

1 Behavior of Dey-Lee estimator in the absence of square root-like behavior

1.1 Notation

We fix notation. The data matrix X is $N \times M$, with N independent samples of some M -dimensional vectors. The distribution of each row of X has covariance matrix,

$$\begin{pmatrix} A & 0 \\ 0 & \Sigma \end{pmatrix} =: T^{1/2} \quad (1.1)$$

where A is $m \times m$ and Σ is $M - m \times M - m$. A is the ‘spike’ and Σ is the ‘noise.’

We denote the eigenvalues of A , the **population spikes** by p_1, \dots, p_m . We denote the spectral measure of the noise Σ by H_M , and assume it converges to some limiting measure H . The **sample eigenvalues** will be denoted by $\lambda_1, \lambda_2, \dots$.

The **sample covariance** matrix $N^{-1}X^*X$ has the block-matrix decomposition,

$$\frac{1}{N}X^*X = \frac{1}{N} \begin{pmatrix} X_1^*X_1 & X_1^*X_2 \\ X_2^*X_1 & X_2^*X_2 \end{pmatrix} \quad (1.2)$$

where X_1 is $N \times m$, consisting of the first m coordinates of each row of X and X_2 is the last $M - m$ coordinates, i.e., $X = (X_1, X_2)$. It is from this block matrix decomposition that all of the formulas for overlaps of sample eigenvectors with population eigenvectors, relation between sample eigenvalues and population eigenvalues, etc., come from. The derivation is summarized in Section 2.

1.2 Marcenko-Pastur quantities

The Dey-Lee estimators all involve Marcenko-Pastur-type quantities related to X_2 through Σ . We now recall these quantities. If H_M converges to a limit H , and then the empirical spectral distribution of $N^{-1}X_2^*X_2$ converges to a measure ρ admitting the following description. First, recall that the eigenvalues of $N^{-1}X_2X_2^*$ are the same as $N^{-1}X_2^*X_2$ up to $|(M - m) - N|$ zeros. We denote the limit of the ESD of $N^{-1}X_2X_2^*$ by $\tilde{\rho}$, and denote the Stieltjes transforms of ρ and $\tilde{\rho}$ by m and \tilde{m} respectively. They are related via,

$$\gamma zm = (1 - \gamma) + z\tilde{m}, \quad (1.3)$$

where

$$\gamma := \lim_{N \rightarrow \infty} \frac{M}{N} \quad (1.4)$$

The function \tilde{m} satisfies the functional equation,

$$\tilde{m}(z) = - \left(z - \gamma \int \frac{\lambda dH(\lambda)}{1 + \lambda \tilde{m}(z)} \right)^{-1}. \quad (1.5)$$

Now, Dey-Lee define the function ψ as,

$$\psi(\alpha) := \alpha + \gamma \alpha \int \frac{\lambda dH(\lambda)}{\alpha - \lambda}. \quad (1.6)$$

Hence, we see the functional relation,

$$\psi(-1/\tilde{m}(z)) = z. \quad (1.7) \quad \{\text{eqn:func}\}$$

For the remainder of the discussion we will assume that the measures ρ and $\tilde{\rho}$ have a single interval of support (on which the density is analytic, which comes from [Choi-Silverstein, “Analysis of the limiting spectral distribution of...”]).

1.3 Formulas for estimators

From the paper [Bai-Yao, “On sample eigenvalues in a generalized spiked population model”], we know that a population spiked eigenvalue p_i gives rise to an outlier iff $p_i > \mathfrak{p}$, where \mathfrak{p} is the point,

$$\mathfrak{p} := \inf_p \{p : \psi'(p') > 0 \text{ for all } p' > p\}. \quad (1.8)$$

When everything is quite regular, usually \mathfrak{p} just turns out to be the largest solution to $\psi'(\mathfrak{p}) = 0$. The location of the resulting spike is at $s_i := \psi(p_i)$.

The equation for the overlap between the population spiked eigenvector and sample eigenvector is,

$$-\frac{\tilde{m}(s_i)}{s_i \tilde{m}'(s_i)} = \frac{p_i \psi'(p_i)}{\psi(p_i)}. \quad (1.9) \quad \{\text{eqn:over}\}$$

Differentiating the functional equation (1.7) we get,

$$\psi'(-1/\tilde{m}(z)) \frac{\tilde{m}'(z)}{(\tilde{m}(z))^2} = 1. \quad (1.10)$$

(Note that this, together with (1.7) explains (1.9)). Hence, we see the point,

$$\mathfrak{s} := \psi(\mathfrak{p}) \quad (1.11)$$

coincides with the right end-point of the measures $\rho, \tilde{\rho}$. Note that this makes sense; to the right of \mathfrak{p} , the function ψ is increasing. Population spikes p_i closer to \mathfrak{p} give rise to sample spikes closer to the right edge of the noise distribution ρ , until $p_i \downarrow \mathfrak{p}$ at which point the corresponding sample spike is swallowed into the noise ρ .

1.4 Regular and irregular cases

I want to examine more closely the behavior of these estimators for spikes close to the right edge \mathfrak{s} in the case where ρ lacks a square-root behavior.

Denote by $[A, B]$ the **convex hull** of the **support** of H (only the location of the right edge B will be important, so it is no matter that A was already in regard to the population spikes above). Then, from the equation,

$$\psi'(\alpha) = 1 - \gamma \int \frac{\lambda^2 dH}{(\alpha - \lambda)^2}, \quad (1.12)$$

we see that there are two cases that have possibly different behavior:

1. The “regular” case:

$$\int \frac{\lambda^2 dH}{(B - \lambda)^2} > \frac{1}{\gamma} \quad (1.13)$$

2. The “irregular” case:

$$\int \frac{\lambda^2 dH}{(B - \lambda)^2} \leq \frac{1}{\gamma}. \quad (1.14) \quad \{\text{eqn:irre}\}$$

In the regular case, the point \mathfrak{p} satisfies $\mathfrak{p} > B$. In the irregular case $\mathfrak{p} = B$ (pretty sure this is always true).

In the regular case, the measure ρ will have a square-root behavior near its right spectral edge \mathfrak{s} . This is proven in [Bao-Pan-Zhou, “Universality of the largest...”] (see their condition 1.1(iii) which is a finite N version of this, assumed to hold uniformly in N), or [Lee-Schnelli, “Tracy-Widom distribution for...”] (which is essentially the same paper as BPZ), or [El Karoui, “Tracy-Widom limit for the largest...”]. The survey paper [Hachem-Hardy-Najim, “A Survey on the Eigenvalues Local Behavior of Large Complex Correlated Wishart Matrices”] contains this as well, perhaps in a more readable fashion.

The irregular case is less studied, and a square root behavior does not necessarily emerge. An example that this could arrive in is when the measure $dH(B - t) \sim t^\beta dt$ for some $\beta > 1$. Note that if $\beta < 1$, then the value of the integral above is $+\infty$ so the regular case always arises. Moreover, γ must be small enough to be in the irregular case.

1.4.1 Kwak-Lee-Park paper

This kind of power-law vanishing was actually studied quite recently in a paper [Kwak-Lee-Park, “Extremal eigenvalues of sample covariance matrices with general population”] (an August 2019 arXiv date!!). They showed that under the assumption $dH(B-t) \sim t^\beta dt$ for $\beta > 1$, the limiting density ρ has a power law with the same exponent β for γ small enough. They have some **incomplete** results on fluctuations (their main theorem, Thm 2.6 is also stated incorrectly). Recall that we are just looking at the behavior of the noise matrix when we have no spikes. Essentially, they show that the sample eigenvalues satisfy

$$\lambda_i = \sigma_i(\Sigma) + \text{error} \quad (1.15) \quad \{\text{eqn:k1p}\}$$

for some error term. Here $\sigma_i(\Sigma)$ is the i th largest eigenvalue of Σ (we introduce this notation to not confuse it with p_k which involved the spike matrix A above).

In the case that Σ is diagonal with iid entries coming from H , then the fluctuations $\sigma_i(\Sigma)$ are coming from order statistics of iid random variables, and in this case are a Weibull distribution. The fluctuations of the $\sigma_i(\Sigma)$ are larger than the error term (1.15), so in this specific iid case, you get Weibull fluctuations for λ_i .

I do not know how relevant this is at all, because I see no reason why the eigenvalues $\sigma_i(\Sigma)$ would be iid random variables in any realistic situation.

1.5 Other papers mentioning irregular case

The aforementioned [Hachem-Hardy-Najim] mentions that they believe that the irregular case “falls outside the RMT universality class” and that they believe that the limiting extremal eigenvalue fluctuations depend on the limiting measure H (see Section 5). An additive version of this model, the deformed GUE was considered in [Capitaine-Peche, “Fluctuations at the edges of the spectrum...”]. In the irregular case, they claim that one can prove that the limiting fluctuations are Gaussian instead of TW, in Remark 1.1. This remark is unproven, and I see no reason to believe it one way or another. If it is Gaussian, it may be heavily dependent on the fact that the matrix entries of the GUE are themselves Gaussian.

1.6 Comparison of estimators of irregular and regular cases near the edge

Now, we know that the squared overlap of the i th sample eigenvector and population eigenvector converges to the function of the i th population spike p_i given by,

$$\eta(p_i) := \begin{cases} 0, & p_i \leq \mathfrak{p} \\ \frac{p_i \psi'(p_i)}{\psi(p_i)}, & p_i > \mathfrak{p} \end{cases} \quad (1.16)$$

In the regular case, this is a **continuous** function of p_i ; that is, $\lim_{p \downarrow \mathfrak{p}} \psi'(p) = 0$. In the irregular case whenever we have a strict inequality in (1.14) this will be a **discontinuous** function of p_i , as we will have

$$\frac{\mathfrak{p} \psi'(\mathfrak{p})}{\psi(\mathfrak{p})} > 0. \quad (1.17)$$

Essentially, I would be cautious in applying the Dey-Lee theory to spikes that are close to the noise distribution when the noise appears to lack any square-root behavior. Consider for the moment the case of pure noise. When ρ has a square root, then since $\lambda_1 = \mathfrak{s} + \mathcal{O}(N^{-2/3})$, one sees that

$$\eta(\lambda_1) \approx N^{-2/3} \quad (1.18)$$

which is still too large by a factor of $N^{1/3}$, but is at least going to 0 as $N \rightarrow \infty$. In the irregular case, as long as λ_1 falls to the right of the rightmost spectral edge \mathfrak{s} , one will estimate a positive quantity not going to 0 as $N \rightarrow \infty$ due to the discontinuity.

2 Derivation of estimators from Bai-Yao and Dey-Lee

The set-up is that we look at,

$$X = ZT^{1/2} \quad (2.1)$$

where Z is an $N \times M$ matrix (N is number of samples, M is number of features/covariates), and T is an $M \times M$ block symmetric matrix in the form,

$$T = \begin{pmatrix} A & 0 \\ 0 & \Sigma \end{pmatrix} \quad (2.2)$$

where A is the “spike” and is dimension $d \times d$ for finite d , and Σ is the population covariance matrix.

The sample covariance matrix that is inspected is then,

$$\frac{1}{N}X^*X = \frac{1}{N}T^{1/2}Z^*ZT^{1/2}. \quad (2.3)$$

Now, X has the form

$$X = (X_1 \quad X_2), \quad (2.4)$$

where the columns of X_1 is $N \times d$ and X_2 is $N \times (M - d)$. Each row of X_1 is found by multiplying the row vector formed by the first d entries of a row of Z on the right by $A^{1/2}$. Each row of X_2 is found by multiplying the row vector formed by the last $M - d$ entries of a row of Z on the right by $\Sigma^{1/2}$. Note that X_1 and X_2 are independent due to our block matrix assumption for T .

The sample covariance matrix $N^{-1}X^*X$ has the form,

$$\frac{1}{N}X^*X = \frac{1}{N} \begin{pmatrix} X_1^*X_1 & X_1^*X_2 \\ X_2^*X_1 & X_2^*X_2 \end{pmatrix}. \quad (2.5)$$

Due to the block matrix structure, the condition for μ to be an eigenvalue of $N^{-1}X^*X$ is,

$$0 = \det(N^{-1}X^*X - \mu) \quad (2.6)$$

$$= \det(N^{-1}X_2^*X_2 - \mu) \det \left[N^{-1}X_1^*X_1 - \mu - (N^{-1}X_1^*X_2) \frac{1}{N^{-1}X_2^*X_2 - \mu} (N^{-1}X_2^*X_1) \right] \quad (2.7)$$

as long as μ is not an eigenvalue of $N^{-1}X_2^*X_2$, but generically $N^{-1}X_2^*X_2$ and $N^{-1}X^*X$ will never share eigenvalues. The term on the RHS is the determinant of a $d \times d$ matrix. Since d is fixed, we can find the pointwise limit of this matrix which will give us the equation for μ to be an eigenvalue, as the determinant will be continuous in the matrix entries.

Now, since d is finite, by the law of large numbers $N^{-1}X_1^*X_1$ converges to A very quickly. Conditioning on X_2 , the second term should be thought of as the matrix

$$N^{-1}X_1^*MX_1. \quad (2.8)$$

with M defined appropriately. This has expectation $A \times N^{-1}\text{tr}M$ and variance bounded by $N^{-2}\text{tr}M^2$ assuming M symmetric. The tracial quantities can all be bounded in terms just the resolvent $(N^{-1}X_2^*X_2 - \mu)^{-1}$ without having to worry about conjugation with X_2 etc.

Now,

$$\frac{1}{N}\text{tr}M = \frac{M - d}{N} + \mu \frac{1}{N} \text{tr} \frac{1}{X_2^*X_2 - \mu} \rightarrow \frac{M}{N} (1 + \mu m(\mu)) \quad (2.9)$$

where m is the Stieltjes transform of the generalized MP law for a sample covariance matrix of N samples with covariance Σ (or a limit, or whatever). Defining $\gamma = M/N$, we get the equation,

$$\det [A - \mu - A\gamma(1 + \mu m(\mu))] = 0. \quad (2.10)$$

Diagonalizing A gives us the relation,

$$p_k \approx \frac{s_k}{1 - \gamma(1 + s_k m(s_k))} \quad (2.11)$$

where p_k are population eigenvalues and s_k are sample eigenvalues. The RHS seems to be the function $-1/\tilde{m}(s_k)$ where \tilde{m} is the limiting spectral distribution $N^{-1}\text{tr}(X_2X_2^* - z)^{-1}$.

Now, Bai-Yao works as follows. Define,

$$\psi(-1/\tilde{m}(z)) = z. \quad (2.12) \quad \{\text{eqn:inv}\}$$

Then, if ψ is of the form,

$$\psi(\alpha) = \alpha + \gamma\alpha \int \frac{\lambda dH(\lambda)}{\alpha - \lambda}, \quad (2.13)$$

then necessarily,

$$\tilde{m} = - \left(z - \gamma \int \frac{\lambda dH(\lambda)}{1 + \lambda \tilde{m}(z)} \right)^{-1}. \quad (2.14)$$

And we use this to find a best fit $dH(\lambda)$.

The cut-off for a “distant spike” is essentially tautological due to (2.12). A point p satisfies $\psi'(p) > 0$ iff

$$1 = \psi'(p) \frac{\tilde{m}'(s)}{\tilde{m}^2(s)}, \quad s = \psi(p). \quad (2.15)$$

which coincides with the right spectral edge of the limiting measure generating \tilde{m} as this is the point where \tilde{m}' becomes infinite.

Let e_k denote the sample eigenvector E_k the population eigenvector. Consider for vectors u, v the quadratic form,

$$u^T e_k e_k^T v. \quad (2.16)$$

Let us find a formula for this quantity. As usual, we relate it to the resolvent via contour integration,

$$u^T e_k e_k^T v = -\frac{1}{2\pi i} \int_{\Gamma} u^T (N^{-1}X^*X - z)^{-1} v dz \quad (2.17)$$

for a contour encircling only the sample eigenvalue s_k . We write each vector $u = (u_1^T, u_2^T)^T$ for vectors u_1 and u_2 of dimensions d and $M - d$ respectively, and the same for v, v_1, v_2 . Define the Schur complement of $X^*X - z$ with respect to $X_1^*X_1 - z$ by

$$S := \frac{1}{N}X_1^*X_1 - z - N^{-1}X_1^*X_2(N^{-1}X_2^*X_2 - z)^{-1}N^{-1}X_2^*X_1. \quad (2.18)$$

This matrix S is $d \times d$ and by the law of large numbers converges to,

$$S \rightarrow A - z - A\gamma(1 + zm(z)), \quad (2.19)$$

very quickly. By the Schur complement formula, with $(N^{-1}X_2^*X_2 - z) := R_2(z)$ we have,

$$(N^{-1}X^*X - z)^{-1} = \begin{pmatrix} S^{-1} & -S^{-1}N^{-1}X_1^*X_2R_2 \\ -R_2X_2^*X_1S^{-1} & R_2 + R_2N^{-1}X_2^*X_1S^{-1}N^{-1}X_1^*X_2R_2 \end{pmatrix} \quad (2.20)$$

In order to analyze it, we need a **key assumption** that within the contour integral, $N^{-1}X_2^*X_2$ has no eigenvalues.

We now look at each component. For the function $S(z)^{-1}$, it converges to the matrix,

$$\sum_{k=1}^d E_k \frac{1}{p_k(1 - \gamma(1 + zm(z))) - z} E_k^T. \quad (2.21)$$

Under the assumption that the contour Γ encircles only a single population eigenvalue under the map $z \rightarrow z/(1 - \gamma(1 + zm(z))) = -1/\tilde{m}(z)$ the contour integral equals,

$$\left(\frac{d}{dz} (z - p_k(1 - \gamma(1 + zm(z)))) \Big|_{z=s_k} \right)^{-1} = (1 + \gamma p(m(s) + sm'(s)))^{-1} \quad (2.22)$$

$$= (1 + p(\tilde{m} + s\tilde{m}'))^{-1} \quad (2.23)$$

$$= -\frac{\tilde{m}(s_k)}{s_k \tilde{m}'(s)} = \frac{p_k \psi'(p_k)}{\psi(p_k)} \quad (2.24)$$

where we used $\tilde{m}(z) = (1 - \gamma)/(-z) + \gamma m(z)$. Hence, the contribution from the top left block is,

$$u_1^T e_k e_k^T v_1 \approx \left(-\frac{\tilde{m}(s_k)}{s_k \tilde{m}'(s_k)} \right) u_1^T E_k E_k^T v_k. \quad (2.25)$$

The contributions from the other blocks are smaller. First, top right block. Here, the relevant vector to inspect is,

$$N^{-1} X_1^* (X_2 R_2 v). \quad (2.26)$$

Each coordinate of this d -dimensional vector is, conditionally on X_2 , mean 0 and asymptotically Gaussian distributioend with variance,

$$\frac{1}{N^2} v_2^T R_2^* X_2^* X_2 R_2 v_2 = N v_2^T R_2^* v_2 + z N^{-2} v_2^T R_2^* R_2 v_2 = \mathcal{O}(N^{-1}). \quad (2.27)$$

So the top right block is $\mathcal{O}(N^{-1/2})$. Same for contribution from bottom left block. For the bottom right, since R_2 is analytic inside Γ , the term

$$v_2^T R_2 u_2 \quad (2.28)$$

vanishes when the contour integral is taken. For the other term, the same argument as above means it is $\mathcal{O}(N^{-1})$.

3 Model from Bloemendal-Knowles-Yau-Yin

We look at the matrix,

$$Q = T X X^* T^* \quad (3.1)$$

where T is $M \times (M+1)$ and of the form $(\mathbf{1}, \lambda v)$ and X is $(M+1) \times N$ and of the form $(Z^T, u^T)^T$ where u is in \mathbb{R}^N and Z is a usual noise matrix, of size $M \times N$.

So note that $TX = Z + \lambda v u^*$, and so $\mathbb{E}_Z[Q] = \mathbf{1} + \lambda^2 \|u\|_2^2 v v^*$. The normalization is then that $\mathbb{E}[(Z_{ij})^2] = N^{-1}$. We also assume $\|u\|_2^2 = 1$. We'll assume $\|v\|_2^2 = 1$.

Now we follow Section 8 of [BKYY2]. We want to find the SVD of T . Since $TT^* = \mathbf{1} + \lambda^2 v v^T$ we see that if,

$$T = O'(\Lambda, 0)O'' \quad (3.2)$$

for O' an orthogonal $M \times M$ and O'' an orthogonal $(M+1) \times (M+1)$ then Λ is diagonal all ones except the first entry is $\sqrt{1 + \lambda^2}$ and O' has first column v then the rest are anything obtained from Gram-Schmidt to extend v to an orthonormal basis of \mathbb{R}^M . Denote it v_2, \dots, v_M .

To find O'' we use that $T^* v_i$ gives eigenvectors of $T^* T$. So the first row of O'' is $(v^*, \lambda)/\sqrt{1 + \lambda^2}$. Then the rest of the rows are just $(v_i^*, 0)$.

Section 8 introduces,

$$\Sigma^{1/2} = O' \Lambda (O')^*, \quad O := \begin{pmatrix} O' & 0 \\ 0 & 1 \end{pmatrix} O''. \quad (3.3)$$

Its easy to see that $\Sigma = \mathbf{1} + \lambda^2 v v^*$. Following Section 8, we have $Q = \Sigma^{1/2} H \Sigma^{1/2}$ where $H = Y Y^*$ and $Y = (\mathbf{1}_M, 0) O X$. From (3.34) of [BKYY2] the main quantity we need to control is then,

$$v^* \frac{1}{H - z} v. \quad (3.4)$$

From Section 8, it looks like this quantity then goes to

$$(v^*, 0) \frac{1}{O X X^* O^* - z} (v^*, 0)^* = [O^*(v^*, 0)^*]^* \frac{1}{X X^* - z} O^*(v^*, 0)^* \quad (3.5)$$

The vector $O^*(v^*, 0)^*$ is $(v^*, \lambda)^*/\sqrt{1 + \lambda^2}$.