# Assignment 3: Computational Methods of Optimization

E0-230

SRNumber

November 20, 2024

## Contents

# Instructions

- This is an individual assignment; all work must be your own.
- Code must be written in Python with appropriate comments.
- Submit two files: `SRNumber.pdf` and `SRNumber.py`.

# 1 Systems of Linear Equations (15 points)

We're given the following matrix and vector and we'd like to calculate the least norm solution that satisfies this linear constraint.

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix} \textbf{ and } b = \begin{bmatrix} 10 \\ -6 \\ -5 \end{bmatrix}$$

## 1.1 Problem 1.1 (2 points)

Show that the given system of equations has an infinite number of solutions.

### Solution

Given the system of linear equation, $\mathbf{Ax = b}$, the augmented matrix $[A|b]$ would be .

$$[A|b] = \begin{bmatrix} 2 & -4 & 2 & -14 & 10 \\ -1 & 2 & -2 & 11 & -6 \\ -1 & 2 & -1 & 7 & -5 \end{bmatrix}$$

We know that if the rank(A) = rank([A—b] $\leq$ Number of variables then we'd have $\infty$ solutions. Using gaussian elimination the Gaussian elimination, we've :

$$A = \begin{bmatrix} 1 & -2 & 0 & -3 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and A} = \begin{bmatrix} 1 & -2 & 0 & -3 & 4 \\ 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus we can see that the rank of both the matrices are 2 that is less than the number of variables, thus there are $\infty$ many solutions

## 1.2 Problem 1.2 (2 points)

Express the problem of finding the least-norm solution as an optimization problem (ConvProb) with convex constraints and a strongly convex objective function.

### Solution

The above problem could be expressed in the form of

$$\arg\min_{x \in R^4} f(x) = \frac{1}{2}\|x\|^2$$

$$Ax = b$$

Now to show that the constraints are convex, Let's suppose we have S $= x : A^T x = b$ and let $x_1, x_2 \in$ S. So we need to prove that $\theta x_1 + (1-\theta)x_2 \in$ S, $\forall\theta \in (0,1)$. So $A^T x_1 = b$ and$A^T x_2 = b$. Let y $= \theta x_1 + (1-\theta)x_2$. Hence

$$A^T y = A^T[\theta x_1 + (1-\theta)x_2] = A^T\theta x_1 + A^T(1-\theta)x_2 = \theta A^T x_1 + (1-\theta)A^T x_2 = \theta b + (1-\theta)b = b$$

So, we can conclude the the set S is convex. Now we'd like to prove that function is strongly convex. Consider function, $x = [x_1, x_2, x_3, x_4]^T$

$$\text{f(x)} = \frac{1}{2}[x_1^2 + x_2^2 + x_3^2 + x_4^2]$$

$$\nabla f = \frac{1}{2}\begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \\ 2x_4 \end{bmatrix}$$

$$H_f(x) = \mathbf{I}$$

So the hessian of the objective function is identity which is P.D., hence the objective function is strongly convex.

## 1.3 Problem 1.3 (3 points)

Use the KKT conditions to solve ConvProb, and derive an expression for $x^*$.

### Solution

Given that the objective function is strongly convex and the constraints are linear, a KKT point is also the global optimiser of the objective function. So, the Lagrange function is defined by

$$\mathcal{L}(\text{x}^*, \gamma) = \frac{1}{2}x^{*T}x^* + \gamma(Ax^* - b)$$

. Let $x^*$ be a KKT point, then

$$\nabla\mathcal{L}(x^*, \gamma) = 0 \implies x^* + A^T\gamma = 0 - (1)$$

$$Ax^* - b = 0 - (2)$$

Now, we multiply $A^T$ in eq(1)

$$A^T x^* + AA^T\gamma = 0 \implies b + AA^T\gamma = 0 \implies \gamma = -(AA^T)^{-1}b$$

So, from equation (1), x$^* = -A^T\gamma = A^T(AA^T)^{-1}b$

## 1.4 Problem 1.4 (3 points)

Derive a projection operator for the constraint set of ConvProb.

## Solution

So, we'd like to derive the projection operator.For x $\in R^4$, P$_c$(x) be the projection operator of the x in the constraint, **C**. So, the problem of find the projection operator could be stated as

$$\arg\min_{x \in R^4} P_c(z) = \frac{1}{2}\|z-x\|^2 \text{and g(x) } = Ax - b = 0$$

Expanding the P$_c$(x), we've:

$$P_c(z) = \frac{1}{2}(x^Tx - 2z^Tx + z^Tz)$$

$$\nabla P_c(z) = \frac{1}{2}(2x - 2z_)$$

$$H_{P_c}(z) = \mathbf{I}$$

So, the optimization problem is convex and as the constraints are linear, we can again try to find the KKT point which also be a global optimum. We define the Lagrange's as follows :

$$\mathcal{L}(x,\gamma) = \frac{1}{2}(x^Tx - 2z^Tx + z^Tz) - \gamma(Ax^* - b)$$

Based on the KKT conditions, we've:

$$\nabla\mathcal{L}(x^*,\gamma) = x^* - z + A^T\gamma = 0 - (1) \text{ and } Ax^* - b = 0 - (2)$$

Now multiplying the equation 1 with A, then we have

$$Ax^* - Az - AA^T\gamma = 0 \implies b - Az + AA^T\gamma = 0 \implies \gamma = (AA^T)^{-1}(Az - b)$$

Now putting the value of $\gamma$ back to equation 1, we've

$$x^* = z + A^T\gamma \implies z + A^T(AA^T)^{-1}(Az - b)$$

So the final expression of projection operator, P$_c$(z) is

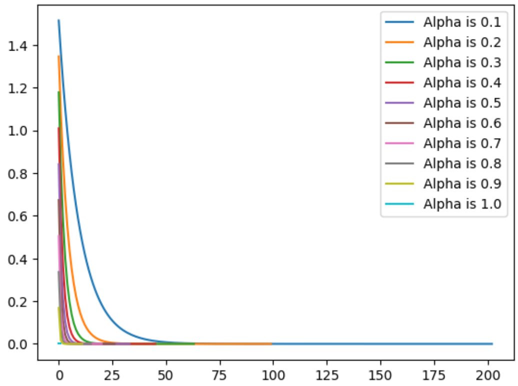
$$P_c(z) = [I - A^T(AA^T)^{-1}A]z + A^T(AA^T)^{-1}b$$

## 1.5 Problem 1.5 (5 points)

Implement projected gradient descent to solve ConvProb and plot $\|x(t) - x^*\|$ for each iteration.

## Solution

Below is the plot for $\|x(t) - x^*\|$ for each iteration and different values of $\alpha$



# 2 Support Vector Machines (15 points)

## 2.1 Problem 2.1 (2 points)

Solve the primal using CVXPY for the given data.

### Solution

After solving through the CVXPY, we find that the optimal objective value is 2.66 and it is attained at w$^* = [1.1547, -2]^T$ and b$^* = 1$

## 2.2 Problem 2.2 (3 points)

Show that the dual function has the given form.

### Solution

We start with the given primal optimization problem:

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

subject to:

$$y_i(w^\top x_i + b) \geq 1, \quad i = 1, \ldots, N.$$

Here, $w \in \mathbb{R}^d$, $b \in \mathbb{R}$, and $y_i \in \{-1, 1\}$ are the labels of the data points $x_i \in \mathbb{R}^d$. The Lagrangian for this problem is:

$$L(w, b, \Lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i \left( y_i(w^\top x_i + b) - 1 \right),$$

where $\Lambda = [\lambda_1, \ldots, \lambda_N]^\top$ are the Lagrange multipliers (dual variables) with $\lambda_i \geq 0$. The dual function $g(\Lambda)$ is defined as:

$$g(\Lambda) = \inf_{w,b} L(w, b, \Lambda).$$

#### Minimization with respect to $w$

The Lagrangian can be rewritten as:

$$L(w, b, \Lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i y_i w^\top x_i - b \sum_{i=1}^{N} y_i \lambda_i + \sum_{i=1}^{N} \lambda_i.$$

Focusing on the $w$-dependent terms:

$$\frac{1}{2} w^\top w - w^\top \left( \sum_{i=1}^{N} \lambda_i y_i x_i \right).$$

Minimizing this with respect to $w$ involves taking the gradient and setting it to zero:

$$\nabla_w L = w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \quad \implies \quad w = \sum_{i=1}^{N} \lambda_i y_i x_i.$$

**Substituting $w$ back**

Substituting $w = \sum_{i=1}^{N} \lambda_i y_i x_i$ into $L(w, b, \Lambda)$, the term $\frac{1}{2}\|w\|^2$ - $w^\top \left(\sum_{i=1}^{N} \lambda_i y_i x_i\right)$becomes:

$$\frac{1}{2}w^\top w - w^\top w = -\frac{1}{2}w^\top w = -\frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i x_i\right)^\top \left(\sum_{j=1}^{N} \lambda_j y_j x_j\right).$$

This simplifies to:

$$\frac{1}{2}\Lambda^\top A \Lambda,$$

where $A_{ij} = -y_i y_j x_i^\top x_j$.
The remaining terms in the Lagrangian become:

$$-b\sum_{i=1}^{N} y_i \lambda_i + \sum_{i=1}^{N} \lambda_i.$$

**Minimization with respect to $b$**

The $b$-dependent term is:

$$-b\sum_{i=1}^{N} y_i \lambda_i$$

Taking the derivative of the b-dependent term w.r.t b, we have

$$\sum_{i=1}^{N} y_i \lambda_i = 0$$

This term is linear in $b$, and since there are no constraints on $b$, it simply contributes $\Lambda^\top b$ to the dual function.

Combining all terms, the dual function is:

$$g(\Lambda) = \Lambda^\top \mathbf{b} + \frac{1}{2}\Lambda^\top A \Lambda,$$

where $\mathbf{b}$ is a vector of ones and $A_{ij} = y_i y_j x_i^\top x_j$.
Value of K is N

## 2.3 Problem 2.3 (3 points)

Prove that $\sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i = \gamma$.

### Solution

We know from above that

$$\sum_{i=1}^{N} y_i \lambda_i = 0$$

and the $\lambda_i \geq 0, \forall I = 1...N$, So this means

$$\sum_{i=1}^{N} y_i \lambda_i = 0 \implies \sum_{i=1,y=1}^{N} \lambda_i - \sum_{i=1,y=-1}^{N} \lambda_i = 0$$

$$\sum_{i=1,y=1}^{N} \lambda_i = \sum_{i=1,y=-1}^{N} \lambda_i = \gamma, \text{for some } \gamma$$

It's been observed that the value of $\gamma = 2.666$

## 2.4 Problem 2.4 (3 points)

Write a program to solve the dual problem and report the optimal value.

### Solution

## 2.5 Problem 2.5 (2 points)

Identify the active primal constraints.

### Solution

In the initial convex optimization question, the primal constraints are given by

$$g_i(x) = y_i(w\top x_i + b) - 1 \geq 0$$

For a constraint to be active, $g_i(x) = 0$ for that i. Hence, for the optimal x*, $\lambda_i * g_i(x^*) = 0, \forall i$. Hence if the $\lambda_i > 0$, then we can conclude that the $g_i(x) = 0$, i.e. the i$^{th}$ constraint is active for x*. For the above dual problem, the constraints are g$_1, g_2, g_7, g_1 0$

## 2.6 Problem 2.6 (2 points)

Plot the classifier, data points, and active constraint points.

### Solution

# References