From the Ground Up: Dispersion Statistics

Taqi

Goal

In this document, we seek to motivate a precise notion for "eigenvalue spacings" in order to study them meaningfully. To do so, we will carefully lay out notation and definitions along the way as our "toolkit". Furthermore, there are small caveats that will be discussed - mostly, those are free parameters we need to be aware of so that we precisely know what we are talking about when we say "eigenvalue spacings".

Spectral Statistics and Random Matrices: What are they?

Random Matrices

What are *spectral statistics*? Do they have to do with rainbows? Sceptres? No, they don't, but they're almost as colorful and regal. *Spectral statistics*, aptly named so, borrows from the spectral-like patterns observed in statistical physics - whether it may be atomic spectra or other quantum mechanical phenomena. The borrowing is loose and not literal, but still somewhat well founded.

Random matrix theory - as a field - was developed extensively in the 1930s by Wigner, a nuclear physicist. He found connections between the deterministic properties of atomic nuclei and their random and stochastic behaviors. The link? – Random matrices.

Definition 0.1 (Random Matrix). *A (homogenous) random matrix is any matrix $M \in \mathbb{F}^{N \times N}$ is a matrix whose entries are i.i.d random variables. So, if a random matrix $M = (m_{ij})$ is $\mathcal{D}$-distributed, this is equivalent to saying $m_{ij} \sim \mathcal{D}$. In the scope of this thesis, we will only work with homogenous random matrices. From thereonafter, assume every random matrix to be homogenously distributed.*

With a *random matrix* intialized, we can study its features by studying a very core type of number that "summarizes" it - the eigenvalue. The study of eigenvalues and eigenvectors are primarily in the scope of Linear Algebra. They are quite important for many reasons. In statistical physics, many processors are represented by operators or matrices, and as such, their behaviours can be described by the eigenvalues of their corrosponding matrices! So, what are *eigenvalues* exactly?

The Quintessential Spectral Statistic: the Eigenvalue

Given any standard square matrix $P \in \mathbb{F}^{N \times N}$, the *eigenvalue* is given by the roots of the characteristic polynomial $\text{char}_P(\lambda) = \det(P - \lambda I)$. By the Fundamental Theorem of Algebra, we always have as many complex eigenvalues $\lambda \in \mathbb{C}$ as there are dimensions to the square matrix (n). When the matrix is distributed a certain way, we can see patterns in the eigenvalue distributions! So, an eigenvalue is a **spectral statistic** of a random matrix!

The Glorious Spectrum

Now that we have defined eigenvalues, we need to talk about them in a concise manner. Since we know an $N \times N$ matrix has exactly N eigenvalues (counting repeated values), we motivate the *eigenvalue spectrum*.

Definition 0.2 (Spectrum). *Given a matrix P, the spectrum of P is defined as the multiset of its eigenvalues, denoted by $\sigma(P) = \{\lambda_i \in \mathbb{C}\}_{i=1}^{N}$. Note that we say multiset since eigenvalues could be repeated (through algebraic multiplicity).*

Interlude: Ensembles

While the spectrum of a matrix provides a good summary of the matrix, in random matrix theory, a matrix is considered a single point or observation. Additionally, simulating large matrices becomes harder and harder as $N \to \infty$. As such, to obtain more eigenvalue statistics efficiently, another dimension is introduced by motiving the *random matrix ensemble*.

Definition 0.3 (Random Matrix Ensemble)**.** *A $\mathcal{D}$-distributed random matrix ensemble $\mathcal{E}$ over $\mathbb{F}^{N \times N}$ of size K is defined as a set of $\mathcal{D}$-distributed random matrices $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N \times N}\}_{i=1}^K$. In simple words, it is simply a collection of iterations of a specified class of random matrix.*

Now that ensembles are well defined, we can motivate a very core object in random matrix theory - the spectrum of a random matrix ensemble. From its name, it is indeed what one might expect it to be.

Definition 0.4 (Ensemble Spectrum)**.** *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\sigma(\mathcal{E})$. To take the spectrum of an ensemble, simply take the union of the spectra of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N \times N}\}_{i=1}^K$, then $\sigma(\mathcal{E}) = \bigcup_{i=1}^K \sigma(P_i)$.*

A common theme in this thesis will be that singleton matrices do not provide insightful information on their own. Rather, it is the collective behavior of a $\mathcal{D}$-distributed ensemble that tells us about how $\mathcal{D}$ impacts our spectral statistics. As such, ensemble statistics are in a way the engine of our research.

A Caveat: Eigenvalue Ordering

When we motivate the idea of matrix dispersion, we will consider order statistics of that matrix's eigenvalues in tandem with its dispersion. However, to do so presupposes that we have a sense of what "ordered" eigenvalues means. Suppose are given a matrix P which has an "unordered" spectrum $\sigma(P) = \{\lambda_j\}$. It is paramount to know what ordering scheme $\sigma(P)$ is using, because otherwise, the indices are meaningless! So, to delineate this, we add an index to σ. Often, the ordering context will be clear and the indexing will be omitted. Consider the two following *ordering schema*:

Classical definitions of an *ordered spectrum* follow the standard ordering in the reals; denote this as the ordering by "sign" scheme. Note that because well-ordering is defined on the reals, we cannot use the sign scheme when $\sigma_P \subset \mathbb{C}^N$. Additionally, if we have a symmetric matrix (of real or complex entries), we have real eigenvalues. In that case, we could either this metric of the other. Without further ado, we write the *sign-ordered spectrum* as follows:

$$\sigma_S(P) = \{\lambda_j : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N\}_{j=1}^N$$

Alternatively, we can motivate a different scheme that properly handles complex eigenvalues. We could sort the spectrum by the norm of its entries; denote this as the ordering by "norm" scheme. Note that when we take the norms of the eigenvalues, we essentially ignore "rotational" features of the eigenvalues. Signs of eigenvalues indicate reflection or rotation, so when we take the norm, we essentially become more concerned with scaling. Without further ado, we write the *norm-ordered spectrum* as follows:

$$\sigma_N(P) = \{\lambda_j : |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|\}_{j=1}^N$$

Pairing Schema

Next up, we introduce a new notation for a pairing scheme denoted Π. What are pairing schemes and why do they matter? Recall that our goal is to study the spacings between eigenvalues. If we are studying spacing, then a priori, we are concerned with pairs of eigenvalues! Spacing, after all, is a binary relationship. So, with the definitions of spectra well motiviated, a natural definition of pairing schema follows. Essentially, Π is just a subset of the Cartesian product of a spectrum with itself. In other words, if we denote $\mathbb{S} := \sigma(P)$, then we say that a pairing scheme is simply a subset $\Pi \subseteq \mathbb{S}^2$. There are a few special pairing schema to consider:

1. Let $\Pi_>$ be the set of unique upper-pair ($>$) combinations of ordered eigenvalues. This will be the standard ordered pair scheme used in lieu of our dispersion metric argument orders (more later).

$$\Pi_> = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i > j\}_{i=1}^{N-1}$$

2. Let $\Pi_<$ be the set of unique lower-pair ($<$) combinations of ordered eigenvalues.

$$\Pi_< = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i < j\}_{i=1}^{N-1}$$

3. Let Π_1 be the largest pair of eigenvalues of a spectrum. Nice and sweet.

$$\Pi_1 = \{(\lambda_2, \lambda_1)\}$$

4. Let Π_C be the consecutive pairs of eigenvalues in a spectrum. This pairing scheme gives us the minimal information needed to express important bounds and spacings in terms of its elements.

$$\Pi_C = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i = j + 1\}_{i=1}^{N-1}$$

Dispersion Metrics

When we define the dispersion of a matrix, we will see that there is a free argument d called the dispersion metric. This function d is a general function whose domain is always two eigenvalues. In set notation, this is the set $\mathbb{C} \times \mathbb{C}$ - a pair of two complex numbers. Its range will often be the positive reals $\mathbb{R}^+$; this is because the dispersion metric often will be substitutable with distance metric. Sometimes, the range will be $\mathbb{C}$. So, the dispersion metric will take the following form:

$$d : \mathbb{C} \times \mathbb{C} \to \{\mathbb{R}^+, \mathbb{C}\}$$

Consider the following dispersion metrics below. Out of those 4 dispersion metrics, only the first one has a range of $\mathbb{C}$. The rest have a range of $\mathbb{R}^+$. Additionally, the second and third metrics are **symmetric** operations while the rest are not. The β-norm is only a symmetric operation when β is even.

1. The identity difference: $d_{id}(z, z') = z' - z$

2. The standard norm: $d_n(z, z') = |z' - z|$

3. The β-norm: $d_\beta(z, z') = |z' - z|^\beta$

4. The difference of absolutes: $d_{ad}(z, z') = |z'| - |z|$

Order of Arguments and the Lower-Pairing Scheme

One of the common pairing schemes we consider is $\Pi_o \in \{\Pi_>, \Pi_<\}$ We can use either the upper or the lower pairing scheme; there is no distinction since combinations are order-invariant. However, if we specify one of the other, the way our dispersion metrics will behave is impacted. For this reason, we choose to simeltanously: (i) choose upper-pair schema $\Pi_>$ and (ii) use the above argument order.

To motivate intuitive mathematical notation, we read $d(z, z')$ as the dispersion of z with respect to (or in reference to) z'. Since we use the upper-pair scheme $\Pi_>$, the smaller eigenvalue is to the left and the larger to the right. This way, we take the dispersion of π_{21} for example, we'd read $d(\pi_{21})$ as saying we're taking the dispersion of λ_2 with respect to λ_1. If this were the norm metric, we'd evaluate $\delta_{21} = |\lambda_1| - |\lambda_2|$.

Matrix & Ensemble Dispersion

Finally, we are able to motivate the definition of a matrix dispersion! Suppose we have a $\mathcal{D}$-distributed random matrix $P \in \mathbb{F}^{N \times N}$ or a random matrix ensemble $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N \times N}\}$. Then we define their dispersion as follows.

Definition 0.5 (Dispersion). *The dispersion of a matrix $P \in \mathbb{F}^{N \times N}$ with respect to a dispersion metric $d : \mathbb{C} \times \mathbb{C} \to \mathbb{F}$ and pairing scheme Π, call it $\Delta_d(P, \Pi)$, is defined as follows. Suppose $\sigma(P) := \mathbb{S}$ is the ordered spectrum of P where $\sigma(P) = \{\lambda_1, \ldots, \lambda_N\}$. Then, let $\Pi = \{\pi_{ij} = (\lambda_i, \lambda_j)\} \subseteq \mathbb{S}^2$ be a subset of eigenvalue ordered pairs. Then, the dispersion of P with respect to d is simply the set $\Delta_d(P, \Pi) = \{\delta_{ij} = d(\pi_{ij}) \mid \pi_{ij} = (\lambda_i, \lambda_j) \in \Pi\}$.*

Definition 0.6 (Ensemble Dispersion). *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\Delta_d(\mathcal{E}, \Pi)$. To take the dispersion of an ensemble, simply take the union of the dispersions of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N \times N}\}_{i=1}^K$, then $\Delta_d(\mathcal{E}, \Pi) = \bigcup_{i=1}^K \Delta_d(P_i, \Pi)$.*

Pairing Schema Continued: A Commentary

Bounds and Considerations

One thing to consider from this dispersion set is what we could learn from it. The dispersion of a matrix P, $\Delta_d(P)$ could tell us quite a few things. One thing it could tell us is the distribution of the dispersion of eigenvalues given that we uniformly and randomly select a pair of eigenvalues; this works when d is a symmetric function. However, if we are considered about spacings, then we must taking into consideration a few facts.

Linear Combinations

The identity difference metric is not necessarily a "bad metric". In fact, it is one of the few metrics which allow transitivity under composition - which brings up an important detail. In the total dispersion $\Delta_d(P)$, there is some redundant information if we are totally concerned with spacings. Namely, consider the following fact:

$$\lambda_1 - \lambda_3 = (\lambda_1 - \lambda_2) - (\lambda_3 - \lambda_2)$$

In other words, if d is the ID dispersion metric, then $d(\pi_{13}) = d(\pi_{12}) - d(\pi_{32})$. We could say the second eigenvalue is a pivot in which we could perform right-cancellation.

Triangle Inequality

However, as we said previously, only the identity difference metric gives us transitive properties. On the other hand, the other metrics provide us bounds when we compose them. They are given by a simple application of the reverse triangle inequality. Namely, consider the following fact:

$$||y| - |x|| \leq |y - x|$$

If we let d_1 be the difference of absolutes metric and d_2 be the standard norm, then we could rephrase this as saying $|d_1(\pi_{ij})| \leq d_2(\pi_{ij})$. This is just a rephrasing of the inequality. However, if we are careful enough, we can remove the absolute value to obtain a more meaningful upper bound. The value of $d_1(\pi_{ij})$ can be interpreted as the difference between the sizes of λ_j and λ_i. However, recall that in the standard definition of a dispersion, we use the **lower** pair combinations so that we get $i > j$, making the value $|\lambda_j|$ always greater than $|\lambda_i|$ by construction. As such, the left-hand value would always be positive, allowing us to drop the absolute value. So, we could say:

$$d_1(\pi_{ij}) \leq d_2(\pi_{ij}) \text{ given } \pi_{ij} \in \Pi_>$$

Sufficiency of Consecutive Pairs

Recall from previously that with respect to the identity difference, some elements of the dispersion $\Delta_d(P)$ can be expressed as a linear combination of other elements. Note that using those linear combinations and the triangle inequality, we could derive the analogous bounds. It is quite useful to impose a condition that none of the values are a linear combination of the other.

So, we motivate a more restrictive definition of matrix dispersion: the consecutive-values dispersion $\Delta_d(P, \Pi_c)$. This dispersion is essentially a special case of the regular dispersion, but with a particular pairing scheme Π_c. To avoid this issue of linear combinations, we simply take the pairs of consecutive eigenvalues. In other words, $\Pi_c = \{(\lambda_i, \lambda_j) \in \sigma_P \times \sigma_P \mid j = i + 1\}_{i=1}^{N-1}$. This pairing scheme enumerates the pairs such that the dispersion metric has a natural mapping - it is simply the "distance" between an eigenvalue and the one that is one rank smaller than it.