Homework 4

6.s955 Applied Numerical Analysis

Pitipat Wongsittikan

November, 2023

Problem 1

a. Since x_k is in a small neighborhood of the fixed point x^*, we can linearize $g(x_k)$ as,

$$g(x_k) \approx g(x^*) + Dg(x^*)(x_k - x^*) \tag{1}$$

Rearranging the term, we will have,

$$Dg(x^*) = \frac{g(x_k) - g(x^*)}{x_k - x^*} \tag{2}$$

$$||Dg(x^*)||_2 = ||\frac{g(x_k) - g(x^*)}{x_k - x^*}||_2 \tag{3}$$

$$||Dg(x^*)||_2 \leq \frac{||g(x_k) - g(x^*)||_2}{||x_k - x^*||_2} \tag{4}$$

Since g is 1-Lipschitz in this neighborhood, therefore, $\forall\ x, y$ in this neighborhood,

$$\frac{||g(x) - g(y)||_2}{||x - y||_2} \leq 1 \tag{5}$$

$$\frac{||g(x_k) - g(x^*)||_2}{||x_k - x^*||_2} \leq 1 \tag{6}$$

$$||Dg(x^*)||_2 \leq 1 \tag{7}$$

Therefore, the norm of $Dg(x^*)$ has to be less than or equal to 1.

b.

$$\begin{align}
E_{k+1} &= ||x^* - x_{k+1}||_2 \tag{8}\\
&= ||x^* - g_(x_k)||_2 \tag{9}\\
&= ||x^* - g(x^*) + Dg(x^*)(x_k - x^*)||_2 \tag{10}\\
&= ||x^* - x^* + Dg(x^*)(x_k - x^*)||_2 \tag{11}\\
&= ||Dg(x^*)(x_k - x^*)||_2 \tag{12}\\
&\leq ||Df(x^*)||_2||x_k - x^*||_2 \tag{13}
\end{align}$$

From a, we know that $||Dg(x^*)|| \leq 1$, therefore,

$$\begin{aligned} E_{k+1} &\leq ||Df(x^*)||_2 ||x_k - x^*||_2 &(14)\\ &\leq (1)||x_k - x^*||_2 &(15)\\ &\leq E_k &(16) \end{aligned}$$

Hence, $E_{k+1} \leq E_k$

c. Since $Dg(x^*) \leq 1$, the largest singular value of Dg must less than or equal to 1.

d. Expand the Taylor's series by another term, we will have the similar expression as the previous question, but with 2nd derivative term instead.

$$g(x_k) \approx g(x^*) + D^2 g(x^*)(x_k - x^*)^2 \tag{17}$$

Here, if we follow what we did in part b, we will arrive at.

$$\begin{aligned} E_{k+1} &\leq ||D^2 f(x^*)||_2 ||x_k - x^*||_2^2 &(18)\\ &\leq ||D^2 f(x^*)||_2 * E_k^2 &(19) \end{aligned}$$

This means, if the first derivative becomes 0, the fixed point iteration method will converge at a quadratic rate.

Problem 2

Rearranging $f(x)$,

$$f(x) = \sum_{ij}^{n} D_{ij}^2 - 2D_{ij}||x_i - x_j||_2 + ||x_i - x_j||_2^2 \tag{20}$$

Since $D_{ij}s$ are constants, we can ignore them in the minization problem.
The last term can be expanded further,

$$\begin{aligned} \sum_{ij} ||x_i - x_j||_2^2 &= \sum_{ij} ||x_i||^2 - 2x_i^T x_j + \sum ||x_j||^2 &(21)\\ &= 2n \sum_{i} ||x_i||^2 - 2 \sum_{ij} x_i^T x_j &(22) \end{aligned}$$

The first term is basically 2n times the sum of the square of each elements in X.
This could be done by take each column of X and dot product it with itself and sum all over them. This is the sum of the diagonals of matrix $X^T X$ multiply by 2n, which comes to $2n(\text{tr}X^T X)$.
*Note that $\text{tr}(AB) = \text{tr}(BA)$, therefore, $\text{tr}(X^T X) = \text{tr}(XX^T)$. We will use this to simplify the term later.

The second term is 2 times the sum of the inner product between every column of X, which is $2*(x_1+x_2+...+x_n)^T(x_1+x_2+x_3+...+x_n)$ or $\sum_i^d(\sum_k x_k^T * \sum_k x_k)$, where d is the dimension of x.

If we look into the matrix product of X and vector of ones, [1 1 1 ...], we will get a vector with row i as the sum of row i of X.

$$X\mathbf{1}_i = \sum_k X_{ik} \tag{23}$$

If we multiply this by transpose of itself, $(X\mathbf{1})^T$, we will have,

$$[X\mathbf{1}*(X\mathbf{1})^T]_{ij} = \sum_k X_{ik} * \sum_l X_{jl} \tag{24}$$

Therefore,

$$[X\mathbf{1}*(X\mathbf{1})^T]_{ii} = \sum_k X_{ik} * \sum_l X_{il} \tag{25}$$

$$\text{tr}(X\mathbf{1}*(X\mathbf{1})^T) = \sum_i(\sum_k X_{ik})^2 \tag{26}$$

Which is what we've shown above.

Therefore, the second term can be written as $2\text{tr}(X\mathbf{1}(X\mathbf{1})^T) = 2\text{tr}X\mathbf{1}\mathbf{1}^TX^T$.

Combining the first and the second term, we will have,

$$2n\text{tr}(XX^T) - 2\text{tr}X\mathbf{1}\mathbf{1}^TX^T = \text{tr}(XVX^T) \tag{27}$$

Where $V = 2n(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$.

For the term $-2D_{ij}||x_i - x_j||_2$, we can multiply by the norm and divide by the norm.

$$-2D_{ij}||x_i - x_j||_2 = -\frac{2D_{ij}||x_i - x_j||^2}{||x_i - x_j||} \tag{28}$$

If we look at the last term, it's similar to what we did with the previous one, but with a constant $-\frac{2D_{ij}}{||x_i-x_j||}$ in front. Following the same procedure as before, we will arrive at the similar structure as $\text{tr}(XV_2X^T)$, but this time, $V_2 = 4(A-B)$, where A is a diagonal matrix, with it's diagonal i is $\sum_k w_{ik}$, where $w_{ij} = -\frac{2D_{ij}}{||x_i-x_j||}$, and $B_{ij} = -\frac{2D_{ij}}{||x_i-x_j||}$. Note that if $||x_i - x_j|| = 0$ that entry ij will already be 0 at the first place, before we multiply/divide with the norm. With this, we will have,

$$\begin{aligned} \sum -2D_{ij}||x_i - x_j||_2 &= \text{tr}(XV_2X^T) &(29)\\ &= \text{tr}(X(4(A-B)X^T) &(30)\\ &= \text{tr}(X(A-B)X^T) &(31) \end{aligned}$$

And if we look at $A - B$, we can see that it can be describe compactly as the matrix B that was given in the problem.

Hence, minimizing f(x) is equivalent to minimizing $\text{tr}(XVX^T) - 4\text{tr}(XBX^T)$.

b. From the Cauchy-Schwartz inequality, we will have,

$$||xi - xj||_2 \geq (xi - xj).(zi - zj)/||zi - zj||_2 \tag{32}$$

From the previous part, we know that,

$$tr(XB(X)X^T) = \sum Dij||xi - xj||_2 \tag{33}$$

Apply the previous inequality that we just derived,

$$\sum D_{ij}||x_i - x_j||_2 \geq \sum D_{ij}(xi_{-}x_j).(z_i - z_j)/||z_i - z_j||_2 \tag{34}$$

Again, from the previous part, we can see that, this is the sum of the sum of the inner product between columns of X and Z. Following the same step in 2a., we will have $\sum D_{ij}(x_i - x_j)(z_i - z_j)/||z_i - z_j||_2 = 4tr(XB(Z)Z^T)$

Therefore,

$$4tr(XB(X)X^T) \geq 4tr(XB(Z)Z^T) \tag{35}$$

$$-4tr(XB(X)X^T) \leq -4tr(XB(Z)Z^T) \tag{36}$$

$$\mathrm{tr}XVX^T) - 4tr(XB(X)X^T) \leq \mathrm{tr}XVX^T) - 4tr(XB(Z)Z^T) \tag{37}$$

Therefore, $\tau(X, X) \leq \tau(X, Z)$.
paragraphc.

$$f(X^k) = \tau(X^k, X^k) \tag{38}$$

since $X^k + 1$ minimize $\tau(X, X^k)$

$$\tau(X^k + 1, X^k) \leq tau(X^k, X^k) \tag{39}$$

From the part b, let $X = X^k + 1$ and $Z = X^k$, we will have,

$$\tau(X^k + 1, X^(k + 1) \leq \tau(X^k + 1, X^k) \tag{40}$$

And since $\tau(X^k + 1, X^k + 1) = f(X^k + 1)$, if we work things in reverse, we will have,

$$f(X^k + 1) = \tau(X^k + 1, X^k + 1) \leq \tau(X^k + 1, X^k) \leq \tau(X^k, X^k) = f(X^k). \tag{41}$$

Therefore, $f(X^k + 1) \leq f(X^k)$

d.

$$\frac{\partial f}{\partial X} = 2X \cdot V - 4 \cdot Z \cdot B^\top \tag{42}$$

Set $\partial f = 0$ and solve for X.

$$X_{k+1} = 2X_k B^T V^{-1} \tag{43}$$

e. Please refer to the .py file.
Note: The function values seem to decrease monotonically at first, but as the values get lower, there a occasionally jumps as you can see in the given plot.

The result from the implemented smacof looks similar to that of scikit-learn, but blurrier eventhough the objective function does not decrease much after 100 iterations.

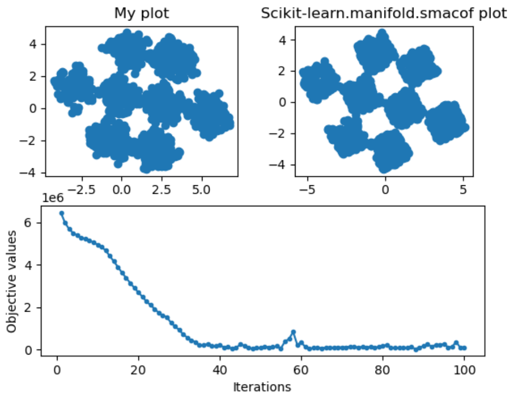

Figure 1: Enter Caption

f Note, I'm using Scikit-learn.manifold.smacof which only provides the last function value, not the entire iterations, therefore, I will look at the overall convergence time instead of wall-clock time.

Sklearn value: 4.337319146347275
Sk iteration till convergence: 74
My optimal value: 119574.54546701656
My iteration till convergence: none, so I ran for 100 iterations

Sk elapse time: 5.831323623657227 s
Sk elapse time per iteration: 0.07880167058996253 s
My elapse time: 28.339877605438232 s
My elapse time per iteration: 0.28339877605438235 s

We can clearly see that the package provided by sklearn has lower objective value with a faster computation time.

Problem 3

0.1 a.

We have to find set of S such that $s \leq \frac{|x|-0}{x-} = \frac{|x|}{x}$. Solve this inequality and we will have, $S \in [-1, 1]$.

0.2 b.

If x^* is the minimizer of $f(x)$, $f(x) \geq f(x^*)$ or $f(x) - f(x^*) \geq 0$. Therfore, s = 0, must be in the subdifferential at x*, otherwise, $f(x) - f(x^*) \geq 0$ wont always holds true, and x* wont be the global minimum.