E0 230 : Computational Methods of Optimization
Assignment 4

Instructions:

- Attempt all questions
- For Numerical Answer questions, insert only the digits **with no leading or trailing spaces** in the MS Teams Form provided
- You can submit the MS Teams Form **only once**. No further changes can be made, so click on *Submit* wisely.
- For the file `a4.csv`, each column is a 3d vector of the form $[x_i, y_i]$; $x_i \in \mathbb{R}^2$, and $y_i \in \{\pm 1\}$.
- Late Submissions will be penalised.

1. Consider the problem

$$x^* = \arg\min \frac{1}{2}\|Cx + d\|^2 \text{ such that } Ax = b$$

(a) Suppose we want to project a point u onto the set $\{x : Ax = b\}$, where A is full rank and b is nonzero. What is the closed form solution to the projection x?

A. $x = u - A^T(AA^T)^{-1}(b - Au)$
B. $x = u - A^T(A^TA)^{-1}(b - Au)$
C. $\boldsymbol{x = u - A^T(AA^T)^{-1}(Au - b)}$
D. $x = u - A^T(A^TA)^{-1}(Au - b)$

Solution: You have to solve an equality constrained minimization problem

$$x^* = \arg\min \ \frac{x^Tx}{2} - u^Tx \quad \text{such that:} \quad Ax = b.$$

We set up the Lagrangian and it's gradient to get

$$\begin{aligned}\mathcal{L}(x,\mu) &= \frac{1}{2}x^Tx - u^Tx + \mu^T(Ax - b)\\ \nabla_x\mathcal{L}(x,\mu) &= x - u + A^T\mu = 0.\end{aligned}$$

We then solve for μ by multiplying the second equation by A (to eliminate x since $Ax = b$) and substitute the solution for μ to get the value x. The correct answer here is C.

(b) Suppose A is of full rank, C^TC is invertible and b is nonzero. What is the closed form solution to this problem?

A. $\boldsymbol{x^* = (C^TC)^{-1}A^T(A(C^TC)^{-1}A^T)^{-1}(b + A(C^TC)^{-1}C^Td) - (C^TC)^{-1}C^Td}$
B. $x^* = (C^TC)^{-1}A(A^T(C^TC)^{-1}A)^{-1}(b - A^T(C^TC)^{-1}Cd) - (C^TC)^{-1}C^Td$
C. $x^* = (C^TC)^{-1}A^T(A^TA)^{-1}b - (C^TC)^{-1}C^Td$
D. $x^* = (C^TC)^{-1}A(AA^T)^{-1}b - (C^TC)^{-1}C^Td$

Solution: We use the same methodology as above:

$$\begin{aligned}\mathcal{L}(x,\mu) &= \frac{1}{2}x^TC^TCx - (C^Td)x + \mu^T(Ax - b)\\ \nabla_x\mathcal{L}(x,\mu) &= C^TCx - C^Td + A^T\mu = 0.\end{aligned}$$

We isolate and remove x, by first premultiplying the second expression by $(C^TC)^{-1}$ and then by A. We then solve for μ, and substitute that value back into the expression for x. The correct answer is A.

(c) Suppose

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \ b = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \ C = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ d = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

Starting at $(0,0,0)$, how many iterations of projeted gradient descent with stepsize $\eta = 0.5$ are required to reach a point z such that $\|z - x^*\| \leq 0.001$?

Solution: $N = 8$

(d) Starting at $(0,0,0)$, how many iterations of projeted gradient descent with stepsize $\eta = 0.25$ are required to reach a point z such that $\|z - x^*\| \leq 0.001$?

> **Solution:** $N = 21$

(e) Starting at $(0,0,0)$, how many iterations of projeted gradient descent with stepsize $\eta = 2.5$ are required to reach a point z such that $\|z - x^*\| \leq 0.001$?

> **Solution:** No convergence.

Consider the sets $S(0,1) := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ and $S(a,r) := \{x \in \mathbb{R}^n : \|x - a\|^2 \leq r^2\}$.

2. (a) What is the projection x of a point y onto $S(0,1)$?

A. $x = \frac{1}{\|y\|} y$ for all y.

B. $\boldsymbol{x = y \frac{1}{\max\{1, \|y\|\}}}$ for all $\boldsymbol{y}$

C. $x = y \frac{1}{\min\{1, \|y\|\}}$ for all y

> **Solution:** This problem can be solved in one of two ways. The first is with KKT conditions, whereby we take objective function to the projection cost. The second is by using geometry: $S(0,1)$ is the unit circle. Therefore, to project a vector onto it, we simply need to divide by the norm if the vector has norm greater than 1 (that is, if the point is outside $S(0,1)$).
>
> To solve this problem with KKT, we do the following:
>
> $$\begin{aligned} L(x, \lambda) &= \|x - u\|^2 + \lambda(\|x\|^2 - 1) \quad \text{(why?)} \\ \nabla_x L(x^*, \lambda^*) &= 2x^* - 2u + 2\lambda^* x^* = 0 \Rightarrow x^*(1 + \lambda^*) = u \\ \lambda^* &\geq 0 \\ \lambda^*(\|x^*\|^2 - 1) &= 0 \\ \|x^*\|^2 - 1 &\leq 0 \end{aligned}$$
>
> If $\|x^*\| < 1$ then $\lambda^* = 0$. If $\|x^*\| = 1$, we have $\lambda^* > 0$. It follows that if $\lambda^* = 0$, $x = u$. Thus, we focus on the latter case. With some algebra, we get $\lambda^* = \langle u, x^* \rangle - 1$. Substituting this, and with some further algebra, we get $x^*(u^T x^*) = u$, which shows that x^* and u are parallel (why?). So we can take $x^* = ku$. We can then solve for k, getting $k = \frac{1}{\|u\|}$. From this, we get answer B.

(b) What is the projection x of a point y onto $S(a,r)$?

A. $x = \frac{r}{\|y-a\|}(y - a)$ for all $y \in \mathbb{R}^n \backslash S(a,r)$

B. $\boldsymbol{x = \frac{r}{\|y-a\|}(y - a) + a}$ for all $\boldsymbol{y \in \mathbb{R}^n \backslash S(a,r)}$

C. $x = \frac{r}{\|y-a\|}(y - a) - a$ for all $y \in \mathbb{R}^n \backslash S(a,r)$

> **Solution:** Substitute $\bar{x} = x - a$, $\bar{y} = y - a$, and solve the projection onto $S(0,r)$ using KKT.

(c) Suppose $S \subset \mathbb{R}^2$, $S := S(0,1) \cap S((-1,1), 1/\sqrt{2})$. Let

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix},$$

and consider the problem

$$x^* = \arg\min_{x \in S} \|Ax + b\|.$$

We'll solve this problem with projected gradient descent. How many iterations are needed to reach a point z such that $\|z - x^*\| \leq 0.0001$ with a stepsize $\alpha = 0.01$, starting at $x_0 = (3, 0)$ and what is the final answer?

Use the following algorithm to project any point $z \in \mathbb{R}^2$ onto S:

Algorithm:

1. Initialize: $C = S(0, 1)$, $D = S((-1, 1), 1/\sqrt{2})$, $z \in \mathbb{R}^2$ $w_0 = P_C(z)$
2. while $\|w_k - y_k\| \geq 0.0001$, Compute:
   (a) $y_k = P_D(w_k)$
   (b) $w_{k+1} = P_C(y_k)$
3. Output $w = w_{k+1} = P_S(z)$

Solution: $N = 6/27/3$ for and $x^* = [-.9557, 2943]^T$.

3. Consider the problem

$$\begin{aligned} \underset{w,b}{\arg\min} \ & \frac{1}{2}\|w\|^2 \\ \text{such that} \ & y_i(w^T x_i + b) \geq 1 \end{aligned}$$

where $y_i \in \{\pm 1\}$, $x_i \in \mathbb{R}^n$ are given scalars. We have provided data $\{(x_i, y_i)\}_{i=1}^{10}$ in `a4.csv`.

(a) Consider the problem stated above. We will use active set methods to solve this problem. Initialize the problem with $w0 = [132.1473, -53.9619, 0]$,. What is the initial working set?

Solution: We check $v = Ax_k - b$. If $v_i = 0$, the ith constraint is active. We have only 2 active constraints. $[2, 6]$.

(b) Find the feasible direction for this set.

Solution: We solve the quadratic

$$\arg\min \ u^T Q u + g_0^T u \quad \text{such that} \quad A_w u = 0.$$

Here, we get $u = [-132.9293, 51.9612, -0.9446]$. However, if $b = 0$ is fixed, then $u = 0$

(c) What is the next step you need to take?

A. Find the blocking constraints

B. Check the feasibility of $x + u$

C. Compute the Lagrange multipliers.

Solution: Case 1 ($b = 0$): Since $u = 0$, we need to check the lagrange multipliers.

Case 2: Check the feasibility of $x + u$

(d) If you chose (a), what is the optimal stepsize? If you chose (b), is the point feasible? If you chose (c), what is the smallest Lagrange multiplier?

> **Solution:** Case 1: Check quadprog or solve directly. 2 values are accepted based on some minor inconsistencies (2e4, $\approx$, 595).
>
> Case 2: no, $x + u$ is not feasible (check with constraints).

(e) What is the working set after you complete this iteration?

> **Solution:** 2 (though I'll accept 6 as well due to ambiguity), or [2,3,6].
>
> To get the [2,3,6], we first solve for α and j as discussed in the lectures, i.e.
>
> $$\alpha = \arg\min_{i:\langle a_i, u\rangle < 0} \frac{b - \langle a_i, x_0\rangle}{\langle a_i, u\rangle}.$$
>
> We see that for all $i \ni W$, $\langle a_i, u\rangle < 0$, and ultimately solve $\alpha = .937$ and $i^* = 3$. Thus, we get the final active set to be $[2, 3, 6]$.

4. Consider the problems

$$m_1 = \arg\min_m \sum_i \|x_i - m\|_1$$

and

$$m_\infty = \arg\min_m \sum_i \|x_i - m\|_\infty.$$

We will use the data given in `a4.csv` to answer the following questions. You may use any LP solver you please, such as `linprog`, `sedumi`, or `cvx/cvxpy`.

(a) Reformulate the problem of finding m_∞ into a linear program, and solve it.

A. $\boldsymbol{m_\infty = (-.116, .167)}$
B. $m_\infty = (0.821, 0.733)$
C. $m_\infty = (.137, .924)$
D. $m_\infty = (-.137, .924)$

> **Solution:** To understand how to set up the LP, see tutorial 4. The optimal value of the LP is 4.244, and is achieved at multiple points. You'll get a different answer based on the solver used.

(b) Now consider the problem

$$x^* = \arg\min \|x\|_1 \text{ such that } Ax = b.$$

We aim to solve this problem with linear programming. Consider the following change of variables: $x = u - v$, where $u_i = \max\{x_i, 0\}$ and $v_i = \max\{-x_i, 0\}$. Using this change of variables, convert the problem into a linear program and solve it. Use the following problem data:

$$A = \begin{bmatrix} 0.8147 & 0.6324 & 0.9575 & 0.9572 \\ 0.9058 & 0.0975 & 0.9649 & 0.4854 \\ 0.1270 & 0.2785 & 0.1576 & 0.8003 \\ 0.9134 & 0.5469 & 0.9706 & 0.1419 \end{bmatrix}, \quad b = \begin{bmatrix} 0.4218 \\ 0.9157 \\ 0.7922 \\ 0.9595 \end{bmatrix}$$

A. $x = (0.1060, 0.5418, 0.0233, 0.5866)$
B. $x = (3.1400, -12.9328, 4.2818, -1.8395)$.
C. $\boldsymbol{x = (17.268, 0.838, -15.893, 1.088)}$

D. $x = (2.8825, -0.5676, 0.9225, 10.0653)$

Solution: Uses $x_i = u_i - v_i$, $|x_i| = u_i + v_i$. The problem becomes

$$\begin{aligned}(u^*, v^*) = \arg\min \;\; & 1^T u + 1^T v \\ \text{s.t.} \;\; & [A, -A][u^T, v^T]^T = b \\ & u \geq 0 \\ & v \geq 0\end{aligned}$$

(c) We will now try and convert the problem of finding m_1 into a linear program as well. Suppose we use the substitution of variables we defined in the previous part. Reformulate the problem of finding m_1 into a linear program, and find the answer.

A. $m_1 = (-0.8407, 0.1580)$

B. $\boldsymbol{m_1 = (0.0032, 0.0264)}$

C. $m_1 = (-0.1660, -0.0313)$

D. $m_1 = (0.0264, 0.0108)$

Solution: Let $t_i = x_i - m = u_i - v - i$, and $|t_{i,j}| = u_{i,j} + v_{i,j}$. Set up the LP similar to the above, and solve. The optimal value for the LP is 6.377, and there are infinitely many points that achieve this. Note that the problem is asking you to find a vector that is the element-wise sample median of the dataset (why?).