DSA3102: Practice Exam Solutions

1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is quadratic and of the form

$$f(x) = \frac{1}{2}x^T Qx - b^T x$$

where $Q \in \mathbf{S}^n_{++}$.

(a) (5 points) Let $\lambda_m(Q)$ be the m-th *largest* eigenvalue of Q so

$$\lambda_1(Q) \geq \lambda_2(Q) \geq \ldots \geq \lambda_n(Q) > 0.$$

Show that the Lipschitz condition

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

is satisfied with $L = \lambda_m(Q)$. Find m, an integer in $\{1, 2, \ldots, n\}$.
Solution: The gradient of f is $\nabla f(x) = Qx$. Consequently we have the bound

$$\|\nabla f(x) - \nabla f(y)\|_2^2 = \|Q(x - y)\|_2^2 = (x - y)^T Q^T Q(x - y) \leq \|x - y\|_2^2 \lambda_{\max}(Q^T Q)$$

Furthermore,

$$\lambda_{\max}(Q^T Q) = \lambda_{\max}(Q^2) = \lambda_{\max}(Q)^2$$

So taking square roots,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq \lambda_1(Q)\|x - y\|_2$$

Hence, the Lipschitz constant is $\lambda_1(Q)$ and $m = 1$.

(b) (5 points) Consider the unconstrained minimization problem

$$\min \ f(x)$$

where the optimization variable is x. What is the optimal solution x^*?
Solution: $x^* = Q^{-1}b$.

(c) (15 points) Consider the gradient method

$$x^{(k+1)} = x^{(k)} - sD\nabla f(x^{(k)})$$

where $D \in \mathbf{S}^n_{++}$.
It is known that the method converges to x^* for every starting point $x^{(0)}$ if and only if the stepsize

$$0 < s < \frac{2}{M}.$$

Find the scalar M in terms of the matrices D and Q.

Hint: You might want to consider a change of variables $y^{(k)} = D^{1/2}x^{(k)}$. Note that DQ need not be symmetric.

Solution: Note that $\nabla f(x) = Qx - b$. The gradient method is equivalent to

$$x^{(k+1)} = x^{(k)} - sD(Qx^{(k)} - b)$$

By introducing $x^* = Q^{-1}b$, we obtain

$$x^{(k+1)} - x^* = (I - sDQ)(x^{(k)} - x^*)$$

Use a change of variables $y^{(k)} = D^{-1/2}x^{(k)}$ and $y^* = D^{-1/2}x^*$, we obtain

$$D^{1/2}(y^{(k+1)} - y^*) = (I - sDQ)D^{1/2}(y^{(k)} - y^*)$$

Left multiplying by $D^{-1/2}$ on both sides, we get

$$y^{(k+1)} - y^* = D^{-1/2}(I - sDQ)D^{1/2}(y^{(k)} - y^*)$$

which means that

$$y^{(k+1)} - y^* = A(y^{(k)} - y^*),$$

where

$$A = I - sD^{1/2}QD^{1/2}.$$

The system converges if and only if

$$|\lambda_{\max}(A)| < 1$$

Note that these eigenvalues are real because $D^{1/2}QD^{1/2}$ is symmetric. This is equivalent to having all the eigenvalues of $sD^{1/2}QD^{1/2}$ be in $(0, 2)$. Thus the condition is

$$0 < s\lambda_{\max}(D^{1/2}QD^{1/2}) = \lambda_{\max}(sD^{1/2}QD^{1/2}) < 2$$

which means we can take $M = \lambda_{\max}(D^{1/2}QD^{1/2})$.

2. (a) (15 points) Solve for the optimal $(x_0, x_1, \dots, x_n)$ using the KKT conditions

$$\begin{aligned}
\min_{(x_0,x_1,\dots,x_n)\in\mathbb{R}^{n+1}} \quad & -\log(1+x_0) - \sum_{i=1}^{n} \log(x_i) \\
\text{s.t.} \quad & x_0 + x_i \le 1, \quad \forall\, i = 1, \dots, n, \\
& x_0, x_1, \dots, x_n \ge 0.
\end{aligned}$$

Hint 1: Can any of the $x_i, i \ge 1$ be 0?
Hint 2: If λ_0 denotes the Lagrange multiplier associated to the constraint $x_0 \ge 0$, consider two cases $\lambda_0 = 0$ and $\lambda_0 > 0$ and show that the first case is infeasible.

Solution: First, we see that none of the x_i for $i \ge 1$ can be 0, since the objective function would be $-\infty$. Set $x = (x_0, x_1, \dots, x_n)$ and $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_n)$. The Lagrangian is

$$L(x, \lambda) = -\log(1+x_0) - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \lambda_i(x_0 + x_i - 1) + \lambda_0(-x_0)$$

The KKT conditions are

$$\begin{aligned}
-\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i - \lambda_0 &= 0 \\
-\frac{1}{x_i} + \lambda_i &= 0, \qquad \forall\, i \ge 1 \\
\lambda_i &\ge 0, \qquad \forall i \ge 0 \\
\lambda_i(x_0 + x_i - 1) &= 0, \qquad \forall\, i \ge 1 \\
\lambda_0 x_0 &= 0,
\end{aligned}$$

First note that $\lambda_i = 1/x_i$ and $0 < x_i \le 1$ for all $i > 0$. Thus, $\lambda_i > 0$ and by complementary slackness $x_i = 1 - x_0$ for all $i > 0$.

Case (A): $\lambda_0 = 0$. In this case, the KKT conditions simplify as follows:

$$\begin{aligned}
\frac{1}{1+x_0} &= \sum_{i=1}^{n} \lambda_i \\
\frac{1}{x_i} &= \lambda_i, \qquad \forall i \ge 1 \\
x_0 + x_i + 1 = 0&, \qquad \forall i \ge 1.
\end{aligned}$$

This implies that

$$-\frac{1}{1+x_0} + \frac{n}{1-x_0} = 0$$

Now, either (i) $x_0 \in (0, 1]$ or (ii) $x_0 = 0$. If (i), this is impossible since

$$\frac{n}{1-x_0} \ge n \ge 1 > \frac{1}{1+x_0}$$

If (ii), $\lambda_0 > 0$ and we go to Case (B).

Case (B): $\lambda_0 > 0$. In this case, the KKT conditions simplify as follows:

$$\begin{aligned}
-\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i - \lambda_0 &= 0 \\
\frac{1}{x_i} &= \lambda_i, \qquad \forall\, i \ge 1 \\
x_0 + x_i - 1 &= 0, \qquad \forall\, i \ge 1 \\
x_0 &= 0
\end{aligned}$$

Thus, we find that

$$\begin{aligned} x_i &= 1 \qquad \forall\, i \geq 1 \\ \lambda_i &= 1 \qquad \forall\, i \geq 1 \\ \lambda_0 &= n-1 \end{aligned}$$

Thus $x^* = (0, 1, 1, \ldots, 1)$ and $\lambda^* = (n-1, 1, 1, \ldots, 1)$ satisfy the first-order optimality conditions. Since the objective function is strictly concave with a convex constraint set, this solution must be the global maximum.

(b) (10 points) TRUE or FALSE: For any LP, the dual of the dual is the primal. Work with the LP

$$\min_{x \in \mathbb{R}^n} \; c^T x \quad \text{s.t.} \quad Ax = b, x \succeq 0.$$

You must substantiate your answer by deriving duals carefully. To make things simpler, eliminate the Lagrange multiplier associated to the constraint $\lambda \succeq 0$ in the first dual.

Solution: The Lagrangian of the LP is

$$L(x, \lambda, \nu) = c^T x + \lambda^T(-x) + \nu^T(Ax - b)$$

Thus the Lagrangian dual is

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & c^T - \lambda^T + \nu^T A = 0 \\ -\infty & \text{else} \end{cases}$$

By eliminating $\lambda \geq 0$, the Lagrangian dual is

$$\min_{\nu} \; b^T \nu \quad \text{s.t.} \quad A^T \nu + c \geq 0.$$

The Lagrangian is

$$\tilde{L}(\nu, x) = b^T \nu + x^T(-c - A^T \nu)$$

Thus the dual function is

$$\tilde{g}(x) = \begin{cases} -c^T x & b^T = A^T x^T \\ -\infty & \text{else} \end{cases}$$

Thus, the dual of the dual is

$$\min_{x} \; c^T x \quad \text{s.t.} \quad b = Ax, x \geq 0$$

which is the original LP.

3. For a convex (not necessarily differentiable) function $f : \mathbb{R}^n \to \mathbb{R}$, a *subgradient* of f at x is defined to be any vector $\delta_x(f) \in \mathbb{R}^n$ satisfying

$$f(x) + (y - x)^T \delta_x(f) \le f(y), \qquad \forall\, x, y \in \mathbf{dom}\, f$$

Note that the usual definition of gradient of f at x is recovered for differentiable convex f. In this case,

$$\nabla f(x) = \delta_x(f)$$

Similarly, for a concave (not necessarily differentiable) function $f : \mathbb{R}^n \to \mathbb{R}$, a *supergradient* of f at x is defined to be any vector $\bar{\delta}_x(f) \in \mathbb{R}^n$ satisfying

$$f(x) + (y - x)^T \bar{\delta}_x(f) \ge f(y), \qquad \forall\, x, y \in \mathbf{dom}\, f$$

(a) (10 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Let $\partial f(x)$ be the set of all the subgradients of f at x. This set is called the subdifferential of f at x. Is the set $\partial f(x)$ convex? Reason very carefully.

Solution: Yes.

$$\partial f(x) = \{v : f(x) + (y - x)^T v \le f(y), \quad \forall\, x, y \in \mathbf{dom}\, f\}$$

This can be written as

$$\partial f(x) = \bigcap_{x,y \in \mathbf{dom}\, f} S_{x,y}$$

where

$$S_{x,y} = \{v : f(x) + (y - x)^T v \le f(y)\}$$

Each of the sets $S_{x,y}$ is a halfspace, hence convex. So $\partial f(x)$ is a convex set since it's the intersection of many convex sets.

(b) (5 points) Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = |x|$$

What is the subdifferential of f at (i) $x < 0$, (ii) $x = 0$ and (iii) $x > 0$?

Hint: Plot $f(x)$ against $x \in \mathbb{R}$. Interpret the subgradient condition geometrically.

Solution: At $x < 0$, $\partial f(x) = \{-1\}$. At $x = 0$, $\partial f(x) = [-1, 1]$ and at $x > 0$, $\partial f(x) = \{1\}$.

(c) (10 points) Let $f_0, f_1 : \mathbb{R} \to \mathbb{R}$ be convex functions. Now consider the convex optimization problem

$$\min_{x \in \mathbb{R}} \; f_0(x) \quad \text{s.t.} \quad f_1(x) \le 0$$

Consider the Lagrange dual function $g(\lambda)$ of this convex optimization problem. Show that a *supergradient* of g at λ is given by $f_1(x_\lambda)$ where x_λ is a minimizer of $f_0(x) + \lambda f_1(x)$, i.e.,

$$\bar{\delta}_\lambda(g) = f_1(x_\lambda), \quad \text{where} \quad x_\lambda = \arg\min\{f_0(x) + \lambda f_1(x)\}.$$

Solution: The Lagrange dual function

$$g(\lambda) = \inf\{f_0(x) + \lambda f_1(x)\}$$

We need to show that

$$g(\lambda) + (\mu - \lambda) f_1(x_\lambda) \ge g(\mu).$$

This is by direct calculation. We have

$$\begin{aligned} g(\mu) &= \inf\{f_0(x) + \mu f_1(x)\} \\ &\le f_0(x_\lambda) + \mu f_1(x_\lambda) \\ &= f_0(x_\lambda) + (\mu - \lambda) f_1(x_\lambda) + \lambda f_1(x_\lambda) \\ &= g(\lambda) + (\mu - \lambda) f_1(x_\lambda). \end{aligned}$$

4. (a) (10 points) Apply Newton's method with a constant stepsize of 1 to the minimization of the function $f(x) = x^3$. In other words, express $x^{(k+1)}$ in terms of $x^{(k)}$. Show that the convergence is order one (linear), and explain why it is not order two (superlinear/quadratic) as was proved in class.

Solution: We can calculate

$$\nabla f(x) = 3x^2, \quad \nabla^2 f(x) = 6x$$

Newton's method with stepsize 1 yields

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - (\nabla^2 f(x^{(k)}))^{-1} \nabla f(x^{(k)}) \\ &= x^{(k)} - \frac{1}{2} x^{(k)} = \frac{1}{2} x^{(k)} \end{aligned}$$

Convergence is linear because the Hessian is not strictly PSD at the optimum. Here $x^* = 0$ but $\nabla^2 f(x^*) = 0$.

(b) In a beamforming design, the signal-to-interference plus noise ratio (SINR) for each user $i = 1, \ldots, n$ can be expressed as

$$\mathrm{SINR}_i(x_1, x_2, \ldots, x_n) = \frac{(h_i^T x_i)^2}{\sigma^2 + \sum_{j \neq i} (h_i^T x_j)^2}$$

where $h_i \in \mathbb{R}^k$ are given gain vectors, $\sigma^2 \in \mathbb{R}_+$ is a given noise variance and the $x_i \in \mathbb{R}^k$ are the beamforming vectors. An important class of optimal downlink beamforming problem involves finding a set of $x_i' s$ that minimizes the total transmit power, while satisfying a given set of SINR constraints

$$\text{(P)} \qquad \min_{x_1, \ldots, x_n} \sum_{i=1}^{n} \|x_i\|^2, \quad \text{s.t.} \quad \mathrm{SINR}_i(x_1, x_2, \ldots, x_n) \geq \gamma_i, \quad \forall\, i = 1, \ldots, n$$

(i) (7 points) Now, define the matrix $X_i = x_i x_i^T \in \mathbf{S}_+^k$ and the given matrices $H_i = h_i h_i^T \in \mathbf{S}_+^k$. Show that Problem (P) can be re-expressed as

$$\begin{aligned} \text{(P')} \qquad & \min_{X_1, \ldots, X_n} \sum_{i=1}^{n} \mathbf{tr}(X_i) \\ \text{s.t.} \quad & g(H_i, X_i) - \gamma_i \sum_{j \neq i} g(H_i, X_j) \geq c_i, \\ & \mathbf{rank}(X_i) = b \quad \forall\, i = 1, \ldots, n \\ & X_1, \ldots, X_n \in \mathbf{S}_+^k \end{aligned}$$

Identify the function $g(\cdot, \cdot)$ and the constants c_i and b. Now is Problem (P') convex?

Solution: The objective can be written as

$$\sum_{i=1}^{n} \|x_i\|^2 = \sum_{i=1}^{n} \mathbf{tr}(X_i)$$

because $\mathbf{tr}(X_i) = \mathbf{tr}(x_i x_i^T) = \mathbf{tr}(x_i^T x_i) = \|x_i\|^2$. Furthermore,

$$(h_i^T x_i)^2 = \mathbf{tr}((h_i^T x_i)^2) = \mathbf{tr}(x_i^T h_i h_i^T x_i) = \mathbf{tr}(x_i^T H_i x_i) = \mathbf{tr}(H_i x_i x_i^T) = \mathbf{tr}(H_i X_i)$$

Thus, the SINR constraints can be written as

$$\mathbf{tr}(H_i X_i) \geq \gamma_i (\sigma^2 + \sum_{j \neq i} \mathbf{tr}(H_i X_j))$$

or

$$g(H_i, X_i) - \gamma_i \sum_{j \neq i} g(H_i, X_j) \geq c_i$$

where

$$g(H, X) = \mathbf{tr}(HX), \quad c_i = \gamma_i \sigma^2.$$

Because $X_i = x_i x_i^T$ has rank 1 for each i,

$$b = 1.$$

Problem (P') is not convex due the the rank constraints on the X_i's.

(ii) (4 points) Now define $X = [x_1, \ldots, x_n] \in \mathbb{R}^{k \times n}$. Show that the SINR constraints in (P) can be written as

$$\beta_i (h_i^T x_i)^2 \geq \left\| \begin{bmatrix} h_i^T X \\ \sigma \end{bmatrix} \right\|^2, \qquad \forall\, i = 1, \ldots, n.$$

for some $\beta_i = f(\gamma_i)$. Identify the function $f(\gamma_i)$.

Solution: The SINR constraints for user i can be written as

$$(h_i^T x_i)^2 \geq \gamma_i \sum_{j \neq i} (h_i^T x_j)^2 + \gamma_i \sigma^2$$

or

$$\frac{1}{\gamma_i} (h_i^T x_i)^2 \geq \sum_{j \neq i} (h_i^T x_j)^2 + \sigma^2$$

or

$$\left(1 + \frac{1}{\gamma_i}\right) (h_i^T x_i)^2 \geq \sum_{j=1}^{n} (h_i^T x_j)^2 + \sigma^2$$

The RHS is exactly

$$\sum_{j=1}^{n} (h_i^T x_j)^2 + \sigma^2 = \left\| \begin{bmatrix} h_i^T X \\ \sigma \end{bmatrix} \right\|^2$$

because

$$h_i^T X = [h_i^T x_1, \ldots, h_i^T x_n]$$

from the definition of X. So $f(\gamma_i) = 1 + 1/\gamma_i$.

(iii) (3 points) Based on formulation in (iii), write down a second-order cone program (SOCP) that is equivalent to (P) by introducing a new variable t which serves as an upper bound for $\sum_{i=1}^n \|x_i\|^2$, i.e., by expressing the problem in epigraphic form. Explain briefly why your optimization problem is an SOCP.

Solution: First introduce another variable $t \in \mathbb{R}_+$. Then the objective $\sum_{i=1}^n \|x_i\|^2$ is exactly

$$\left\| \mathbf{diag}(0, I, \ldots, I) \tilde{x} \right\|^2. \tag{1}$$

Minimizing this is the same as minimizing

$$\left\| \mathbf{diag}(0, I, \ldots, I) \tilde{x} \right\|. \tag{2}$$

As such we have

$$\min_{t \in \mathbb{R}_+, x_1, \ldots, x_n \in \mathbb{R}^k} \quad t$$

subject to

$$\begin{aligned} \sqrt{1+\frac{1}{\gamma_i}}h_i^T x_i \geq \left\| \begin{bmatrix} h_i^T X \\ \sigma \end{bmatrix} \right\|, \quad \forall\, i \\ \|\mathbf{diag}(0, I, \ldots, I)\tilde{x}\| \leq t \end{aligned}$$

The objective is linear in $\tilde{x} := (t, x_1, \ldots, x_n) \in \mathbb{R}^{1+nk}$. Both constraint are of the form

$$\|A_i \tilde{x} + b_i\|_2 \leq c_i^T \tilde{x} + d_i, \quad i = 1, \ldots, m.$$

for some appropriate A_i, b_i, c_i, d_i so we have reformulated the problem as an SOCP.