Q1

$A = U\Sigma V^T \in \mathbb{R}^{m\times n}$ $\quad$ σ_i - largest singular value of A

$u_i \in U \in \mathbb{R}^m$

$v_i \in V \in \mathbb{R}^n$ $\quad$ $\operatorname{rank}(A) = r$

Theorem: $x_{LS} = V\Sigma^{+}U^T = A^{+}b = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \|Ax-b\|_2^2$

Proof:

Let the SVD of A is $U\Sigma V^T$

$\Rightarrow$ $\underline{U\cdot U^T = I_{m\times m}}$ & $\underline{V\cdot V^T = I_{n\times n}}$ & $\underline{\Sigma \text{ is diagonal with } \sigma_i}$

STEP 1: Lets find x_{LS} by $\nabla\|Ax-b\|_2^2 = 0$.

$A \in \mathbb{R}^{m\times n} \Rightarrow x \in \mathbb{R}^n \Rightarrow Ax \in \mathbb{R}^m \Rightarrow b \in \mathbb{R}^m$

$$\|Ax-b\|_2^2 = (Ax-b)^T(Ax-b) = (Ax)^T\cdot Ax - (Ax)^T b - b^TAx + b^Tb =$$

$$= x^TA^TAx - x^TA^Tb - b^TAx + b^Tb =$$

$$= x^T(U\Sigma V^T)^T(U\Sigma V^T)x - \underbrace{x^T(U\Sigma V^T)^Tb}_{b^T(U\Sigma V^T)x} - b^T(U\Sigma V^T)x + b^Tb =$$

$$= x^TA^TAx - 2b^TAx + b^Tb$$

$$2x^TA^Tb$$

$$\longrightarrow \|Ax-b\|_2^2 = x^TAAx - 2b^TAx + b^Tb$$

/

to minimize this term:

(I) $$\frac{d\|Ax-b\|_2^2}{dx} = \frac{d}{dx}\left[\underbrace{x^TA^TAx}_{(*)} - \underbrace{2x^TA^Tb}_{(**)} + b^Tb\right] = 0$$

Notice:

1. $$\frac{\partial(x^Ty)}{\partial x} = y$$

2. $$\frac{\partial(xy^T)}{\partial x} = \frac{\partial(x^Ty)}{\partial x} = y$$

(*). chain rule, where $y = A^TAx = x^TA^TA$

$$\frac{d(x^TA^TAx)}{dx} = \frac{d(x^Ty)}{dx} + \frac{d(y(x))}{dx}\cdot\frac{\partial(x^Ty)}{\partial y} = y + \frac{d(A^TAx)}{dx}\cdot x =$$

$$= y + AA^T\cdot x = A^TAx + A^TAx = A^TAx + A^TAx = 2A^TAx$$

(I): $2A^TAx - 2A^Tb = 0 \quad \backslash \cdot \frac{1}{2}$

$(U\Sigma V^T)^T\cdot(U\Sigma V^T)\cdot x = (U\Sigma V^T)^T\cdot b \quad \backslash (\cdot)^T$

$x^T\cdot(U\Sigma V^T)^T\cdot(U\Sigma V^T) = b^T\cdot U\Sigma V^T$

$x^T(V\ \Sigma^TU^T)(U\Sigma V^T) = b^T\cdot U\Sigma V^T$

$$x^T V \Sigma^T \Sigma V^T = b^T U \Sigma V^T \quad \Big| \cdot V$$

$$x^T V \Sigma^T \Sigma = b^T \Sigma U \quad \Big| \cdot (\cdot)^T$$

$$\Sigma^T \Sigma V^T x = U^T \Sigma^T b \quad \Big| \cdot [V \cdot \Sigma^{+} \cdot (\Sigma^{+})^T]$$

$$V \Sigma^{+} \cdot (\Sigma^{+})^T \cdot \Sigma^T \cdot \Sigma \cdot V^T x = V \Sigma^{+} \cdot (\Sigma^{+})^T U^T \cdot \Sigma^T b$$

$$\longrightarrow \quad x_{LS} = V \Sigma^{+} U^T b$$

$$\boxed{\text{notice: } \quad A^T \cdot A = (U \Sigma V^T)^T (U \Sigma V^T) = V \Sigma^T U^T \cdot U \Sigma V^T = V \Sigma^T \Sigma V^T}$$

STEP 2: Show that x_{LS} is minimizing the norm

SVD properties of Σ & V

$$\nabla^2 \|Ax - b\|_2^2 = 2A^T A = 2V \Sigma^T \Sigma V^T \geq 0$$

$\Rightarrow$ the norm is a convex function.

Hence,

x_{LS} is a minimum

/

STEP 3 : Lets show that $V\Sigma^{+}U^{T}b = A^{+}b$

$$\text{let} \quad A \overset{\text{SVD}}{=} \sum_{i}^{n} \sigma_i u_i v_i^T = U \Sigma V^T$$

$$A^{+} \triangleq (A^T A)^{-1} \cdot A^T = (V \Sigma^T \Sigma V^T)^{-1} (U \Sigma V^T)^T =$$

$$= (V \Sigma^T \Sigma V^T)^{-1} \cdot V \Sigma^T U^T = V\left[\Sigma^T \Sigma\right]^{-1} \Sigma^T U^T \underset{\substack{\uparrow \\ \Sigma^{+} = \left[\Sigma^T \Sigma\right]^{-1} \cdot \Sigma^T}}{=} V \Sigma^{+} U^T$$

$$\Rightarrow A^{+} = \sum_{i}^{n} \sigma_i^{-1} u_i v_i^T = V \Sigma^{+} U^T$$

$$\Rightarrow \quad A^{+} b = V \Sigma^{+} U^T \cdot b$$

$\boxtimes$

/

1

Q2 **Theorem:** $\operatorname{rank}(A) = \operatorname{rank}(\Sigma)$

Proof:

$A \in \mathbb{R}^{m\times n}$ & $A = U\Sigma V^T$

$\Rightarrow U \in \mathbb{R}^m$, $V \in \mathbb{R}^n$

I $\Rightarrow \Sigma \in \mathbb{R}^{m\times n}$

$\longrightarrow \operatorname{rank}(\Sigma) \leq \min\{m, n\}$

II $\operatorname{rank}(A) = \operatorname{rank}(U\Sigma V^T) \leq \min\{\operatorname{rank}(U), \operatorname{rank}(\Sigma), \operatorname{rank}(V)\}$

Denote,
u and v are orthogonal vectors, Hence they are linear independent. which means $\det(U) \neq 0$ & $\det(V) \neq 0$ $\longrightarrow$ non-Singular
$\longrightarrow$ U & V has full rank

$\Rightarrow$ rank $U = m$, $U \in \mathbb{R}^m$
rank $V = n$, $V = \mathbb{R}^n$

l

according to I & II

if $\operatorname{rank}(\Sigma) \leq \min\{m, n\}$

So we can say that

$$\min\{\underset{\underset{m}{\leq}}{\operatorname{rank}(U)}, \underset{\underset{\min\{m,n\}}{\leq}}{\operatorname{rank}(\Sigma)}, \underset{\underset{n}{\leq}}{\operatorname{rank}(V^T)}\} = \operatorname{rank}(\Sigma)$$

Since Σ is diagonal matrix, its rank will be its non zero element

III

Q_3

a. **Theorem**: $\text{range}(A) = \text{Span}(u_1, \dots, u_r)$

Proof:

Let $A \in \mathbb{R}^{m \times n}$, then range(A) is the set of all vectors Ax, $x \in \mathbb{R}^n$.

Furthermore, Ax can be written:

$$Ax = a_1 x_1 + a_2 x_2 + \dots a_n x_n = \sum_{i=1}^{n} a_i x_i$$

Therefore, it is represents all the linear combination of the column vector a_i, $i=1,\dots,n$.

That is

$$\text{range}(A) = \{Ax : x \in \mathbb{R}^n\} = \text{span}\{a_1, \dots a_n\}$$

Now, we need to find a specific linea combination for $\text{Span}(u_1, \dots u_r)$

Define,

$$y \triangleq Ax = U \Sigma V^T x$$

$$h \triangleq V^T x \quad , \quad r \triangleq \text{rank}(\Sigma) \quad , \quad \sigma_{ii} \triangleq \sigma_i$$

$$\longrightarrow \quad y = U\Sigma h = U\sum_{i=1}^{r}\sigma_i h_i = \sum_{i=1}^{r}(\sigma_i h_i)u_i =$$

$$= \sigma_1 h_1 \cdot u_1 + \sigma_2 h_2 \cdot u_2 + \dots + \sigma_r h_r \cdot u_r =$$

$$= span(u_1, \dots u_r)$$

we found a linear combinatio that defines the span of U by A, which mean that the vector u defines the space of the given Matrix A.

b. Theorem: $Null(A) = Span(V_{r+1}, \dots, V_n)$

proof:

Let $b \; \mathbb{R}^{n \times 1} \neq 0$

$A \cdot b = 0$

$\longrightarrow \quad U\Sigma V^T \cdot b = \bar{0} \quad \big| \cdot U^T$

$\Sigma V^T \cdot b = \bar{0}$

$\Sigma \in \mathbb{R}^r \neq \bar{0} \rightarrow b \in \mathbb{R}^r = \bar{0}$

$\Sigma \in \mathbb{R}^{[r+1, n]} = \bar{0} \rightarrow b \in \mathbb{R}^{[r+1, n]} \neq \bar{0}$

$\Rightarrow \quad b = (0, \dots 0, b_{r+1} \dots b_n)$

$h \mathrel{\hat{=}} V^T b \quad \rightarrow \quad b = V \cdot h = \sum_{i=r+1}^{n} v_i h_i = span(V_{r+1}, \dots V_n)$

1

Q4 Theorem : $\inf_{\substack{B\in\mathbb{R}^{m\times n}\\ \mathrm{rank}(B)\le s}} \|A-B\|_2 = \sigma_{s+1}$

Proof :

Denote

1. The strategy for proof is contradiction.

2. $S \le \mathrm{rank}(A) = r$

① Suppose for every $0 < s < \mathrm{rank}(A)$ there is the best approximation matrix

$$A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^T \quad , \text{ which } \mathrm{rank}(A_s) = s$$

② Suppose, for contradiction, that there is some matrix B with $\mathrm{rank}(B) \le s$.

$$\text{s.t } \|A-B\|_2 < \|A-A_s\|_2 = \sigma_{s+1} .$$

③
$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$$B = U\Sigma V^T = \sum_{k=1}^{s} \sigma_i u_i v_i^T$$

1

$$A-B = \sum_{i=1}^{r} \sigma_i u_i v_i^T - \sum_{k=1}^{s} \sigma_i u_i v_i^T = \bar{0} + \sum_{i=s+1}^{r} \sigma_i u_i v_i^T$$

$$\longrightarrow A-B = \sum_{i=s+1}^{r} \sigma_i u_i v_i^T$$

we know that $\sigma_1 > \sigma_2 > \dots > \sigma_n$

and the 2-norm of sum of 1-rank matrices is the first σ_i.

$\Rightarrow$ in our case σ_{s+1} is the largest singular value of $A-B$.

4 Since, $n = \text{rank}(B) + \text{null}(B)$

we can say, based on previous questions that:

$$n \leq s + \text{null}(B) \longrightarrow \text{null}(B) \geq n-s$$

so we defined the null space of B:

$\text{null}(B)$ s.t its Dim is bigger than $n-s$

5. Let $x \in X$ be a unity vector from null(B) space.

s.t: $B \cdot x = 0$, $\|x\|_2 = 1 \neq 0$

$$\rightarrow \|Ax\|_2 = \|(A-B)x\|_2 \leq$$

according to 3 $\leq \|A-B\|_2 \cdot \|x\|_2$

$$< \sigma_{s+1} \|x\|_2 = \sigma_{s+1}$$

6. Now consider the space span $(v_1, \dots, v_{s+1})$.
Since σ_{v+1} is the smallest singular value associated with $v_1 \dots v_{r+1}$,
we know that

$$y \in \mathrm{Span}(v_1, \dots v_{s+1})$$

s.t $\|Ay\|_2 \geq \sigma_{s+1} \|y\|_2$

Summary:

I for non-zero $x \in \ker B$, we know

$$\|Ax\|_2 < \sigma_{s+1} \|x\|_2 = \sigma_{s+1}$$

II for $y \in \mathrm{span}(v_1, \dots, v_{s+1})$

$$\|Ay\|_2 \geq \sigma_{+1} \|y\|_2$$

/

$\underline{\text{III}}$

$\text{Dim}(\text{null}(B)) \geq n - s$

$\text{Dim}(\text{span}(v_1, \dots v_{s+1})) = s+1$

The sum of those 2 dimensions will exceed n.
So, there must be non-zero vectors common to both spaces.

But there can't be a single non zero vector that satisfies both the inequalities above.

This is the contradiction

Prove that

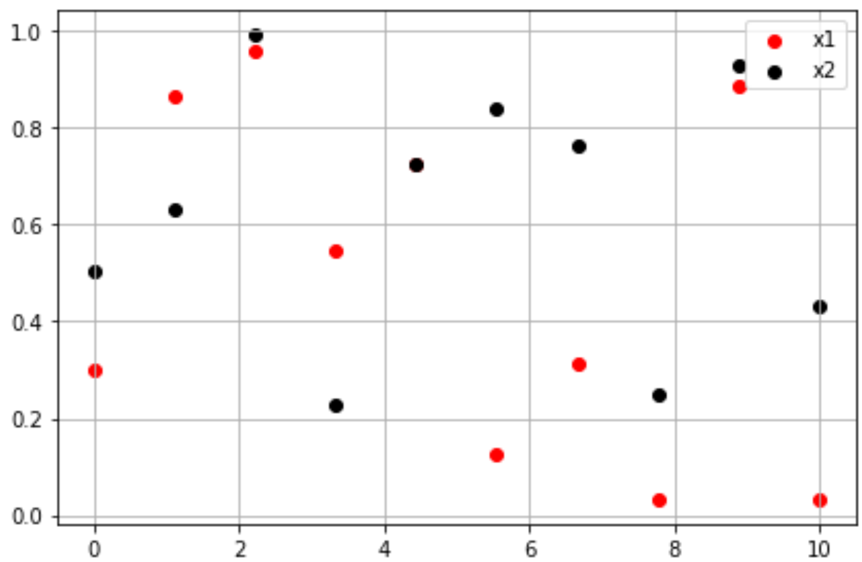
$$\inf_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq s}} \|A - B\|_2 = \sigma_{s+1}$$

$\square$

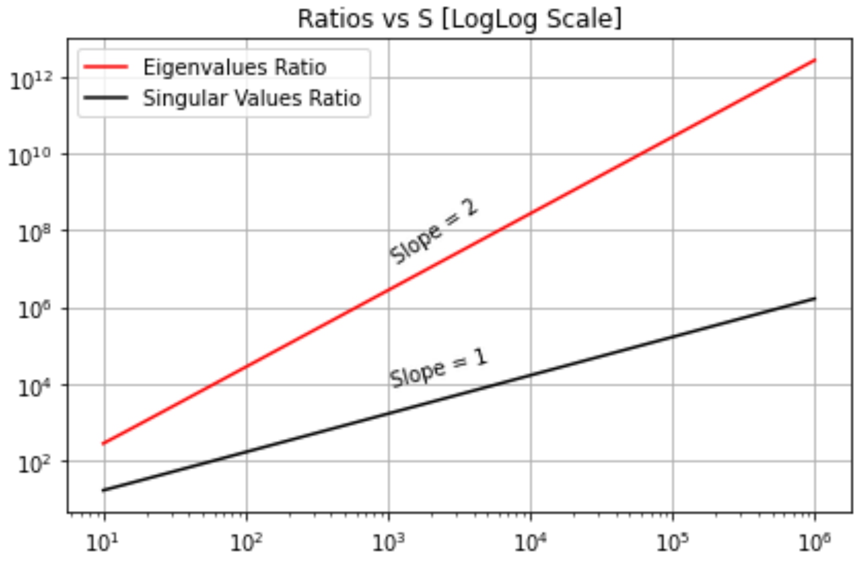
Mathematical Method for DS and PS

HW 1 - Question 5

```
import numpy as np
import matplotlib.pyplot as plt

# Drawing 2 random signals x1, x2 with dimention p

p = 10

x1 = np.random.rand(1,p)
x2 = np.random.rand(1,p)

x = np.linspace(0,p,p)

plt.scatter(x, x1.transpose(), c='r', label='x1')
plt.scatter(x, x2.transpose(), c='k', label='x2')

plt.grid()
plt.legend()
plt.tight_layout()
plt.show()
```



```
# Generate n observations of y

Y = []
n = 1000 # observations
mu = 0
sig = 1

a = np.random.normal(mu, sig, size = (n,1))
b = np.random.normal(mu, sig, size = (n,1))

pts = n//p
S = np.array(np.logspace(1,6,num=pts))
eigen_ratio = np.zeros(pts)
sing_ratio = np.zeros(pts)

for i,s in enumerate(S):

  Y = s * a.dot(x1) + b.dot(x2)

  # Covariance, Eigenvalues, Singular Values
  covar = np.cov(Y, rowvar = False)
  eigen_vals = np.sort(np.linalg.eigvals(covar))[::-1]
  sing_vals = np.linalg.svd(Y, compute_uv=False)

  # Ratio No.1
  eigen_ratio[i] = eigen_vals.real[0] / eigen_vals.real[1]

  # Ratio No.2
  sing_ratio[i] = sing_vals[0] / sing_vals[1]

plt.loglog(S, eigen_ratio, 'r', label='Eigenvalues Ratio')
plt.loglog(S, sing_ratio, 'k', label='Singular Values Ratio')

slope1, intercept1 = np.polyfit(np.log(S), np.log(eigen_ratio), 1)
slope2, intercept2 = np.polyfit(np.log(S), np.log(sing_ratio), 1)

plt.text(10e2, 10e7, "Slope = %.0f" %slope1, size=10, rotation=17*slope1,
         ha="left", va="center")

plt.text(10e2, 2.5*10e3, "Slope = %.0f" %slope2, size=10, rotation=15*slope2,
         ha="left", va="center")

plt.title('Ratios vs S [LogLog Scale]')
plt.grid()
plt.legend()
plt.tight_layout()
plt.show()
```



Conclusions

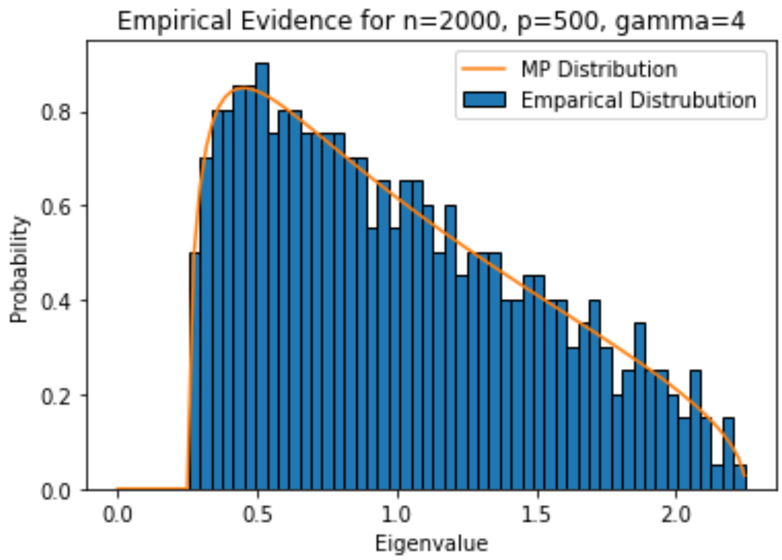
1. We can see that the slope of the ratio of the eigenvalues is 2 times greater than the ratio of the singular values. This result makes sense because a singular value is also known as the positive square root of the eigenvalue. Since our graph is on a logarithmic scale, this square root is represented as a multiple of 2.
2. The eigenvalues are very large and could be affected by numerical errors, so that it will be better to use the SVD.

   Note that the slopes of the graphs are rounded but very close to the values represented.

HW 1 - Question 6 - a

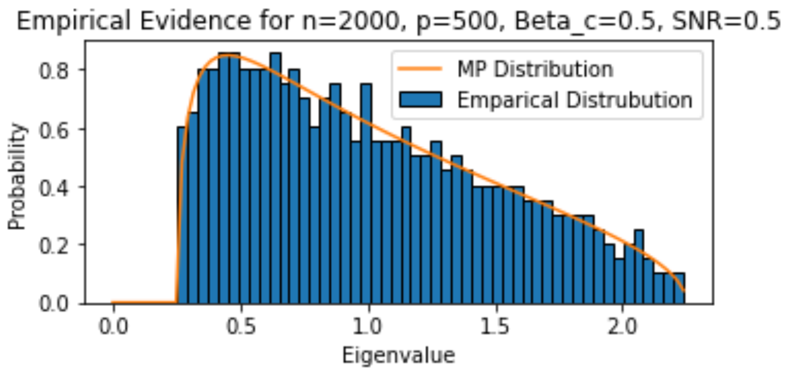
```
import numpy as np

def marchenko_pastur_mu(x, gamma, sigma2=1):
    x = np.atleast_1d(x).astype(float)
    gamma_p = sigma2 * (1 + np.sqrt(gamma)) ** 2
    gamma_m = sigma2 * (1 - np.sqrt(gamma)) ** 2
    mu = np.zeros_like(x)
    is_nonzero = (gamma_m < x) & (x < gamma_p)
    x_valid = x[is_nonzero]
    factor = 1 / (2 * np.pi * sigma2 * gamma)
    mu[is_nonzero] = factor / x_valid
    mu[is_nonzero] *= np.sqrt((gamma_p - x_valid) * (x_valid - gamma_m))
    if gamma > 1:
        mu[x == 0] = 1 - 1 / gamma
    return mu
```

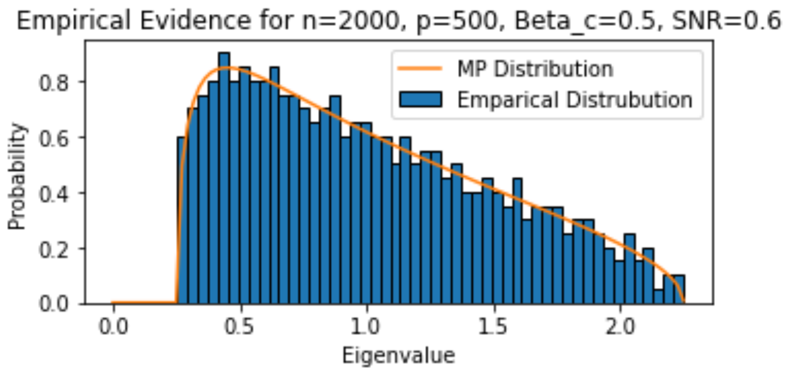
```
import numpy as np

p = 500   # feature length
n = 2000  # examples (observations)

mu = np.asarray([0] * p)
I = np.identity(p)

X = np.random.multivariate_normal(mu, I, size=n)

covar = np.cov(X, rowvar=False)
eigen = np.sort(np.linalg.eigvals(covar)).real
x_axis = np.linspace(0, np.max(eigen), 100)

plt.hist(eigen, 50, density=True, edgecolor='k', label='Emparical Distrubution')
plt.plot(x_axis, marchenko_pastur_mu(x_axis, gamma=p/n), label='MP Distribution')

plt.grid(False)
plt.legend()
plt.title('Empirical Evidence for n={}, p={}, gamma={}'.format(n,p,n//p))
plt.xlabel('Eigenvalue')
plt.ylabel('Probability')


plt.show()
```

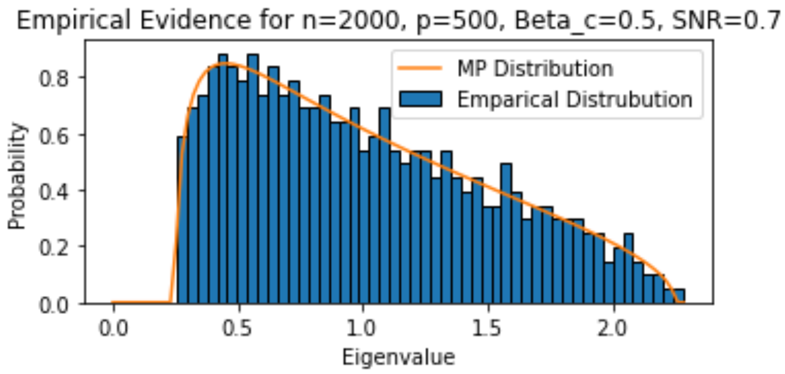

Empirical Evidence for n=2000, p=500, gamma=4
MP Distribution
Emparical Distrubution
Probability
Eigenvalue
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0

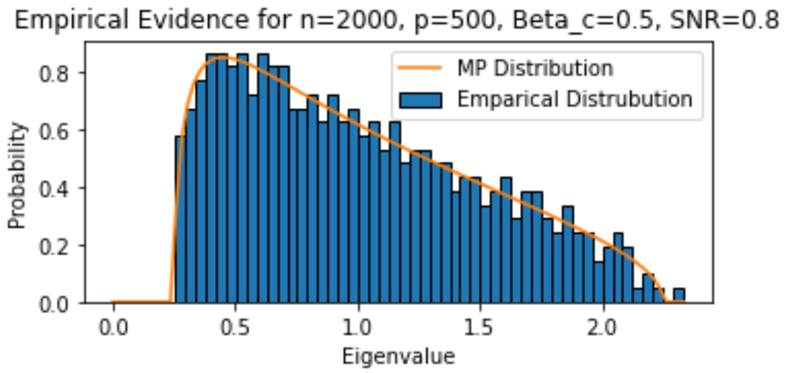
HW 1 - Question 6 - b

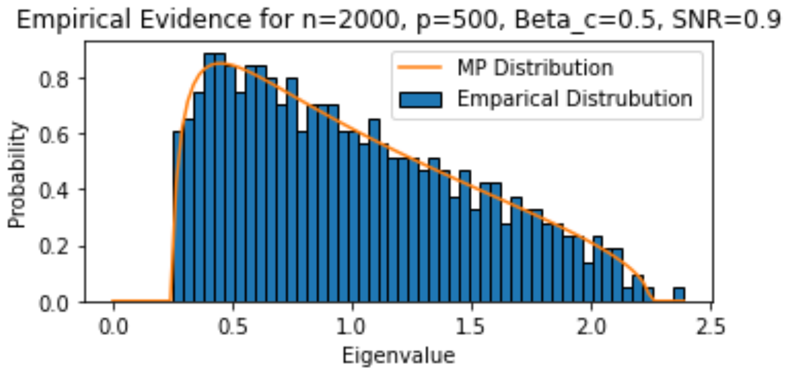
```
n, p = 2000, 500

beta_c = np.sqrt(p/n)

Beta = np.linspace(beta_c,0.9,5)

fig, ax = plt.subplots(len(Beta), figsize=(5.5, 13))

g = np.random.normal(0, 1, [1, n])
g = np.tile(g, (p, 1))
g0 = np.random.normal(0, 1, [p, n])
u = np.random.normal(0, 1, [p, 1])
u = u / np.linalg.norm(u)
u_n = np.tile(u, (1, n))

for i,beta in enumerate(Beta):

  X_spike = np.sqrt(beta) * g * u + g0
  X_spike = X_spike.T

  covar_spike = np.cov(X_spike, rowvar=False)
  eigen_spike = np.sort(np.linalg.eigvals(covar_spike))

  x_axis = np.linspace(0, np.max(eigen_spike), 100)

  ax[i].hist(eigen_spike, 50, density=True, edgecolor='k', label='Emparical Distrub
  ax[i].plot(x_axis, marchenko_pastur_mu(x_axis, gamma=p/n), label='MP Distribution

  ax[i].grid(False)
  ax[i].legend()
  ax[i].set_title('Empirical Evidence for n={}, p={}, Beta_c={}, SNR={}'.format(n,p
  ax[i].set_xlabel('Eigenvalue')
  ax[i].set_ylabel('Probability')

plt.tight_layout()
plt.show()
```


Empirical Evidence for n=2000, p=500, Beta_c=0.5, SNR=0.5
MP Distribution
Emparical Distrubution
Probability
Eigenvalue


Empirical Evidence for n=2000, p=500, Beta_c=0.5, SNR=0.6
MP Distribution
Emparical Distrubution
Probability
Eigenvalue


Empirical Evidence for n=2000, p=500, Beta_c=0.5, SNR=0.7
MP Distribution
Emparical Distrubution
Probability
Eigenvalue


Empirical Evidence for n=2000, p=500, Beta_c=0.5, SNR=0.8
MP Distribution
Emparical Distrubution
Probability
Eigenvalue


Empirical Evidence for n=2000, p=500, Beta_c=0.5, SNR=0.9
MP Distribution
Emparical Distrubution
Probability
Eigenvalue

Conclusions

The minimal SNR (β) value which we can see the signal is ~0.7-0.8 which is greater than the theorical value (βc) that we expecterd to see in the theorical case and this is probably because in our case n is not infinity.

3s completed at 10:37 AM