# ECE 602 – Introduction to Optimization

## Solutions to Home Assignment 1

**Exercise 1**

First, we rewrite the cost function as

$$f(x) = \left\langle \mathbf{1}, \sqrt{(Ax)^2 + \epsilon} \right\rangle,$$

with $\mathbf{1} = [1, 1, \ldots, 1]^T \in \mathbf{R}^n$, where the functions $(\cdot)^2$ and $\sqrt{\cdot}$ are applied element-wise. In this case,

$$df(x) = f(x + dx) - f(x) = \left\langle \mathbf{1}, \sqrt{(A(x + dx))^2 + \epsilon} \right\rangle - f(x) =$$

$$= \left\langle \mathbf{1}, \sqrt{(Ax + Adx)^2 + \epsilon} \right\rangle - f(x) = \left\langle \mathbf{1}, \sqrt{(Ax)^2 + 2(Ax) \cdot (Adx) + (Adx)^2 + \epsilon} \right\rangle - f(x) \approx$$

$$\approx \left\langle \mathbf{1}, \sqrt{\left((Ax)^2 + \epsilon\right) + 2(Ax) \cdot (Adx)} \right\rangle - f(x),$$

where we discard the high-order term $(Ax)^2$ due to its relative smallness. Note that, in the expressions above, the dot $\cdot$ denotes an element-wise product of two vectors.

Next, we use the following (1st-order) Taylor approximation around $\left((Ax)^2 + \epsilon\right)$

$$\sqrt{\left((Ax)^2 + \epsilon\right) + 2(Ax) \cdot (Adx)} \approx \sqrt{(Ax)^2 + \epsilon} + \frac{1}{\sqrt{(Ax)^2 + \epsilon}} \cdot (Ax) \cdot (Adx) =$$

$$= \sqrt{(Ax)^2 + \epsilon} + \frac{Ax}{\sqrt{(Ax)^2 + \epsilon}} \cdot (Adx),$$

where the vector division is assumed to be element-wise. Thus, we have

$$df(x) = \left\langle \mathbf{1}, \sqrt{(Ax)^2 + \epsilon} + \frac{Ax}{\sqrt{(Ax)^2 + \epsilon}} \cdot (Adx) \right\rangle - f(x) =$$

$$= \left\langle \mathbf{1}, \sqrt{(Ax)^2 + \epsilon} \right\rangle + \left\langle \mathbf{1}, \frac{Ax}{\sqrt{(Ax)^2 + \epsilon}} \cdot (Adx) \right\rangle - f(x) = \left\langle \mathbf{1}, \frac{Ax}{\sqrt{(Ax)^2 + \epsilon}} \cdot (Adx) \right\rangle =$$

$$= \langle \frac{Ax}{\sqrt{(Ax)^2+\epsilon}}, Adx \rangle = \langle A^T \Big( \frac{Ax}{\sqrt{(Ax)^2+\epsilon}} \Big), dx \rangle.$$

Therefore,

$$\nabla f(x) = A^T \Big( \frac{Ax}{\sqrt{(Ax)^2+\epsilon}} \Big).$$

**Exercise 2**

PART A

a) Let $x_1, x_2 \in C_\alpha$. Let's show that, for any $0 \le \theta \le 1$, $\theta x_1 + (1-\theta)x_2 \in C_\alpha$, meaning $f(\theta x_1 + (1-\theta)x_2) \le \alpha$. Indeed, due to the convexity of $f$, we have

$$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2) \le \theta\alpha + (1-\theta)\alpha = \alpha.$$

b) Let $Q_1, Q_2 \in \mathbf{S}^n_+$. Let's show that, for any $0 \le \theta \le 1$, $Q = \theta Q_1 + (1-\theta) Q_2 \in \mathbf{S}^n_+$, meaning $x^T Q x \ge 0$, for any $x \in \mathbf{R}^n$. Indeed, we have

$$x^T Q x = x^T \Big( \theta Q_1 + (1-\theta) Q_2 \Big) x = \theta\, x^T Q_1 x + (1-\theta)\, x^T Q_2 x \ge 0,$$

since both $x^T Q_1 x \ge 0$ and $x^T Q_2 x \ge 0$.

PART B

a) The Hessian of $f(x) = 1/2 x^T Q x + c^T x$ is equal to $Q$, which is a positive-definite matrix. Therefore, the function is convex (albeit, not strictly).

b) Let $x_1, x_2 \in \mathbf{R}$. To prove the convexity of $f(x) = g(h(x))$, we need to show that, for any $\theta \in [0,1]$, we have

$$f(\theta\, x_1 + (1-\theta)\, x_2) \le \theta\, f(x_1) + (1-\theta) f(x_2).$$

First, we observe that, by the convexity of $h$, we have

$$h(\theta\, x_1 + (1-\theta)\, x_2) \le \theta h(x_1) + (1-\theta) h(x_2).$$

Consequently, due to the monotonicity of $g$, we have

$$g(h(\theta\, x_1 + (1-\theta)\, x_2)) \le g(\theta h(x_1) + (1-\theta) h(x_2)).$$

At the same time, due to the convexity of $g$, we can write

$$g(\theta h(x_1) + (1-\theta)h(x_2)) \leq \theta\, g(h(x_1)) + (1-\theta)g(h(x_2)).$$

In summary,

$$f(x) = g(\theta\, h(x_1)+(1-\theta)\, h(x_2)) \leq \theta\, g(h(x_1))+(1-\theta)\, g(h(x_2)) = \theta\, f(x_1)+(1-\theta)\, f(x_2).$$

**Exercise 3**

Let $p^\star$ be the global minimum of $f$. Then the set of all global minimizer of $f$ can be defined as

$$C_{p^\star} = \{x \in \mathbf{R}^n \mid f(x) \leq p^\star\}.$$

This is a sublevel set of a convex function, which makes it convex.

**Exercise 4**

a) To show that $\|z\|_\infty = \sup_{\|x\|_1 \leq 1} x^T z$, we first notice that

$$x^T z \leq |x^T z| \leq \sum_{i=1}^{n} |x_i||z_i| \leq \max_i |z_i| \sum_{i=1}^{n} |x_i| = \|z\|_\infty \|x\|_1 \Big|_{\|x\|_1 = 1} = \|z\|_\infty.$$

However, it is not clear yet if the bound is tight. To show that, we need to find a vector $x$ (with $\|x\|_1 = 1$) for which $x^T z = \|z\|_\infty$. Such $x$ can be defined as

$$x_i = \begin{cases} \text{sign}(z_{i^\star}), & \text{if } i = i^\star \\ 0, & \text{otherwise} \end{cases},$$

with $i^\star = \arg\max_{1 \leq i \leq n} |z_i|$.

b) To show that $\|z\|_2 = \sup_{\|x\|_2 \leq 1} x^T z$, we simply use the Cauchy–Schwarz inequality

$$x^T z \leq |x^T z| \leq \|z\|_2 \|x\|_2 \Big|_{\|x\|_2 = 1} = \|z\|_2,$$

which is known to be tight for $x = z/\|z\|_2$.