Homework 5

6.s955 Applied Numerical Analysis

Pitipat Wongsittikan

December, 2023

Problem 1

a. We start from the given expression,

$$x_{k+1} = \text{argmin}_x[f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2\alpha}||x - x_k||_2^2] \tag{1}$$

Since we are minimizing the term, take derivative of the term with respect to x and set it equal 0, we will have

$$\nabla f(x_k) + \frac{1}{\alpha}(x - x_k) = 0 \tag{2}$$

Therefore,

$$x = x_k - \alpha \nabla f(x_k) \tag{3}$$

Which equals to the given expression for x_{k+1}.

b. From $x_{k+1} = \text{argmin}_x[f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2\alpha}||x - x_k||_2^2 + g(x)]$, we can drop the term that does not relate to x and complete the square as follow,

$$= f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2\alpha}||x - x_k||_2^2 + g(x) \tag{4}$$

$$= \nabla f(x_k)^T(x - x_k) + \frac{1}{2\alpha}||x - x_k||_2^2 + g(x) \tag{5}$$

$$= \frac{1}{2\alpha}[2\alpha \nabla f(x_k)^T(x - x_k) + ||x - x_k||_2^2] + g(x) \tag{6}$$

$$= \frac{1}{2\alpha}[||\alpha \nabla f(x_k)||^2 + 2(\alpha \nabla f(x_k)^T)(x - x_k) + ||x - x_k||_2^2 - ||\alpha \nabla f(x_k)||^2] + g(x)] \tag{7}$$

$$= \frac{1}{2\alpha}||\alpha \nabla f(x_k) + (x - x_k)||_2^2 - ||\alpha \nabla f(x_k)||^2 + g(x) \tag{8}$$

Again, the term that's not related to x can be dropped, and now we are left with

$$= \frac{1}{2\alpha}||\alpha\nabla f(x_k) + (x - x_k)||_2^2 + g(x) \tag{9}$$
$$= \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 + g(x) \tag{10}$$
$$= g(x) + \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 \tag{11}$$

c. Substitute $g(x) = \lambda|x|$

$$= g(x) + \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 \qquad = \lambda|x| + \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 \tag{12}$$

We will separate the term into 3 cases, $x > 0$, $x = 0$, and $x < 0$ then take derivative of those terms and equal it to 0.

Case: $\mathbf{x > 0}$

$$f(x) = \lambda x + \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 \tag{13}$$
$$\frac{\partial f(x)}{\partial x} = \lambda + \frac{1}{\alpha}(x - (x_k - \alpha\nabla f(x_k)) \tag{14}$$
$$0 = \lambda + \frac{1}{\alpha}(x - (x_k - \alpha\nabla f(x_k)) \tag{15}$$
$$x = x_k - \alpha(\nabla f(x_k) + \lambda) \tag{16}$$

Recall that $x > 0$. Therefore,

$$0 \leq x_k - \alpha(\nabla f(x_k) + \lambda) \tag{17}$$
$$\alpha \leq \frac{x_k}{\nabla f(x_k) + \lambda} \tag{18}$$

Therefore, for the first case, if eq.18 is true,we proceed with $x_{k+1} = x_k - \alpha(\nabla f(x_k) + \lambda)$

Case: $\mathbf{x < 0}$

$$f(x) = -\lambda x + \frac{1}{2\alpha}||x - (x_k - \alpha\nabla f(x_k)||_2^2 \tag{19}$$
$$\frac{\partial f(x)}{\partial x} = -\lambda + \frac{1}{\alpha}(x - (x_k - \alpha\nabla f(x_k)) \tag{20}$$
$$0 = -\lambda + \frac{1}{\alpha}(x - (x_k - \alpha\nabla f(x_k)) \tag{21}$$
$$x = x_k - \alpha(\nabla f(x_k) - \lambda) \tag{22}$$

Recall that $x < 0$. Therefore,

$$0 \geq x_k - \alpha(\nabla f(x_k) - \lambda) \tag{23}$$
$$\alpha \geq \frac{x_k}{\nabla f(x_k) - \lambda} \tag{24}$$

Similar to what we did earlier, but in this case,
if eq.24 is true, we proceed with $x_{k+1} = x_k - \alpha(\nabla f(x_k) - \lambda)$.

Finally, the last case is when $x = 0$, which implies $x_{k+1} = x_k - \alpha \nabla f(x_k)$.

Problem 2

a. Take a derivative with respect to t.

$$\frac{\partial E(t)}{\partial t} = \theta'\theta'' + \sin\theta\theta' \tag{25}$$
$$= \theta'(\theta'' + \sin\theta) \tag{26}$$

We were given a relationship, $\theta'' + \sin\theta = 0$, therefore,

$$\frac{\partial E(t)}{\partial t} = \theta' 0 \tag{27}$$
$$= 0 \tag{28}$$

Since the derivative of E with respect to t is 0, E is a constant function in terms of t.

b. Note that h is a small number and some useful Taylor's approximation of sin and cos.

$$\cos(\theta_k + h\omega_k) = \cos\theta_k - h\omega_k \sin\theta_k + O(h^2) \tag{29}$$
$$\sin(\theta_k + h\omega_k) = \sin\theta_k + h\omega_k \cos\theta_k + O(h^2) \tag{30}$$

Now, we work on E_{k+1},

$$E_{k+1} = \frac{1}{2}\omega_{k+1}^2 - \cos\theta_{k+1} \tag{31}$$
$$= \frac{1}{2}(\omega_k - h\sin\theta_{k+1})^2 - \cos(\theta_k + h\omega_k) \tag{32}$$
$$= \frac{1}{2}\omega_k^2 - h\omega_k \sin\theta_{k+1} + \frac{1}{2}h^2 \sin^2\theta_{k+1} - \cos(\theta_k + h\omega_k) \tag{33}$$
$$= \frac{1}{2}\omega_k^2 - h\omega_k \sin\theta_{k+1} + \frac{1}{2}h^2 \sin^2\theta_{k+1} - \cos\theta_k + h\omega_k \sin\theta_k + O(h^2) \tag{34}$$
$$= [\frac{1}{2}\omega_k^2 - \cos\theta_k] - h\omega_k \sin\theta_{k+1} + \frac{1}{2}h^2 \sin^2\theta_{k+1} + h\omega_k \sin\theta_k + O(h^2) \tag{35}$$
$$= [\frac{1}{2}\omega_k^2 - \cos\theta_k] - h\omega_k(\sin\theta_k + h\omega_k \cos\theta_k + O(h^2)) + \frac{1}{2}h^2 \sin^2\theta_{k+1} + h\omega_k \sin\theta_k + O(h^2) \tag{36}$$
$$= [\frac{1}{2}\omega_k^2 - \cos\theta_k] - h\omega_k \sin\theta_k - h^2\omega_k^2 \cos\theta_k + \frac{1}{2}h^2 \sin^2\theta_{k+1} + h\omega_k \sin\theta_k + O(h^2) \tag{37}$$
$$= [\frac{1}{2}\omega_k^2 - \cos\theta_k] - (h^2\omega_k^2 \cos\theta_k + \frac{1}{2}h^2 \sin^2\theta_{k+1} + O(h^2)) \tag{38}$$

The first group of terms is E_k and the rest is equivalent to $O(h^2)$.
Therefore, $E_{k+1} = E_k + O(h^2)$.

c.

$$\begin{aligned}
\theta_{k+1} &= \theta_k + h\omega_k && (39)\\
\omega_{k+1} &= \omega_k - h\theta_{k+1} && (40)\\
&= \omega_k - h(\theta_k + h\omega_k) && (41)\\
&= \omega_k - h\theta_k - h^2\omega_k && (42)\\
&= \omega_k(1-h^2) - h\theta_k && (43)
\end{aligned}$$

Therefore, we can write the whole relationship in the form of,

$$\begin{bmatrix}\theta_{k+1}\\ \omega_{k+1}\end{bmatrix} = \begin{bmatrix}1 & h\\ -h & 1-h^2\end{bmatrix}\begin{bmatrix}\theta_k\\ \omega_k\end{bmatrix} \tag{44}$$

Therefore, $A = \begin{bmatrix}1 & h\\ -h & 1-h^2\end{bmatrix}$.

d.

$$\begin{aligned}
E_k &= \omega_k^2 + h\omega_k\theta_k + \theta_k^2 && (45)\\
E_{k+1} &= \omega_{k+1}^2 + h\omega_{k+1}\theta_{k+1} + \theta_{k+1}^2 && (46)\\
& && (47)
\end{aligned}$$

Substitute the relationship on part c to E_{k+1}, we will have,

$$\begin{aligned}
E_{k+1} &= (\omega_k - h\theta_{k+1})^2 + h(\omega_k - h\theta_{k+1})\theta_{k+1} + \theta_{k+1}^2 && (48)\\
&= \omega_k^2 - 2h\omega_k\theta_{k+1} + h^2\theta_{k+1}^2 + h\omega_k\theta_{k+1} - h^2\theta_{k+1}^2 + \theta_{k+1}^2 && (49)\\
&= \omega_k^2 - 2h\omega_k\theta_{k+1} + h\omega_k\theta_{k+1} + \theta_{k+1}^2 && (50)\\
&= \omega_k^2 - h\omega_k\theta_{k+1} + \theta_{k+1}^2 && (51)\\
&= \omega_k^2 - h\omega_k(\theta_k + h\omega_k) + (\theta_k + h\omega_k)^2 && (52)\\
&= \omega_k^2 - h\omega_k\theta_k - h^2\omega_k^2 + \theta_k^2 + 2h\theta_k\omega_k + h^2\omega_k^2 && (53)\\
&= \omega_k^2 + h\omega_k\theta_k + \theta_k^2 && (54)\\
&= E_k && (55)
\end{aligned}$$

Therefore, $E_{k+1} = E_k$

Problem 3

a. For the given expression of $E[f]$, we can rewrite $E = ||DF||_2^2$ where F is a vector with $F_i = f(v_i)$, including the constraints $f(v_j) = f_0(v_j)$ for v_j in V_0 and D is a matrix with each row represents an edge. Each row of D is mostly 0s except where the connected nodes are, which will be 1 and -1 (in any order in this case).

Following that new reconstruction of E we can further see that E is a PSD matrix that can be rewritten as $F^T D^T DF = F^T LF$. We can solve this by taking the derivative of the term with respected to F and equal it to 0. We will have

$$LF = 0 \tag{56}$$

However, this equation is useless since L is not invertible since it has property that its diagonals are the sum of its other indices on the row, i.e., the rows are linearly **dependent**. However, with that property, if we remove any pair of rows and columns on the same index, the new matrix, L_{11} would have linearly independent columns/rows, thus invertible. We will use this fact to modify the equation so it's solvable.

Since there are known values in the F vector, i.e., the constraints, we can group those indices corresponding to the constraints vertices both on F and on rows/columns of L and separate them into block of matrices, where unknown and known parts of F are called $F1$ and $F2$ respectively. For L, it will be sectioned into 4 parts L_{11}, L_{12}, L_{21} ,and L_{22}. L_{11} corresponding to the unknown/known indices of F. Now, we can rewrite the last expression as,

$$\begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = 0 \tag{57}$$

Moreover, since L_{11} is the sliced version of L, we can be sure that L_{11} has independent rows and columns, which mean it's invertible! Now, since $F2$ is known, we will only have to solve for the first part, which is

$$L_{11}F_1 + L_{12}F_2 = 0 \tag{58}$$

$$F_1 = -F_{11}^{-1} L_{12} F_2 \tag{59}$$

Which can be linearly solved.
Therefore, E can be minimized using a linear solve.

b. Since the graph is connected, we can say that the "information" of the nodes are shared. This means, each node tries to stay relatively "similar" to its neighbors as much as it can. Again, since they are connected, each node can trace to all of the given constraint values. Therefore, we can intuitively says that for the optimal case of f, values of $f(v_i)$ have to be balanced and stay inbound of the given constraints, otherwise it won't be the minimum case, optimal case of f. With this tuition, we can say that the maximum value of f will be capped by the maximum values of f_0.

(A little bit more rigorous way to prove this, maybe?)
If we look at each row of the equation $L_{11}F_1 + L_{12}F_2 = 0$, we can see that it follows the form of (number of edges on v_i) x $f(v_i)$ - $\sum f(v_j)$ connected to that node = $\sum f_0(v_0)$ connected to that node i.

Here, we can see that $f(v_i) = (\sum f_0(v_0) + \sum f(v_j))/n$, where n is the number of the edges on the node i. We can see that the node i is basically the average of its neighbours' values.

Therefore, if each node is the weight average of its neighbours, the values of those nodes can't be more than the highest of their neighbours. Propagating this property through out the graph, we can see that max f = max f_0.

In terms of the physical interpretation of this problem, we can see that E measures how different each nodes are compare to their neighbors, or how "smooth" your graph is. Laplace's equation also measures how different the function is on the continuous space. Therefore, these two things are related by they both measuring differences for a given structures (graphs or functions), but on a different domain (discrete and continuous).

c. Please see the attached .py file.

d. Please see the attached .py file.

e. Please see the attached .py file. k-nearest neighbor graphs
Note: I could not fit 1000 numbers on the plot, but I can assured you that, based on the color gradient, each value lies between -1 and 1 of the constrainted node values.

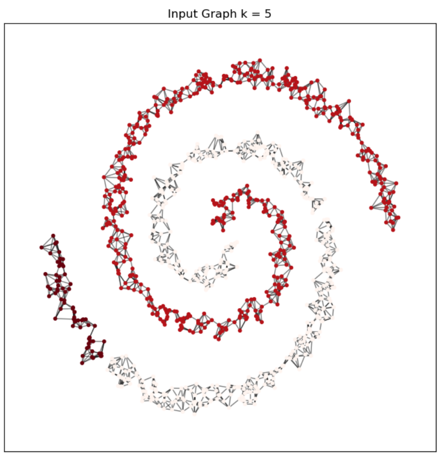


Figure 1: k = 5

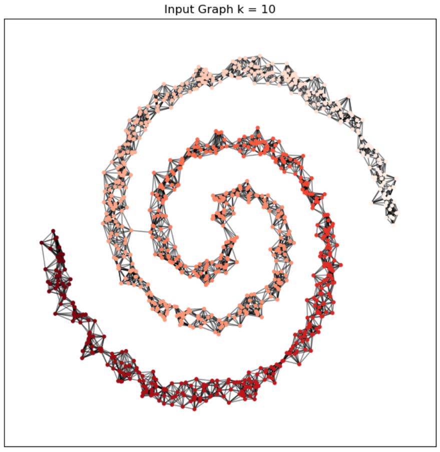


Figure 2: k = 10

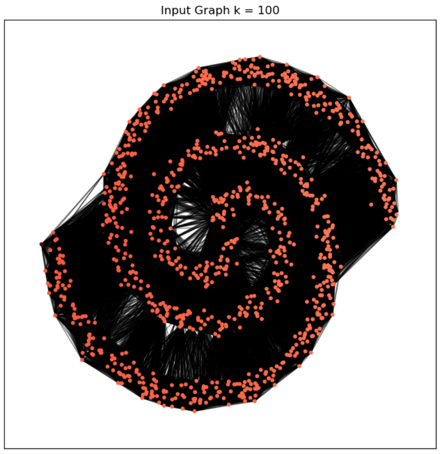


Figure 3: k = 100

We can see that, if k is too small, k = 5, the graph is disconnected, so the ”information” can’t be spread through out the system. Therefore, we ended up with separated graphs that does not share information with each other. a bunch of constants values between those graphs. If k is too large, k = 100, we can see that the nodes are connected too much, even with a very far one, resulting in a useless optimal node values. The result graph is kind of the average of the constraint values, which is not very insightful.

problem 4

Please see the attached scanned document.

①

1.a) Not all matrices $A \in \mathbb{R}^{n \times n}$ can be factored $A = LU$.
- an example would be the case where one of the diagonals is 0, the forward substitution would have to divided by $0 \to$ fail.
(permulation will be needed $\to$ LUP
permutation.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 1 \\ 5 & 7 & 9 \end{pmatrix}$$

b) If A admits $A = LL^T$

$$x^T A x = x^T L L^T x$$
$$= (L^T x)^T (L^T x)$$
$$= \|(L^T x)\|_2^2 \geq 0 \quad \forall x \in \mathbb{R}^n$$

Therefore A must be positive semi definite.
Positive Definite is when $x^T A x > 0 \quad \forall x \in \mathbb{R}^n - \{0\}$
So, this will fail when $x^T A x = 0 \quad \forall x \in \mathbb{R}^n - \{0\}$
$\|L^T x\|_2^2 = 0$ which can only happen either $x = 0$, which is not allowed, or $L^T x = 0 \quad x \neq 0$ which implies L^T is not linearly independent, which is also not true for lower triagular matrix with positive diagonals.

c) $Ax = b \Rightarrow \min_x \|Ax - b\|_2^2 \Rightarrow$ take derivative
$$2A^T A x - 2A^T b = 0$$
$$A^T A x - A^T b = 0$$
$$A^T A x = A^T b \;\#$$

This is numerically ill-advice to solve becau we are dealing with $\text{cond } A^T A \approx (\text{cond } A)^2$

If columns of A are "similar", $a_1 \approx a_2 \to$ we will have poor conditiony for A
column1 column2

and even ~~worse~~ worse for $A^T A$ #

① (cont)

d) Newton's method for root finding

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

If we pick x_0 such that $f(x_0) = 0$, x_{k+1} will equal x_k → No updates. So, this won't converge.

For example $f(x) = x - x^2$

$$x_{k+1} = x_k - \frac{x - x^2}{1 - 2x}$$

$$x_0 = 0 \Rightarrow x_1 = x_0 - \frac{x_0 - x_0^2}{1 - 2x_0}$$

$$= 0 - \frac{0 - 0}{1}$$

$$= 0$$

$\therefore x_0 = x_1 = x_2 = \dots = 0$ and will not converge (quadratically)

(2)

2.2) G has rank $\underline{d}$ (from $X \in \mathbb{R}^{d \times n}$)

for $G = Q \Lambda Q^T$, where $Q \in \mathbb{R}^{n \times r}$, $\Lambda \in \mathbb{R}^{r \times r}$

Apply SVD to G (eigenvalue decompositions)
we can write

$$G x_i = \lambda_i x_i \, , \quad x_i \in \mathbb{R}^n$$

Since G is symmetric, rank $d = r$
It has r orthogonal vectors (eigenvalues)

Therefore, we will have $x_1, x_2, \ldots, x_r$ correspondy to $G x_i = \lambda_i x_i$, $\forall \lambda_i$ positive.
Stack them together and x_i are orthogonal

$$G \begin{pmatrix} | & & & | \\ x_1 & x_2 & x_3 \cdots & x_r \\ | & & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{pmatrix}$$

Define $Q = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_r \\ | & | & & | \end{pmatrix}$; $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{pmatrix}$

$$G Q = Q \Lambda \Rightarrow G = Q \Lambda Q^{-1}$$

but $Q^{-1} = Q^T$

$\therefore G = Q \Lambda Q^T$

② (cont.)

2.3) $\hat{X} = \begin{bmatrix} | & | & & | \\ \hat{x}_1 & \hat{x}_2 & \cdots & \hat{x}_r \\ | & | & & | \end{bmatrix}$ $\hat{X}^T\hat{X} = \begin{bmatrix} \hat{x}_1^T\hat{x}_1 & \hat{x}_1^T\hat{x}_2 & \cdots & \hat{x}_1^T\hat{x}_r \\ \hat{x}_2^T\hat{x}_1 & \cdots & \cdots & \\ \cdots & & & \end{bmatrix}$

$\hat{X}^T\hat{X} = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^T$
$= Q\Lambda Q^T$
$= G$

$\boxed{\hat{X}^T\hat{X} = X^TX} \rightarrow \boxed{\hat{x}_i^T\hat{x}_j = x_i^Tx_j} \rightarrow$ Therefore, $\hat{P}_{ij} = \|\hat{x}_i\|_2^2 - 2\hat{x}_i^T\hat{x}_j + \|\hat{x}_j\|_2^2$

$= \hat{x}_i^T\hat{x}_i - 2\hat{x}_i^T\hat{x}_j + \hat{x}_j^T\hat{x}_j$

$= x_i^Tx_i - 2x_i^Tx_j + x_j^Tx_j$

$= \|x_i - x_j\|_2^2$

$= P_{ij}$

$\therefore \hat{P}_{ij} = P_{ij}$

$\therefore \hat{P} = P$ #

(3)

3.1) 1. Not all $\sqrt[3]{x}$ can be represented in a floating-point form, therefore, $\hat{y}$ would have to be ~~some~~ the approximation to some precision of $\sqrt[3]{x}$
$(\hat{y} \approx \sqrt[3]{x})$

2) Since we will use iterative algorithm, let say, Newton's method, the solution would depend on the Taylor's approximation $\left(f(x) \approx f(a) + \frac{f'(a)}{1}(x-a)\right)$
Therefore, the solution will have some small error $O(h^2)$ which makes $\hat{y} \neq \sqrt[3]{x}$

3.2) Forward error: $\sqrt[3]{x} - \hat{y}$

Backward error: $x - \hat{y}^3$

3.3) check how the solution changes when we perturb small h.

$$\lim_{h\to 0} \frac{\sqrt[3]{x+h} - \sqrt[3]{x}}{(x+h)-(x)} = \lim_{h\to 0} \frac{\sqrt[3]{x+h} - \sqrt[3]{x}}{h} \cdot \frac{\left(\sqrt[3]{(x+h)^2} + \sqrt[3]{x+h}\cdot\sqrt[3]{x} + \sqrt[3]{x^2}\right)}{\sqrt[3]{(x+h)^2} + \sqrt[3]{x+h}\sqrt[3]{x} + \sqrt[3]{x^2}}$$

$$= \lim_{h\to 0} \frac{\cancel{(x+h)-x}}{h\left(\sqrt[3]{(x+h)^2} + \sqrt[3]{x+h}\cdot\sqrt[3]{x} + \sqrt[3]{x^2}\right.}$$

$$= \lim_{h\to 0} \frac{1}{\sqrt[3]{(x+h)^2} + \sqrt[3]{x+h}\cdot\sqrt[3]{x} + \sqrt[3]{x^2}} = \frac{1}{3\sqrt[3]{x^2}} \#$$

④

4.2) $\min_{x} \frac{1}{2}\|Ax-b\|_2^2$

s.t. $\|x\|_2^2 = 1$

Lagrange multiplier

$$\frac{1}{2}\|Ax-b\|_2^2 + \lambda(\|x\|^2-1)$$

expand

$$f(x) = \frac{1}{2}\left(x^TA^TAx - \cancel{2x^T(A^Tb)}^{0} + \|b\|_2^2 \right) + \lambda(\|x\|^2-1)$$

$$\boxed{\begin{aligned} \frac{\partial f(x)}{\partial x} &= A^TAx + 2\lambda x = 0 \\ \frac{\partial f(x)}{\partial \lambda} &= \|x\|^2 - 1 = 0 \end{aligned}}$$

$\rightarrow A = U\Sigma V^T$

$(V\Sigma \cancel{U^T U}\Sigma V^T)x = -2\lambda x$

$V\Sigma^2 V^T x = -2\lambda x$

$\|x\|^2 = 1$

This is now an eigenvalue problem $\boxed{\begin{aligned} &Ax = \lambda x \\ &\text{s.t. } \|x\|^2 = 1 \end{aligned}}$

Finding eigenvector x that $\|x\|^2 = 1$

We can use KKT Conditions & SQP to solve this problem.

4.3) $\mathcal{L}(x,\lambda) = \frac{1}{2}\|Ax-b\|_2^2 + \lambda(\|x\|_2^2-1)$

$$\frac{\partial L}{\partial x} = A^TAx - A^Tb + 2\lambda x = 0$$

$$A^TAx + 2\lambda x = A^Tb$$

$$(A^TA + \lambda I)x = A^Tb$$

A^TA is invertible $\rightarrow (A^TA + \lambda I)$ is invertible

λI is diagonal for λ $\therefore x = (A^TA + \lambda I)^{-1}A^Tb$ #

④ cont.

4.4

From the previous, we can see that $x = (A^T A - \lambda I_{nn})^{-1} A^T b$

is a solution to the
$$\min_x \frac{1}{2}\|Ax-b\|_2^2$$
$$\text{s.t. } \|x\|_2 = 1$$

also, $\frac{\partial \mathcal{L}(x,\lambda)}{\partial \lambda} = \|x\|_2^2 - 1$

root of $f(\lambda) = \|x(\lambda)\|_2^2 - 1 = 0$

$$\|x(\lambda)\|_2^2 = 1$$

$$\|(A^T A - \lambda I_{nn})^{-1} A^T b\|_2^2 = 1$$

Therefore, the solution to the problem must satisfy both

$$x(\lambda) = (A^T A - \lambda I_{nn})^{-1} A^T b$$

and

$$\|x\|_2^2 = 1$$

where it can be written as

the root of $f(\lambda) = \|x(\lambda)\|_2^2 - 1$ #

(4.5)

since $A^TA = U\Sigma V^T$ (4.2) with $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots \sigma_n)$

Considers the denominator term : $A^TA - \lambda I_{nn}$.

$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$

if λ is close to $\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \end{bmatrix}$ it will make the diagonal term $\approx 0 \Rightarrow$ ill-condition #

(4.6)

$$\lim_{\lambda \to \infty} f(\lambda) = \lim_{\lambda \to \infty} \left\| \left(A^TA - \lambda I_{nn}\right)^{-1} A^T b \right\|_2^2 - 1$$

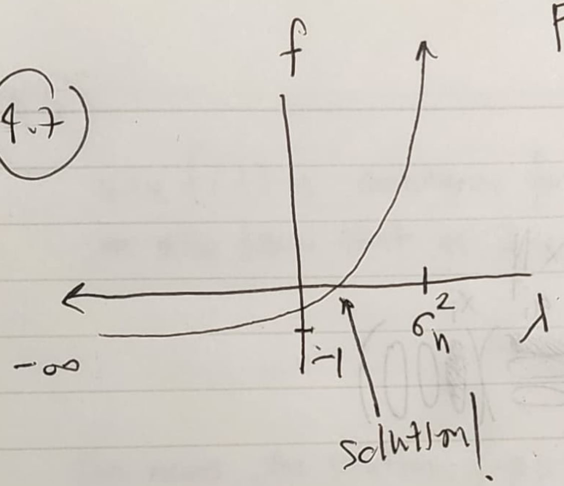
if λ is really large, the denominator will be dominated by the λI term.

$\hookrightarrow$ $A^TA - \lambda I_{nn} \Rightarrow -\lambda I_{nn}$

The term become $\left\| -\lambda^{-1} (I_{nn})^{-1} A^T b \right\|_2^2 - 1 = \left\| \frac{A^Tb}{\lambda} \right\|_2^2 - 1$

$$\lim_{\lambda \to \infty} \left\| \frac{A^Tb}{\lambda} \right\|_2^2 - 1 = 0 - 1$$
$$= -1 \#$$

(4.7)



From the limits $\lambda \to -\infty$, $f \to -1$

and

$\lambda \to \sigma_n^2$, $f \to \infty$

~~we can see~~ and f is continuous on $\lambda < \sigma_n^2$

we can be sure that there is a solution to $f \approx 0$ for $\lambda < \sigma_n^2$

Consider $f = \|(A^TA - \lambda I)^{-1} A^T b\|_2^2 - 1$, we can see that it's also differentiable on $\lambda < \sigma_n^2$, therefore, we might use Newton's method to iteratively find the root.