Problem Set 3

Due: Apr 1st 2025 11:59pm ET

Instructions. This problem set is due after the midterm, but solving it will be helpful for the midterm. While we will release the solutions after the midterm, you are encouraged to go to the office hours to verify your solutions. The total number of points is 100.

Collaboration policy. We encourage working together whenever possible: in the recitations, problem sets, and general discussion of the material and assignments. Keep in mind, however, that for the problem sets the solutions you hand in should reflect your own understanding of the class material, and should be written solely by you. It is not acceptable to copy (in whole or in part) a solution that somebody else has written.

1. KKT conditions and Farkas lemma [20pts]

Consider a general optimization problem with functional constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0 \qquad i = 1, ..., s \\ & h_j(x) = 0 \qquad j = 1, ..., r \end{aligned}$$

where f and all g_i, h_j are differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$.

(1.a) [15pts] Let x^* be a minimizer of the problem and define the index set of the active constraints $I(x^*) = \{i \in \{1, ..., m\} : g_i(x) = 0\}$. Show that there exists $\lambda^* \in \mathbb{R}^s$ and $\mu^* \in \mathbb{R}^r$ that satisfies KKT conditions:

$$\begin{aligned} \lambda_i^* \ge 0, \qquad & i = 1, ..., s \\ \lambda_i^* g_i(x^*) = 0, \qquad & i = 1, \cdots, s \end{aligned}$$

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{r} \mu_i^* \nabla h_j(x^*) = 0$$

if and only if

$$\begin{aligned} & \{d \in \mathbb{R}^n : \langle \nabla f(x^*), d \rangle < 0\} \\ \cap & \{d \in \mathbb{R}^n : \langle \nabla g_i(x^*), d \rangle \le 0, i \in I(x^*)\} \\ \cap & \left\{d \in \mathbb{R}^n : \left\langle \nabla h_j(x^*), d \right\rangle = 0, j = 1, ..., r\right\} = \emptyset \end{aligned}$$

- *Hint:* Use the Farkas Lemma.

Solution. For simplicity, we assume $I(x^*) = \{1, 2, ..., s\}$ (since for $i \notin I(x^*)$, we can simply set $\lambda_i^* = 0$). We denote I_k as the identity matrix of size k and $0_{k_1 \times k_2}$ as the $k_1 \times k_2$ matrix with all entries being zeros. We further define

$$A = \begin{pmatrix} — \nabla g_1(x^\star)^\top — \\ \vdots \\ — \nabla g_s^\top(x^\star) — \end{pmatrix} \in \mathbb{R}^{s \times n}$$

$$B = \begin{pmatrix} — \nabla h_1(x^\star)^\top — \\ \vdots \\ — \nabla h_r(x^\star)^\top — \end{pmatrix} \in \mathbb{R}^{r \times n}$$

Note that the KKT conditions can be equivalently written as

$$\begin{pmatrix} A^\top & B^\top \\ -A^\top & -B^\top \\ -I_s & 0_{s \times r} \end{pmatrix} \begin{pmatrix} \lambda^\star \\ \mu^\star \end{pmatrix} \le \begin{pmatrix} -\nabla f(x^\star) \\ \nabla f(x^\star) \\ 0 \end{pmatrix}$$

By Farkas Lemma, we know that one of the following options is true: the KKT conditions hold; or there exists a vector $y = (y_1, y_2, y_3) \geq 0$ such that

$$\begin{pmatrix} A & -A & -I_s \\ B & -B & 0_{s\times r} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$-\nabla f(x^\star)^\top y_1 + \nabla f(x^\star)^\top y_2 < 0$$

Note that $y_2 - y_1$ can be any vector in $\mathbb{R}^n$, and any vector in $\mathbb{R}^n$ can be written as the difference between two positive vectors. We set $z = y_2 - y_1$ and reformulate the previous condition as: there exists $z \in \mathbb{R}^n$ such that

$$Az = -y_3 \leq 0, \qquad Bz = 0, \qquad \nabla f(x^*)^\top z < 0$$

By the definitions of matrices A and B, this is clearly equivalent to

$$\begin{aligned} z \in \{d \in \mathbb{R}^n \mid \langle \nabla f(x^*), d\rangle < 0\} \\ \cap \{d \in \mathbb{R}^n \mid \langle \nabla g_i(x^*), d\rangle \leq 0, i \in I(x^*)\} \\ \cap \left\{d \in \mathbb{R}^n \mid \left\langle \nabla h_j(x^*), d\right\rangle = 0, j = 1, \cdots, r\right\} \end{aligned}$$

Therefore, we obtain that the KKT conditions hold if and only if

$$\begin{aligned} \{d \in \mathbb{R}^n \mid \langle \nabla f(x^*), d\rangle < 0\} \\ \cap \{d \in \mathbb{R}^n \mid \langle \nabla g_i(x^*), d\rangle \leq 0, i \in I(x^*)\} \\ \cap \left\{d \in \mathbb{R}^n \mid \left\langle \nabla h_j(x^*), d\right\rangle = 0, j = 1, \cdots, r\right\} = \emptyset \end{aligned}$$

◀

(1.b) [5pts] Suppose all the $g_i(x)$, $i \in \{1, ..., s\}$, are *concave* functions, and all $h_j(x)$, $j \in \{1, ..., r\}$, are *affine* functions. Let x^* be a minimizer of the problem. Use Problem (1.a) to prove that the KKT conditions hold at x^*.

Solution. For any vector d that satisfies the conditions

$$\begin{aligned} \langle \nabla g_i(x^*), d\rangle \leq 0, \qquad & i \in I(x^*) \\ \left\langle \nabla h_j(x^*), d\right\rangle = 0, \qquad & j = 1, \cdots, r. \end{aligned}$$

The point $x(t) = x^* + td$ is feasible for $t > 0$, since

$$g_i(x^* + td) \leq g_i(x^*) + t\langle \nabla g_i(x^*), d\rangle \leq 0,$$

because g_i is a concave function and

$$h_j(x^* + td) = h_j(x^*) + \left\langle \nabla h_j(x^*), d\right\rangle = 0$$

because h_j is an affine function. Since x^* is a minimizer of the problem, we have

$$\langle \nabla f(x^*), d \rangle = \lim_{t \to 0^+} \frac{f(x^* + td) - f(x^*)}{t} \geq 0.$$

Therefore,

$$\begin{aligned} &\{d \in \mathbb{R}^n \mid \langle \nabla f(x^*), d \rangle < 0\} \\ \cap &\{d \in \mathbb{R}^n \mid \langle \nabla g_i(x^*), d \rangle \leq 0, i \in I(x^*)\} \\ \cap &\left\{d \in \mathbb{R}^n \mid \left\langle \nabla h_j(x^*), d \right\rangle = 0, j = 1, \cdots, r\right\} = \emptyset \end{aligned}$$

and it follows from Problem (1.a) that the KKT conditions hold. ◂

2. SOCP formulation [20pts]

A set $S \subseteq \mathbb{R}^n$ is said to have a SOCP formulation if there exists matrices A_i and it can be written as

$$S = \left\{ x \in \mathbb{R}^n : \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \text{for } i = 1, 2, ..., m \right\}$$

where $A_i \in \mathbb{R}^{p_i \times n}$ are given matrices, $b_i \in \mathbb{R}^{p_i}$ and $c_i \in \mathbb{R}^n$ are given vectors, and $d_i \in \mathbb{R}$ are given scalars. A set $S \in \mathbb{R}^n$ is said to have a lifted SOCP formulation if there exists $m \geq 0$ and $S' \in \mathbb{R}^{n+m}$ with SOCP formulation such that

$$S = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^m \ \text{ s.t. } \ (x; y) \in S'\}$$

i.e., S is the projection of S' onto the first n dimensions.

(2.a) [5pts] Prove that for any $G \in \mathbb{R}^{m \times n}$ and $f \in \mathbb{R}^m$, the linear constraints

$$S = \{x \in \mathbb{R}^n : Gx \leq f\}$$

has a SOCP formulation.

Solution. Take $A_i = 0$ and $b_i = 0$, we obtain the linear constraints

$$0 \leq c_i^T x + d_i \quad \text{for} \ \ i = 1, 2, ..., m.$$

Hence, the set of linear constraints has a SOCP formulation. ◂

(2.b) [5pts] Show that the cone

$$S = \left\{ (x, y, z) \in \mathbb{R}^3_{\geq 0} : x^2 \leq yz \right\}$$

has an SOCP formulation.

Solution. Note that $yz = \frac{(y+z)^2 - (y-z)^2}{4}$. Thus,

$$\begin{aligned} S &= \left\{ (x, y, z) \in \mathbb{R}^3_+ | \ x^2 \leq yz \right\} \\ &= \left\{ (x, y, z) \in \mathbb{R}^3_+ \mid 4x^2 + (y-z)^2 \leq (y+z)^2 \right\} \\ &= \left\{ (x, y, z) \in \mathbb{R}^3_+ \mid \sqrt{4x^2 + (y-z)^2} \leq y + z \right\} \\ &= \left\{ (x, y, z) \in \mathbb{R}^3 \mid \sqrt{4x^2 + (y-z)^2} \leq y + z, \ 0 \leq x, \ 0 \leq y, \ 0 \leq z \right\} \end{aligned}$$

which gives the desired SOCP form as $A_i = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$ for $i = 1, 2, 3$, $A_4 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}$, $b_i = 0$ for $i = 1, 2, 3, 4$, $c_1^T = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, $c_2^T = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$, $c_3^T = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$, $c_4^T = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$ and $d_i = 0$ for $i = 1, 2, 3, 4$. ◂

(2.c) [5pts] If both $S_1, S_2 \subseteq \mathbb{R}^n$ have lifted SOCP formulations, show that $S_1 \cap S_2$ also has a lifted SOCP formulation

Solution. If S_1 and S_2 have lifted SOCP formulations, then there exists $m_1, m_2 \geq 0$ and $S_1' \in \mathbb{R}^{n+m_1}$ and $S_2' \in \mathbb{R}^{n+m_2}$ such that

$$S_1 = \{x \in \mathbb{R}^n | \ \exists y \in \mathbb{R}^{m_1} \ \text{ s.t. } \ (x; y) \in S_1'\}$$
$$S_2 = \{x \in \mathbb{R}^n | \ \exists z \in \mathbb{R}^{m_2} \ \text{ s.t. } \ (x; z) \in S_2'\}.$$

$$S_1 \cap S_2 = \{x \in \mathbb{R}^n | \ \exists y \in \mathbb{R}^{m_1}, z \in \mathbb{R}^{m_2} \text{ s.t. } \ (x; y) \in S_1', (x; z) \in S_2'\}$$

Define

$$S_1'' = \{(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} | \ (x; y) \in S_1'\}$$
$$S_2'' = \{(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} | \ (x; z) \in S_2'\}.$$

With this definition, the intersection of S_1 and S_2 can be rewritten as:

$$S_1 \cap S_2 = \{x \in \mathbb{R}^n | \ \exists (y, z) \in \mathbb{R}^{m_1+m_2} \ \text{ s.t. } \ (x, y, z) \in S_1'' \cap S_2''\}.$$

Hence $S_1 \cap S_2$ has SOCP formulation. ◂

(2.d) [5pts] If $k = 2^t$ with some $t \geq 1$, prove the set

$$S := \left\{(x, y_1, ..., y_k) \in \mathbb{R}^{k+1}_{\geq 0} : x^k \leq y_1 y_2 ... y_k\right\}$$

has a lifted SOCP formulation.

Solution. We show the result by strong induction. We showed the base of induction for $t = 1$ in (2.b). Given

$$S_l = \left\{(x, y_1, ..., y_k) \in \mathbb{R}^{k+1}_+ \mid x^k \leq y_1 ... y_k\right\},$$

and

$$S_r = \left\{(x, y_{k+1}, ..., y_{2k}) \in \mathbb{R}^{k+1}_+ \mid x^k \leq y_{k+1} ... y_{2k}\right\},$$

we define

$$\begin{aligned} S = \{(x, y_1, ..., y_{2k}, a, b) \in \mathbb{R}^{k+1}_+ \mid\ & (a, y_1, ..., y_k) \in S_l, \\ & (b, y_{k+1}, ..., y_{2k}) \in S_r, \\ & x^2 \leq ab\}. \end{aligned}$$

The set S has a lifted SOCP formulation and since

$$x^2 \leq ab,$$
$$a^k \leq y_1, ...y_k, \text{and}$$
$$b^k \leq y_{k+1}...y_{2k},$$

hold, it must be that $x^{2k} \leq y_1...y_{2k}$.

Now if we lift the last two elements, we end up with a lifted SOCP formulation for $t + 1$, concluding the proof. ◂

3. Dual cones [20pts]

Recall that for a closed convex cone $K \subseteq \mathbb{R}^n$, its dual cone is defined as

$$K^* = \{u \in \mathbb{R}^n : \langle u, x \rangle \geq 0, \quad \forall x \in K\}.$$

(3.a) [10pts] Let K be the Lorentz cone, *i.e.* $K = \{(x, z) \in \mathbb{R}^{n-1} \times \mathbb{R} : \|x\|_2 \leq z\}$. Prove that $K^* = K$.

Solution.

Direction 1: $K \subseteq K^*$ For any $(x, z), (y, w) \in K$, we can write

$$\langle (x, z), (y, w) \rangle = \langle x, y \rangle + zw \geq -\|x\|_2 \|y\|_2 + zw$$

from Cauchy-Schwarz inequality. Note that since $(x, z), (y, w) \in K$, $\|x\|_2 \leq z$ and $\|y\|_2 \leq w$. Hence,

$$\langle (x, z), (y, w) \rangle = \langle x, y \rangle + zw \geq -\|x\|_2 \|y\|_2 + zw \geq 0.$$

Thus, every $(y, w) \in K$ is also in K^*.

Direction 2: $K^* \subseteq K$

Let $(y, w) \in K^*$. Then from the definition of dual cones, $\forall (x, z) \in K$,

$$\langle (y, w), (x, z) \rangle = \langle y, x \rangle + zw \geq 0.$$

Observe that $(0, 1) \in K$ since it satisfies $\|0\|_2 = 0 \leq 1$. If $y = 0$, then $\langle y, x \rangle + zw = w \geq 0$ from the definition of the dual cone. Hence $w \geq \|y\|_2 = 0$. If $y \neq 0$, pick $(x, z) = \left(-\frac{y}{\|y\|_2}, 1\right)$. Since this choice of (x, z) satisfies the Lorentz cone definition, we must have the following for (y, w) to be in the dual cone of K.

$$\left\langle (y, w), \left(-\frac{y}{\|y\|_2}, 1\right) \right\rangle = -\|y\|_2 + w \geq 0.$$

And thus, $w \geq \|y\|_2$. Then every (y, w) in K^* is also in K. ◀

(3.b) [10pts] Let K be the semidefinite cone, *i.e.* $K = \{A \in \mathbb{S}^n : a^\top A a \geq 0 \quad \forall a \in \mathbb{R}^n\}$. Prove that $K^* = K$.

▸ *Hint:* The inner product on matrices is defined as the trace of the matrix product *i.e.*,

$$\langle A, B \rangle = \text{tr}(AB^\top).$$

Solution. Observe that

$$\langle A, B \rangle = \text{tr}(AB^\top) = \text{tr}(AB),$$

Since all matrices we deal with are symmetric.

Direction 1: $K \subseteq K^*$ Given any $X \in K$, for all $Y \in K$ we have that $X = U^\top U$ and $Y = V^\top V$ for some matrix U and V. In particular, U is the root of the eigenvalue matrix times the eigenvector matrix or the Cholesky decomposition. This implies that

$$\langle X, Y \rangle = \operatorname{tr}(U^\top U V^\top V) = \operatorname{tr}(U V^\top V U^\top) = \langle U V^\top, U V^\top \rangle \geq 0.$$

Thus $X \in K^*$. Note that we used the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Direction 2: $K^* \subseteq K$ Suppose $X \in K^*$, let $U^\top(\operatorname{Diag} \lambda X)U$ be the spectral decomposition of X. Since $Y_i = U^\top \operatorname{Diag} e_i U$ is PSD, it must be that $\langle X, Y_i \rangle = (\lambda X)_i$ is non negative which implies that $X \in K$ as all of the eigenvalues of X must be nonnegative.

◂

4. Equilibrium configuration of a mechanical system [20pts]

We consider a mechanical system that consists of N nodes at positions $x_1, ..., x_N \in \mathbb{R}^2$ with node i connected to node $i+1$ for $i = 1, 2, ..., N-1$, by a nonlinear spring. The nodes x_1 and x_N are fixed at given values $a \in \mathbb{R}^2$ and $b \in \mathbb{R}^2$ respectively. The tension T_i in spring i is given by

$$T_i = k\Big(\|x_i - x_{i+1}\|_2 - l_0\Big)_+$$

for some constants $k > 0$ and $l_0 > 0$ with $z_+ = \max\{z, 0\}$. Each node has a mass of w_i attached to it. The problem is to compute the equilibrium configuration of the system, *i.e.*, to solve the following optimization problem:

$$\begin{array}{ll} \displaystyle\min_{(x_1,...,x_N)\in\mathbb{R}^2} & \sum_{i=1}^N w_i e_2^\top x_i + \sum_{i=1}^{N-1} \frac{k}{2}\left(\Big(\|x_i - x_{i+1}\|_2 - l_0\Big)_+\right)^2 \\ \text{s.t.} & x_1 = a, \qquad x_N = b, \end{array}$$

where $e_2 := (0, 1)$.

(4.a) [10pts] Express the problem as a conic optimization problem.

▸ *Hint:* You need to write a minimization problem where the objective is linear and has
- linear *e.g.* $Ax \geq b$, or
- second order conic *e.g.*, $c^\top t + d \geq \|Ax + b\|_2$

constraints.

Solution. Observe that $t \geq (y)_+$ is equivalent to $t \geq y$ and $t \geq 0$. Furthermore, note that $s \geq t^2$ is equivalent to $(s+1)^2 \geq (s-1)^2 + (2t)^2$ and $s \geq 0$ or $s + 1 \geq \|(s-1, 2t)^\top\|_2$.

So we have

$$\begin{array}{lll} \displaystyle\min_{x,t,s} & \sum_{i=1}^N w_i e_2^\top x_i + \sum_{i=1}^{N-1} \frac{k}{2} s_i & \\ \text{s.t.} & t_i \geq \|x_i - x_{i-1}\|_2 - l_0 & \forall i \in [N-1] \\ & s + 1 \geq \|(s-1, 2t)^\top\|_2 & \forall i \in [N-1] \\ & x_i \in \mathbb{R}^2 & i \in [N] \\ & t \in \mathbb{R}^{N-1}_{\geq 0} & \\ & s \in \mathbb{R}^{N-1}_{\geq 0} & \end{array}$$

◂

(4.b) [10pts] Solve the problem numerically on an example with $N = 20$. You may generate the data w_i, a, b, l_0 and k by yourself. You may use any language or algorithm you are familiar with. We suggest to use `CVXPY` with Python. `CVXPY` is an open-source Python-embedded modeling language for convex optimization problems. You can find resources at https://www.cvxpy.org/. We also provide a template in Jupyter Notebook to help you test your solution. Please submit your code and output in a PDF file onto Gradescope. Since Gradescope only allows for one PDF submission, please combine multiple documents into one single PDF.

5. The Motzkin polynomial [20pts]

The Motzkin polynomial is the bivariate polynomial

$$M(x,y) := x^4y^2 + x^2y^4 - 3x^2y^2 + 1.$$

(5.a) [10pts] In Lecture 10, we mentioned that M is nonnegative on $\mathbb{R}^2$, and claimed that M cannot be expressed as a sum of squares. Prove the latter claim.

▸ *Hint:* If M was a sum of squares, then there would exist polynomials $p_1, p_2, \ldots$ of degree at most 3 such that

$$x^4y^2 + x^2y^4 - 3x^2y^2 + 1 = p_1^2(x,y) + p_2^2(x,y) + \ldots$$

What monomials can appear in the p_i and why? Furthermore, $M(x,y) = 0$ at the points $(-1,-1)$, $(-1,1)$, $(1,-1)$, and $(1,1)$. This further constraints what the coefficients of the p_i can be.

Solution.

Each $p_i^2(x,y)$ has the form

$$p_i^2(x,y) = \left(ax^2y + bxy^2 + cxy + dx + ey + f\right)^2$$

for appropriate constants $a, \ldots, f \in \mathbb{R}$. The p_i^2 need not have x^2 or y^2 terms as they would create x^4 and y^4 terms with nonnegative coefficients. Since we cannot have x^3 (as that would create x^6), we cannot have x^2. Since p_i is zero in the four points $(-1,-1), (-1,1), (1,-1), (1,1)$ we find that

$$\begin{aligned} -a - b + c - d - e + f &= 0 \qquad (i) \\ a - b - c - d + e + f &= 0 \qquad (ii) \\ -a + b - c + d - e + f &= 0 \qquad (iii) \\ a + b + c + d + e + f &= 0 \qquad (iv) \end{aligned}$$

Summing (i) and (iv) we find $c + f = 0$. Summing (ii) and (iii) we find $c - f = 0$. Hence, $c = f = 0$. But then, there is no way that the sum of polynomials p_i would be able to generate the constant term $+1$ in Motzkin's polynomial.◂

(5.b) [10pts] Write a semidefinite program to certify that the polynomial

$$p(x,y) := (x^2+1)M(x,y)$$

belongs to the cone of SOS polynomials, and solve it using a conic optimization solver such as `CVXPY`. Take the semidefinite matrix returned by the solver and explicitly construct the sum-of-squares decomposition of the polynomial p.

Please submit your code and output in a PDF file onto Gradescope. Since Gradescope only allows for one PDF submission, please combine multiple documents into one single PDF.

Solution. We find the factorization

$$(x^2+1)M(x,y) = (x^3y - xy)^2 + (x^2y^2 - 1)^2 + (xy^2 - x)^2$$

◂

MIT OpenCourseWare
https://ocw.mit.edu

6.7220 Nonlinear Optimization
Spring 2025

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms