Introduction to Optimization for Computer Science — Assignment 2

Idan Drori
Ohad Sapir

Question #1

(a) Let $f : S \times S \to \mathbb{R}$ be convex in each variable, meaning $f(\cdot, y)$ is convex for all $y \in S$, and $f(x, \cdot)$ is convex for all $x \in S$. Then f is convex.

Solution: We'll disprove: Let $f(x,y) = xy, \forall x, y \in S$, where S is convex. f is convex in each variable because it's simply linear in each variable. Let $u = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, then:

$$0 = f(0,0) = f\left(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot(-1), \frac{1}{2}\cdot(-1) + \frac{1}{2}\cdot 1\right) = f\left(\frac{1}{2}u_1 + \frac{1}{2}u_2, \frac{1}{2}v_1 + \frac{1}{2}v_2\right)$$
$$\not\leq \frac{1}{2}f(u_1,u_2) + \frac{1}{2}f(v_1,v_2) = \frac{1}{2}f(1,-1) + \frac{1}{2}f(-1,1) = \frac{1}{2}(-1) + \frac{1}{2}(-1) = -1$$

□

(b) If $f : S \to \mathbb{R}$ is convex, then the sublevel sets $L_\alpha := \{x \in S \mid f(x) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$.

Solution: We'll prove this. Let $\alpha \in \mathbb{R}, x, y \in L_\alpha$. From convexity of f we get for any $\lambda \in (0,1)$:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$$

Meaning $(\lambda x + (1-\lambda)y) \in L_\alpha$ and so L_α is convex.

□

(c) Assume $f : S \to \mathbb{R}$ is such that the sublevel sets $L_\alpha := \{x \in S \mid f(x) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$. Then f is convex.

Solution: We'll disprove: Let $f : \mathbb{R} \to \mathbb{R}, f(x) = x^3$. For any $\alpha \in \mathbb{R}$, $L_\alpha = \{x \in \mathbb{R} | x^3 \leq \alpha\} = \{x \leq \sqrt[3]{\alpha}\}$. For any $x, y \in L_\alpha$ and $\lambda \in (0,1)$:

$$\lambda x + (1-\lambda)y \leq \lambda\sqrt[3]{\alpha} + (1-\lambda)\sqrt[3]{\alpha} = \sqrt[3]{\alpha} \Rightarrow \lambda x + (1-\lambda)y \in L_\alpha$$

So the sublevel sets L_α is convex, but f is obviously not a convex function.

□

(d) Let $f : S \to \mathbb{R}$ be a strictly convex function: $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in S, x \neq y, \lambda \in (0,1)$. Then it has at most one minimum in S.

Solution: We'll prove this. We'll assume for the sake of contradiction that $x^*, y^* \in S$ are both minima for f in S. Meaning for all $x \in S \backslash \{x^*, y^*\}, f(x) \geq f(x^*) = f(y^*)$. Then from strict convexity of f:

$$f\left(\frac{1}{2}x^* + \frac{1}{2}y^*\right) < \frac{1}{2}f(x^*) + \frac{1}{2}f(y^*) = \frac{1}{2}f(x^*) + \frac{1}{2}f(x^*) = f(x^*)$$

In contradiction to the fact that x^* is a minimum.

$\square$

(e) Assume $X, Y \subseteq \mathbb{R}^d$ are convex sets and $g : X \times Y \to \mathbb{R}$ is a convex function. Note this means that $g(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq \lambda g(x_1, y_1) + (1-\lambda)g(x_2, y_2)$ for all $x_1, x_2 \in X, y_1, y_2 \in Y, \lambda \in (0,1)$. Assume further that for all $x \in X, \arg\min_{y\in Y} g(x,y) \neq \emptyset$. Then $f : X \to \mathbb{R}, f(x) := \min_{y\in Y} g(x,y)$ is convex.

Solution: We'll prove this. Let $\lambda \in (0,1), x_1, x_2 \in X, y_1, y_2 \in Y$ such that $f(x_1) = g(x_1, y_1), f(x_2) = g(x_2, y_2)$, then:

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &= \min_{y\in Y} g(\lambda x_1 + (1-\lambda)x_2, y) \leq g(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\ &\leq \lambda g(x_1, y_1) + (1-\lambda)g(x_2, y_2) = \lambda f(x_1) + (1-\lambda)f(x_2) \end{aligned}$$

Meaning f is convex, as required. $\square$

Question #2

(a) Assume $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable and both convex and concave. Prove that f is affine.

Solution: Let $b = f(0), h(x) = -b, g(x) = f(x) + h(x)$.
g is a sum of two convex and concave, and therefore g is also convex and concave. we'll prove that g is linear.
Homogeneity: Let $x \in \mathbb{R}^d$.
Let $\lambda \in [0,1]$.

$$g(0) = f(0) + h(0) = b - b = 0$$

Since g is both convex and concave,

$$g(\lambda x + (1-\lambda)0) = \lambda g(x) + (1-\lambda)g(0)$$

$$g(\lambda x) = \lambda g(x)$$

Let $\lambda > 1$. $\lambda^{-1} \in [0,1]$

$$\lambda g(x) = \lambda g(\lambda^{-1}\lambda x) = \lambda\lambda^{-1}g(\lambda x) = g(\lambda x)$$

Let $\lambda < 0$. Let $y \in \mathbb{R}^d$.

$$0.5g(y) + 0.5g(-y) = g(0.5y - 0.5y) = g(0) = 0$$

$$-g(y) = g(-y)$$

$$\lambda g(x) = -(-\lambda)g(x) = -g(-(\lambda x)) = -(-g(\lambda x)) = g(\lambda x)$$

Additivity: Let $x, y \in \mathbb{R}^d$.

$$g(x+y) = g(0.5(2x) + 0.5(2y) = 0.5g(2x) + 0.5g(2y) = g(x) + g(y)$$

So g is linear, meaning there exists $a \in \mathbb{R}^d$, such that $g(x) = a^\top x$. And now:

$$f(x) = g(x) - h(x)$$

$$f(x) = a^\top x + b$$

$\square$

(b) Prove that $f : \mathbb{R}^d \to \mathbb{R}$ is convex iff the one dimensional function $h_{x,u}(t) = f(x + tu)$ is convex for any $x, u \in \mathbb{R}^d$.

Solution:

$(\implies)$ Let $\lambda \in [0,1], x, u \in \mathbb{R}^d, t, s \in \mathbb{R}, a = x + tu, b = x + su$.

$$\begin{aligned} h_{x,u}(\lambda t + (1-\lambda)s) &= f(x + (\lambda t + (1-\lambda)s)u) \\ &= f(\lambda(x+tu) + (1-\lambda)(x+su)) = f(\lambda a + (1-\lambda)b) \\ &\leq \lambda f(a) + (1-\lambda)f(b) = \lambda f(x+tu) + (1-\lambda)f(x+su) \\ &= \lambda h_{x,u}(t) + (1-\lambda)h_{x,u}(s) \end{aligned}$$

$(\impliedby)$ Let $\lambda \in [0,1], x, y \in \mathbb{R}^d, u = y - x$.

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= f(x + (1-\lambda)(y-x)) = f(x + (1-\lambda)u) \\ &= f(x + (\lambda \cdot 0 + (1-\lambda)\cdot 1)u) = h_{x,u}(\lambda \cdot 0 + (1-\lambda)\cdot 1) \\ &\leq \lambda h_{x,u}(0) + (1-\lambda)h_{x,u}(1) = \lambda f(x + 0\cdot u) + (1-\lambda)f(x + 1\cdot u) \\ &= \lambda f(x) + (1-\lambda)f(x+y-x) = \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

$\square$

(c) Show that the set $S = \{(x,y) \in \mathbb{R}^2 \mid x + y \leq 4, x^2 + y^2 \leq 10\}$ is convex.

Solution: Let $\lambda \in [0,1], a, b \in S, c = \lambda a + (1-\lambda)b$.

$$c = \begin{pmatrix} \lambda a_1 + (1-\lambda)b_1 \\ \lambda a_2 + (1-\lambda)b_2 \end{pmatrix}$$

$c_1 + c_2 = \lambda a_1 + (1-\lambda)b_1 + \lambda a_2 + (1-\lambda)b_2 = \lambda(a_1+a_2) + (1-\lambda)(b_1+b_2) \leq \lambda 4 + (1-\lambda)4 = 4$

Since $a_1^2 + a_2^2 \leq 10$ and $b_1^2 + b_2^2 \leq 10$, $|a_1|, |a_2|, |b_1|, |b_2| \leq \sqrt{10}$.

$$\begin{aligned} c_1^2 + c_2^2 &= (\lambda a_1 + (1-\lambda)b_1)^2 + (\lambda a_2 + (1-\lambda)b_2)^2 \\ &= \lambda^2 a_1^2 + \lambda(1-\lambda)a_1 b_1 + (1-\lambda)^2 b_1^2 + \lambda^2 a_2^2 + \lambda(1-\lambda)a_2 b_2 + (1-\lambda)^2 b_2^2 \\ &= \lambda^2(a_1^2 + a_2^2) + (1-\lambda)^2(b_1^2 + b_2^2) + \lambda(1-\lambda)(a_1 b_1 + a_2 b_2) \\ &\leq (10\lambda^2 + 10(1-\lambda)^2 + \lambda(1-\lambda)(10+10) = 10(\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)) \\ &= 10(\lambda^2 + 1 - 2\lambda + \lambda^2 - 2\lambda^2 + 2\lambda) = 10 \end{aligned}$$

$\square$

(d) Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \sqrt{2x^2 - xy + y^2}$. Prove that f is convex.

Solution: Let $A = \begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix}$ $a = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}$.

$$a^\top A a = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x & -x + y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2x^2 - xy + y^2$$

Meaning that $f(x) = \sqrt{a^\top A a}$.

$$Tr(A) = 3 > 0$$

$$Det(A) = 2 > 0$$

So A is PD, and f is a norm and therefore convex.

□

(e) Let $A \in \mathbb{R}^{m \times n}; b \in \mathbb{R}^m, c \in \mathbb{R}^n; d \in \mathbb{R}$, and $S = \{x \in \mathbb{R}^n \mid c^\top x + d > 0\}$. Show that $f : S \to \mathbb{R}, f(x) = \frac{\|Ax+b\|^2}{c^\top x + d}$ is convex over S.

Solution: Let $g(x) = \|Ax + b\|, h(x) = c^\top x + d$.
h is affine and therefore both convex and concave. From S definition, $h > 0$.
g is a composition of a convex function with an affine function and therefore convex, and $g < 0$.

$$\begin{aligned}
f(\lambda x + (1 - \lambda)y) &= \frac{g^2(\lambda x + (1 - \lambda)y)}{h(\lambda x + (1 - \lambda)y)} \\
&= \frac{g^2(\lambda x + (1 - \lambda)y)}{\lambda h(x) + (1 - \lambda)h(y)} \leq \frac{(\lambda g(x) + (1 - \lambda)g(y))^2}{\lambda h(x) + (1 - \lambda)h(y)} \\
&= \frac{\lambda^2 g^2(x) + (1 - \lambda)^2 g^2(y) + 2\lambda(1 - \lambda)g(x)g(y)}{\lambda h(x) + (1 - \lambda)h(y)} \\
&\leq \frac{\lambda^2 g^2(x)}{\lambda h(x) + (1 - \lambda)h(y)} + \frac{(1 - \lambda)^2 g^2(y)}{\lambda h(x) + (1 - \lambda)h(y)} \\
&\leq \frac{\lambda^2 g^2(x)}{\lambda h(x)} + \frac{(1 - \lambda)^2 g^2(y)}{(1 - \lambda)h(y)} = \lambda \frac{g^2(x)}{h(x)} + (1 - \lambda)\frac{g^2(y)}{h(y)} \\
&= \lambda f(x) + (1 - \lambda)f(y)
\end{aligned}$$

□

(f) Fix $x \in \mathbb{R}^d$ and let $q : \mathbb{R}^{dK} \to \mathbb{R}^K$ be the softmax-over-linear classifier. Let $y \in [K]$ fixed. Prove that $f(W) = -\log(q(W)_y)$ is convex.

Solution:

$$
\begin{aligned}
f(W) = -\log(q(W)_y) = -\log\left(\frac{e^{x^\top w_i}}{(\sum_j e^{x^\top w_j})}\right) \\
= \log\left(\sum_j e^{x^\top w_j}\right) - \log(e^{x^\top w_i}) \\
= \log\left(\sum_j e^{x^\top w_j}\right) - x^\top w_i
\end{aligned}
$$

$-x^\top w_i$ is affine and hence convex.
$\log\left(\sum_j e^{x^\top w_j}\right)$ is a composition of a convex function (log-sum-exp) with an affine function and therefore convex.
So f is a sum of convex functions, meaning it is convex.

□