Rank 1 real Wishart spiked model

M. Mo: arXiv:1011.5404

- Real Wishart spiked models
- Contour integral formula for j.p.d.f.
- Asymptotic analysis and phase transition

Real Wishart spiked models

- Real Wishart matrix S in $W_{\mathbb{R}}(\Sigma, M)$:

1. X: $N \times M$ $(M \geq N)$ and columns of X are i.i.d. N-variate normal variables with zero mean.

2. Σ: covariance matrix $\Sigma_{ij} = E(X_{i1}X_{j1})$.

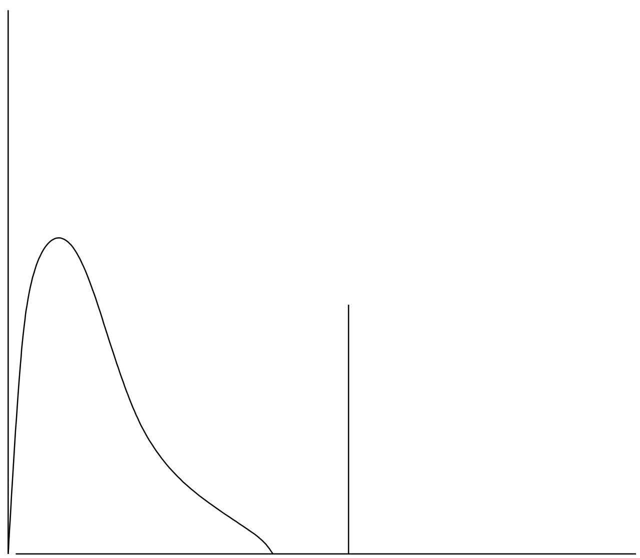
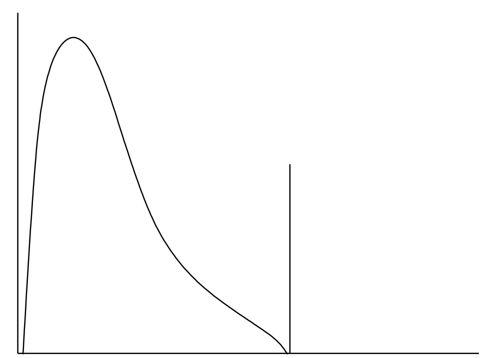
- Wishart matrix:

$$S = \frac{1}{M} X X^T,$$

- Think of each column of X as a sample from a N-variate normal variables with zero mean. Then S is the sample covariance matrix.

• Good model for sample covariance matrix in many situations, e.g. finance, climate data, wireless communications (complex), genetic data.

• In many cases, when both N, M are large, eigenvalues of sample covariance matrix matches very well with Marchenko Pastur law, but there are often large eigenvalues outside of the bulk.

- Spiked model: All but finitely many eigenvalues of Σ are 1. The finitely many non-trivial eigenvalues account for the spikes. Rank 1: 1 non-trivial eigenvalue.

- Phase transition: When the first spike is leaving the bulk.

Previous results

• For complex and quartonionic Wishart matrices, the phase transition in rank 1 spiked model was studied by Baik, Ben-Arous and Péché (complex) and Wang (quartonionic).

• Let $1 + \tau$ be the non-trivial eigenvalue and $\gamma = M/N$. Then for $-1 < \tau < \gamma^{-1}$, the largest eigenvalue distribution are same as the ones with $\Sigma = I$. i.e. Tracy-Widom distribution for the respective symmetry.

• Phase transition occurs at $\tau = \gamma^{-1}$. Largest eigenvalue distribution: Tracy-Widom GOE for both the complex and quartonionic case.

$$P\left(\left(\lambda_{max} - \left(\frac{\gamma+1}{\gamma}\right)^2\right)\frac{\gamma(2M)^{\frac{2}{3}}}{(1+\gamma)^{\frac{4}{3}}} < \zeta\right) = GOE(\zeta)$$

- Real case is more complicated. Very recently, Bloemendal and Virág characterized the distribution at the phase transition by a boundary value problem. They use stochastic operator method that is very different from ours.

- Our work uses orthogonal polynomial method to find the distribution at the phase transition. The result takes the following form

$$P\left(\left(\lambda_{max}-\left(\frac{\gamma+1}{\gamma}\right)^2\right)\frac{\gamma(2M)^{\frac{2}{3}}}{(1+\gamma)^{\frac{4}{3}}}<\zeta\right)=F(\zeta),$$

$F(\zeta)$ is of the form

$$F(\zeta)=\sqrt{GUE(\zeta)}\int_{\Gamma}F_1(T)\sqrt{\det\left(\delta_{ij}-(\alpha_i,\beta_j)\right)_{1\leq i,j\leq 3}}dT$$

$F_1(T)$: integrals of the Airy function. (α_i,β_j): expressed in terms of Painlevé II solution.

- For example, the first column can be written as

$$(\alpha_1, \beta_1) = \mathcal{R}_+ - \mathcal{R}_0, \quad (\alpha_1, \beta_2) = \mathcal{P}_{-,0}, \quad (\alpha_1, \beta_3) = \mathcal{P}_{-,1}.$$

and these functions satisfy the following system of first order linear ODEs

$$\frac{\partial \mathcal{R}_0}{\partial \zeta} = \frac{\partial \log\left(\phi_0 \psi^{-1}\right)}{\partial \zeta} \mathcal{R}_0 - \frac{\tilde{\tau}\sigma(\zeta)}{\psi(\zeta)\phi_0} \mathcal{P}_{-,0},$$

$$\frac{\partial \mathcal{R}_+}{\partial \zeta} = \frac{\tilde{\tau}}{2\psi^2} \mathcal{R}_0 - \frac{\phi_0}{\psi(\zeta)} \mathcal{P}_{+,0}, \quad \frac{\partial \mathcal{P}_{\pm,0}}{\partial \zeta} = \phi_0 \psi^{\pm}(\zeta)\left(1 - \mathcal{R}_{\pm}\right),$$

while $\mathcal{P}_{-,1}$ can be computed from $\mathcal{R}_0$.

$$\mathcal{P}_{-,1} = \tilde{\tau}^{-1} W\left(\mathcal{R}_0, \phi_1 \psi^{-1}\right) / W\left(\phi_0 \psi^{-1}, \phi_1 \psi^{-1}\right)$$

W: Wronskian.

• The functions ϕ_0, σ, ϕ_1 and ψ are all known, ϕ_0 σ and ϕ_1 are expressible in terms of the Hastings-McLeod solution to Painlevé II, and ψ is

$$\psi = (T - \tilde{\tau}\zeta)^{\frac{1}{2}}, \quad \tilde{\tau} = \frac{\tau}{2(1+\tau)},$$

$$\phi_0''(\zeta) = \zeta\phi_0(\zeta) + 2\phi_0^3(\zeta),$$

$$\phi_0(\zeta) \sim -\mathrm{Ai}(\zeta), \quad \zeta \to +\infty,$$

$$\sigma(\zeta) = \int_\zeta^\infty \phi_0^2(u)du$$

Contour integral formula for j.p.d.f.

- Difficulty in finding the j.p.d.f.

$$P(\lambda) = \frac{1}{Z_{M,N}} |\Delta(\lambda)| \prod_{j=1}^{N} \lambda_j^{\frac{M-N-1}{2}} \int_{O(N)} e^{-\frac{M}{2}\mathrm{Tr}(\Sigma^{-1} g S g^{-1})} g^T dg,$$

If Σ has only one non-trivial eigenvalue $1+\tau$, then the integral is given by

$$\int_{\Gamma} e^{Mt} \prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \left(t - \frac{\tau}{2(1+\tau)} \lambda_j \right)^{-\frac{1}{2}} dt$$

This is similar to a formula by Bergére and Eynard

$$\int_{O(N)} e^{-\mathrm{Tr}\left(X g Y g^{-1}\right)} g^T dg \propto \int \frac{e^{\mathrm{Tr}(S)}}{\prod_{j=1}^{N} \det(S - y_j X)} dS$$

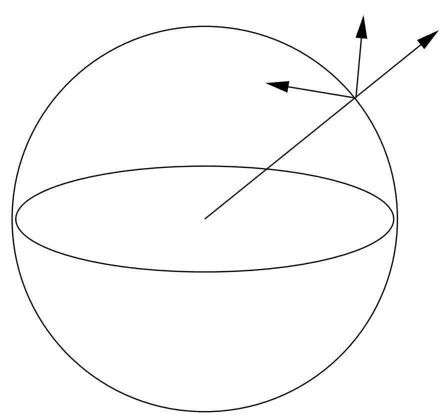
over i times real symmetric matrices.

- First, a simple computation shows

$$\int_{O(N)} e^{-\mathrm{Tr}\left(XgYg^{-1}\right)} g^T dg \propto \prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \int_{SO(N)} e^{\frac{\tau M}{2(1+\tau)} \sum_{j=1}^{N} \lambda_j g_{jN}^2} g^T dg$$

where g_{jN} is the last column of g.

- We can think of $SO(N)$ as the space of orthonormal frames.



• Can identify the last column of g as a point in S^{N-1} and the rest of the columns as a point in $SO(N-1)$. E.g. if $Gg_N = (0, 0, \ldots, 1)$, then Gg is of the form

$$Gg = \begin{pmatrix} V & 0 \\ 0 & 1 \end{pmatrix}$$

$V \in SO(N-1)$. V and g_N can then be used as coordinates of $SO(N)$.

• The group action on g then decompose into an action on S^{N-1} and an action on $SO(N-1)$.

$$G : g_N \mapsto Gg_N, \quad G : V \mapsto \tilde{G}V$$

where $\tilde{G}$ depends on g_N. The product of the volume form on S^{N-1} and the Haar measure on V is then invariant under the $SO(N)$ action and hence the Haar measure can be written as

$$g^T dg = dg_N \wedge V^T dV$$

- We can use this to compute the integral

$$\int_{O(N)} e^{-\mathrm{Tr}\left(XgYg^{-1}\right)} g^T dg \;\propto\; \prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \int_{SO(N)} e^{\frac{\tau M}{2(1+\tau)}\sum_{j=1}^{N}\lambda_j g_{jN}^2} g^T dg$$

$$\propto\; \prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \int_{S^{N-1}} e^{\frac{\tau M}{2(1+\tau)}\sum_{j=1}^{N}\lambda_j g_{jN}^2} dg_N$$

dg_N: volume form in S^{N-1}.

- Can be simplified by treating it as a measure on $\mathbb{R}^N$, but with a delta function to restrict it on S^{N-1}.

$$\prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \int_{\mathbb{R}^N} e^{\frac{\tau M}{2(1+\tau)}\sum_{j=1}^{N}\lambda_j x_j^2} \delta\left(\sum_{j=1}^{N} x_j^2 - 1\right) dX$$

- To evaluate it, consider its Laplace transform.

$$I(\Sigma,\Lambda,t) \propto \prod_{j=1}^{N} e^{-\frac{M}{2}\lambda_j} \int_{\mathbb{R}^N} e^{\frac{\tau M}{2(1+\tau)}\sum_{j=1}^{N}\lambda_j x_j^2} \delta\left(\sum_{j=1}^{N} x_j^2 - t\right) dX,$$

then

$$\int_0^\infty e^{-st} I(\Sigma, \Lambda, t) dt \propto \prod_{j=1}^N e^{-\frac{M}{2}\lambda_j} \int_{\mathbb{R}^N} e^{\sum_{j=1}^N \left(-s+\frac{\tau M}{2(1+\tau)}\lambda_j\right) x_j^2} dX$$

Can be evaluated as

$$\int_0^\infty e^{-st} I(\Sigma, \Lambda, t) dt \propto \prod_{j=1}^N e^{-\frac{M}{2}\lambda_j} \left(s - \frac{\tau M}{2(1+\tau)}\lambda_j\right)^{-\frac{1}{2}}$$

- Taking the inverse Laplace transform, we obtain the contour integral formula

$$I(\Sigma, \Lambda) \propto \int_\Gamma e^{Mt} \prod_{j=1}^N e^{-\frac{M}{2}\lambda_j} \left(t - \frac{\tau}{2(1+\tau)}\lambda_j\right)^{-\frac{1}{2}} dt,$$

- To derive the formula, we first decompose the Haar measure into two parts, integrate out the part we do not need and then use the Laplace transform to 'flatten' to measure on S^{N-1}.

- The idea is similar to Bergére and Eynard, in which the whole Haar measure is 'flatten' by an integral transform to obtain an integral formula over the space of $N \times N$ symmetric matrices.

$$\int_{O(N)} e^{-\mathrm{Tr}\left(XgYg^{-1}\right)} g^T dg \propto \int \frac{e^{\mathrm{Tr}(S)}}{\prod_{j=1}^{N} \det(S - y_j X)} dS$$

Asymptotic analysis

- The contour integral expression reduces the problem to the analysis of the orthogonal ensemble with weight w

$$w(x) = e^{-\frac{M}{2}x} x^{\frac{M-N-1}{2}} \left(t - \frac{\tau}{2(1+\tau)} x \right)^{-\frac{1}{2}}.$$

- In fact, we have

$$\mathbb{P}(\lambda_{max} < z) = \int_{\lambda_1 \le \ldots \le \lambda_N \le z} \cdots \int P(\lambda) d\lambda_1 \ldots d\lambda_N,$$

$$= \tilde{Z}_{M,N}^{-1} \int_{\Gamma} e^{Mt} \int_{\lambda_1 \le \ldots \le \lambda_N \le z} \cdots \int |\Delta(\lambda)| \prod_{j=1}^{N} w(\lambda_j) d\lambda_1 \ldots d\lambda_N dt$$

- This can be written as the integral of Fredholm determinant.

$$\mathbb{P}(\lambda_{max} < z),$$
$$= \tilde{Z}_{M,N}^{-1} \int_{\Gamma} e^{Mt} \sqrt{\det \mathcal{U}} \sqrt{\det(I - \chi_{[z,\infty)} K \chi_{[z,\infty)})} dt$$

$\mathcal{U}$ is the moment matrix with entries

$$\frac{1}{2}\int_0^\infty \int_0^\infty r_j(x)\text{sgn}(x-y)r_k(y)w(x)w(y)dxdy,$$

r_k: monic polynomials of degree k. The operator K has matrix kernel

$$K(x,y) = \begin{pmatrix} S_1(x,y) & -\frac{\partial}{\partial y}S_1(x,y) \\ IS_1(x,y) & S_1(y,x) \end{pmatrix}$$

and S_1 is the kernel

$$\begin{aligned} S_1(x,y) &= -\textstyle\sum_{j,k=0}^{N-1} r_j(x)w(x)\mu_{jk}\epsilon(r_k w)(y), \\ IS_1(x,y) &= -\textstyle\sum_{j,k=0}^{N-1} \epsilon(r_j w)(x)\mu_{jk}\epsilon(r_k w)(y) - \frac{1}{2}\text{sgn}(x-y) \end{aligned}$$

and $\epsilon(f) = (1/2)\int_0^\infty \text{sgn}(x-y)f(y)dy$.

- The kernel S_1 can be written in terms of Laguerre polynomials.

$$S_1 = K_1 + K_2,$$

$$K_2 = \left(\frac{y(t-\tilde{\tau}y)}{x(t-\tilde{\tau}x)}\right)^{\frac{1}{2}} w_0^{\frac{1}{2}}(x) w_0^{\frac{1}{2}}(y) \frac{L_N(x)L_{N-1}(y) - L_N(y)L_{N-1}(x)}{h_{N-1,0}(x-y)},$$

$$K_1 = \epsilon \begin{pmatrix} \pi_{N+1,1}w & \pi_{N,1}w \end{pmatrix}(y) \begin{pmatrix} 0 & -\frac{M\tilde{\tau}}{2h_{N-1,0}} \\ -\frac{M\tilde{\tau}}{2h_{N-2,0}} & \frac{Mt-\tilde{\tau}(M+N)}{2h_{N-1,0}} \end{pmatrix} \begin{pmatrix} L_{N-2}(x) \\ L_{N-1}(x) \end{pmatrix} w$$

L_N: monic Laguerre polynomials orthogonal to the weight $x^{M-N}e^{-Mx}$.

- Asymptotics of S_1 can be computed using the asymptotics of the Laguerre polynomials. The moment matrix is also related to S_1.

$$\frac{\partial}{\partial t} \log \det \mathcal{U} = -\int_{\mathbb{R}_+} \frac{S_1(x,x)}{t - \tilde{\tau} x} dx,$$

- This gives us the following representation for the largest eigenvalue distribution.

$$\mathbb{P}(\lambda_{max} < z) \propto \int_{\Gamma} \exp\left(Mt - \frac{1}{2}\int_{c_0}^{t}\int_{\mathbb{R}_+} \frac{S_1(x,x)}{s - \tilde{\tau} x} dx ds\right) \sqrt{\det\left(I - K\chi_{[z,\infty)}\right)} dt$$

Phase transition

- In the large N limit, both the Fredholm determinant and the integral

$$\int_{c_0}^{t}\int_{\mathbb{R}_+}\frac{K_1(x,x)}{s-\tilde{\tau}x}dxds$$

remains finite. The t integral in the largest eigenvalue distribution can therefore be computed using saddle point analysis for

$$Mt-\int_{c_0}^{t}\int_{\mathbb{R}_+}\frac{K_2(x,x)}{s-\tilde{\tau}x}dxds$$

Note that $K_2(x,x)$ is the same as the CD kernel for the Laguerre polynomials.

$$K_2(x,x)=w_0^{\frac{1}{2}}(x)w_0^{\frac{1}{2}}(x)\frac{L_N'(x)L_{N-1}(x)-L_N(x)L_{N-1}'(x)}{h_{N-1,0}}$$

and we have

$$K_2(x,x) \sim N\rho, \quad \int_{c_0}^{t}\int_{\mathbb{R}_+} \frac{K_2(x,x)}{s-\tilde{\tau}x}dxds \sim N\int_{\mathbb{R}_+}\rho\log(t-\tilde{\tau}x)dx.$$

- Saddle point analysis gives (e.g. for $\gamma = 1$)

$$1-\frac{1}{4\tilde{\tau}}\left(1-\sqrt{\frac{t-4\tilde{\tau}}{t}}\right)=0.$$

Then the saddle point is at

$$\frac{1}{2-4\tilde{\tau}}, \quad \tilde{\tau}=\frac{\tau}{2(1+\tau)}, \quad -\infty<\tilde{\tau}<\frac{1}{2}$$

and

$$\frac{t}{\tilde{\tau}}=\frac{1}{(2-4\tilde{\tau})\tilde{\tau}}.$$

- The function $t/\tilde{\tau}$ decreases from 0 to $-\infty$ for $\tilde{\tau}<0$ and is greater than or equal to 4 for $\tilde{\tau}>0$.

- When $\tilde{\tau} = 1/4$, $\tau = 1$ we have $t/\tilde{\tau} = 4$ and the saddle point coincide with the edge point of spectrum. This gives a phase transition.

- When $\tilde{\tau} \to \infty$, $t/\tilde{\tau} \to 0$, $\Rightarrow$ different behavior for smallest eigenvalue.

- When $\tau \neq 1$, the effect of the factor $\left(t - \frac{\tau}{2(1+\tau)}x\right)^{-\frac{1}{2}}$ in the weight

$$w(x) = e^{-\frac{M}{2}x} x^{\frac{M-N-1}{2}} \left(t - \frac{\tau}{2(1+\tau)}x\right)^{-\frac{1}{2}}.$$

is small near the edge point.

$$\begin{aligned} K_2 &\to \frac{Ai(\xi_1)Ai'(\xi_2) - Ai'(\xi_1)Ai(\xi_2)}{\xi_1 - \xi_2}, \\ K_1 &\to \frac{1}{2}Ai(\xi) \int_{-\infty}^{\xi_2} Ai(u)du \end{aligned}$$

gives Tracy-Widom distribution.

• When $\tau = 1$, the singularity $t/\tilde{\tau}$ in the weight appears near the edge point and the kernel at the edge point changes significantly.

$$K_1 \to \left(\frac{T}{2} H_0(\xi_1) \int_{-\infty}^{\xi_2} H_0 du + \frac{\tilde{\tau}}{2} H_1(\xi_1) \int_{-\infty}^{\xi_2} H_1(u) du \right.$$
$$+ B \left(H_1(\xi_1) \int_{-\infty}^{\xi_2} H_0(u) du - H_0(\xi_1) \int_{-\infty}^{\xi_2} H_1(u) du \right)$$
$$\left. - \tilde{\tau} H_0(\xi_1) \int_{-\infty}^{\xi_2} H_2(u) du + H_0(\xi_1) C \right).$$

$$H_j(u) = \frac{Ai^{(j)}(u)}{(T - \tilde{\tau} u)^{\frac{1}{2}}}, \qquad B = -\frac{\tilde{\tau} \int_{-\infty}^{\infty} H_1(u) du + \sqrt{\tilde{\tau}}}{2 \int_{-\infty}^{\infty} H_0(u) du},$$
$$C = -\frac{B}{\sqrt{t}} - \frac{T}{2} \int_{-\infty}^{\infty} H_0 du + B \int_{-\infty}^{\infty} H_1 du + \tilde{\tau} \int_{-\infty}^{\infty} H_2 du,$$

$$K_2 \to \left(\frac{T-\tilde{\tau}\xi_2}{T-\tilde{\tau}\xi_1}\right)^{\frac{1}{2}} \frac{Ai(\xi_1)Ai'(\xi_2)-Ai'(\xi_1)Ai(\xi_2)}{\xi_1-\xi_2}$$

- Contribution of each term to asymptotics:

$$\det(I-\chi K\chi) = \det((I-K_2\chi)^{-1})\det(I-K_c)$$

first term gives Tracy-Widom GUE independent on T. Second term is a determinant of 3×3 matrix.

$$\det\mathcal{U} = F_1^2(T)$$

This gives

$$F(\zeta) = \sqrt{GUE(\zeta)}\int_\Gamma F_1(T)\sqrt{\det\left(\delta_{ij}-(\alpha_i,\beta_j)\right)_{1\le i,j\le 3}}dT$$