Regression Analysis
2023 Fall HW1

Tuorui Peng[1]

Problem 1

The problem is solved with the following algorithm:

Algorithm *Solve $Rx = b$ with R being upper-triangular*

Given: $R = \{r_{ij}\}_{0 \le i \le j \le n} \in \mathbb{R}^{n \times n}$ upper-triangular as stated, $b \in \mathbb{R}^n$

Initialize $x_{n:(n+1)} = []$, for $i = n : 1$

1. We already got $\vec{x}_{n:(i+1)}$
2. Focus on $R_{i,:}$, we have
$$\sum_{j=i}^{n} r_{ij} x_j = b_i \Rightarrow x_i = \frac{b_i - \sum_{j=i+1}^{n} r_{ij} x_j}{r_{ii}}$$
3. Solve x_i from $x_{n:(i+1)}$, and now we have $x_{n:i}$

Return: $x_{n:1} = x$

Problem 2

Question 2.(a)

Since we have $m < n$ and A has full row rank, we have its right inverse matrix

$$A_r := A^T (AA^T)^{-1} \in \mathbb{R}^{n \times m}$$

in this way $\mathcal{S}$ could be represented as

$$\mathcal{S} = \{x | x = A^T (AA^T)^{-1} b + \xi,\, \xi \in \mathrm{Null}(A)\} \subset \mathbb{R}^n$$

and now the optimization problem is formulated as

$$\begin{aligned}
\pi_{\mathcal{S}}(y) &:= \arg\min_{x \in \mathcal{S}} \|y - x\|^2 \\
&= \arg\min_{\xi \in \mathrm{Null}(A)} \left\| y - A^T (AA^T)^{-1} b - \xi \right\|^2
\end{aligned}$$

[1] TuoruiPeng2028@u.northwestern.edu

into a constraint optimization problem, and the Lagrangian function is

$$
\begin{aligned}
\mathcal{L}(\xi,\lambda) = \left\|y - A^T(AA^T)^{-1}b - \xi\right\|^2 + \lambda^T A\xi, \quad \lambda \in \mathbb{R}^m \\
\Rightarrow & \begin{cases} y - A^T(AA^T)^{-1}b - \xi = A^T\lambda \\ A\xi = 0 \end{cases} \\
\Rightarrow & Ay - b = AA^T\lambda \Rightarrow \lambda = (AA^T)^{-1}(Ay - b) \\
\Rightarrow & \xi = (I - A^T(AA^T)^{-1}A)y \\
\Rightarrow & \pi_{\mathcal{S}}(y) = A^T(AA^T)^{-1}b + (I - A^T(AA^T)^{-1}A)y
\end{aligned}
$$

Question 2.(b)

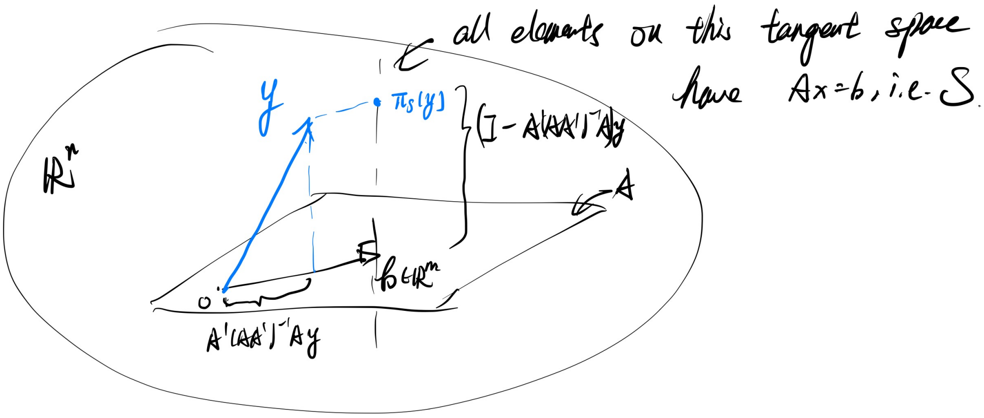

图 1: Illustration of $\pi_{\mathcal{S}}(y)$

Problem 3

Question 3.(a)

Verify directly:

$$
\begin{aligned}
H_u' &= (I - 2uu')' = I - 2uu' = H_u \\
H_u'H_u &= (I - 2uu')(I - 2uu') = I - 4uu' + 4uu'u'u = I
\end{aligned}
$$

Question 3.(b)

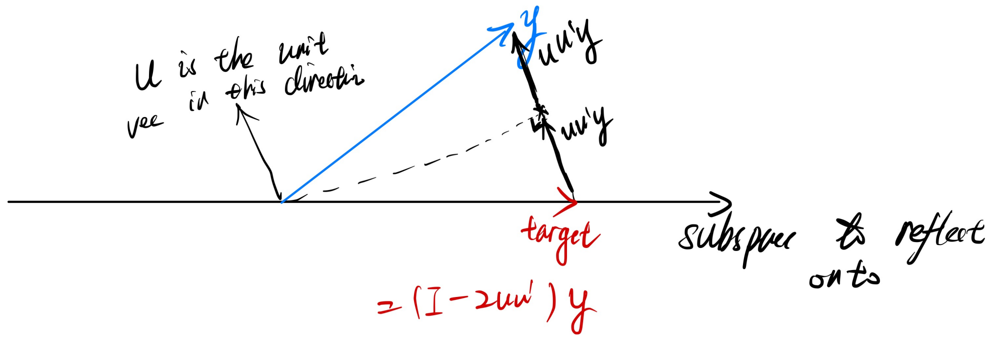

图 2: Illustration of the reflection process

Question 3.(c)

So here e_1 is the 'direction of subspace' we want to reflect onto. We can see that here v is just the angular bisector of $\langle x, e_1\rangle$. Denote $e_x = x/\|x\|$, the unit vector in x direction. Then we have

$$\begin{aligned}
H_v x &= \left(I - 2\frac{(e_x+e_1)(e_x+e_1)'}{\|e_x+e_1\|^2}\right)x \\
&= \frac{\|x\|}{\|e_x+e_1\|^2}\left(\|e_x+e_1\|^2 e_x - 2(e_x+e_1)(e_x+e_1)'e_x\right) \\
&= \frac{\|x\|}{\|e_x+e_1\|^2}\left((2+2e_x\cdot e_1)e_x - 2(1+e_x\cdot e_1)(e_x+e_1)\right) \\
&= \frac{\|x\|}{2+2e_x\cdot e_1} \times -2(1+e_x\cdot e_1)e_1 \\
&= -\|x\|\, e_1
\end{aligned}$$

Question 3.(d)

i.e. to prove that operator H is norm-preserving. We have

$$\begin{aligned}
\|Hx\|_2^2 &= x'H'Hx \\
&= x'H'_{u_k}H'_{u_{k-1}}\cdots H'_{u_1}H_{u_1}\dots H_{u_{k-1}}H_{u_k}x \\
&= x'x = \|x\|_2^2
\end{aligned}$$

Question 3.(e)

as definition

$$u = \frac{a_1/\|a_1\| + e_1}{\|a_1/\|a_1\| + e_1\|}, \qquad H_u = I - 2uu'$$

Question 3.(f)

Notation: the l^{th} unit vector in $\mathbb{R}^m$ is denoted $e_{m,l}$. We have

$$\underset{k+1}{u} = \frac{b_1/\|b_1\| + e_{k+1,1}}{\|b_1/\|b_1\| + e_{k+1,1}\|}, \quad F = I - 2uu'$$

$$\text{Reflection: } S = \begin{bmatrix} \underset{(n-k-1)\times(n-k-1)}{I} & 0 \\ 0 & \underset{(k+1)\times(k+1)}{F} \end{bmatrix}$$

Question 3.(g)

Use the process in the previous subproblem, say it could reflect the i^{th} column, denoted by S_i. Then we could use

$$\Xi = S_{n-1}S_{n-2}\dots S_1$$

to reflect all n columns. i.e.

$$\Xi A = S_{n-1}S_{n-2}\dots S_1 A = \text{upper triangular}$$

we can write down the Q and R

$$A = QR, \quad Q = S_1S_2\dots S_{n-1}, \quad R = S_{n-1}S_{n-2}\dots S_1 A$$

Problem 4

Notation: in this section I use v_i in place of s_i, i.e. the eigenvector of A. And further I use the convention that we have S is orthonormal (which can always be valid with some transform to Λ).

Question 4.(a)

To $\operatorname{sgn}(v_1'x^0)\lambda_1 v_1$, i.e. $\pm\lambda_1 v_1$.

Question 4.(b)

Note that in each step $\dfrac{1}{\|x^k\|}$ just acts as a scalar. So doing n iteration with each step a scalar multiplication IS EQUIVALENT TO doing n iteration with only the last step a scalar (to unitary). So we could express a $n+1$ steps iteration as

$$x^n = \xi A^n x^0, \quad x^{n+1} = \frac{1}{\|x^n\|}Ax^n, \quad \xi \in \mathbb{R}$$

in which we denote the 'un-scaled' $A^n x^0$ as $\tilde{x}^n$, i.e. $x^n = \xi\tilde{x}^n$. Then

$$\tilde{x}^n = A^n x^0 = (S\Lambda S^{-1})^n x^0 = S\Lambda^n S^{-1}x^0, \quad \forall n$$

Then we have

$$x^{n+1} = \frac{1}{\|S\Lambda^n S^{-1}x^0\|} S\Lambda^{n+1}S^{-1}x^0$$

in which spectrum of $\frac{1}{\|S\Lambda^n S^{-1}x^0\|} S\Lambda^{n+1}S^{-1}$ is simply

$$\begin{aligned}
\frac{1}{\|S\Lambda^n S^{-1}x^0\|} S\Lambda^{n+1}S \overset{(i)}{\sim=} \frac{1}{\|\Lambda^n\|} S\Lambda^{n+1}S = S\mathrm{diag}\{\lambda_1, \lambda_2\left(\frac{\lambda_2}{\lambda_1}\right)^n, \ldots, \lambda_n\left(\frac{\lambda_n}{\lambda_1}\right)^n\}S^{-1} \\
\to & S\mathrm{diag}\{\lambda_1, 0, \ldots, 0\}S^{-1} = \lambda_1 v_1 v_1'
\end{aligned}$$

in which (i) means only up to a constant $\sim O(1)$

And finally

$$\begin{aligned}
x^{n+1} \to & \propto \lambda_1 v_1 v_1' x^0 \propto v_1 \mathrm{sgn}(v_1' x^0) \\
x^{n+2} \to & A\frac{x^{n+1}}{\|x^{n+1}\|} = Av_1\mathrm{sgn}(v_1'x^0) = \mathrm{sgn}(v_1'x^0)\lambda_1 v_1
\end{aligned}$$

Problem 5

Question 5.(a)

$$\begin{cases}
x = (A - BD^{-1}C)^{-1}(a - BD^{-1}b) \\
y = (D - CA^{-1}B)^{-1}(b - CA^{-1}a)
\end{cases}$$

Question 5.(b)

Using the above fomula, we have

$$\begin{aligned}
y = & (C^{-1} + V'A^{-1}U)^{-1}V'A^{-1}z \Rightarrow x = A^{-1}(z - Uy) = \left(A^{-1} - A^{-1}U(C^{-1} + V'A^{-1}U)^{-1}V'A^{-1}\right)z \\
\Rightarrow & (A + UCV')^{-1} = A^{-1} - A^{-1}U(C^{-1} + V'A^{-1}U)^{-1}V'A^{-1}
\end{aligned}$$

Problem 6

Question 6.(a)

Prove it directly

$$\text{L.H.S.} = (a_2 + (a_1 - a_2))b_1 + (a_1 - (a_1 - a_2))b_2 = a_2b_1 + a_1b_2 + (a_1 - a_2)(b_1 - b_2) \geq a_2b_1 + a_1b_2 = \text{R.H.S.}$$

Question 6.(b)

Denote $P_{\mathrm{swap}(i,j)}$ as the operator to swap the i^{th} and j^{th} element.

1. Since $P_{\text{swap}(i,j)}$ keep all other $n-2$ elements unchanged, we can make use of the above 2-dimensional inequality, we have the following statement:

    If $P_{\text{swap}(i,j)}$ is **moving the larger element to a higher larger index, then the inequality is preserved**. Say swapping $v_n@n \leftrightarrow v_{n-1}@n-1$ then the inequality perserve, while $v_n@n-1 \leftrightarrow v_{n-1}@n$ does not.

    i.e.

    $$u'\tilde{v} \geq u'P_{\text{swap}(i,j)}\tilde{v} \Leftrightarrow \tilde{v}_i \geq \tilde{v}_j \& i > j$$

2. Then we prove that: any permutation matrix $P \in \mathcal{P}_n$ could be represented as products of $P_{\text{swap}(i,j)}$s, in which each $P_{\text{swap}(i,j)}$ is moving the larger element to a higher larger index for the declared $u, \tilde{v}$.

    We do it by the following iteration process:

Algorithm *Decomposition $Pv = \prod P_{\text{swap}(i,j)}v$ which perserce inequality*

For $\imath = n : 1$:

- Move v_k to the desired position as in Pv by repeatedly swap it with 'the first element on its left with subscript smaller than itself'. e.g. to move v_{n-1} which is now at position n to position $n-3$, with v_n already at position $n-2$, i.e.

    $$\ldots, v_{n-3}, v_n, v_{n-2}, v_{n-1}$$

    we do by following:

    $$\begin{aligned} &\ldots, v_{n-3}, v_n, v_{n-1}, v_{n-2} && \text{swap } v_{n-1} \text{ with } v_{n-2} \\ &\ldots, v_{n-1}, v_n, v_{n-3}, v_{n-2} && v_n \text{ is skipped, then swap } v_{n-1} \text{ with } v_{n-3} \end{aligned}$$

Repeat the above operation, which preserve the inquality, until $v_{n:1}$ are all at the right position to become Pv.

Now we have $u'v \geq u'Pv$

Question 6.(c)

According to the above algorithm, we know that $u'v \geq u'Pv$, $\forall P \in \mathcal{P}_n$. Then we have

$$\begin{aligned} u'v =& u'(\sum_{l=1}^{k}\lambda_l)v \geq u'\sum_{l=1}^{k}\lambda_l P_l, \quad w.r.t. \lambda_l \geq 0, \sum_{l=1}^{k}\lambda_l = 1 \\ =& u'Sv, \quad \forall S \in \mathcal{S}_n \end{aligned}$$

And also $I \in \mathcal{S}$, thus

$$u'v = \max_{S \in \mathcal{S}_n} u'Sv$$

Problem 7

Question 7.(a)

Say we denote the SVD of A as $A = \tilde{U}\tilde{\Sigma}\tilde{V}$, then we have

$$\text{L.H.S.} = tr(AB) = tr(\tilde{U}\tilde{\Sigma}\tilde{V}) = tr(\tilde{\Sigma}\tilde{V}B\tilde{U})$$

and here we can define $\tilde{B} \leftarrow \tilde{V}B\tilde{U}$. And on the other hand, singular values of a matrix preserve under unitary transformation, thus

$$\sigma(\tilde{B}) = \sigma(\tilde{V}B\tilde{U}) = \sigma(B)$$

so vNti is now

$$tr(\tilde{\Sigma}\tilde{B}) \leq \sum_{i=1}^{n} \sigma_i(\tilde{\Sigma})\sigma_i(\tilde{B}) =$$

Now we can re-define $B \leftarrow \tilde{B}$, and $A \leftarrow \tilde{\Sigma}$. So its equivalent to prove

$$tr(AB) \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B), \quad \text{with } A \text{ being diagonal}$$

Question 7.(b)

$$tr(AB) = tr(AU\Sigma V') = tr(V'AU\Sigma) = \sum_{i=1}^{n} \sigma_i v_i' A u_i = \sum_{i,j\leq n} \sigma_i v_{ij} a_j u_{ij}$$

Question 7.(c)

Using $xy \leq \frac{1}{2}(x^2 + y^2)$, we have

$$tr(AB) = \sum_{i,j\leq n} \sigma_i v_{ij} a_j u_{ij} \leq \frac{1}{2} \sum_{i,j\leq n} \sigma_i a_j (v_{ij}^2 + u_{ij}^2)$$

Question 7.(d)

Since U and V are both orthonormal bases, we have

$$\sum_{j=1}^{n} v_{ij} v_{kj} = \delta_{ik}, \quad \sum_{j=1}^{n} u_{ij} u_{kj} = \delta_{ik}$$

so matrix $\{v_{ij}^2\}_{i,j=1}^n$ has row sum 1 & column sum 1 (similar for $\{u_{ij}^2\}_{i,j=1}^n$). Then using notation in the previous problem, in which $\mathcal{S}_n$ is the doubly stochastic matrices, we have

$$\{v_{ij}^2\}_{i,j=1}^n, \{u_{ij}^2\}_{i,j=1}^n \in \mathcal{S}_n \Rightarrow \{\frac{1}{2}(v_{ij}^2 + u_{ij}^2)\} \in \mathcal{S}_n$$

then using the inequality $\tilde{u}'\tilde{v} = \max_{S \in \mathcal{S}_n} \tilde{u}' S \tilde{v}$, we can place $\sigma_i(A), \sigma_i(B)$ as the $\tilde{u}, \tilde{v}$, respectively, and then we have

$$\begin{aligned} tr(AB) &\leq \sum_{i,j \leq n} \sigma_i \frac{1}{2}(v_{ij}^2 + u_{ij}^2) a_j \\ &\leq \sum_{i=1}^{n} \sigma_i(B) \sigma_i(A) \end{aligned}$$

$\underline{\sum_{i=1}^{n}}$

$\frac{\pi}{6}$

$$\frac{\pi}{6}$$