AI 512 Assignment #1

SHOUNAK SHIRODKAR
IMT2019083

Solution 1

The optimal p in terms of distance to a is also written as $argmin_x \; ||(1-\theta)a + \theta b - x||$.

Consider $y \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that $y(\theta) = ||(1-\theta)a + \theta b - x||^2$. Also let $g \colon \mathbb{R} \to \mathbb{R}_{n\times 1}$ be such that $g(\theta) = (1-\theta)a + \theta b - x$.

Thus, $y(\theta) = (g(\theta))^T(g(\theta))$, from the formulation $||A||^2 = A^T A$, for all matrices A.

Equivalently, $y(\theta) = (g(\theta))^T I(g(\theta))$, where I is the $n \times n$ identity matrix.

Claim 1.1. *Consider a scalar λ defined as $\lambda = x^T Ax$, where x, A are matrices of dimensions $n \times 1$, $n \times n$ respectively, and A is independent of x. Then*

$$\frac{\partial \lambda}{\partial x} = x^T(A + A^T)$$

Proof. By definition, we have -

$$\lambda = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

Differentiating with respect to x_k, the k^{th} element of x,

$$\frac{\partial \lambda}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

for $k = 1, 2, 3 \ldots n$. Combining all x_k, we get

$$\sum_{k=1}^{n} \frac{d\lambda}{dx_k} = \sum_{k=1}^{n} \sum_{j=1}^{n} a_{kj} x_j + \sum_{k=1}^{n} \sum_{i=1}^{n} a_{ik} x_i$$

This equation may be alternatively presented as

$$\frac{\partial \lambda}{\partial x} = x^T(A + A^T)$$

as in the stated claim. $\square$

From the above result,

$$\begin{aligned}
y'(\theta) &= \frac{\partial y(\theta)}{\partial g(\theta)} \frac{\partial g(\theta)}{\partial \theta} \\
&= (g(\theta))^T (I + I^T) g'(\theta) \\
&= (g(\theta))^T (2I) g'(\theta), \quad \because I = I^T \\
&= 2g(\theta)^T g'(\theta) \\
&= 2((1-\theta)a^T + \theta b^T - x^T)(b - a), \quad \because g'(\theta) = b - a
\end{aligned}$$

Investigating the second-order derivative

$$\begin{aligned} y''(\theta) &= 2(b^T - a^T)(b-a) \\ &= 2(b-a)^T(b-a) \\ &= 2||b-a||^2 \geq 0 \end{aligned}$$

The function is concave in the neighbourhood of the inflection point, therefore the inflection point(s) is a minima.

To obtain the inflection point, we solve

$$\begin{aligned} y'(\theta) &= 0 \\ 2g(\theta)\cdot(b-a) &= 0 \\ (b-a)\cdot((1-\theta)a+\theta b - x) &= 0 \\ (b-a)\cdot(\theta(b-a)+(a-x)) &= 0 \\ \theta &= \frac{(b-a)\cdot(x-a)}{||b-a||^2} \end{aligned}$$

The optimal point closest to x and along vector $a-b$ is

$$\begin{aligned} p &= (1-\theta)a+\theta b \\ &= \frac{(b-a)}{||b-a||^2}\cdot((b-x)\cdot a + (x-a)\cdot b) \end{aligned}$$

To show $(p-x)\perp(a-b)$, it is sufficient to demonstrate $(p-x)\cdot(a-b)=0$.

$$\begin{aligned} (p-x)\cdot(a-b) &= \frac{(b-a)}{||b-a||^2}((b-x)\cdot a + (x-a)\cdot b - (b-a)\cdot x)\cdot(a-b) \\ &= 0 \end{aligned}$$

Diagrammatically:

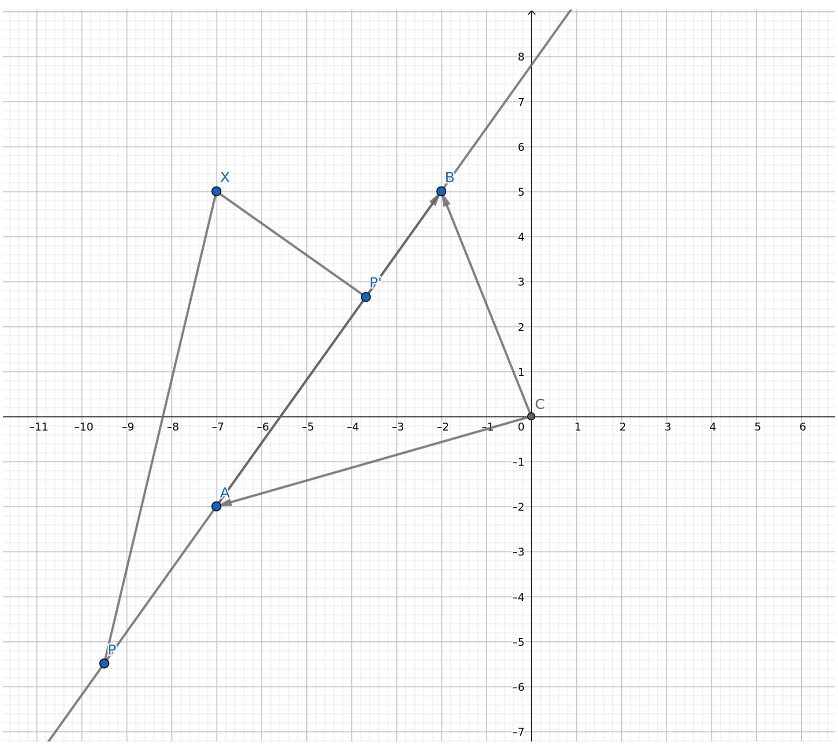

Figure 1: P slides along vector A - B; P' is the optimal choice for closeness to X, and lies on the perpendicular to A - B through X.

Solution 2

(1) $x = [-sin7\frac{1}{2}^{\circ}, 1 + cos7\frac{1}{2}^{\circ}], z_1 = [sin7\frac{1}{2}^{\circ}, 1 + cos7\frac{1}{2}^{\circ}], z_2 = [0\ 1];$
$\angle(x, z_1) = 2\angle(x, z_2) = 7\frac{1}{2}^{\circ}$ and $||x - z_2|| = 1 > ||x - z_1|| = 2sin7\frac{1}{2}^{\circ}$.

(2) The *distance* nearest neighbour of x is $argmin_{z_i}||x - z_i||$, for $i \in \{1, \ldots, m\}$.

The expression for distance may be written as

$$\begin{aligned} ||x - z_i|| &= \sqrt{(x - z_i)^T (x - z_i)} \\ &= \sqrt{||x||^2 - 2x^T z_i + ||z_i||^2}, \quad \because A^T A = ||A||^2 \end{aligned}$$

The distance *argmin* may be simplified as

$$argmin_{z_i}||x - z_i|| = argmax_{z_i}(x^T z_i)$$

given that the rest of the quantities are non-parametric.

The *angle* nearest neighbour is $argmin_{z_i}\angle(x, z_i)$, for $i \in \{1, \ldots, m\}$.

The angle between the two vectors may be evaluated via the dot product as

$$\begin{aligned} \angle(x, z_i) &= arccos\left(\frac{x^T z_i}{||x|| \cdot ||z_i||}\right) \\ &= arccos\left(\frac{x^T z_i}{||x||}\right) \quad \because ||z_i|| = 1,\ z_i \text{ are normalised} \end{aligned}$$

The angle *argmin* may be simplified as

$$argmin_{z_i}\angle(x, z_i) = argmax_{z_i}(x^T z_i)$$

given that *arccos* monotonically decreases across its domain.

Since both *argmins* are equivalent, the *distance* and *angle* nearest neighbours of x are the same when all z_i are normalised.

Solution 3

(1) The Richardson algorithm performs the operation $x^{(k+1)} = x^{(k)} - \mu A^T(Ax^{(k)} - b)$, where k is the iteration number. It is assured that A is tall and with linearly independent columns, so that $A^T A$ is invertible, by the preconditions of the Least Squares Problem.

$$
\begin{aligned}
\mu A^T(Ax^{(k)} - b) &= 0, \quad \because x^{(k+1)} = x^{(k)} \\
A^T A x^{(k)} - A^T b &= 0 \\
(A^T A)^{-1} A^T A x^{(k)} &= (A^T A)^{-1} A^T b, \quad \text{left-multiplying with } (A^T A)^{-1} \\
x^{(k)} &= (A^T A)^{-1} A^T b
\end{aligned}
$$

The optimal solution to the Least Squares Problem, $\hat{x}$, is given by the expression $\hat{x} = A^\dagger b$, where $A^\dagger$ is the Moore-Penrose Pseudo Inverse. $A^\dagger$ is determined from the formulation $A^\dagger = (A^T A)^{-1} A^T$, which is also the prefix to matrix b in our expression for $x^{(k)}$.

Therefore, $x^{(k)} = \hat{x}$ if $x^{(k+1)} = x^{(k)}$.

(2) The program was written using Python3.8 with the aid of numpy and matplotlib libraries.

RICHARDSON ALGORITHM

In [37]:

```
import numpy as np
import matplotlib.pyplot as plt

#initialise A and b
A = np.random.rand(20, 10)
b = np.random.rand(20,1)

#closed-form solution to least squares problem
x_hat = np.dot(np.linalg.pinv(A), b)

#store norms to plot convergence
norm_array = np.zeros(500)

#initialise x_k=1
x_k = np.zeros((10, 1))
#set parameter mu to 1 / ||A||^2
mu = 1 / (np.linalg.norm(A))**2

#run Richardson for 500 iterations
for i in range(500):
    x_k -= mu*np.dot(np.transpose(A), (np.dot(A, x_k) - b))
    norm_array[i] = np.linalg.norm(x_k - x_hat)

#plot ||x_k - x||
plt.figure(figsize=(30,10))
plt.plot(norm_array, 'b')
plt.plot([0 for i in range(500)], 'r')
plt.show()
```

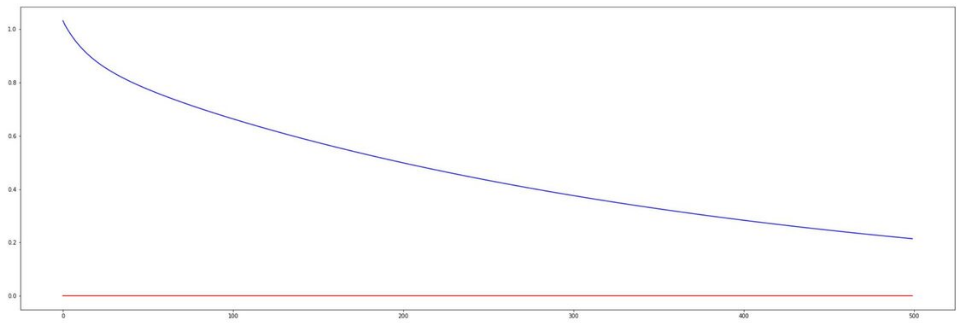

Figure 2: Jupyter Notebook for problem 3.2