Solution to Assignment 0

BY XIAONAN PENG

Part I

Problem 1. Skipped.

Problem 2.

a) $N(A^TA)=N(A)$

Prove: We want to prove two matrices have the same nullspace is equivalent to prove the following two equations have same solution set:

$$A^TAx=0, Ax=0$$

One side: $Ax=0 \Rightarrow A^TAx=0 \Rightarrow N(A)\subset N(A^TA)$

The other side: $A^TAx=0 \Rightarrow x^TA^TAx=0 \Rightarrow \|Ax\|_2^2=0 \Rightarrow Ax=0 \Rightarrow N(A^TA)\subset N(A)$

Thus, we proved they have the same nullspace.

b) $C(A^TA)=C(A^T)$

Prove: Suppose $y\in C(A^TA)$, then $\exists x$ s.t. $y=A^TAx=A^T(Ax)$, this implies that $y\in C(A^T)\Rightarrow C(A^TA)\subset C(A^T)$.

From (a), we know $n-r(A^TA)=n-r(A)\Rightarrow r(A^TA)=r(A)=r(A^T)$, thus the two column spaces have the same dimension. We have $C(A^TA)=C(A^T)$.

c) $r(A)=r(A^TA)=r(AA^T)$

Prove: From (b) we know $r(AA^T)=r(A)=r(A^T)=r(A^TA)$.

Problem 3. A is SPD

a) A is non-singular.

Prove: If A is singular, $\exists x\neq 0$ s.t. $Ax=0\Rightarrow x^TAx=0$, contradicting to positive definite.

b) All eigenvalues of A are positive.

Prove: $Ax=\lambda x\Rightarrow x^TAx=\lambda\|x\|_2^2>0\Rightarrow \lambda>0$.

c) $\exists$ full column rank matrix R s.t. $A=R^TR$

Prove: A is real symmetric, it has an eigenvalue decomposition: $A=UDU^T$, from (b) we know that all entries of D are positive, then we can define $\sqrt{D}=\mathrm{diag}(\sqrt{\lambda_i})_{i=1}^n$. Then we have

$$A=U\sqrt{D}\sqrt{D}U^T=(\sqrt{D}U^T)^T(\sqrt{D}U^T)=R^TR$$

Problem 4.

Prove: A is symmetric, then we have $A=UDU^T=\sum_i\lambda_iu_iu_i^T$, where $U=[u_1,\ldots,u_n]$.

$$\begin{aligned}
x^TAx &= x^T\sum_i\lambda_iu_iu_i^Tx\\
&= \sum_i\lambda_ix^Tu_iu_i^Tx\\
&\leqslant \sum_i\lambda_{\max}x^Tu_iu_i^Tx\\
&= \lambda_{\max}x^T\left(\sum_i u_iu_i^T\right)x\\
&= \lambda_{\max}x^TUU^Tx\\
&= \lambda_{\max}x^Tx
\end{aligned}$$

Thus, we have $\frac{x^TAx}{x^Tx}\leqslant\lambda_{\max}$, when x is the eigenvector corresponding to $\lambda_{\max}$, the inequality attains equal. Thus we have

$$\lambda_{\max}=\sup_x\frac{x^TAx}{x^Tx}=\sup_{\|x\|_2=1}x^TAx$$

Part II

Problem 5.

a)#nonzero eigenvalues of $A^TA = r(A)$

Prove: Suppose $A^TA = VDV^T$, since V is invertible, $r(A^TA) = r(D) =$ #nonzero eigenvalues, and we know that $r(A) = r(A^TA)$, so #nonzero eigenvalues of $A^TA = r(A)$.

b) A^TA and AA^T share the same nonzero eigenvalues.

Prove: From (a), we know that A^TA and AA^T have the same number of nonzero eigenvalues, we just need to show each eigenvalue of A^TA is also an eigenvalue of AA^T.

Suppose $A^TAx = \lambda x$, then $AA^T(Ax) = \lambda(Ax)$, λ is also an eigenvalue of AA^T with eigenvector Ax.

The spectrum of $A^TA = \{\lambda_1, \ldots, \lambda_r, 0_{r+1}, \ldots, 0_n\}$

The spectrum of $AA^T = \{\lambda_1, \ldots, \lambda_r, 0_{r+1}, \ldots, 0_m\}$

Since A^TA is semi-positive definite, there is no negative eigenvalues. Thus we can write $\lambda_i = \sigma_i^2$.

c) Suppose $A^TAv_i = \sigma_i^2 v_i, \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, let $u_i = \frac{Av_i}{\sigma_i}, i = 1, \ldots, r$.

$$AA^Tu_i = AA^TAv_i\frac{1}{\sigma_i} = \sigma_i Av_i = \sigma_i^2 u_i$$

Thus, u_i is eigenvector of AA^T corresponding to eigenvalue σ_i^2. We call v_i, u_i as singular vectors.

Up to now, we have an orthonormal set $\{v_1, \ldots, v_r\}$ in the rowspace of A, orthonormal set $\{u_1, \ldots, u_r\}$ in the columnspace of A. We need extend the set to a basis.

Let $v_{r+1}, \ldots, v_n$ be the last $n-r$ eigenvectors of A^TA corresponding to eigenvalue 0, $u_{r+1}, \ldots, u_m$ be the last $m-r$ eigenvectors of AA^T corresponding to eigenvalue 0.

d)**Algorithm:**

$\lambda, V = \text{eigen}(A^TA); \lambda, U = \text{eigen}(AA^T); r = \text{rank}(A)$

for $i = 1, \ldots, r$:

  $\sigma_i = \text{sqrt}(\lambda_i), u_i = \frac{Av_i}{\sigma_i}$

if $m > n$:

  $\Sigma = [\text{diag}(\sigma_1, \ldots, \sigma_n), 0_{m-n})^T$

else:

  $\Sigma = [\text{diag}(\sigma_1, \ldots, \sigma_m), 0_{n-m})^T$

Problem 6.

a)

$$\begin{aligned}
\frac{\partial}{\partial\beta}\left(\frac{1}{2}\|X\beta - y\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2\right) &= X^T(X\beta - y) + \lambda\beta \\
&= (X^TX + \lambda I)\beta - X^Ty
\end{aligned}$$

b)

$$\begin{aligned}
\left(\frac{\partial \text{tr}(AX)}{\partial X}\right)_{i,j} &= \frac{\partial \text{tr}(AX)}{\partial x_{ij}} \\
&= \frac{\partial \sum_k \sum_i a_{ik}x_{ki}}{\partial x_{ij}} \\
&= a_{ji} \\
\Rightarrow \frac{\partial \text{tr}(AX)}{\partial X} &= A^T
\end{aligned}$$

c)

$$\begin{aligned}
\frac{\partial \mathrm{tr}(AXB)}{\partial X} &= \frac{\partial \mathrm{tr}(BAX)}{\partial X} \\
&= A^T B^T
\end{aligned}$$

d)

$$\begin{aligned}
\left(\frac{\partial \mathrm{tr}(AX^{-1}B)}{\partial X}\right)_{i,j} &= \frac{\partial \mathrm{tr}(AX^{-1}B)}{\partial x_{ij}} \\
&= \frac{\partial \mathrm{tr}(BAX^{-1})}{\partial x_{ij}} \\
&= \mathrm{tr}\left(BA\frac{\partial X^{-1}}{\partial x_{ij}}\right) \\
&= \mathrm{tr}(BA(-X^{-1}e_i e_j^T X^{-1})) \\
&= -\mathrm{tr}(e_j^T X^{-1}BAX^{-1}e_i) \\
&= -e_j^T X^{-1}BAX^{-1}e_i \\
&= -(X^{-1}BAX^{-1})_{ji} \\
\Rightarrow \quad \frac{\partial \mathrm{tr}(AX^{-1}B)}{\partial X} &= -(X^{-1}BAX^{-1})^T
\end{aligned}$$

Problem 7.

$$\begin{aligned}
l(\mu, \Sigma | D) &= \ln\left(\prod_{i=1}^{n} p(X_i)\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(\frac{-1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)\right) \\
&= \sum_{i=1}^{n} -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\log(\det(\Sigma)) - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) \\
\Rightarrow \frac{\partial l}{\partial \mu} &= \sum_{i=1}^{n} \Sigma^{-1}(x_i-\mu) \\
&= \Sigma^{-1}\left(\sum_{i=1}^{n} x_i\right) - n\Sigma^{-1}\mu \\
\Rightarrow \hat{\mu}^{(\mathrm{MLE})} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
\Rightarrow \frac{\partial l}{\partial \Sigma} &= \sum_{i=1}^{n} -\frac{1}{2}\frac{\partial}{\partial \Sigma}(\log(\det(\Sigma))) - \frac{1}{2}\frac{\partial}{\partial \Sigma}\{(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\Sigma^{-1} - \frac{1}{2}\frac{\partial}{\partial \Sigma}\{\mathrm{tr}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\Sigma^{-1} + \frac{1}{2}(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\Sigma^{-1}) \\
\Rightarrow \hat{\Sigma}^{(\mathrm{MLE})} &= \frac{1}{n}\sum_{i=1}^{n} (x_i-\mu)(x_i-\mu)^T
\end{aligned}$$

Problem 8.

$$\begin{aligned}
\mathbb{E}(\|x\|_2^2) &= \mathbb{E}(x^T x) \\
&= \mathbb{E}(\mathrm{tr}(x^T x)) \\
&= \mathbb{E}(\mathrm{tr}(x x^T)) \\
&= \mathrm{tr}(\mathbb{E}(x x^T)) \\
&= \mathrm{tr}(\mathrm{Var}(x) + \mathbb{E}(x)\mathbb{E}(x)^T) \\
&= \mathrm{tr}(\Sigma) + \|\mathbb{E}(x)\|_2^2
\end{aligned}$$