Solution to Math4230 Tutorial 2

1. (a) Let C be a nonempty subset of $\mathbf{R}^n$, and let λ_1 and λ_2 be positive scalars. Show that if C is convex, then $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$. Show by example that this need not be true when C is not convex.
(b) Show that the intersection $\cap_{i\in I} C_i$ of a collection $\{C_i \mid i \in I\}$ of cones is a cone.
(c) Show that the image and the inverse image of a cone under a linear transformation is a cone.
(d) Show that the vector sum $C_1 + C_2$ of two cones C_1 and C_2 is a cone.
(e) Show that a subset C is a convex cone if and only if it is closed under addition and positive scalar multiplication, i.e., $C + C \subset C$, and $\gamma C \subset C$ for all $\gamma > 0$.

Solution.
(a) We always have $(\lambda_1 + \lambda_2)C \subset \lambda_1 C + \lambda_2 C$, even if C is not convex. To show the reverse inclusion assuming C is convex, note that a vector x in $\lambda_1 C + \lambda_2 C$ is of the form $x = \lambda_1 x_1 + \lambda_2 x_2$, where $x_1, x_2 \in C$. By convexity of C, we have

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} x_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_2 \in C,$$

and it follows that

$$x = \lambda_1 x_1 + \lambda_2 x_2 \in (\lambda_1 + \lambda_2)C,$$

so $\lambda_1 C + \lambda_2 C \subset (\lambda_1 + \lambda_2)C$.

For a counterexample when C is not convex, let C be a set in $\mathbf{R}^n$ consisting of two vectors, 0 and $x \neq 0$, and let $\lambda_1 = \lambda_2 = 1$. Then C is not convex, and $(\lambda_1 + \lambda_2)C = 2C = \{0, 2x\}$, while $\lambda_1 C + \lambda_2 C = C + C = \{0, x, 2x\}$, showing that $(\lambda_1 + \lambda_2)C \neq \lambda_1 C + \lambda_2 C$.

(b) Let $x \in \cap_{i\in I} C_i$ and let α be a positive scalar. Since $x \in C_i$ for all $i \in I$ and each C_i is a cone, the vector αx belongs to C_i for all $i \in I$. Hence, $\alpha x \in \cap_{i\in I} C_i$, showing that $\cap_{i\in I} C_i$ is a cone.

(c) First we prove that $A \cdot C$ is a cone, where A is a linear transformation and $A \cdot C$ is the image of C under A. Let $z \in A \cdot C$ and let α be a positive scalar. Then, $Ax = z$ for some $x \in C$, and since C is a cone, $\alpha x \in C$. Because $A(\alpha x) = \alpha z$, the vector αz is in $A \cdot C$, showing that $A \cdot C$ is a cone.

Next we prove that the inverse image $A^{-1} \cdot C$ of C under A is a cone. Let $x \in A^{-1} \cdot C$ and let α be a positive scalar. Then $Ax \in C$, and since C is a cone, $\alpha Ax \in C$. Thus, the vector $A(\alpha x) \in C$, implying that $\alpha x \in A^{-1} \cdot C$, and showing that $A^{-1} \cdot C$ is a cone.

(d) Let $x \in C_1 + C_2$ and let α be a positive scalar. Then, $x = x_1 + x_2$ for some $x_1 \in C_1$ and $x_2 \in C_2$, and since C_1 and C_2 are cones, $\alpha x_1 \in C_1$ and $\alpha x_2 \in C_2$. Hence, $\alpha x = \alpha x_1 + \alpha x_2 \in C_1 + C_2$, showing that $C_1 + C_2$ is a cone.

(e) Let C be a convex cone. Then $\gamma C \subset C$, for all $\gamma > 0$, by the definition of cone. Furthermore, by convexity of C, for all $x, y \in C$, we have $z \in C$, where

$$z = \frac{1}{2}(x + y).$$

Hence $(x + y) = 2z \in C$, since C is a cone, and it follows that $C + C \subset C$.

Conversely, assume that $C + C \subset C$, and $\gamma C \subset C$. Then C is a cone. Furthermore, if $x, y \in C$ and $\alpha \in (0, 1)$, we have $\alpha x \in C$ and $(1 - \alpha)y \in C$, and $\alpha x + (1 - \alpha)y \in C$ (since $C + C \subset C$). Hence C is convex.

2. Let C be a nonempty convex subset of $\mathbf{R}^n$. Let also $f = (f_1, \ldots, f_m)$, where $f_i : C \mapsto \Re$, $i = 1, \ldots, m$, are convex functions, and let $g : \mathbf{R}^m \mapsto \mathbf{R}$ be a function that is convex and monotonically nondecreasing over a convex set that contains the set $\{f(x) \mid x \in C\}$, in the sense that for all u_1, u_2 in this set such that $u_1 \leq u_2$, we have $g(u_1) \leq g(u_2)$. Show that the function h defined by $h(x) = g(f(x))$ is convex over C. If in addition, $m = 1$, g is monotonically increasing and f is strictly convex, then h is strictly convex.

Solution.

Let $x, y \in \mathbf{R}^n$ and let $\alpha \in [0, 1]$. By the definitions of h and f, we have

$$\begin{aligned}
h(\alpha x + (1 - \alpha)y) &= g(f(\alpha x + (1 - \alpha)y)) \\
&= g(f_1(\alpha x + (1 - \alpha)y), \ldots, f_m(\alpha x + (1 - \alpha)y)) \\
&\leq g(\alpha f_1(x) + (1 - \alpha)f_1(y), \ldots, \alpha f_m(x) + (1 - \alpha)f_m(y)) \\
&= g(\alpha(f_1(x), \ldots, f_m(x)) + (1 - \alpha)(f_1(y), \ldots, f_m(y))) \\
&\leq \alpha g(f_1(x), \ldots, f_m(x)) + (1 - \alpha)g(f_1(y), \ldots, f_m(y)) \\
&= \alpha g(f(x)) + (1 - \alpha)g(f(y)) \\
&= \alpha h(x) + (1 - \alpha)h(y) \qquad (1)
\end{aligned}$$

where the first inequality follows by convexity of each f_i and monotonicity of g, while the second inequality follows by convexity of g.

If $m = 1$, g is monotonically increasing, and f is strictly convex, then the first inequality is strict whenever $x \neq y$ and $\alpha \in (0, 1)$, showing that h is strictly convex.

3. Show that the following functions from $\mathbf{R}^n$ to $(-\infty, \infty]$ are convex:
(a) $f_1(x) = \ln(e^{x_1} + \cdots + e^{x_n})$.
(b) $f_2(x) = \|x\|^p$ with $p \geq 1$.
(c) $f_3(x) = e^{\beta x'Ax}$, where A is a positive semidefinite symmetric $n \times n$ matrix and β is a positive scalar.
(d) $f_4(x) = f(Ax + b)$, where $f : \mathbf{R}^m \mapsto \mathbf{R}$ is a convex function, A is an $m \times n$ matrix, and b is a vector in $\mathbf{R}^m$.

Solution.
(a) We show that the Hessian of f_1 is positive semidefinite at all $x \in \mathbf{R}^n$. Let $\underline{\ }(x) = e^{x_1} + \cdots + e^{x_n}$. Then a straightforward calculation yields

$$z'\nabla^2 f_1(x) z = \frac{1}{\underline{\ }(x)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{(x_i + x_j)} (z_i - z_j)^2 \geq 0, \qquad \forall\, z \in \mathbf{R}^n.$$

Hence by the previous problem, f_1 is convex.

(b) The function $f_2(x) = \|x\|^p$ can be viewed as a composition $g(f(x))$ of the scalar function $g(t) = t^p$ with $p \geq 1$ and the function $f(x) = \|x\|$. In this case, g is convex and monotonically increasing over the nonnegative axis, the set of values that f can take, while f is convex over $\mathbf{R}^n$ (since any vector norm is convex). From problem 2, it follows that the function $f_2(x) = \|x\|^p$ is convex over $\mathbf{R}^n$.

(c) The function $f_3(x) = e^{\underline{X}'Ax}$ can be viewed as a composition $g(f(x))$ of the function $g(t) = e^{\underline{t}}$ for $t \in \mathbf{R}$ and the function $f(x) = x'Ax$ for $x \in \mathbf{R}^n$. In this case, g is convex and monotonically increasing over $\mathbf{R}$, while f is convex over $\mathbf{R}^n$ (since A is positive semidefinite). From problem 2, it follows that f_3 is convex over $\mathbf{R}^n$.

(d) This part is straightforward using the definition of a convex function.

4. Let C be a nonempty convex subset of $\mathbb{R}^n$. Show that:

$$\mathrm{cone}(C) = \cup_{x \in C} \{\gamma x | \gamma \geq 0\}.$$

Solution

Let $y \in \mathrm{cone}(C)$. If $y = 0$, then $y \in \cup_{x \in C} \{\gamma x \mid \gamma \geq 0\}$ and we are done. If $y \neq 0$, then by definition of $\mathrm{cone}(C)$, we have

$$y = \sum_{i=1}^{m} \lambda_i x_i,$$

for some positive integer m, nonnegative scalars λ_i, and vectors $x_i \in C$. Since $y \neq 0$, we cannot have all λ_i equal to zero, implying that $\sum_{i=1}^m \lambda_i > 0$. Because $x_i \in C$ for all i and C is convex, the vector

$$x = \sum_{i=1}^{m} \frac{\lambda_i}{\sum_{i=1}^m \lambda_i} x_i$$

belongs to C. For this vector, we have

$$y = \left(\sum_{i=1}^{m} \lambda_i \right) x,$$

with $\sum_{i=1}^m \lambda_i > 0$, implying that $y \in \cup_{x \in C} \{\gamma x \mid \gamma \geq 0\}$ and showing that

$$\mathrm{cone}(C) \subset \cup_{x \in C} \{\gamma x \mid \gamma \geq 0\}.$$

The reverse inclusion follows directly from the definition of $\mathrm{cone}(C)$.

5. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a differentiable function. We say that f is strongly convex with coefficient α if

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \alpha\|x - y\|^2, \forall x, y \in \mathbb{R}^n,$$

where α is some positive scalar.

(a) Show that if f is strongly convex with coefficient α, then f is strictly convex.

(b) Assume that f is twice continuously differentiable. Show that strongly convexity of f with coefficient α is equivalent to the positive semi definiteness of $\nabla^2 f(x) - \alpha I$ for every $x \in \mathbb{R}^n$, where I is the identity matrix.

Solution

(a) Fix some $x, y \in \Re^n$ such that $x \neq y$, and define the function $h : \Re \mapsto \Re$ by $h(t) = f\big(x + t(y - x)\big)$. Consider scalars t and s such that $t < s$. Using the chain rule and the equation

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) \geq \alpha\|x - y\|^2, \qquad \forall\, x, y \in \Re^n, \tag{1.8}$$

for some $\alpha > 0$, we have

$$\begin{aligned}
&\left(\frac{dh(s)}{dt} - \frac{dh(t)}{dt}\right)(s - t) \\
&\quad = \Big(\nabla f\big(x + s(y - x)\big) - \nabla f\big(x + t(y - x)\big)\Big)'(y - x)(s - t) \\
&\quad \geq \alpha(s - t)^2\|x - y\|^2 > 0.
\end{aligned}$$

Thus, dh/dt is strictly increasing and for any $t \in (0, 1)$, we have

$$\frac{h(t) - h(0)}{t} = \frac{1}{t}\int_0^t \frac{dh(\tau)}{d\tau}\, d\tau < \frac{1}{1 - t}\int_t^1 \frac{dh(\tau)}{d\tau}\, d\tau = \frac{h(1) - h(t)}{1 - t}.$$

Equivalently, $th(1) + (1 - t)h(0) > h(t)$. The definition of h yields $tf(y) + (1 - t)f(x) > f\big(ty + (1 - t)x\big)$. Since this inequality has been proved for arbitrary $t \in (0, 1)$ and $x \neq y$, we conclude that f is strictly convex.

(b) Suppose now that f is twice continuously differentiable and Eq. (1.8) holds. Let c be a scalar. We use Prop. 1.1.13(b) twice to obtain

$$f(x + cy) = f(x) + cy'\nabla f(x) + \frac{c^2}{2}y'\nabla^2 f(x + tcy)y,$$

and

$$f(x) = f(x + cy) - cy'\nabla f(x + cy) + \frac{c^2}{2}y'\nabla^2 f(x + scy)y,$$

for some t and s belonging to $[0, 1]$. Adding these two equations and using Eq. (1.8), we obtain

$$\frac{c^2}{2}y'\big(\nabla^2 f(x + scy) + \nabla^2 f(x + tcy)\big)y = \big(\nabla f(x + cy) - \nabla f(x)\big)'(cy) \geq \alpha c^2\|y\|^2.$$

We divide both sides by c^2 and then take the limit as $c \to 0$ to conclude that $y'\nabla^2 f(x)y \geq \alpha\|y\|^2$. Since this inequality is valid for every $y \in \Re^n$, it follows that $\nabla^2 f(x) - \alpha I$ is positive semidefinite.

For the converse, assume that $\nabla^2 f(x) - \alpha I$ is positive semidefinite for all $x \in \Re^n$. Consider the function $g : \Re \mapsto \Re$ defined by

$$g(t) = \nabla f\big(tx + (1 - t)y\big)'(x - y).$$

Using the Mean Value Theorem (Prop. 1.1.12), we have

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) = g(1) - g(0) = \frac{dg(t)}{dt}$$

for some $t \in [0,1]$. On the other hand,

$$\frac{dg(t)}{dt} = (x-y)'\nabla^2 f\big(tx + (1-t)y\big)(x-y) \geq \alpha\|x-y\|^2,$$

where the last inequality holds because $\nabla^2 f\big(tx+(1-t)y\big) - \alpha I$ is positive semidefinite. Combining the last two relations, it follows that f is strongly convex with coefficient α.