Practice Exam: Continuous Optimization

1. Consider the problem $\min_x\{x^2 : x \geq 1\}$. For a parameter $\rho > 0$, this problem can be approximated by the unconstrained optimization problem [3 points]

$$\begin{aligned} \min_x \quad & x^2 - \rho \ln(x-1) \\ \text{s. t.} \quad & x > 1. \end{aligned} \tag{A}$$

Find the optimal solutions to (A) as a function of $\rho > 0$, and find the limit of these optimal solutions as $\rho \to 0^+$.

> **Solution:** Let $h_\rho : \mathbb{R}_{++} \to \mathbb{R}$ be given by $h_\rho(x) = x^2 - \rho \ln(x-1)$. We then have
>
> $$\begin{aligned} h'_\rho(x) &= 2x - \rho(x-1)^{-1} = (x-1)^{-1}\left(2x^2 - 2x - \rho\right), \\ h''_\rho(x) &= 2 + \rho(x-1)^{-2} > 0. \end{aligned}$$
>
> Therefore h_ρ is a convex function, and thus is minimized at $x_\rho \in \mathbb{R}_{++}$ when $h'(x_\rho) = 0$, or equivalently $0 = 2x_\rho^2 - 2x_\rho - \rho$.
>
> Therefore $x_\rho = \frac{1}{2} \pm \frac{1}{2}\sqrt{1+2\rho}$. As $x_\rho \in \mathbb{R}_{++}$, we have that the optimal solution to (A) is given by $x_\rho = \frac{1}{2} + \frac{1}{2}\sqrt{1+2\rho}$.
>
> We then have $\lim_{\rho \to 0^+} x_\rho = 1$.

2. Consider a closed nonempty set $\mathcal{C} \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$ defined to be the distance to the set for some given norm, i.e. [3 points]

$$f(\mathbf{x}) = \min_{\mathbf{y}}\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \mathcal{C}\}.$$

Prove that if $\mathcal{C}$ is a convex set then f is a convex function.
[You may assume that the minimum defining f is attained.]

> **Solution:** Consider arbitrary $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and arbitrary $\theta \in [0,1]$. There exists $\mathbf{y}, \mathbf{z} \in \mathcal{C}$ such that $f(\mathbf{u}) = \|\mathbf{u} - \mathbf{y}\|$ and $f(\mathbf{v}) = \|\mathbf{v} - \mathbf{z}\|$. We then have
>
> $$\begin{aligned} \theta f(\mathbf{u}) + (1-\theta) f(\mathbf{v}) &= \theta\|\mathbf{u} - \mathbf{y}\| + (1-\theta)\|\mathbf{v} - \mathbf{z}\| \\ &\geq \|\theta(\mathbf{u} - \mathbf{y}) + (1-\theta)(\mathbf{v} - \mathbf{z})\| \\ &= \|(\theta\mathbf{u} + (1-\theta)\mathbf{v}) - (\theta\mathbf{y} + (1-\theta)\mathbf{z})\| \\ &\geq f(\theta\mathbf{u} + (1-\theta)\mathbf{v}) \end{aligned}$$
>
> The last inequality follows from the definition of f and the fact that if $\mathcal{C}$ is a convex set then $\theta\mathbf{y} + (1-\theta)\mathbf{z} \in \mathcal{C}$.
>
> NB A common error made was to say that for two functions $g, h : \mathbb{R}^n \to \mathbb{R}$ we have $\min_{\mathbf{x}}\{g(\mathbf{x}) + h(\mathbf{x})\} \leq \min_{\mathbf{x}}\{g(\mathbf{x})\} + \min_{\mathbf{x}}\{h(\mathbf{x})\}$. This is incorrect, and a simple counter example to this is given by $g, h : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$,

$h(x) = (x-2)^2$. We then have $\min\{g(x) + h(\mathbf{x})\} = \min\{2(x-1)^2 + 2\} = 2$, $\min_{\mathbf{x}}\{g(\mathbf{x})\} = 0 = \min_{\mathbf{x}}\{h(\mathbf{x})\}$.

We do always have the inequality $\min_{\mathbf{x}}\{g(\mathbf{x})+h(\mathbf{x})\} \geq \min_{\mathbf{x}}\{g(\mathbf{x})\}+\min_{\mathbf{x}}\{h(\mathbf{x})\}$, but this does not help in this problem.

3. For a fixed parameter $\alpha \in \mathbb{R}$, consider the function $f_\alpha(\mathbf{x}) = \exp(x_1+x_2)+\alpha x_1^2+x_2^4$.
   (a) For what values of the parameter $\alpha \in \mathbb{R}$ is f_α a convex function? [3 points]

From now on consider having $\alpha = 1$ (for which we have that f_α is a convex function).

   (b) By considering the function at $\mathbf{x} = \mathbf{0} = \begin{pmatrix}0\\0\end{pmatrix}$, show that $f_1(\mathbf{y}) \geq 1 + y_1 + y_2$ for all $\mathbf{y} \in \mathbb{R}^2$. [2 points]
   (c) Give the direction of steepest descent of f_1 at $\mathbf{x} = \mathbf{0}$. [1 point]
   (d) Give the Newton direction of f_1 at $\mathbf{x} = \mathbf{0}$. [2 points]

[These directions do not need to be normalised.]

Solution:

(a)

$$\nabla f(\mathbf{x}) = \begin{pmatrix}\exp(x_1+x_2)+2\alpha x_1\\ \exp(x_1+x_2)+4x_2^3\end{pmatrix}$$
$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix}\exp(x_1+x_2)+2\alpha & \exp(x_1+x_2)\\ \exp(x_1+x_2) & \exp(x_1+x_2)+12x_2^2\end{pmatrix}$$

If $\alpha \geq 0$ then $\nabla^2 f(\mathbf{x})$ is a diagonally dominant matrix, and thus it is also positive semidefinite, and f is a convex function.

If $\alpha < 0$ then for $x_1 + x_2 < \ln(-2\alpha)$ we have $(\nabla^2 f(\mathbf{x}))_{11} < 0$, and thus $\nabla^2 f(\mathbf{x})$ is not positive semidefinite when $x_1 + x_2 < \ln(-2\alpha)$, and f is not a convex function over $\mathbb{R}^2$.

(b) We have $f(\mathbf{0}) = 1$ and $\nabla f(\mathbf{0}) = \begin{pmatrix}1\\1\end{pmatrix}$, and by Theorem 1.27 we get

$$f(\mathbf{y}) \geq f(\mathbf{0}) + \nabla f(\mathbf{0})^\mathsf{T}(\mathbf{y}-\mathbf{0}) = 1 + y_1 + y_2.$$

(c) The direction of steepest descent is $\mathbf{d}_s = -\nabla f(\mathbf{0}) = \begin{pmatrix}-1\\-1\end{pmatrix}$.

(d) Letting $\mathbf{d}_n$ be the Newton direction, we have

$$\nabla^2 f(\mathbf{0}) = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}, \qquad [\nabla^2 f(\mathbf{0})]^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix},$$

$$\mathbf{d}_n = -[\nabla^2 f(\mathbf{0})]^{-1} \nabla f(\mathbf{0}) = -\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

4. Consider the problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & 4x_1 + x_2^2 \\ \text{s.t.} \quad & x_2 \geq x_1^2 \\ & \mathbf{x} \in \mathbb{R}^2. \end{aligned} \tag{B}$$

(a) Show that problem (B) is a convex problem. [2 points]

(b) Does Slater's condition hold for problem (B)? (You must justify your answer.) [1 point]

(c) Find the KKT point(s) for problem (B). [3 points]

(d) What is the global minimizer for problem (B), and prove that this minimizer is a local minimizer of order 2. [3 points]

(e) Formulate and solve the Langrangian Dual problem to problem (B). [4 points]

Solution:

(a) We have

$$f(\mathbf{x}) = 4x_1 + x_2^2, \quad \nabla f(\mathbf{x}) = \begin{pmatrix} 4 \\ 2x_2 \end{pmatrix}, \quad \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \succeq O,$$

$$g(\mathbf{x}) = x_1^2 - x_2, \quad \nabla g(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}, \quad \nabla^2 g(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \succeq O.$$

(b) Yes, for example at the point $\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ we have $g(\mathbf{x}) = -1 < 0$.

(c) We require $\mathbf{x} \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$ such that

$$\lambda \geq 0, \tag{1}$$
$$g(\mathbf{x}) \leq 0, \tag{2}$$
$$\lambda\, g(\mathbf{x}) = 0, \tag{3}$$
$$\nabla f(\mathbf{x}) = -\lambda \nabla g(\mathbf{x}) \tag{4}$$

Equation (4) is equivalent to $\begin{pmatrix} 4 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} -2x_1\lambda \\ \lambda \end{pmatrix}$, or $4 = -2x_1\lambda$ and $2x_2 = \lambda$. From (1), this implies that $\lambda > 0$ (and $x_1 < 0$). From (3) this implies that $g(\mathbf{x}) = 0$, or equivalently $x_2 = x_1^2$. Therefore $\lambda = 2x_2 = 2x_1^2$ and $4 = -2x_1\lambda = -4x_1^3$. Thus $x_1^3 = -1$, $x_1 = -1$, $x_2 = x_1^2 = 1$, $\lambda = 2x_2 = 2$.

Thus the only KKT point is $\mathbf{x}^* = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, with dual multiplier $\lambda = 2$.

(d) As we have a convex problem, any KKT point is a global minimizer, thus $\mathbf{x}^* = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is a global minimizer.

Considering the multiplier $\lambda = 2$ at $\mathbf{x}^*$, we have

$$\nabla^2 f(\mathbf{x}) + \lambda \nabla^2 g(\mathbf{x}) = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

This is a positive definite matrix, and thus

$$\mathbf{d}^\mathsf{T} \left(\nabla^2 f(\mathbf{x}) + \lambda \nabla^2 g(\mathbf{x}) \right) \mathbf{d} > 0$$

for all $\mathbf{d} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$, and by Theorem 5.14 we have that $\mathbf{x}^*$ is a strict local minimizer of order 2.

(e) We have

$$\begin{aligned} L(\mathbf{x}; y) &= f(\mathbf{x}) + y\, g(\mathbf{x}) \\ &= 4x_1 + x_2^2 + y(x_1^2 - x_2) \\ &= yx_1^2 + 4x_1 + x_2^2 - yx_2, \\ \psi(y) &= \inf_{\mathbf{x}} L(\mathbf{x}; y). \end{aligned}$$

If $y = 0$ then $L(\mathbf{x}; 0) = 4x_1 + x_2^2$, and considering $x_1 \to -\infty$ we get $\psi(0) = -\infty$.

If $y > 0$ then

$$\begin{aligned} L(\mathbf{x}; y) &= y \left(x_1 + \frac{2}{y} \right)^2 + \left(x_2 - \frac{y}{2} \right)^2 - \frac{4}{y} - \frac{y^2}{4}, \\ \psi(y) &= -\frac{4}{y} - \frac{y^2}{4}. \end{aligned}$$

The dual problem is thus

$$\max_y \quad -\frac{4}{y} - \frac{y^2}{4} \qquad \text{s.t.} \quad y > 0.$$

The Lagrangian dual is always a convex optimisation problem.

We have $\psi'(y) = 4y^{-2} - \frac{1}{2}y = \frac{1}{2}y^{-2}(8 - y^3)$, and the problem is maximised when $\psi'(y) = 0$.

Therefore the solution to the dual problem is $y = 2$, giving an optimal value of -3.

5. Let $M(x,y) := \begin{pmatrix} x & 1 \\ 1 & y \end{pmatrix}$ for $x, y \in \mathbb{R}$ and let $\lambda(M(x,y)) := \max\{|\lambda_1(M(x,y))|, |\lambda_2(M(x,y)|\}$ be the absolute value of the eigen value of $M(x,y)$ of largest absolute value.

   (a) Formulate a semidefinite program that solves the problem of finding $x, y \in \mathbb{R}$ minimizing $\lambda(M(x,y))$. [1 point]

   (b) Formulate the corresponding dual semidefinite program. [3 points]

   (c) Show that $x = y = 0$ is the optimal solution by exhibiting a dual solution whose value is equal to $\lambda(M(0,0))$. [2 points]

Solution:

(a) We recall that $\lambda(M(x,y)) \leq t$ iff the eigen values of $M(x,y)$ are in the range $[-t,t]$ iff $-tI_2 \preceq M(x,y) \preceq tI_2$, where I_2 is the 2×2 identity. Therefore, the semidefinite program minimizing $\lambda(M(x,y))$ is

$$\begin{aligned} &\min_{x,y,t} \; t \\ &\begin{pmatrix} t-x & 0 \\ 0 & t-y \end{pmatrix} \succeq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &\begin{pmatrix} t+x & 0 \\ 0 & t+y \end{pmatrix} \succeq \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \end{aligned}$$

For the purpose of taking duals, the above program can be equivalently written as

$$\begin{aligned} &\min_{x,y,t} \; 0 \cdot x + 0 \cdot y + 1 \cdot t \\ &x \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} + y \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} + t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \succeq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &x \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + y \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \succeq \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \end{aligned}$$

(b) Letting $X^1, X^2 \succeq 0$ denote the "multipliers" for the first and second constraint respectively, the corresponding dual semidefinite program can be expressed as

$$\begin{aligned} &\max \langle X^1, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \rangle + \langle X^2, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \rangle \\ &\langle X^1, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \rangle + \langle X^2, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \rangle = 0 \quad \text{(coefficient of } x) \\ &\langle X^1, \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \rangle + \langle X^2, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle = 0 \quad \text{(coefficient of } y) \\ &\langle X^1, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rangle + \langle X^2, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rangle = 1 \quad \text{(coefficient of } t) \\ &X^1 \succeq 0, X^2 \succeq 0. \end{aligned}$$

Let $X^1 = \begin{pmatrix} a_1 & b_1 \\ b_1 & c_1 \end{pmatrix}$ and $X^2 = \begin{pmatrix} a_2 & b_2 \\ b_2 & c_2 \end{pmatrix}$, note that the first equality constraint enforces $a_1 = a_2$, the second enforces $c_1 = c_2$, and the last enforces $a_1 + c_1 + a_2 + c_2 = 2(a_1 + c_1) = 1$. The dual can therefore be simplified to

$$\begin{aligned} \max_{a,c,b_1,b_2} \; & 2(b_1 - b_2) \\ & a + c = 1/2 \\ & \begin{pmatrix} a & b_1 \\ b_1 & c \end{pmatrix} \succeq 0, \begin{pmatrix} a & b_2 \\ b_2 & c \end{pmatrix} \succeq 0 \end{aligned}$$

(c) For $x = y = 0$, we see that the spectral decomposition is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} + -1 \cdot \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Therefore, the eigen values of $M(0,0)$ are -1 and 1. In particular, $\lambda(M(0,0)) = 1$. The dual solution of same value is $a = c = 1/4$, $b_1 = 1/4$, $b_2 = -1/4$. The value of this solution is $2(b_1 - b_2) = 1$. Furthermore it is feasible, since $\begin{pmatrix} 1/4 & \pm 1/4 \\ \pm 1/4 & 1/4 \end{pmatrix} \succeq 0$, since the diagonal is non-negative and $(1/4)^2 \succeq (\pm 1/4)^2$.

6. Let $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^n$. Examine the optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^{N} \|\mathbf{x} - \mathbf{y}_i\|_2^2$$

(a) Prove that $\mathbf{x}^* = \sum_{i=1}^N \mathbf{y}_i / N$ is the optimal solution. [2 points]

(b) Show that $\mathbf{x}^*$ is a local minimum of order 2. [1 point]

Solution: Letting $f(\mathbf{x}) = \sum_{i=1}^N \|\mathbf{x} - \mathbf{y}_i\|_2^2$, and $\mathbf{x}^* = \sum_{i=1}^N \mathbf{y}_i/N$, we see that

$$\begin{aligned}
f(\mathbf{x}) &= \sum_{i=1}^N \|(\mathbf{x} - \mathbf{x}^*) - (\mathbf{x}^* - \mathbf{y}_i)\|_2^2 \\
&= \sum_{i=1}^N (\|\mathbf{x} - \mathbf{x}^*\|_2^2 - 2(\mathbf{x}^* - \mathbf{y}_i)^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) + \|\mathbf{x}^* - \mathbf{y}_i\|_2^2) \\
&= N\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \sum_{i=1}^N \|\mathbf{x}^* - \mathbf{y}_i\|_2^2 - 2(\sum_{i=1}^N (\mathbf{x}^* - \mathbf{y}_i))^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) \\
&= N\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \sum_{i=1}^N \|\mathbf{x}^* - \mathbf{y}_i\|_2^2 - 2N(\mathbf{x}^* - \mathbf{x}^*)^\mathsf{T}(\mathbf{x} - \mathbf{x}^*) \\
&= N\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \sum_{i=1}^N \|\mathbf{x}^* - \mathbf{y}_i\|_2^2.
\end{aligned}$$

(a) By the representation above, we see that $f(\mathbf{x}) - f(\mathbf{x}^*) = N\|\mathbf{x} - \mathbf{x}^*\|_2^2$, and hence $\mathbf{x}^*$ is clear the unique global minimum.

(b) We clearly also have $f(\mathbf{x}) - f(\mathbf{x}^*) \geq \alpha\|\mathbf{x} - \mathbf{x}^*\|_2^2$ for $\alpha = N$, and hence $\mathbf{x}^*$ is a local minimum of order 2 as well.

7. (Automatic additional points) [4 points]

Question:	1	2	3	4	5	6	7	Total
Points:	3	3	8	13	6	3	4	40

A copy of the lecture-sheets may be used during the examination. You may use any results from the lecture slides in your answers (Lemmas, Theorems, Corollaries, Exercises, etc.), however you should reference the result.

Hints:

1. $\begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} = \dfrac{1}{ac - b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$
2. $\begin{pmatrix} a & b \\ b & c \end{pmatrix} \succeq 0$ *iff* $a, c \geq 0$ *and* $ac \geq b^2$.
3. *A norm* $\| \bullet \|$ *on* $\mathbb{R}^n$ *has the following properties:*

(a) $\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$ *for all* $\lambda \in \mathbb{R}$*,* $\mathbf{x} \in \mathbb{R}^n$*;*

(b) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*;*

(c) $\|\mathbf{x}\| > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.