# Dispersion Statistics

## Taqi

## Matrix & Ensemble 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 0.1** (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 $\Pi$, 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 $\Delta_d(P) = \{\delta_{ij} = d(\lambda_i, \lambda_j) \mid \pi_{ij} \in \Pi\}$.*

**Definition 0.2** (Ensemble Dispersion). *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\Delta_d(\mathcal{E})$. 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

In our definition, we introduce a new notation for a pairing scheme denoted $\Pi$. In essence, $\Pi$ is simply a subset of the Cartesian product of a spectrum $\mathbb{S} = \sigma(P)$ with itself. In other words, $\Pi \subseteq \mathbb{S}^2$. There are a few special pairing schema to consider:

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

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

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

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

3. Let $\Pi_1$ be the largest pair of eigenvalues of a spectrum.

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

4. Let $\Pi_C$ be the consecutive pairs of eigenvalues in a spectrum.

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

## A Caveat: Eigenvalue Ordering

In our definition of matrix dispersion, we are given a matrix $P$ which has an "ordered" spectrum $\sigma_P$. It is paramount to know what ordering scheme $\sigma(P)$ is using, because otherwise, the indices are meaningless! So, consider the two following 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 $\sigma_P \subset \mathbb{C}^N$. 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. We write the norm-ordered spectrum as follows:

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

## Dispersion Metrics

In the definition of matrix dispersions, we see that there is a free argument $d$ called the dispersion metric. This function $d$ is a general function whose domain is always $\mathbb{C} \times \mathbb{C}$ - a pair of two eigenvalues. Its range, strangely, could be any of the two fields $\mathbb{R}$ or $\mathbb{C}$. Most often, the dispersion metric will be a standard distance function so its range will often be the positive reals $\mathbb{R}^+$. 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.

1. The identity difference: $d(x, y) = y - x$
2. The standard norm: $d(x, y) = |y - x|$
3. The $\beta$-norm: $d_\beta(x, y) = |y - x|^\beta$
4. The difference of absolutes: $d(x, y) = |y| - |x|$

## 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 $\lambda_j$ and $\lambda_i$. 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 $\Pi_c$. 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.