian independent random variables. This is the Gaussian Orthogonal Ensemble.*

[1] With this, the number of vertices of degree k are proportional to $k^{-\beta}$

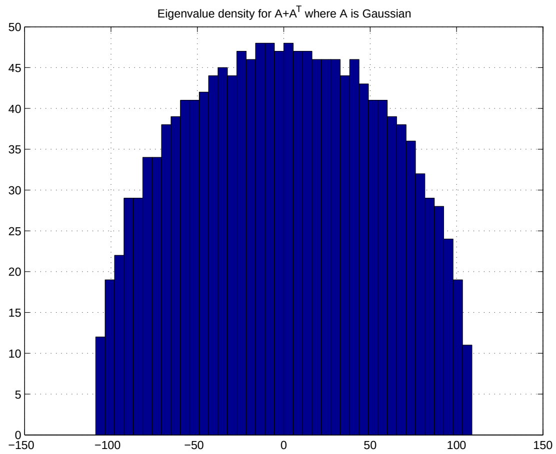


Figure 1: An illustration of the semicircle law; eigenvalue density of a matrix with N=1000

We assume that $a(x,y)$ with $x < y$ is identically distributed with $E[a(x,y)] = 0$ and $E[a^2(x,y)] = v^2$; similarly, we assume that $a(x,x)$ are identically distributed with $E[a(x,x)] = 0$ and $E[a^2(x,x)] = 2v^2$.

Let $\{\lambda_j^{(N)}\}$ be the set of eigenvalues of A_N and let $\sigma(\lambda, A_N) = \frac{1}{N}\#\{\lambda_j^{(N)} \leq \lambda\}$.

We are now in a position to state Wigner's semicircle law:

Theorem 1 *There exists a distribution $\sigma(\lambda)$ with the property*

$$lim_{N\rightarrow\infty} \int_R \phi(\lambda) d\sigma(\lambda; A_N) = \int_R \phi(\lambda) d\sigma(\lambda)$$

i.e. $\sigma(\lambda; A_N) \rightarrow \sigma(\lambda)$ weakly as $N \rightarrow \infty$ and

$$\sigma'(\lambda) = \rho(\lambda)$$

where $\rho(\lambda)$ has been defined in (2).

Sketch of Main Idea in the Proof: The proof is essentially a demonstration that the Stitljes transforms of the distributions converge. We give a brief definition of the Stiltjes transform and follow it up with highlights from the proof.

Given a probability measure F, its Stiltjes transform m_F is de£ned to be

$$m_F(z) = \int \frac{1}{x-z} dF$$

A useful property of the Stiltjes transform is that if G_N is the resolvent of A_N,

$$G_N = (A_N - Iz)^{-1}$$

then

$$m_{\sigma(\lambda; A_N)} = \frac{1}{N} Tr G_N$$

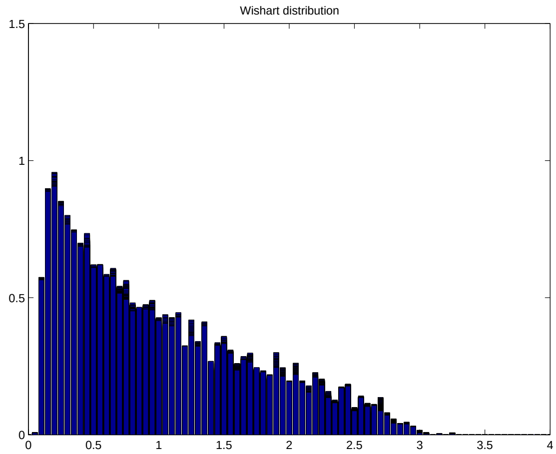


Figure 2: A realization of the wishart distribution

The proof then follows roughly along the following lines: let A_N be the random matrix of dimension N, G_N be its resolvent as above, and let

$$f(z) = lim_{N\to\infty} \frac{1}{N} TrG_N$$

It can be shown that $f(z)$ satis£es the functional equation

$$f(z) = \frac{1}{-z - v^2 f(z)}$$

This equation, however, has a unique solution and is satis£ed by $m_{\rho(\lambda)}(z)$; this shows that the Stiltjes transforms of the distribution of eigenvalues converge to ρ. It can then be shown that the distributions converge in the weak sense, as described by the theorem.

2.2 Eigenvalues of Other Ensembles

In this section, we are particularly interested in computing the distribution of the eigenvalues of the Wishart ensemble, de£ned as follows. The presentation of this section follows [4]

De£nition 2 *Let A be an MxN random matrix with independent identically distributed $N(0,1)$ elements. Let $W = AA^T$. We say W belongs to $W(M,N)$, the Wishart ensemble.*

We now state some de£nitions and facts before proceeding further.

Let Q be a real orthogonal *mxn* matrix, i.e. $QQ^T = I$. The set of all such matrices $V_{m,n}$ is called the Stiefel manifold. The following identity for the volume of the Stiefel manifold holds

$$\int_{V_{m,n}} dQ = \frac{2^m \pi^{mn/2}}{\Gamma_m(n/2)}$$

where $\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$.

A similar identity holds for complex matrices. Let $\tilde{Q}$ be an mxn unitary matrix with $\tilde{Q}\tilde{Q}^H = I$. Then the volume of these matrices is

$$\int_{V_{\tilde{m},n}} d\tilde{Q} = \frac{2^m \pi^{mn}}{\tilde{\Gamma}(n)}$$

where $\tilde{\Gamma}$ is the complex-valued Gamma function.

Next, we state (without proof) some facts on transformation and densities.

Consider the Cholesky factorization $H = LL^H$. It can be viewed as a change of variables: from the $n(n+1)/2$ independent elements of H to the $n(n+1)/2$ potentially nonzero elements of L. As such, it has a Jacobian. The same goes for the $H = LQ$ factorization: viewed as a change of variables, it has a Jacobian.

Theorem 2 *Let $H = LL^T$. The Jacobian of the transformation $H \rightarrow L$ is*

$$dH = 2^m \prod_{i=1}^{m} l_i i^{2m-2i+1} dL$$

Let $A = L\tilde{Q}$. The Jacobian of this factorization is

$$dA = \prod_{i=1}^{m} l_i i^{2n-2i+1} dL d\tilde{Q}$$

Theorem 3 *The join density of the elements of the Wishart matrix is*

$$\frac{1}{2^{mn/2}\Gamma_m(n/2)} e^{-\frac{1}{2}W} det(W)^{(n-m-1)/2}$$

Idea of Proof: If $W = AA^T$ where A is a member of the Gaussian Orthogonal Ensemble, factor A as $A = LQ$ and then note that $W = LL^T$. Integrate over the Stiefel manifold to get the density of L and apply the Jacobian formula from the previous theorem to get the density of W.

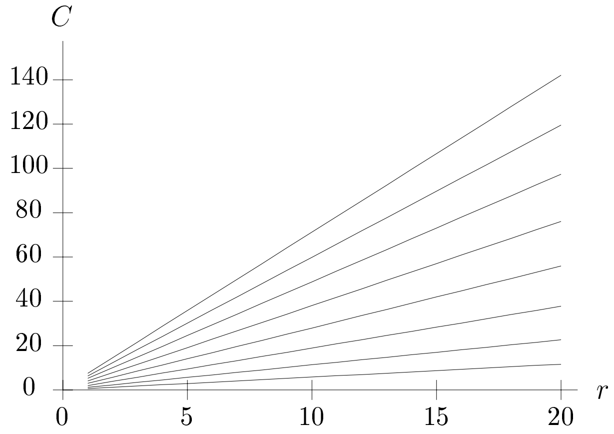
Theorem 4 *Let a Hermitian matrix S have joint density function $f(S)$ and let $f(S)$ be invariant under unitary similarity transformations, then the density of the eigenvalues is*

$$\frac{\pi^{m(m-1)}}{\tilde{\Gamma_m}(m)} f(\Delta) \prod_{i<j} (\lambda_i - \lambda_j)^2$$

Idea of Proof: Essentially repeat the arguments of the previous theorem by integrating the Q component over the Stiefel manifold.

We therefore obtain that the density of the eigenvalues of $W(m,n)$ is

$$\frac{2^{-mn}\pi^{m(m-1)}}{\tilde{\Gamma_m}(n)\tilde{\Gamma_m}(m)} e^{-\frac{1}{2}\Sigma_i \lambda_i^2} \prod \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j) \quad (3)$$



The value of the capacity (in nats) as found from (11) vs. r for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments.

Figure 3: Capacity for various input power constrains, with $r = t$

2.3 Back to Multiple Antennas

In the previous section, we have computed the density of the distribution of the eigenvalues of the Wishart ensemble. However, this is precisely the problem we had reduced the capacity of multiple-antenna systems to. Indeed, if H is Gaussian, then the performance of the system is determined by the eigenvalues of HH^T which is Wishart. Plugging equation 3 into equation 1 and simplifying, we get the following expression for capacity:

$$\int_0^\infty \log(1+P\lambda/t) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} (L_k^{n-m}(\lambda))^2 \lambda^{n-m} e^{-\lambda} d\lambda$$

where $m = min\{r,t\}$ and $n = max\{r,t\}$ and L_j^i is the Laguerre polynomial. This formula is the core result of [10]. The integral can be computed numerically to give a numerical estimate for the capacity.

2.4 Back to Power Law Graphs

The question posed in the motivation question - what are the distributions of eigenvalues of a power law graph - is essentially unsolved today. However, some progress has been made towards its solution and we describe the results here.

We £rst describe the result of [9], which is roughly the following: the largest eigenvalues of a power-law graphs are also distributed with a power law, *provided the exponent is large enough*.

We now state this result formally. Note that we use d and w interchangeably to refer to degrees.

Theorem 5 *Let $(w_1, \ldots, w_n)$ be a vector of degrees with $w_i = \frac{w_1}{i^\alpha}$, with $\alpha \in (\frac{1}{2}, 1)$; this yields a power law distribution with $\beta > 3$; then, the eigenvalues of A_{G_w} satisfy*

$$\sqrt{w_i}(1 - o(1)) \leq \lambda_i \leq \sqrt{d_i}(1 + o(1)$$

for $i = 1, \ldots, k' = O(n^\beta)$.

Sketch of Main Idea in Proof: One constructs a subgraph made entirely of stars (a star on n vertices is made up of a root and $n-1$ leaves, with an edge between the root and each of the leaves, but no edges between the leaves). This subgraph is constructed as follows: star S_i has node i at its center and leaves those nodes among $i+1, \ldots, n$ that are not adjacent to any node in $1, \ldots, i$.

It can then be shown by by taking expectation that the expected degree, s_i of star S_i, satisfies

$$d_i(1 - n^{\beta(1-2\alpha)}) \leq s_i \leq d_i$$

The eigenvalues of a star on n vertices are $\sqrt{n-1}$ and $-\sqrt{n-1}$. Approximating the graph G as the union of the stars, and recalling that the set of eigenvalues of a graph G *is* the union of eigenvalues of its components, it follows that λ_i behaves as $\sqrt{d_i}$. Since d_i follows a power law, so does λ_i.

We should note that to formalize the above argument, it must be shown that the non-star components of the graph do not contribute to the largest eigenvalues.

As the reader may remember, real graphs have decay coefficients much smaller than 3. The graph of internet routers, for example, has an exponent between 0.85 and 0.93; measurement of the decay exponent of the internet web page graph typically have given results between 2.1 and 2.4 [9]. Hence an adequate understanding of these case is not available. Extending the above results to lower exponents is still an unsolved problem.

2.5 Experimental Results for Some More Realistic Models

Real graphs, however, cannot be modelled very well with the random graph models described above. While such models insert edges into the graphs between vertices in an uncorrelated way, in reality the nodes are usually connected in a correlated manner.

Such differences have great impact on the eigenvalues. Consider, for example,figure 4 which shows convergence to the semicircle law for a simple random graph model in which each edge is inserted with probability p. All figures are taken from [5].

However, we know that the internet has been observed to have eigenvalues that follow that power law - and not the semi-circle! Such discrepancies motivate the need for better models.

One feature that real-life graphs have that uncorrelated models have is a typically small distance between any two nodes. To remedy this, two models of random graphs have been recently introduced [5].

2.5.1 Small World Graphs

This is a class of graphs constructed as follows:first draw the vertices $1, 2, \ldots, N$ on a circle in ascending order. Then for every i connect verte i to the k vertices lying closest to it on the circle:

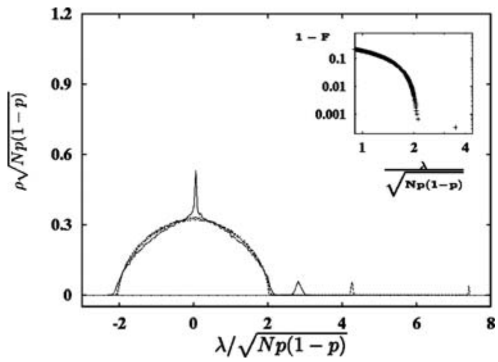


Figure 4: Convergence to the semi-circle for the random graph

$i-k/2,\ldots,i-1,i+1,\ldots,i+k/2$. Next, for vertex 1, consider the £rst forward connection, e.g from 1 to 2. With probability p, reconnect vertex 1 to another randomly chosen vertex instead (without allowing for multiple edges). Proceed forward and do the same for all of vertex 1's forward connections. Then, repeat this for every vertex.

2.5.2 Scale Free Graphs

Construct a scale-free graph as follows: starting from a set of m isolated vertices, at each step add another vertex and k connections. Pick any one of the existing vertices for a connections with probability $\frac{k_i}{\sum_j k_j}$, where k_i is the degree of vertex i. Such a model can be proven to converge to a power law distribution.

The behavior of the spectra of these graphs is not theoretically understood, but [5] reports the results of a variety of simulations which we include here. In short, the behavior of the eigenvalues follows neither the power law nor the semicircle law and is too complex to be understood via current methods.

The eigenvalues of small-world graphs follow the semicircle law for $p = 1$; however, if $0 < p < 1$ then the behavior of the eigenvalues is highly irregular and consists of a number of spikes whose positions are not understood.

The eigenvalue distribution of the scale free graphs tends to look triangular with high N - see £gure 5. This phenomenon is not understood and its explanation is an open question.

References

[1] Y. Azar, A. Fiat, A. Karlin, F. McSherry, J. Saia, *Spectral analysis for data dining*, Proc. of 33rd STOC (2001), 619-626.

[2] F. Chung, L. Lu, V. Vu, *The spectra of random graphs with given expected degrees*, Proceedings of National Academy of Sciences, 100, no. 11, (2003), 6313-6318.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley, 1991.

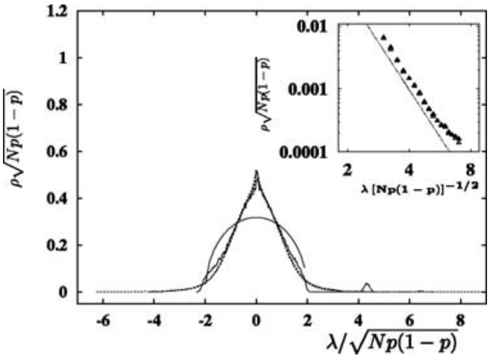


Figure 5: Eigenvalue distribution of the scale free graphs

[4] A. Edelman, Eigenvalues and Condition Numbers of Random Matrices, MIT PhD Dissertation, 1989.

[5] I.J. Farkas, I. Derenyi, A.L. Barabasi, T. Vicsek, *Spectra of "real-world" graphs: Beyond the semicircle law*, Physical Review E, 64, 2001.

[6] A. Fabrikant, E. Koutsoupias, C. Papadimitriou, *Heuristically Optimized Tradeoffs*, available at http://www.cs.berkeley.edu/ christos/hot.ps

[7] J. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of the ACM 46 (1999), 604-632.

[8] A. de Monvel and A. Khorunzhy, Some Elementery Results around the Wigner Semicircle Law.

[9] M. Mihail and C. Papadimitriou, *On the Eigenvalue Power Law*, Procedings of RANDOM '02

[10] E.Telatar, *Capacity of Multi-antenna Gaussian Channels*, European Transactions on Telecommunications, Vol42, 2000

[11] E. Wigner, *Characteristic vector of bordered matrices with in£nite dimensions*, Annals of Mathematics, 62, 1955, 548-564.