EECS 551 Homework 6 YUZHAN JIANG

P1:

(a) We are given that A has full column rank, let $A \in F^{m\times n}$

let $A = U_r \Sigma_r V_r'$ by SVD of A

$$A'A = (U_r \Sigma V_r')'(U_r \Sigma_r V_r')$$
$$= V_r \Sigma' U_r' U_r \Sigma_r V'$$
$$= V_r \Sigma_r' \Sigma_r V_r'$$

$\therefore \sigma_i(A'A) = \sigma_i^2(A)$

Condition number of $A'A$ cond$(A'A)$:

$$\frac{\sigma_1(A'A)}{\sigma_n(A'A)}$$
$$= \frac{\sigma_1^2(A)}{\sigma_n^2(A)}$$
$$= \frac{\sigma_1^2}{\sigma_n^2}$$

where $\sigma_1, \sigma_2, \ldots \ldots \sigma_n$ are the singular values of A

σ_1 and σ_n are the maximum and minmum singular value of A

(b)

$$\hat{x} = (A'A + \beta I)^{-1} A'y$$

Based on part (a):

$$A'A + \beta I = V_r \Sigma_r' \Sigma_r V_r' + \beta I$$
$$= V_r (\Sigma_r' \Sigma_r + \beta I) \cdot V_r'$$

$\therefore \sigma_i(\Sigma_r' \Sigma_r + \beta I) = \sigma_i^2(A) + \beta^2$

Thus, condition number of $A'A + \beta I = \dfrac{\sigma_1^2 + \beta^2}{\sigma_n^2 + \beta^2}$

Now, let Cond$(A'A)$ - Cond$(A'A + \beta I)$

$$= \frac{\sigma_1^2}{\sigma_n^2} - \frac{\sigma_1^2 + \beta^2}{\sigma_n^2 + \beta^2} = \frac{\sigma_1^2 \sigma_n^2 + \sigma_1^2 \beta^2 - \sigma_1^2 \sigma_n^2 - \sigma_n^2 \beta^2}{\sigma_n^2(\sigma_n^2 + \beta^2)} = \frac{(\sigma_1^2 - \sigma_n^2)\beta^2}{\sigma_n^2(\sigma_n^2 + \beta^2)} > 0 \quad (\text{Since } \sigma_1^2 > \sigma_n^2)$$

$\therefore$ Condition number of $A'A$ is greater than Condition number of $A'A + \beta I$

Thus, the regularized solution has a "better" condition number

P2:

(a) Consider $A_1 = [0]$, $N \le M$, and A_1 is not a frame

Consider $A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$,

$A_1 A_1' = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$

$\sigma_1 = 2$ and $\sigma_2 = 1$ $\quad \sigma_1 \neq \sigma_2 \Rightarrow \alpha \neq \beta$

$\therefore$ A_2 is a frame but not a tight frame

Consider $A_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

$\sigma_1 = \sigma_2 = 2$, and $\alpha = \beta = \sigma_1^2 = 4$ and $\alpha = \beta \neq 1$

$\therefore$ A_3 is a tight frame

Consider $A_4 = [1 \quad 1]/\sqrt{2}$

$A_4 \cdot A_4' = [1 \quad 1]/\sqrt{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \sqrt{2} = [1] = I_1$

$\therefore$ A_4 is a Parseval frame but A_4 is not unitary

(b) The necessary and sufficient conditions for a diagonal Matrix $D = diag(d_{11}, d_{22} \dots d_{nn})$ to be matrix is that

$d_{11}, d_{22}, d_{33} \cdots d_{nn}$ should have unit norm

P3:

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|_2}$$

(a) We are given that $A \in F^{n \times n}$ has n different eigenvalue (in magnitude).

$$\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_N$$

Since λ_1 is known, given $Ax = \lambda x$ and λ_1 is the largest eigenvalue, then we have

$[A - \lambda_1 I] x = \alpha x$ where α is the eigenvalue of the shifted matrix $A - \lambda_1 I$, which is $0, \lambda_2 - \lambda_1, \lambda_3 - \lambda_1, \cdots, \lambda_n - \lambda_1$

Consider $\alpha_n = \lambda_N - \lambda_1$

Then, we can use one run of power method to get λ_N

(b) The "sign ambiguity" does not affect the calculation of λ_N

P4.

We are given that matrix $X = [x_1 \cdots\cdots x_N]$ has full row rank.

For finding W to minimize the average loss over the training data:

$$\hat{W} = \underset{W}{\arg\min}\ L(W) \text{ where } L(w) = \frac{1}{N}\sum_{n=1}^{N} L(y_n, \hat{y}_n) \quad \hat{y} = Wx_n$$

$$L(y_n, \hat{y}_n) = \frac{1}{2}(y_n - \hat{y}_n)^2$$

$$\Rightarrow \quad \hat{W} = \underset{W}{\arg\min}\ \frac{1}{N}\sum_{n=1}^{N} \frac{1}{2}(y_n - W'x_n)^2$$

let $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in R^N$

$$\hat{W} = \underset{W}{\arg\min}\ \frac{1}{N}\sum_{n=1}^{N} \|y_n - W'x_n\|_2^2$$

$$= \frac{1}{N}\|y' - W'X\|_2^2$$

$$= (XX')^{-1}Xy \quad \text{Since } X \text{ is a full row rank matrix}$$

Since the training data feature correlation matrix $K_X = \frac{1}{N}\sum_{n=1}^{N} x_n x_n'$ and the cross-correlation between the training data features and responses $K_{yx} = \frac{1}{N}\sum_{n=1}^{N} y_n x_n'$

$$\therefore \hat{X} = (XX')^{-1}Xy$$

$$= K_X^{-1} K_{yX}$$

Conditions needed. $N \geq M$ So that K_X is invertible

P5:

(a) We assume that there exist a non-zero X such that

$$\|BX\|_2 = 0 = X'B'BX = 0$$

$\Rightarrow$ $BX = 0$ which is contradicted that B has full col rank

$\therefore$ X doesn't exist

$\therefore$ $B'B \succ 0$

(b) By the def of positive definite,

if $A \succ 0$, then all eigenvalues of $A > 0$

$\therefore$ $\lambda \neq 0 \Rightarrow A$ has full rank $\Rightarrow A$ is invertible

(c) show that $A \succeq 0$ and $B \succeq B \Rightarrow A+B \succeq 0$

Since there exist $x \neq 0$ such that

$$X'AX + X'BX = X'(A+B)X \geq 0$$

$\therefore$ $A+B \succeq 0$

(d) Show that $A \succ 0$ and $B \succeq 0 \Rightarrow A+B \succ 0$

Since there exist non-zero X Such that

$$X'AX + X'BX = X'(A+B)X > 0$$

$\Rightarrow$ $A+B \succ 0$

(e) If B has full column rank, then based on part (a) $B'B \succ 0$

And we know that $A'A \succeq 0$, based on part (d)

$A'A + B'B > 0$

$\therefore$ $A'A + B'B$ is invertible by part (b)

5 (f) If $N(A) \cap N(B) = \{0\}$, then $A'A + B'B$ is invertible

Assume $A'A + B'B \geq 0$ so there exist non-zero x, such that

$x'(A'A + B'B)x = 0$

$x'A'Ax + x'B'Bx = 0$

$\|Ax\|_2 + \|Bx\|_2 = 0$

$\therefore$ $Ax = Bx = 0$

$\therefore$ x is nullspace for both A and B

$\therefore$ $N(A) \cap N(B) \neq \{0\}$ contradict. with $N(A) \cap N(B) = \{0\}$

$\therefore$ x doesn't exist. and $A'A + B'B > 0$ and it is invertible

P6:

(a) $\hat{x} = \underset{x=(x_1,\cdots x_K)\in F^K}{\operatorname{argmin}} \left\| \left(\sum_{k=1}^{K} x_k A_k \right) - B \right\|_F$

$= \underset{x}{\operatorname{argmin}} \left\| vec\left(\sum_{k=1}^{K} x_k A_k \right) - vec(B) \right\|_2$

$= \underset{x}{\operatorname{argmin}} \| vec(A_1)x_1 + Vec(A_2)x_2 + \cdots + Vec(A_K)x_K - vec(B) \|_2$

$= \operatorname{argmin} \| [vec(A_1) + vec(A_2) + \cdots + Vec(A_K)]x - vec(B) \|_2$

$= \underset{x}{\operatorname{argmin}} \| \tilde{A}x - \hat{B} \|_2$ where $A = vec(A_1) + vec(A_2) + \cdots + Vec(A_K)$

$\hat{B} = vec(B)$

$\therefore \hat{x} = \tilde{A}^+ b$

(b) $\tilde{A}$ need to have full column rank to make sure a unique solution.

(c) When $k=3$, Julia code should be:

$\hat{x} = Pinv([vec(A_1)\ vec(A_2)\ \ vec(A_3)]) * vec(B)$

P7.

(a) $Z = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and let $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

$Zx = \begin{bmatrix} 0 \\ x_2 \\ 2x_3 \\ 0 \end{bmatrix} = 0 \quad \Rightarrow \quad x_2 = 0$ and $x_3 = 0$

$\therefore \quad X = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

$\therefore \quad \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is the orthonormal basis for the $N(Z)$

(b) $N^{\perp}(Z) = R(Z') = R(Z) = \text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$

(Z is diagonal matrix)

$\therefore \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is the orthonormal basis for the $N^{\perp}(Z)$

(c)

$U = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

$P^{\perp}_{N(Z)} X = UU'x = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$

$= \begin{bmatrix} 0 \\ 2 \\ 3 \\ 0 \end{bmatrix}$

(d) $Y = J_3 J_3'$ and $W = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

By using Julia:

The orthonormal basis for the nullspace of Y is $\left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} / \sqrt{2}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} / \sqrt{6} \right\}$

the orthonormal basis for the orthogonal complement of the nullspace of Y is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{3} \right\}$ and the projection is $\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$

7(e)

orthcompnull.jl

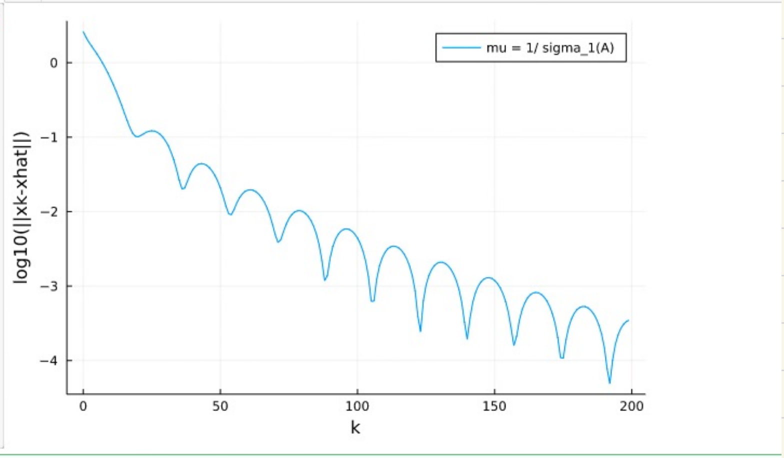
```julia
"""
R = orthcompnull(A, X)
Project each column of `X` onto the orthogonal complement of the null space
of the input matrix `A`.
In:
* `A` `M × N` matrix
* `X` vector of length `N`, or matrix with `N` rows and many columns
Out:
* `R` : vector or matrix of size ??? (you determine this)
For full credit, your solution should be computationally efficient!
"""

using LinearAlgebra
function orthcompnull(A, X)
    r = rank(A)
    u, s, v = svd(A)
    vr = v[:, 1:r]
    R = vr * (vr' * X)
    return R
end
```

P8:

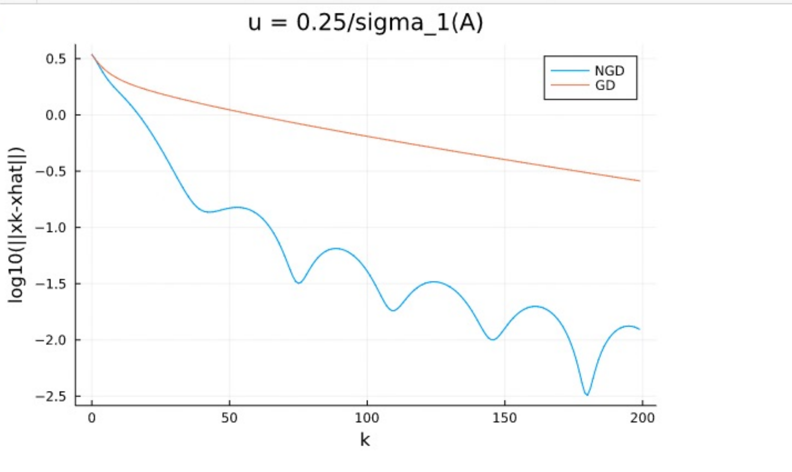
(b)

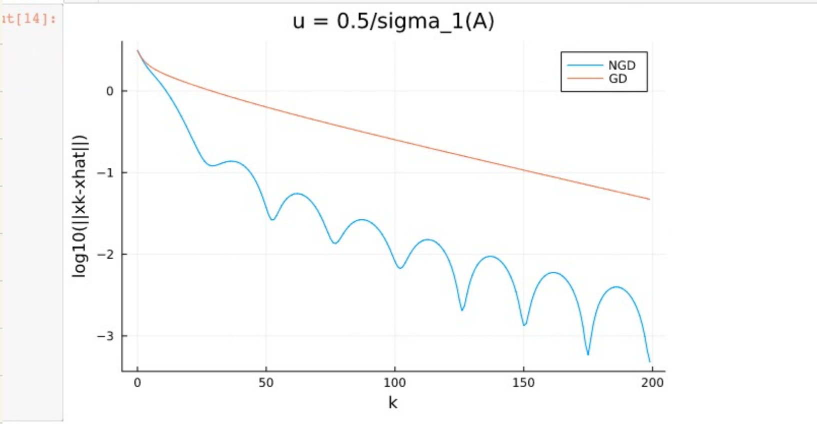
Out[9]:

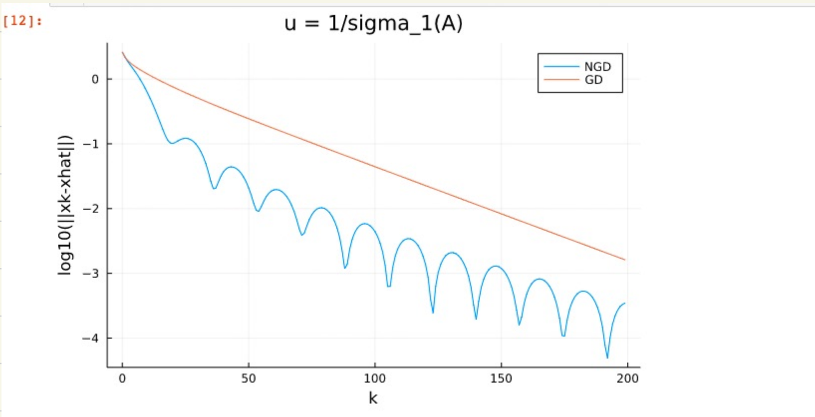


(c)

Out[13]:







As the graph shown, I take $u = \frac{0.25}{\sigma_1^2(A)}$, $\frac{0.5}{\sigma_1^2(A)}$ and $\frac{1}{\sigma_2^2(A)}$

It is clearly to conclude that NGD converges faster than Standard GD