# Self consistent equations (Dyson Equation, Silverstein Equation) and limiting spectrums

Binxu Wang

July 1, 2025

Kempner Institute, Harvard University

## Contents

# 1 Motivation

Why are we interested in *Random Matrix Theory* (RMT)? The eigenvalue and eigenvector of large matrices are so prevalent and critical in many domains.

- Physics.
  - In Quantum Mechanics, the Harmiltonian is a linear Hermitian operator $H$. For a given system, the energy levels correspond to the eigenvalues of $H$.
  - Historically, Wigner developed RMT to understand energy levels of large atom nucleus.
- **Statistics**,
  - **Covariance of data**. $\Sigma = \frac{1}{n}X^\top X$, esp. when the data have high dimensional features.
  - **Kernel matrix**. The kernel matrix $K$, $K_{ij} = \Phi(\mathbf{x}_i, \mathbf{x}_j)$ (or representation similarity matrix), esp. for a large number of data points.
  - **Linear regression,** the solution reads $\Sigma_{xx}^{-1}\Sigma_{xy}$, (Ridge regression reads $(\Sigma_{xx} + \lambda I)^{-1}\Sigma_{xy}$).
    * Under gradient learning, the convergence speed of linear regression is related to the eigenvalues of $\Sigma_{xx}$, faster convergence on higher variance eigenmodes.
    * The generalization error of linear regression is also determined by $\Sigma_{xx}$ structure and its alignment with $\Sigma_{xy}$. \cite
  - **Kernel matrix and regression**. , kernel regression
- **Machine learning,**
  - **Linear RNN** has structure $\dot{x} = -x + Ax$, the stability of the system is determined by the eigenvalues of $A$, esp. real part $< 1$.
  - **Gradient descent, or gradient flow dynamics** is $\dot{w} = -\eta\nabla_w\mathcal{L}$. For quadratic loss, the learning dynamics $\dot{w} = -\eta H w$ is determined by the Hessian of loss function, it will converge faster on dimensions with higher eigenvalues
  - **Denoising**. Denoising of white Gaussian noise under Gaussian prior $\mathcal{N}(\mu, \Sigma)$ is $\mathbf{D}(\mathbf{x}) = (\Sigma + \sigma^2 I)^{-1}\Sigma\,(\mathbf{x} - \mu) + \mu$.
  - **Jacobian of neural network and Dynamical isometry.**
- **Graph theory**,
  - for a large graph, the connectivity matrix $A$, and the Laplacian $L = D - A$, the eigenvectors of top eigenmodes are used in spectral clustering.

# 2 Analysis Tools

## 2.1 Resolvent

For a symmetric / Hermitian matrix $A$, we define the resolvent as

$$G_A(z) = (A - zI)^{-1}$$

Dual perspective on matrix resolvent,

- Resolvent itself is a matrix or a linear operator on a vector space. It shifts the operator $A$ by $z$, and applied the inverse.
- Resolvent is also a matrix-valued complex function $\mathbb{C} \to \mathbb{R}^{d\times d}$ (or $\mathbb{C} \to \mathbb{C}^{d\times d}$ for Hermitian).
  - This function is analytical everywhere, except at the poles $\mathbb{C}\backslash\{\sigma(A)\}$. So we can consider its complex derivative.

**Spectral decomposition** If the matrix $A$ has eigendecomposition as $A = V\Lambda V^\top$, then the resolvent has the decomposition

$$\begin{aligned} G_A(z) &= V(\Lambda - zI)^{-1}V^\top \\ &= \sum_{k=1}^{d} \frac{1}{\lambda_k - z}\mathbf{v}_k\mathbf{v}_k^\top \end{aligned}$$

**Remark:**

- The poles are simple poles!
- Behavior of resolvent around each pole is projecting input onto *the whole eigenspace* of that eigenvalue. $\lim_{z\to\lambda} G_A(z) = \frac{1}{\lambda - z}\Pi_k$

**Entry structure of resolvent** Consider individual entries

$$[G(z)]_{ii} = \sum_{k=1}^{d} \frac{(\mathbf{v}_k[i])^2}{\lambda_k - z}$$

the summation trace is

$$\frac{1}{d}\text{Tr}[G_A(z)] = \frac{1}{d}\sum_{k=1}^{d}\frac{1}{\lambda_k - z}$$

Let $z = s + i\eta$,

$$\begin{aligned} \frac{1}{\lambda_k - z} &= \frac{1}{\lambda_k - s - i\eta} \\ &= \frac{\lambda_k - s + i\eta}{(\lambda_k - s)^2 + \eta^2} \end{aligned}$$

So the imaginary part encodes the smoothed density,

$$\begin{aligned} \text{Im}\Big[\frac{1}{d}\text{Tr}[G_A(s + i\eta)]\Big] &= \text{Im}\Big[\frac{1}{d}\sum_{k=1}^{d}\frac{1}{\lambda_k - z}\Big] \\ &= \frac{1}{d}\sum_{k=1}^{d}\frac{\eta}{(\lambda_k - s)^2 + \eta^2} \end{aligned}$$

Other linear algebraic property

$$\begin{aligned} A(A - zI)^{-1} &= AG_A(z) \\ &= (A - zI)G_A(z) + zG_A(z) \\ &= I + zG_A(z) \end{aligned}$$

$$A^2(A - zI)^{-2} = (I + zG_A(z))^2$$

## 2.2 Stieltjes transform

Consider the Stieltjes transform of a real symmetric matrix $A$, i.e. trace of resolvent [1]

$$\begin{aligned} m_A(z) &= \frac{1}{d}\text{Tr}[(A - zI)^{-1}] \\ &= \frac{1}{d}\text{Tr}[G_A(z)] \end{aligned}$$

Some properties

[1] Note our definition here is the same as the convention in [4], but differ by a sign from $g_A$ in [1]. In some paper it's defined as $\frac{1}{d}\text{Tr}[(A+\lambda I)^{-1}]$ so make sure to match the definition before applying the formula.

- Stieltjes transform itself is a complex function $\mathbb{C} \to \mathbb{C}$. It's analytical except for the eigenvalues or poles.
  - Note the Stieltjes transform $m_A(z)$ maps upper plane $\mathbb{C}_+ = \{s + i\eta \mid \eta > 0\}$ to upper plane $\mathbb{C}_+$. This property is useful to determine the correct solution branch.
- In our convention, at infinity, the Stieltjes transform scales as $\lim_{z\to\infty} m_A(z) \sim -\frac{1}{z}$. This is also useful to determine the correct branch.
- Stieltjes transform encodes the empirical eigenvalue distribution of $A$. We can consider its eigenvalues, this transform is encoding the distribution

$$\begin{aligned}
m_A(z) &= \frac{1}{d}\sum_k \frac{1}{\lambda_k - z} \\
&= \int dF(s)\frac{1}{s-z}
\end{aligned}$$

where for a fixed matrix, this measure is $F(s) = \frac{1}{d}\sum_k \delta(s - \lambda_k)$.

  - The imaginary part of Stieltjes encodes the smoothed version of eigenvalue density, where the resolution depends on $\eta$.

$$\begin{aligned}
&\mathrm{Im}\Big[m_A(z)\Big] \\
&= \mathrm{Im}\Big[m_A(s+i\eta)\Big] \\
&= \mathrm{Im}\Big[\frac{1}{d}\mathrm{Tr}[G_A(s+i\eta)]\Big] \\
&= \frac{1}{d}\sum_{k=1}^{d}\frac{\eta}{(\lambda_k - s)^2 + \eta^2}
\end{aligned}$$

- The conjugate of Stieltjes transform $\bar{m_A}(u+iv)$ can be seen as a *gradient vector field* of a *logrithmic potential*, which is pointing towards the eigenvalues as "point charges". Note this is non analytical complex function.

$$\begin{aligned}
m_A(u+iv) &= \frac{1}{d}\sum_{k=1}^{d}\frac{1}{\lambda_k - u - iv} \\
&= \frac{1}{d}\sum_{k=1}^{d}\frac{\lambda_k - u + iv}{(\lambda_k - u)^2 + v^2} \\
\bar{m_A}(u+iv) &= \frac{1}{d}\sum_{k=1}^{d}\frac{\lambda_k - u - iv}{(\lambda_k - u)^2 + v^2} \\
&= \frac{1}{d}\sum_{k=1}^{d}\frac{\lambda_k - z}{|\lambda_k - z|^2} \\
&= -\frac{\partial}{\partial u}\Phi(u,v) - i\frac{\partial}{\partial v}\Phi(u,v)
\end{aligned}$$

$$\begin{aligned}
\Phi(u,v) &= \frac{1}{d}\sum_{k=1}^{d}\log|\lambda_k - u - iv| \\
&= \frac{1}{d}\sum_{k=1}^{d}\log\sqrt{(\lambda_k - u)^2 + v^2}
\end{aligned}$$

  - The logarithmic potential $\Phi(u,v)$ satisfies the Poisson equation on 2d plane! so it can be regarded as a form of electric / Columb potential.
  - Moving along the conjugate Stieltjes as vector field, the gradient lines results in the following equation, which will converge towards the point charges or the eigenvalues.

$$\frac{d}{d\tau}z = \bar{m_A}(z)$$

**Columb gas perspective** Consider a density or measure $\mu$ on 1d space, two particles have their energy potential $\log|x-y|$, which is a form of Columb potential in 1d; and single particle has potential $V(x)$,

$$I[\mu] = \iint_{\mathbb{R}^2} \log|x-y| \; d\mu(x)d\mu(y) + \int_{\mathbb{R}} V(x)d\mu(x)$$

Differentiating the variational functional yields,

$$\int \frac{d\mu(t)}{x-t} = -\frac{1}{2}V'(x)$$

on the support of $\mu$, $x \in S$ .
This equation has clear connection to the Stieltjes transform and self consistent equation of random matrix, e.g. quadratic potential is linked to Wigner ensemble.

$$\text{P.V.} \int_a^b \frac{\rho(t)}{x-t}\, dt = \frac{x}{2}, \quad \rho(t) = \frac{1}{2\pi}\sqrt{4-t^2}, \quad S = [-2, 2].$$

# 3 Self consistent equations of resolvent

## 3.1 Dyson or Pastur equation for Wigner matrix

Consider Wigner matrix $H$, $H_{ij} \sim \mathcal{N}(0, \frac{1}{d})$ for $i \leq j$ it's symmetric $H_{ij} = H_{ji}$. The scaling is chosen to ensure eigenvalues have expected value around 1.

**Derivation of self consistent equation** This derivation can be found in Song Mei's lectureLecture note (Sec. 3) and [3].

The resolvent is

$$G_H(z) = (H - zI)^{-1}$$

Consider one diagonal entry $[G_H(z)]_{11}$

Recall on Schur complement

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & \star \\ \star & \star \end{bmatrix}$$

In our case

$$H - zI = \begin{bmatrix} h_{11} - z & \mathbf{h}_{:1}^\top \\ \mathbf{h}_{:1} & H_{[1]} - zI_{d-1} \end{bmatrix}$$

$$G_H(z) = \begin{bmatrix} h_{11} - z & \mathbf{h}_{:1}^\top \\ \mathbf{h}_{:1} & H_{[1]} - zI_{d-1} \end{bmatrix}^{-1}$$

Then

$$\begin{aligned} [G_H(z)]_{11} &= (h_{11} - z - \mathbf{h}_{:1}^\top (H_{[1]} - zI_{d-1})^{-1}\mathbf{h}_{:1})^{-1} \\ &= \frac{1}{h_{11} - z - \mathbf{h}_{:1}^\top G_{H_{[1]}}(z)\mathbf{h}_{:1}} \end{aligned}$$

Now consider the trace part

$$\mathbf{h}_{:1}^\top G_{H_{[1]}}(z)\mathbf{h}_{:1} = \text{Tr}[\mathbf{h}_{:1}\mathbf{h}_{:1}^\top G_{H_{[1]}}(z)]$$

We could take expectation over the independent variables $\mathbf{h}_{:1}$

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{h}_{:1}}[\mathbf{h}_{:1}^\top G_{H_{[1]}}(z)\mathbf{h}_{:1}] \\
&= \mathbb{E}_{\mathbf{h}_{:1}}[\mathrm{Tr}[\mathbf{h}_{:1}\mathbf{h}_{:1}^\top G_{H_{[1]}}(z)]] \\
&= \frac{1}{d-1}\mathrm{Tr}[I_{d-1}G_{H_{[1]}}(z)] \\
&= \frac{1}{d-1}\mathrm{Tr}[G_{H_{[1]}}(z)] \\
&= m_{H_{[1]}}(z)
\end{aligned}
$$

Thus

$$
[G_H(z)]_{11} \approx \frac{1}{h_{11} - z - m_{H_{[1]}}(z)}
$$

$$
\begin{aligned}
\frac{1}{d}\sum_k [G_H(z)]_{kk} &\approx \frac{1}{d}\sum_k \frac{1}{h_{kk} - z - m_{H_{[k]}}(z)} \\
m_H(z) &\approx \frac{1}{d}\sum_k \frac{1}{h_{kk} - z - m_{H_{[k]}}(z)} \\
&\approx -\frac{1}{z + m_{H_{[k]}}(z)}
\end{aligned}
$$

More generally if we scale the entry by $\sigma$ we have

$$
\begin{aligned}
m_H(z) &\approx -\frac{1}{z + \sigma^2 m_{H_{[k]}}(z)} \\
0 &= \sigma^2 m_{sc}^2(z) + z\ m_{sc}(z) + 1
\end{aligned}
$$

In the large number limit $h_{kk} \sim \mathcal{N}(0, 1/d)$ which has negilible effect. We assume $\lim_{d\to\infty} m_H(z) = m_{sc}(z)$ , and also effect of adding one row small $m_{H_{[k]}}(z) \to m_H(z)$, then we expect at this limit

$$
m_{sc}(z) = -\frac{1}{z + m_{sc}(z)}
$$

We have the Stijies transform satisfying this quadratic equation

$$
m_{sc}^2(z) + z\ m_{sc}(z) + 1 = 0
$$

Solving the quadratic equation yields two solutions,

$$
m_{sc}(z) = \frac{-z \pm \sqrt{z^2 - 4}}{2}
$$

On the top half plane, we can choose the sign to ensure $z \to \infty$, $\frac{-z+\sqrt{z^2-4}}{2} = \frac{z}{2}(\sqrt{1 - \frac{4}{z^2}} - 1) \approx \frac{z}{2}(1 - \frac{2}{z^2} - 1) = \frac{1}{z}$ this is the correct scaling from the self consistent equation.

$$
m_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}
$$

**Derivation of spectral density from Stieltjes transform** Note a valid Stijies transform should be a mapping from $\mathbb{C}^+ \to \mathbb{C}^+$ so we should make is such that $\mathrm{Im}[-z \pm \sqrt{z^2 - 4}] > 0 \quad iff \quad \mathrm{Im}[z] > 0$

$$
\begin{aligned}
m_{sc}(s + i\eta) &= \frac{-s - i\eta + \sqrt{(s + i\eta)^2 - 4}}{2} \\
&= \frac{-s - i\eta + \sqrt{(s - \eta)^2 - 4 + 2is\eta}}{2}
\end{aligned}
$$

Taking the imaginary part we have

$$
\begin{aligned}
\mathrm{Im}[m_{sc}(s+i\eta)] &= \frac{1}{2}\Big(-\eta + \mathrm{Im}\sqrt{(s-\eta)^2-4+2is\eta}\Big) \\
&= \frac{1}{2}\Big(-\eta + \mathrm{sgn}(s\eta)\sqrt{\frac{\sqrt{((s-\eta)^2-4)^2+(2s\eta)^2}-(s-\eta)^2+4}{2}}\Big)
\end{aligned}
$$

using the following identity

$$
\Im\sqrt{u+iv} = \mathrm{sgn}(v)\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}.
$$

Now taking limit of $\eta$

$$
\begin{aligned}
\lim_{\eta\to 0^+}\mathrm{Im}[m_{sc}(s+i\eta)] &= \lim_{\eta\to 0^+}\frac{1}{2}\sqrt{\frac{\sqrt{((s-\eta)^2-4)^2+(2s\eta)^2}-(s-\eta)^2+4}{2}} \\
&= \frac{1}{2}\sqrt{\frac{\sqrt{(s^2-4)^2}-s^2+4}{2}}
\end{aligned}
$$

At least there are two cases is $|s|\geq 2$, $s^2-4\geq 0$ then $\sqrt{(s^2-4)^2}=s^2-4$ else $\sqrt{(s^2-4)^2}=4-s^2$

$$
\lim_{\eta\to 0^+}\mathrm{Im}[m_{sc}(s+i\eta)] = \begin{cases} \frac{1}{2}\sqrt{\frac{s^2-4-s^2+4}{2}} = 0 & |s|\geq 2 \\ \frac{1}{2}\sqrt{\frac{-s^2+4-s^2+4}{2}} = \frac{1}{2}\sqrt{4-s^2} & |s|<2 \end{cases}
$$

which yields the semicircle law.

$$
\rho(s) = \frac{1}{\pi}\lim_{\eta\to 0^+}\mathrm{Im}[m_{sc}(s+i\eta)] = \begin{cases} 0 & |s|\geq 2 \\ \frac{1}{2\pi}\sqrt{4-s^2} & |s|<2 \end{cases}
$$

### 3.2 Silverstein equation for Wishart-like matrix (classic MP law)

Consider a sample covariance matrix from $n$ samples in $p$ dimension, $\mathbf{x}_k \sim \mathcal{N}(0, I_p)$ samples are i.i.d.

$$
S_n = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^\top
$$

Consider the resolvent

$$
G_n(z) = (S_n - zI)^{-1}
$$

and the normalized trace

$$
m_n(z) = \frac{1}{p}\mathrm{Tr}[G_n(z)], \quad z\in\mathbb{C}^+
$$

Note $\lim_{n\to\infty} p/n = \gamma$

**Derivation of self consistent equation** Due to the rank one decomposition, we'd first decompose the resolvent

$$
\begin{aligned}
S_n &= \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^\top \\
S_n - zI + zI &= \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^\top \\
I + z(S_n - zI)^{-1} &= \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^\top(S_n - zI)^{-1}
\end{aligned}
$$

Next using the linear algebra identity for any $B, \mathbf{v}$ (c.f. Eq. 2.1 from [5]) also as a consequence of Woodbery identity.

$$
\begin{aligned}
\mathbf{v}^\top B^{-1}(B+\mathbf{v}\mathbf{v}^\top) &= v^\top + \mathbf{v}^\top B^{-1}\mathbf{v}\mathbf{v}^\top \\
&= (1+\mathbf{v}^\top B^{-1}\mathbf{v})\mathbf{v}^\top
\end{aligned}
$$

$$\frac{\mathbf{v}^\top B^{-1}}{1+\mathbf{v}^\top B^{-1}\mathbf{v}} = \mathbf{v}^\top (B+\mathbf{v}\mathbf{v}^\top)^{-1}$$

Then we can let $B := S_n^{[k]} - zI$, $\mathbf{v} = \frac{1}{\sqrt{n}}\mathbf{x}_k$, then $B + \mathbf{v}\mathbf{v}^\top = S_n - zI$
Thus

$$\begin{aligned}
\mathbf{x}_k\mathbf{x}_k^\top (S_n - zI)^{-1} &= n\mathbf{v}\mathbf{v}^\top (B+\mathbf{v}\mathbf{v}^\top)^{-1} \\
&= \frac{n\mathbf{v}\mathbf{v}^\top B^{-1}}{1+\mathbf{v}^\top B^{-1}\mathbf{v}} \\
&= \frac{\mathbf{x}_k\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}}{1+\frac{1}{n}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k}
\end{aligned}$$

Thus we have identity

$$I + z(S_n - zI)^{-1} = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{1+\frac{1}{n}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k}\mathbf{x}_k\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}$$

This is the matrix equation, we can turn it into scalar equation via normalized trace

$$\begin{aligned}
\frac{1}{p}\mathrm{Tr}[I + z(S_n - zI)^{-1}] &= \frac{1}{n}\sum_{k=1}^{n} \frac{\frac{1}{p}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k}{1+\frac{1}{n}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k} \\
1 + z\frac{1}{p}\mathrm{Tr}[(S_n - zI)^{-1}] &= \frac{1}{p}\sum_{k=1}^{n} (1 - \frac{1}{1+\frac{1}{n}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k}) \\
1 + z\ m_n(z) &= \frac{n}{p} - \frac{1}{p}\sum_{k=1}^{n} \frac{1}{1+\frac{1}{n}\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k}
\end{aligned}$$

Finally, we need this approximation step

$$\mathbf{x}_k^\top (S_n^{[k]} - zI)^{-1}\mathbf{x}_k \approx \mathrm{Tr}[(S_n^{[k]} - zI)^{-1}] \approx p\ m_{n-1}(z)$$

Thus

$$\begin{aligned}
1 + z\ m_n(z) &\approx \frac{n}{p} - \frac{1}{p}\sum_{k=1}^{n} \frac{1}{1+\frac{p}{n}\ m_{n-1}(z)} \\
&= \frac{n}{p} - \frac{n}{p}\frac{1}{1+\frac{p}{n}\ m_{n-1}(z)}
\end{aligned}$$

At the limit, $m_n(z) \approx m_{n-1}(z)$ we have equation (with $p/n \to \gamma$)

$$\begin{aligned}
\frac{p}{n}(1 + zm_{sc}(z)) &= 1 - \frac{1}{1+\frac{p}{n}m_{sc}(z)} \\
\gamma(1 + zm_{sc}(z)) &= 1 - \frac{1}{1+\gamma m_{sc}(z)} \\
1 + zm_{sc}(z) &= \frac{m_{sc}(z)}{1+\gamma m_{sc}(z)}
\end{aligned}$$

Now we obtain the 2nd order equation for $m_{sc}$

$$\begin{aligned}
\gamma(1 + zm_{sc}(z))(1+\gamma m_{sc}(z)) &= 1 + \gamma m_{sc}(z) - 1 \\
(1 + zm_{sc}(z))(1+\gamma m_{sc}(z)) &= m_{sc}(z) \\
\gamma z m_{sc}^2(z) + (\gamma + z - 1)m_{sc}(z) + 1 &= 0
\end{aligned}$$

We obtain the limiting function

$$m_{sc}(z) = \frac{-(\gamma + z - 1) \pm \sqrt{(\gamma + z - 1)^2 - 4\gamma z}}{2\gamma z}$$

To ensure the correct scaling when $z \to \infty$, $m_{sc}(z) \to -1/z$.

$$m_{sc}(z) = \frac{-(\gamma + z - 1) + \sqrt{(\gamma + z - 1)^2 - 4\gamma z}}{2\gamma z}$$

**Derivation of Marchenko Pastur Law from Stijeiel transform (optional)** Final step is to take the limit to get the spectral distribution $z = x + i\eta$

$$\begin{aligned}
\lim_{\eta\to 0} \mathrm{Im}[m_{sc}(x+i\eta)] &= \lim_{\eta\to 0} \mathrm{Im}[\frac{-(\gamma+x+i\eta-1)+\sqrt{(\gamma+x+i\eta-1)^2-4\gamma(x+i\eta)}}{2\gamma(x+i\eta)}] \\
&= \lim_{\eta\to 0} \mathrm{Im}[\frac{-(\gamma+x+i\eta-1)(x-i\eta)+(x-i\eta)\sqrt{(\gamma+x+i\eta-1)^2-4\gamma(x+i\eta)}}{2\gamma(x^2+\eta^2)}]
\end{aligned}$$

Focus on numerator
The first term can be found to vanish.

$$\begin{aligned}
\mathrm{Im}[-(\gamma+x-1+i\eta)(x-i\eta)] &= -\mathrm{Im}[i\eta x - i\eta(\gamma+x-1)] \\
&= \eta(\gamma-1)
\end{aligned}$$

Thus

$$\begin{aligned}
\lim_{\eta\to 0} \mathrm{Im}[\frac{-(\gamma+x-1+i\eta)(x-i\eta)}{2\gamma(x^2+\eta^2)}] &= \lim_{\eta\to 0} \frac{\eta(\gamma-1)}{2\gamma(x^2+\eta^2)} \\
&= 0
\end{aligned}$$

For the 2nd term, note that the square root have roots

$$\begin{aligned}
(z+\gamma-1)^2 - 4\gamma z &= z^2 + 2(\gamma-1)z + (\gamma-1)^2 - 4\gamma z \\
&= z^2 - 2(\gamma+1)z + (\gamma-1)^2 \\
&= (z-\lambda_+)(z-\lambda_-)
\end{aligned}$$

$$\begin{aligned}
\lambda_\pm &= \frac{2(\gamma+1) \pm \sqrt{4(\gamma+1)^2 - 4(\gamma-1)^2}}{2} \\
&= (\gamma+1) \pm 2\sqrt{\gamma} \\
&= (\sqrt{\gamma} \pm 1)^2
\end{aligned}$$

Consider the complex representation

$$\begin{aligned}
(z-\lambda_+)(z-\lambda_-) &= (x+i\eta-\lambda_+)(x+i\eta-\lambda_-) \\
&= (x-\lambda_+)(x-\lambda_-) - \eta^2 + i\eta(2x-\lambda_+-\lambda_-)
\end{aligned}$$

Recall that

$$\Im\sqrt{u+iv} = \mathrm{sgn}(v)\sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}$$

$$\Re\left[\sqrt{u+iv}\right] = \sqrt{\frac{\sqrt{u^2+v^2}\ +\ u}{2}}.$$

$$\begin{aligned}
\lim_{v\to 0^+} \Im\sqrt{u+iv} &= \sqrt{\frac{\sqrt{u^2}-u}{2}} \\
&= \sqrt{\frac{|u|-u}{2}} \\
&= \begin{cases} \sqrt{-u} & u < 0 \\ 0 & u \geq 0 \end{cases}
\end{aligned}$$

$$
\begin{aligned}
\lim_{v\to 0^+} \Re\big[\sqrt{u+iv}\big] &= \sqrt{\frac{\sqrt{u^2}+u}{2}} \\
&= \sqrt{\frac{|u|+u}{2}} \\
&= \begin{cases} 0 & u<0 \\ \sqrt{u} & u\geq 0 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{\eta\to 0^+} \mathrm{Im}\Big[(x-i\eta)\sqrt{(z-\lambda_+)(z-\lambda_-)}\Big] &= x\lim_{\eta\to 0^+}\mathrm{Im}\Big[\sqrt{(z-\lambda_+)(z-\lambda_-)}\Big] - \lim_{\eta\to 0^+}\eta\mathrm{Re}\Big[\sqrt{(z-\lambda_+)(z-\lambda_-)}\Big] \\
&= x\lim_{\eta\to 0^+}\mathrm{Im}\Big[\sqrt{(x-\lambda_+)(x-\lambda_-)-\eta^2+i\eta(2x-\lambda_+-\lambda_-)}\Big] \\
&= \begin{cases} x\sqrt{(\lambda_+-x)(x-\lambda_-)} & (x-\lambda_+)(x-\lambda_-)<0 \\ 0 & (x-\lambda_+)(x-\lambda_-)\geq 0 \end{cases}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\lim_{\eta\to 0}\mathrm{Im}[m_{sc}(x+i\eta)] &= \lim_{\eta\to 0}\mathrm{Im}[\frac{-(\gamma+x+i\eta-1)+\sqrt{(\gamma+x+i\eta-1)^2-4\gamma(x+i\eta)}}{2\gamma(x+i\eta)}] \\
&= \lim_{\eta\to 0}\frac{\eta(\gamma-1)+x\sqrt{(\lambda_+-x)(x-\lambda_-)}}{2\gamma(x^2+\eta^2)} \\
&= \begin{cases} \frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{2\gamma x} & (x-\lambda_+)(x-\lambda_-)<0 \\ 0 & (x-\lambda_+)(x-\lambda_-)\geq 0 \end{cases}
\end{aligned}
$$

Thus the density is

$$
\rho(x) = \frac{1}{\pi}\lim_{\eta\to 0}\mathrm{Im}[m_{sc}(x+i\eta)] = \begin{cases} \frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{2\pi\gamma x} & \lambda_-<x<\lambda_+ \\ 0 & x\geq\lambda_+; x\leq\lambda_- \end{cases}
$$

Where do we need $\gamma<1$?
The full measure looks like

$$
\mu_{MP}(x) = (1-\frac{1}{\gamma})_+\delta(x) + \frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{2\pi\gamma x}1_{[\lambda_-,\lambda_+]}dx
$$

### 3.3 Silverstein equation for General Colored Wishart (according to Francis Bach)

Consider a set of data $\mathbf{x}_i\in\mathbb{R}^d$, $X=[\mathbf{x}_1,...,x_n]$, distribution of each vector is $\mathbf{x}_i\sim\mathcal{N}(0,\Sigma)$

$$
\begin{aligned}
\hat{\Sigma} &= \frac{1}{n}X^\top X \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^\top
\end{aligned}
$$

**A few general linear algebra identities** Resolvent and its related quantity

$$\begin{aligned}A(A-zI_d)^{-1} &= (A-zI_d+zI_d)(A-zI_d)^{-1}\\ &= I_d + z(A-zI_d)^{-1}\\ \mathrm{Tr}[A(A-zI)^{-1}] &= d + z\mathrm{Tr}[(A-zI_d)^{-1}]\end{aligned}$$

Difference between Matrix inverse

$$B^{-1}(A-B)A^{-1} = B^{-1} - A^{-1} \tag{1}$$

Rank one update on matrix inverse (Sherman–Morrison / Woodbury)

$$(A+\mathbf{u}\mathbf{v}^\top)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^\top A^{-1}}{1+\mathbf{v}^\top A^{-1}\mathbf{u}} \tag{2}$$

A corollary of Sherman

$$\begin{aligned}\mathbf{u}\mathbf{v}^\top(A+\mathbf{u}\mathbf{v}^\top)^{-1} &= \mathbf{u}\mathbf{v}^\top A^{-1} - \frac{\mathbf{u}(\mathbf{v}^\top A^{-1}\mathbf{u})\mathbf{v}^\top A^{-1}}{1+\mathbf{v}^\top A^{-1}\mathbf{u}}\\ &= \mathbf{u}\mathbf{v}^\top A^{-1}(1-\frac{\mathbf{v}^\top A^{-1}\mathbf{u}}{1+\mathbf{v}^\top A^{-1}\mathbf{u}})\\ &= \frac{\mathbf{u}\mathbf{v}^\top A^{-1}}{1+\mathbf{v}^\top A^{-1}\mathbf{u}}\end{aligned}$$

Woodbury matrix identity

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \tag{3}$$

a simplified version when $A=\lambda I_d$

$$\begin{aligned}(\lambda I_n + UCV)^{-1} &= \frac{1}{\lambda}I_n - \frac{1}{\lambda^2}U(C^{-1}+\frac{1}{\lambda}VU)^{-1}V\\ &= \frac{1}{\lambda}I_n - \frac{1}{\lambda}U(\lambda C^{-1}+VU)^{-1}V\end{aligned}$$

$$U(\lambda C^{-1}+VU)^{-1}V = I_n - \lambda(\lambda I_n + UCV)^{-1}$$

Thus

$$X(\lambda I_d + X^\top X)^{-1}X^\top = I_n - \lambda(\lambda I_n + XX^\top)^{-1} \tag{4}$$

Resolvent push through identity

$$X(\lambda I_d + X^\top X)^{-1} = (\lambda I_n + XX^\top)^{-1}X$$

Push through identity

$$A(\lambda I_d + BA)^{-1} = (\lambda I_n + AB)^{-1}A$$

**First move, leave one sample out decomposition** First move, let's study [2]

$$\begin{aligned}\hat\Sigma(\hat\Sigma - zI)^{-1} &= \frac{1}{n}X^\top X(\frac{1}{n}X^\top X - zI)^{-1}\\ &= X^\top X(X^\top X - nzI)^{-1}\\ &= \sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j^\top(\sum_{k=1}^n \mathbf{x}_k\mathbf{x}_k^\top - nzI)^{-1}\end{aligned}$$

[2] A equivalent and more naturally motivated construction is in the Sec.3.4. Where the first move is to start from the $n$-by-$n$ kernel matrix and then removing one sample is equivalent to remove one row and one column. Then relating the resolvent of $n$ dim matrix ($[1,1]$ element) and $n-1$ dim sub-matrix with Schur complement is key.

After this, the derivation of self consistent equation is rather similar.

Denoting $\hat{\Sigma}_{-j} = \frac{1}{n}\sum_{k,k\neq j} \mathbf{x}_k\mathbf{x}_k^\top$

$$\begin{aligned}
\hat{\Sigma}(\hat{\Sigma} - zI)^{-1} &= \sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^\top (\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k^\top - nzI)^{-1} \\
&= \sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^\top (\mathbf{x}_j\mathbf{x}_j^\top + n\hat{\Sigma}_{-j} - nzI)^{-1} \\
&= \sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^\top \frac{(n\hat{\Sigma}_{-j} - nzI)^{-1}}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} \qquad (5)
\end{aligned}$$

Note

$$\begin{aligned}
\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI)^{-1}] &= \sum_{j=1}^{n} \frac{\mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} \\
&= n - \sum_{j=1}^{n} \frac{1}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} \qquad (6)
\end{aligned}$$

One way to express

$$\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI)^{-1}] = d + z\mathrm{Tr}[(\hat{\Sigma} - zI)^{-1}]$$

Another way connect it to the resolvent of the other side

$$\begin{aligned}
\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI_d)^{-1}] &= \mathrm{Tr}[X^\top X(X^\top X - nzI_d)^{-1}] \\
&= \mathrm{Tr}[X(X^\top X - nzI_d)^{-1}X^\top] \\
&= \mathrm{Tr}[I_n + nz(-nzI_n + XX^\top)^{-1}] \\
&= n + z\mathrm{Tr}[(\frac{1}{n}XX^\top - zI_n)^{-1}]
\end{aligned}$$

We can see the identity the Stijies transform of the covariance and kernel matrix are connected by such

$$\begin{aligned}
d + z\mathrm{Tr}[(\hat{\Sigma} - zI)^{-1}] = \mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI_d)^{-1}] &= n + z\mathrm{Tr}[(\frac{1}{n}XX^\top - zI_n)^{-1}] \qquad (7) \\
d + dz\ m_{\hat{\Sigma}}(z) &= n + nz\ m_{\frac{1}{n}XX^\top}(z)
\end{aligned}$$

Their difference is only the pole / eigenvalues at zero.
Now we define this Stijies transform of the kernel matrix (instead of covariance)

$$\begin{aligned}
\hat{\varphi}(z) :&= \frac{1}{n}\mathrm{Tr}[(\frac{1}{n}XX^\top - zI_n)^{-1}] \\
&= \mathrm{Tr}[(XX^\top - nzI_n)^{-1}] \\
&= m_{\frac{1}{n}XX^\top}(z) \qquad (8)
\end{aligned}$$

Then observe that

$$\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI_d)^{-1}] = n + nz\hat{\varphi}(z)$$

From the 6, we have the identity expressing $\hat{\varphi}(z)$

$$\hat{\varphi}(z) = -\frac{1}{z}\frac{1}{n}\sum_{j=1}^{n} \frac{1}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} \qquad (9)$$

**Second move, prove of convergence** Second move, prove of convergence of esp. $(\hat{\Sigma} - zI)^{-1}$ and $(-z\hat{\varphi}(z)\Sigma - zI)^{-1}$, i.e. the convergence of two resolvent. Thus we will focus on the difference of two inverse.

Using the difference of inverse formula 1,

$$\begin{aligned}
(-z\hat{\varphi}(z)\Sigma - zI)^{-1} - (\hat{\Sigma} - zI)^{-1} &= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\Big[\hat{\Sigma} - zI - (-z\hat{\varphi}(z)\Sigma - zI)\Big](\hat{\Sigma} - zI)^{-1} \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\Big[\hat{\Sigma} + z\hat{\varphi}(z)\Sigma\Big](\hat{\Sigma} - zI)^{-1} \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\Big[\hat{\Sigma}(\hat{\Sigma} - zI)^{-1} + z\hat{\varphi}(z)\Sigma(\hat{\Sigma} - zI)^{-1}\Big]
\end{aligned}$$

Note the first term enjoys the rank one decomposition 5, and with the expression of $\hat{\varphi}(z)$ 9, we have

$$
\begin{aligned}
&(-z\hat{\varphi}(z)\Sigma - zI)^{-1} - (\hat{\Sigma} - zI)^{-1} \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\Big[\hat{\Sigma}(\hat{\Sigma} - zI)^{-1} + z\hat{\varphi}(z)\Sigma(\hat{\Sigma} - zI)^{-1}\Big] \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\Big[\sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j^\top \frac{(n\hat{\Sigma}_{-j} - nzI)^{-1}}{1 + \mathbf{x}_j^\top(n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} - \frac{1}{n}\sum_{j=1}^{n}\frac{1}{1 + \mathbf{x}_j^\top(n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j}\Sigma(\hat{\Sigma} - zI)^{-1}\Big] \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\frac{1}{n}\sum_{j=1}^{n}\Big[\frac{\mathbf{x}_j\mathbf{x}_j^\top(\hat{\Sigma}_{-j} - zI)^{-1} - \Sigma(\hat{\Sigma} - zI)^{-1}}{1 + \mathbf{x}_j^\top(n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j}\Big] \\
&= (-z\hat{\varphi}(z)\Sigma - zI)^{-1}\square
\end{aligned}
$$

Thus

$$(\hat{\Sigma} - zI)^{-1} = (-z\hat{\varphi}(z)\Sigma - zI)^{-1}(I - \square)$$

Where we abbreviate,

$$\square := \frac{1}{n}\sum_{j=1}^{n}\Big[\frac{\mathbf{x}_j\mathbf{x}_j^\top(\hat{\Sigma}_{-j} - zI)^{-1} - \Sigma(\hat{\Sigma} - zI)^{-1}}{1 + \mathbf{x}_j^\top(n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j}\Big]$$

To show $\square$ is small, the intuition is

$$\mathrm{Tr}\Big[\mathbf{x}_j\mathbf{x}_j^\top(\hat{\Sigma}_{-j} - zI)^{-1}\Big] \approx \mathrm{Tr}\Big[\Sigma(\hat{\Sigma}_{-j} - zI)^{-1}\Big] \approx \mathrm{Tr}\Big[\Sigma(\hat{\Sigma} - zI)^{-1}\Big]$$

The rigorous proof require certain inequality for concentration. More details could be found in [2, 5].
As long as, we can show $\square$ is small in the trace sense, then we have the desired equivalence,

$$(\hat{\Sigma} - zI)^{-1} \sim (-z\hat{\varphi}(z)\Sigma - zI)^{-1} \tag{10}$$

The equivalence should be interpreted in the sense that their trace with regular test matrix B are equal.
So we can further take trace and say,

$$\frac{1}{d}\mathrm{Tr}[(\hat{\Sigma} - zI)^{-1}] \sim \frac{1}{d}\mathrm{Tr}[(-z\hat{\varphi}(z)\Sigma - zI)^{-1}] \tag{11}$$

**Third move, equivalence to self consistent equation** As the third step, self consistent equation can be derived from the trace equivalence

$$
\begin{aligned}
RHS = \mathrm{Tr}[(-z\hat{\varphi}(z)\Sigma - zI)^{-1}] &= -\frac{1}{z}\frac{1}{d}\mathrm{Tr}[(\hat{\varphi}(z)\Sigma + I)^{-1}] \\
&= -\frac{1}{z}\frac{1}{d}\sum_{k}\frac{1}{\hat{\varphi}(z)\lambda_k + 1}\mathrm{Tr}[(\hat{\varphi}(z)\Sigma + I)^{-1}] \\
&= -\frac{1}{z}\int\frac{1}{\hat{\varphi}(z)\lambda + 1}dF_{\Sigma}(\lambda)
\end{aligned}
$$

Due to 7, the two Stieltjes transforms of covariance and kernel are related by following identity.

$$d + dz\ m_{\hat{\Sigma}}(z) = n + nz\ m_{\frac{1}{n}XX^\top}(z)$$

So we can express $\hat{\varphi}(z) := m_{\frac{1}{n}XX^\top}(z)$ as a function of $m_{\hat{\Sigma}}(z)$ and vice versa.

$$
\begin{aligned}
m_{\hat{\Sigma}}(z) &= \frac{n-d}{zd} + \frac{n}{d}\hat{\varphi}(z) \\
&= \frac{1}{z}(\frac{1}{\gamma} - 1) + \frac{1}{\gamma}\hat{\varphi}(z) \\
\hat{\varphi}(z) &= \frac{d-n}{nz} + \frac{d}{n}m_{\hat{\Sigma}}(z) \\
&= \frac{\gamma - 1}{z} + \gamma m_{\hat{\Sigma}}(z)
\end{aligned}
$$

Note on the left hand side we have

$$LHS = \frac{1}{d}\mathrm{Tr}[(\hat{\Sigma} - zI)^{-1}] = m_{\hat{\Sigma}}(z)$$
$$= \frac{1}{z}(\frac{1}{\gamma} - 1) + \frac{1}{\gamma}\hat{\varphi}(z)$$

Now the equality reads,

$$m_{\hat{\Sigma}}(z) = \frac{1}{z}(\frac{1}{\gamma} - 1) + \frac{1}{\gamma}\hat{\varphi}(z) = -\frac{1}{z}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau)$$

Here we've already obtained the self consistent equation for $\hat{\varphi}(z)$,

$$(\frac{1}{\gamma} - 1) + \frac{1}{\gamma}z\hat{\varphi}(z) = -\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau)$$
$$1 - \gamma + z\,\hat{\varphi}(z) = -\gamma\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau)$$
$$\hat{\varphi}(z) = \frac{\gamma - 1}{z} - \frac{\gamma}{z}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau) \tag{12}$$

for each $z$ this equation has one unique solution.
Note we can also write the self consistent equation for the Stijiel transform $m_{\hat{\Sigma}}(z)$

$$\begin{aligned} m_{\hat{\Sigma}}(z) &= -\frac{1}{z}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau) \\ &= -\frac{1}{z}\int \frac{1}{(\gamma m_{\hat{\Sigma}}(z) + \frac{\gamma-1}{z})\tau + 1} dF_{\Sigma}(\tau) \\ &= -\int \frac{1}{\tau\ (z\gamma m_{\hat{\Sigma}}(z) + \gamma - 1) + z} dF_{\Sigma}(\tau) \end{aligned} \tag{13}$$

**Implication: Self-energy induces ridge regularization** Here we have the self consistent integral equation for the function $\hat{\varphi}(z)$.

Note the identities

$$\mathrm{Tr}[A(A - zI)^{-1}] = d + z\mathrm{Tr}[(A - zI_d)^{-1}]$$

Then

$$\mathrm{Tr}[\Sigma(\Sigma + \kappa I)^{-1}] = d - \kappa\mathrm{Tr}[(\Sigma + \kappa I_d)^{-1}]$$

$$\begin{aligned} \hat{\Sigma}(\hat{\Sigma} - zI)^{-1} &= I_d + z(\hat{\Sigma} - zI_d)^{-1} \\ &\sim I_d + z(-z\hat{\varphi}(z)\Sigma - zI)^{-1} \\ &= I_d - \frac{1}{\hat{\varphi}(z)}(\Sigma + \frac{1}{\hat{\varphi}(z)}I)^{-1} \\ &= \Sigma(\Sigma + \frac{1}{\hat{\varphi}(z)}I)^{-1} \end{aligned}$$

In more common notations, if $\lambda$ is the regularization strength for Ridge regression,

$$\hat{\Sigma}(\hat{\Sigma} + \lambda I)^{-1} \sim \Sigma(\Sigma + \frac{1}{\hat{\varphi}(-\lambda)}I)^{-1} = \Sigma(\Sigma + \kappa(\lambda)I)^{-1} \tag{14}$$

where $\kappa(\lambda) := \frac{1}{\hat{\varphi}(-\lambda)} = \frac{1}{m_{\frac{1}{n}XX^\top}(-\lambda)}$ this is the exact relation we need to understand the self-induced regularization effect of the under-sampling regime.

**Derivative of the fixed point equation** Since we have the fixed point equation of $\hat{\varphi}(z)$, we can run it as a function of $z$

$$\begin{aligned} \hat{\varphi}(z) &= \frac{\gamma - 1}{z} - \frac{\gamma}{z}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau) \\ \frac{d\hat{\varphi}(z)}{dz} &= -\frac{\gamma - 1}{z^2} + \frac{\gamma}{z^2}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau) - \frac{\gamma}{z}\int \frac{d}{dz}\Big[\frac{1}{\hat{\varphi}(z)\tau + 1}\Big] dF_{\Sigma}(\tau) \\ &= -\frac{\gamma - 1}{z^2} + \frac{\gamma}{z^2}\int \frac{1}{\hat{\varphi}(z)\tau + 1} dF_{\Sigma}(\tau) + \frac{\gamma}{z}\int \frac{d\hat{\varphi}(z)}{dz}\frac{\tau}{(\hat{\varphi}(z)\tau + 1)^2} dF_{\Sigma}(\tau) \end{aligned}$$

Further at fixed point we have equality 12

$$\frac{\gamma}{z}\int\frac{1}{\hat{\varphi}(z)\tau+1}dF_\Sigma(\tau)=\frac{\gamma-1}{z}-\hat{\varphi}(z)$$

Thus, the derivative satisfies,

$$\frac{d\hat{\varphi}^*(z)}{dz}=-\frac{\hat{\varphi}^*(z)}{z}+\frac{\gamma}{z}\frac{d\hat{\varphi}^*(z)}{dz}\int\frac{\tau}{(\hat{\varphi}^*(z)\tau+1)^2}dF_\Sigma(\tau)$$

We have the forward equation

$$\left(1-\frac{\gamma}{z}\int\frac{\tau}{(\hat{\varphi}^*(z)\tau+1)^2}dF_\Sigma(\tau)\right)\frac{d\hat{\varphi}^*(z)}{dz}=-\frac{\hat{\varphi}^*(z)}{z}$$

$$\begin{aligned}\frac{d\hat{\varphi}^*(z)}{dz}&=-\frac{\hat{\varphi}^*(z)}{z}\frac{1}{1-\frac{\gamma}{z}\int\frac{\tau}{(\hat{\varphi}^*(z)\tau+1)^2}dF_\Sigma(\tau)}\\&=-\frac{\hat{\varphi}^*(z)}{z-\gamma\int\frac{\tau}{(\hat{\varphi}^*(z)\tau+1)^2}dF_\Sigma(\tau)}\end{aligned}$$

Thus we arrived at this equation for evolving the fixed point.
Similarly, we can write the equations for regularization strength $\lambda$ and effective regularization $\kappa(\lambda)$

$$\kappa(\lambda)=\frac{1}{\hat{\varphi}(-\lambda)}$$

Taking derivatives

$$\begin{aligned}\frac{d\kappa(\lambda)}{d\lambda}&=\frac{d}{d\lambda}\frac{1}{\hat{\varphi}(-\lambda)}\\&=-\frac{1}{\hat{\varphi}^2(-\lambda)}\frac{d\hat{\varphi}(-\lambda)}{d\lambda}\\&=\frac{1}{\hat{\varphi}^2(-\lambda)}\frac{d\hat{\varphi}(-\lambda)}{d(-\lambda)}\\&=\frac{1}{\hat{\varphi}^*(-\lambda)^2}\times\left(-\frac{\hat{\varphi}^*(-\lambda)}{-\lambda-\gamma\int\frac{\tau}{(\hat{\varphi}^*(-\lambda)\tau+1)^2}dF_\Sigma(\tau)}\right)\\&=\frac{1}{\hat{\varphi}^*(-\lambda)}\times\frac{1}{\lambda+\gamma\int\frac{\tau}{(\hat{\varphi}^*(-\lambda)\tau+1)^2}dF_\Sigma(\tau)}\\&=\kappa(\lambda)\times\frac{1}{\lambda+\gamma\int\frac{\tau}{(\frac{1}{\kappa(\lambda)}\tau+1)^2}dF_\Sigma(\tau)}\\&=\kappa(\lambda)\times\frac{1}{\lambda+\gamma\kappa^2(\lambda)\int\frac{\tau}{(\tau+\kappa(\lambda))^2}dF_\Sigma(\tau)}\end{aligned}$$

$$\frac{d}{d\lambda}\log\kappa(\lambda)=\frac{1}{\lambda+\gamma\kappa^2(\lambda)\int\frac{\tau}{(\tau+\kappa(\lambda))^2}dF_\Sigma(\tau)}$$

**Numerically Solving self consistent equation and limiting spectrum** Numerical solving the self consistent equation for $\hat{\varphi}(z)$

$$\hat{\varphi}(z)=\frac{\gamma-1}{z}-\frac{\gamma}{z}\int\frac{1}{\hat{\varphi}(z)\tau+1}dF_\Sigma(\tau)$$

We can build a equivalent functional $H(\varphi;z)$ and solve for $H(\varphi;z)=0$

$$H(\varphi;z)=z\varphi-(\gamma-1)+\gamma\int\frac{1}{\varphi\tau+1}dF_\Sigma(\tau)$$

For more efficient root finding, we can compute the derivative to $\varphi$ and leverage Newton's method

$$\partial_\varphi H(\varphi;z)=z-\gamma\int\frac{\tau}{(\varphi\tau+1)^2}dF_\Sigma(\tau)$$

Basically, the workflow is like,

$$
\begin{aligned}
\hat{\varphi}(z) &= \mathrm{Newton}\Big(H(.,z), \partial_\varphi H(.,z), \phi_0\Big) \\
m_{\hat{\Sigma}}(z) &= \frac{1}{z}(\frac{1}{\gamma} - 1) + \frac{1}{\gamma}\hat{\varphi}(z) \\
\rho(x) &= \frac{1}{\pi}\lim_{\eta \to 0} m_{\hat{\Sigma}}(x + i\eta)
\end{aligned}
$$

### 3.4 Silverstein equation for Kernel spectrum (similar derivation)

Consider the kernel matrices $\hat{K} = \frac{1}{p}XX^\top$ where $X \in \mathbb{R}^{n\times p}$, where $K_{ij} = \frac{1}{p}\mathbf{x}_i^\top \mathbf{x}_j$

The theoretical covariance is $\Sigma$, empirical covariance is $\hat{\Sigma} = \frac{1}{n}X^\top X$

Resolvent looks like

$$\mathrm{Tr}[(\hat{K} - zI)^{-1}]$$

The first move will be to consider the structure of the resolvent, and remove one row and column from the kernel and see the change in resolvent. and using Schur decomposition

Consider removing a row and a column

$$(\hat{K} - zI)^{-1} = \begin{bmatrix} k_{11} - z & \mathbf{k}_1^\top \\ \mathbf{k}_1 & \hat{K}_{-1} - zI_{n-1} \end{bmatrix}^{-1}$$

The first element reads

$$\begin{aligned}
[(\hat{K} - zI)^{-1}]_{11} &= (k_{11} - z - \mathbf{k}_1^\top(\hat{K}_{-1} - zI_{n-1})^{-1}\mathbf{k}_1)^{-1} \\
&= \frac{1}{\frac{1}{p}\mathbf{x}_1^\top\mathbf{x}_1 - z - \frac{1}{p^2}\sum_{jk=2}^n \mathbf{x}_1^\top\mathbf{x}_j[(\hat{K}_{-1} - zI_{n-1})^{-1}]_{jk}\mathbf{x}_k^\top\mathbf{x}_1} \\
&= \frac{1}{\frac{1}{p}\mathbf{x}_1^\top\mathbf{x}_1 - z - \frac{1}{p^2}\mathbf{x}_1^\top X_{-1}^\top(\hat{K}_{-1} - zI_{n-1})^{-1}X_{-1}\mathbf{x}_1}
\end{aligned}$$

$\hat{K}_{-i} := \frac{1}{p}X_{-i}X_{-i}^\top$

For the trace we observed the almost self consistent equation, however conneting to the resolvent of the covariance

$$\begin{aligned}
\mathrm{Tr}[(\hat{K} - zI)^{-1}] &= \sum_i^n [(\hat{K} - zI)^{-1}]_{ii} \\
&= \sum_i^n \frac{1}{\frac{1}{p}\mathbf{x}_i^\top\mathbf{x}_i - \frac{1}{p^2}\mathbf{x}_i^\top X_{-i}^\top(\hat{K}_{-i} - zI_{n-1})^{-1}X_{-i}\mathbf{x}_i - z} \\
&= \sum_i^n \frac{1}{\frac{1}{p}\mathbf{x}_i^\top[I_p - \frac{1}{p}X_{-i}^\top(\frac{1}{p}X_{-i}X_{-i}^\top - zI_{n-1})^{-1}X_{-i}]\mathbf{x}_i - z} \\
&= \sum_i^n \frac{1}{\frac{1}{p}\mathbf{x}_i^\top[-pz(X_{-i}^\top X_{-i} - pzI_p)^{-1}]\mathbf{x}_i - z} \\
&= -\frac{1}{z}\sum_i^n \frac{1}{\mathbf{x}_i^\top(X_{-i}^\top X_{-i} - pzI_p)^{-1}\mathbf{x}_i + 1} \\
&= -\frac{1}{z}\sum_i^n \frac{1}{\mathbf{x}_i^\top(n\hat{\Sigma}_{-i} - pzI_p)^{-1}\mathbf{x}_i + 1}
\end{aligned}$$

Intuitively, the denominator is similar to

$$\mathbf{x}_i^\top(X_{-i}^\top X_{-i} - pzI_p)^{-1}\mathbf{x}_i \sim \mathrm{Tr}[\Sigma(X_{-i}^\top X_{-i} - pzI_p)^{-1}] \sim \mathrm{Tr}[\Sigma(n\hat{\Sigma} - pzI_p)^{-1}]$$

To scale it properly we'd need to consider the redifined variable

$$\begin{aligned}
-z\frac{1}{p}\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}] &= -z\frac{1}{p}\cdot -\frac{1}{\frac{n}{p}z}\sum_i^n \frac{1}{\mathbf{x}_i^\top(X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i + 1} \\
&= \frac{1}{n}\sum_i^n \frac{1}{\mathbf{x}_i^\top(X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i + 1}
\end{aligned}$$

**White covariance case** We can proceed similar to the Wigner case, if $\mathbf{x}_i \sim \mathcal{N}(0, I_p)$ then

$$\mathbf{x}_i^\top (X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i \approx \mathrm{Tr}[(X_{-i}^\top X_{-i} - nzI_p)^{-1}] = \frac{1}{n}\mathrm{Tr}[(\frac{1}{n}X_{-i}^\top X_{-i} - zI_p)^{-1}] = \frac{1}{n}\mathrm{Tr}[(\hat{\Sigma}_{-i} - zI_p)^{-1}]$$

Now we have equation

$$\begin{aligned}
-z\frac{1}{p}\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}] &= \frac{1}{n}\sum_i^n \frac{1}{\mathbf{x}_i^\top (X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i + 1} \\
&\approx \frac{1}{n}\sum_i^n \frac{1}{\frac{1}{n}\mathrm{Tr}[(\hat{\Sigma}_{-i} - zI_p)^{-1}] + 1} \\
&\approx \frac{1}{\frac{1}{n}\mathrm{Tr}[(\hat{\Sigma} - zI_p)^{-1}] + 1}
\end{aligned}$$

Then the final steps is to relate the two Stijiels transforms: that of kernel $\frac{1}{p}\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}]$ and that covariance matrix $\frac{1}{n}\mathrm{Tr}[(\hat{\Sigma} - zI_p)^{-1}]$. We have the identity

$$\begin{aligned}
\frac{1}{n}\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI_p)^{-1}] &= \frac{1}{n}\mathrm{Tr}[\hat{K}(\hat{K} - \frac{n}{p}zI)^{-1}] \\
\frac{1}{n}\mathrm{Tr}[I_p + z(\hat{\Sigma} - zI_p)^{-1}] &= \frac{1}{n}\mathrm{Tr}[I_n + \frac{n}{p}z(\hat{K} - \frac{n}{p}zI)^{-1}] \\
\frac{p}{n} + \frac{1}{n}z\mathrm{Tr}[(\hat{\Sigma} - zI_p)^{-1}] &= 1 + \frac{1}{p}z\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}]
\end{aligned}$$

Because

$$\begin{aligned}
\frac{1}{n}\mathrm{Tr}[\hat{\Sigma}(\hat{\Sigma} - zI_p)^{-1}] &= \frac{1}{n}\mathrm{Tr}[\frac{1}{n}X^\top X(\frac{1}{n}X^\top X - zI_p)^{-1}] \\
&= \frac{1}{n}\mathrm{Tr}[X(X^\top X - nzI_p)^{-1}X^\top] \\
&= \frac{1}{n}\mathrm{Tr}[XX^\top(XX^\top - nzI_n)^{-1}] \\
&= \frac{1}{n}\mathrm{Tr}[\frac{1}{n}XX^\top(\frac{1}{n}XX^\top - zI_n)^{-1}] \\
&= \frac{1}{n}\mathrm{Tr}[\frac{1}{p}XX^\top(\frac{1}{p}XX^\top - \frac{n}{p}zI_n)^{-1}] \\
&= \frac{1}{n}\mathrm{Tr}[\hat{K}(\hat{K} - \frac{n}{p}zI)^{-1}]
\end{aligned}$$

Thus we have self consistent equation that leads to MP law

$$1 - \frac{p}{n} - \frac{1}{n}z\mathrm{Tr}[(\hat{\Sigma} - zI_p)^{-1}] = \frac{1}{\frac{1}{n}\mathrm{Tr}[(\hat{\Sigma} - zI_p)^{-1}] + 1}$$

$$-z\frac{1}{p}\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}] = \frac{1}{1 + \frac{1}{z}(1 - \frac{p}{n} + \frac{1}{p}z\mathrm{Tr}[(\hat{K} - \frac{n}{p}zI)^{-1}])}$$

**General colored covariance case** Now the approximation is like

$$\mathbf{x}_i^\top (X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i \sim \mathrm{Tr}[\Sigma(X_{-i}^\top X_{-i} - nzI_p)^{-1}] \sim \frac{1}{n}\mathrm{Tr}[\Sigma(\hat{\Sigma} - zI_p)^{-1}]$$

Also note that

$$\hat{\Sigma}(\hat{\Sigma} - zI_p)^{-1} = \sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j^\top \frac{(n\hat{\Sigma}_{-j} - nzI)^{-1}}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j}$$

We try to construct a quantity from fixed matrix $\Sigma, I$ that $(\hat{\Sigma} - zI_p)^{-1}$ converges to

$$
\begin{aligned}
&(-z\frac{1}{p}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}]\Sigma - zI_p - \hat{\Sigma} + zI_p)(\hat{\Sigma} - zI_p)^{-1} \\
=&(-z\frac{1}{p}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}]\Sigma - zI_p)(\hat{\Sigma} - zI_p)^{-1} - I_p \\
=&(-z\frac{1}{p}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}]\Sigma - \hat{\Sigma})(\hat{\Sigma} - zI_p)^{-1} \\
=&\frac{1}{n}\sum_i^n \frac{1}{\mathbf{x}_i^\top (X_{-i}^\top X_{-i} - nzI_p)^{-1}\mathbf{x}_i + 1}\Sigma(\hat{\Sigma} - zI_p)^{-1} - \hat{\Sigma}(\hat{\Sigma} - zI_p)^{-1} \\
=&\frac{1}{n}\sum_j^n \frac{1}{\mathbf{x}_j^\top (X_{-j}^\top X_{-j} - nzI_p)^{-1}\mathbf{x}_j + 1}\Sigma(\hat{\Sigma} - zI_p)^{-1} - \sum_{j=1}^n \mathbf{x}_j\mathbf{x}_j^\top \frac{(n\hat{\Sigma}_{-j} - nzI)^{-1}}{1 + \mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI)^{-1}\mathbf{x}_j} \\
=&\sum_j^n \frac{1}{\mathbf{x}_j^\top (X_{-j}^\top X_{-j} - nzI_p)^{-1}\mathbf{x}_j + 1}\left[\frac{1}{n}\Sigma(\hat{\Sigma} - zI_p)^{-1} - \frac{1}{n}\mathbf{x}_j\mathbf{x}_j^\top (\hat{\Sigma}_{-j} - zI_p)^{-1}\right]
\end{aligned}
$$

Then we can bound

$$
\left[\frac{1}{n}\Sigma(\hat{\Sigma} - zI_p)^{-1} - \mathbf{x}_j\mathbf{x}_j^\top (n\hat{\Sigma}_{-j} - nzI_p)^{-1}\right] \sim 0
$$

Thus we have

$$
\begin{aligned}
(-z\frac{1}{n}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}]\Sigma - zI_p)(\hat{\Sigma} - zI_p)^{-1} - I_p \sim 0 \\
(\hat{\Sigma} - zI_p)^{-1} \sim (-z\frac{1}{p}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}]\Sigma - zI_p)^{-1}
\end{aligned}
$$

Again we get the scaling relationship.

$$
\begin{aligned}
\frac{1}{p}\mathrm{Tr}[(\hat{K}-\frac{n}{p}zI)^{-1}] &= \frac{1}{p}\mathrm{Tr}[(\frac{1}{p}XX^\top - \frac{n}{p}zI)^{-1}] \\
&= \frac{1}{p}\mathrm{Tr}[\frac{p}{n}(\frac{1}{n}XX^\top - zI)^{-1}] \\
&= \frac{1}{n}\mathrm{Tr}[(\frac{1}{n}XX^\top - zI)^{-1}] \\
&=: \hat{\varphi}(z)
\end{aligned}
$$

$$
(\hat{\Sigma} - zI_p)^{-1} \sim (-z\hat{\varphi}(z)\Sigma - zI_p)^{-1}
$$

# 4 Further reading and Useful resources

**Symmetric matrix: Wigner, Wishart**

- **Francis Bach, 2023,** `https://arxiv.org/abs/2303.01372`
  - Derivation of Silverstein equation in Appendix A. Very clean and intuitive derivation (Woodbury + leave one out approach). Have the same spirit as the Silverstein 1995 paper, but much easier to follow.
  - Nice derivation of two point equivalence using similar leave one out method.
  - Catalogue the known asymptotic results for double descent for regression.
- **Alex, Jacob, Cengiz, 2024, Scaling and renormalization in high-dimensional regression** `https://arxiv.org/abs/2405.00592` ***
  - Very elegant exposition of the Free probability approach for deriving the self-consistent equation (Sec. II-IV) and deterministic equivalence for empirical covariance matrix. and they used it for regression theory (Sec. V).
- Jacob 2024, Lecture note `https://jzv.io/assets/pdf/am226_generalization_in_ridge_regression_lecture_notes.pdf`

- a simpler and shorter version of the review above.

- **Alex Blake Jacob et al. 2025** Two point deterministic equivalence and its application in linear regression
  - `https://arxiv.org/abs/2408.04607`
  - `https://arxiv.org/abs/2502.05074`
  - Most notes above just relied on one point equivalence, this result goes beyond that.
- **Geoff Pleiss, Topics in Deep Learning Theory (stat547u)** `https://geoffpleiss.com/teaching/stat547u/`
  - Geoff Pleiss's nice and fast RMT Intro. `https://geoffpleiss.com/static/media/stat547u_lecture04.2af15a85.pdf`
    * Introduced for computing spectrum and properties for Ridge Regression, covering Silverstein equation. (though the derivation of Silverstein is a bit jumpy.)
- **Pennington et al. RMT tutorial at ICML 2021** `https://random-matrix-learning.github.io/`
  - Part II mentioned some derivation of MP law and Silverstein equation (*the Woodbury approach*), but still quite jumpy... `https://random-matrix-learning.github.io/#presentation2`
- **Ryan Tibshirani, Advanced Topics in Statistical Learning (stat241B)** `https://www.stat.berkeley.edu/~ryantibs/statlearn-s23/`
  - Very nice note covering the theory for Ridge and Ridgeless case. *** `https://www.stat.berkeley.edu/~ryantibs/statlearn-s23/lectures/ridge.pdf` `https://www.stat.berkeley.edu/~ryantibs/statlearn-s23/lectures/ridgeless.pdf`
    * In the Ridge case, nicely stated the RMT approach and the known results for Ridge regression (though the Silverstein and MP results were cited instead of derived.)
- **Song Mei, Mean Field Asymptotics in Statistical Learning (STAT260)** `https://www.stat.berkeley.edu/~songmei/Teaching/STAT260_Spring2021/schedule.html`
  - Note on random matrices and Stieltjes transforms, *** `https://www.stat.berkeley.edu/~songmei/Teaching/STAT260_Spring2021/Lecture_notes/scribe_lecture17.pdf`
    * Very comprehensively cataloguing many known properties of Stieltjes. Very detailed in covering the derivation of Semicircle law of **Wigner matrix**, using Schur complement + leave one out approach.
  - Note on Replica method for GOE+spike, derivation of BBP phase transition. `https://www.stat.berkeley.edu/~songmei/Teaching/STAT260_Spring2021/Lecture_notes/scribe_lecture8.pdf` `https://www.stat.berkeley.edu/~songmei/Teaching/STAT260_Spring2021/Lecture_notes/scribe_lecture9.pdf`
- **Laszlo Erdos, The matrix Dyson equation** `https://arxiv.org/abs/1903.10060`
  - Nice derivation of the **Wigner** semicircle law in Sec. 3, (using resolvent, leave-one-out and Schur complement approach). The author mentioned this as a special case of the matrix Dyson equation, i.e. scalar Dyson equation.
- Recent blog on spectral density empirical covariance with power-law population spectrum `https://aaronjhf.github.io/blog/power-law-spec/`
  - Code `https://github.com/aaronjhf/Power-Law-Data`
- Other classic works [*not recommended*]
  - Marcheko Pastur, 1967, Russian version, English version. Fundamental work, but a bit hard to read, due to notation differences... The derivation is more general than later works.
  - Silverstein, Bai et al. $\sim$ 1995 seminal works, but the proofs are a bit hard to read due to notations and the results are distributed across a few papers ...
    * On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices$\to$ self consistent equation for the kernel like matrix $XTX^*$
    * Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices`https://jack.math.ncsu.edu/strong.pdf` $\to$ Silverstein equation, for the covariance like matrix $XX^*T$. The leave one out approach is great, though the notation is not super clean ...
    * Proof in their 2010 book is also not easy to read either ...
  - Girko 1985 Russian version, English version
    * Generalize to non symmetric case

**Asymmetric matrix**

- **Terence Tao 2010, RMT Teaching notes** `https://terrytao.wordpress.com/category/teaching/254a-random-ma`
  - Esp. nice for the asymmetric non-Hermitian case, esp. Circular law `https://terrytao.wordpress.com/2010/03/14/254a-notes-8-the-circular-law/`
  - Using the Potential energy
- **Sommers & Haim 1988, Spectrum of Large Random Asymmetric Matrices** `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.60.1895` **PDF link**
  - Classic seminal work on non-symmetric case using Greens function / Stieljes + Replica. Generalizing the circular law to the ellipse law when there is correlation between $J_{ij}$ and $J_{ji}$.
  - Mentioning the nice connection of Stieljes to the electric potential in 2d space.
- **Rajan & Abbott 2006, Random matrix with some all positive and all negative (exc and inh) columns** `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.97.188104` **PDF link**Similar approach as above Greens function / Stieljes + Replica.

# References


[1] Alexander Atanasov, Jacob A Zavatone-Veth, and Cengiz Pehlevan. Scaling and renormalization in high-dimensional regression. *arXiv preprint arXiv:2405.00592*, 2024.

[2] Francis Bach. High-dimensional analysis of double descent for linear regression with random projections. *SIAM Journal on Mathematics of Data Science*, 6(1):26–50, 2024.

[3] Laszlo Erdos. The matrix dyson equation and its applications for random matrices. *arXiv preprint arXiv:1903.10060*, 2019.

[4] Fabian Pedregosa, Courtney Paquette, Tom Trogdon, and Jeffrey Pennington. Random matrix theory for machine learning.

[5] Jack W Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.