Optimization Methods - Convex Functions

Bolei Zhang

December 7, 2022

1 Problems

Problem 1: A function $f(\cdot)$ is convex $\iff$ $\forall x, v \in \mathrm{dom} f,\ g(t) = f(x + tv)$ is convex.

Solution: "$\implies$": For any $t_1, t_2 \in \mathrm{dom} g$, $x, v \in \mathrm{dom} f$, there is

$$\begin{aligned} g(\theta t_1 + (1-\theta)t_2) &= f(x + \theta t_1 v + (1-\theta)t_2 v) \\ &= f(\theta(x + t_1 v) + (1-\theta)(x + t_2 v)) \\ &\le \theta f(x + t_1 v) + (1-\theta) f(x + t_2 v)) \\ &= \theta g(t_1) + (1-\theta) g(t_2) \end{aligned} \tag{1}$$

"$\impliedby$": For any $x_1, x_2 \in \mathrm{dom} f$, there is $x_1 + tx_2 \in \mathrm{dom} g$. We have

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &= f(x_1 + (1-\theta)x_2 - (1-\theta)x_1) \\ &= f(x_1 + (1-\theta)(x_2 - x_1)) \\ &= g(0 + (1-\theta)\cdot 1) \\ &\le \theta g(0) + (1-\theta) g(1) \\ &= \theta f(x_1) + (1-\theta) f(x_2) \end{aligned} \tag{2}$$

Theorem 1: First order condition: For a differentiable function f, if dom f is a convex set, then f is a convex function $\iff$ $f(y) \ge f(x) + \nabla f(x)^T (y - x), \forall x, y \in \mathrm{dom} f$.

Proof: First consider the case $n = 1$: We show that a differentiable function $f : R \to R$ is convex if and only if

$$f(y) \ge f(x) + f'(x)(y - x) \tag{3}$$

for all x and y in $\mathrm{dom} f$.

Assume first that f is convex and $x, y \in \mathrm{dom} f$. Since $\mathrm{dom} f$ is convex, we conclude that for all $0 < t \le 1, x + t(y - x) \in \mathrm{dom} f$, and by convexity of f,

$$f(x + t(y - x)) \le (1 - t) f(x) + t f(y). \tag{4}$$

If we divide both sides by t, we obtain

$$f(y) \ge f(x) + \frac{f(x + t(y - x)) - f(x)}{t}, \tag{5}$$

and taking the limit as $t \to 0$ yields (3).

To show sufficiency, assume the function satisfies (3) for all x and y in $\mathrm{dom} f$ (which is an interval). Choose any $x \neq= y$, and $0 \le \theta \le 1$, and let $z = \theta x + (1-\theta) y$. Applying (3) twice yields

$$f(x) \ge f(z) + f'(z)(x - z), f(y) \ge f(z) + f'(z)(y - z). \tag{6}$$

Multiplying the first inequality by θ, the second by $1 - \theta$, and adding them yields

$$\theta f(x) + (1-\theta) f(y) \ge f(z), \tag{7}$$

which proves that f is convex.

Now we can prove the general case, with $f: R^n \to R$. Let $x, y \in R^n$ and consider f restricted to the line passing through them, i.e., the function defined by $g(t) = f(ty + (1-t)x)$, so $g'(t) = \nabla f(ty + (1-t)x)^T (y - x)$.

First assume f is convex, which implies g is convex, so by the argument above we have $g(1) \geq g(0) + g'(0)$, which means

$$f(y) \geq f(x) + \nabla f(x)^T (y - x). \tag{8}$$

Now assume that this inequality holds for any x and y, so if $ty + (1-t)x \in \mathrm{dom} f$, and $\tilde{t}y + (1-\tilde{t})x \in \mathrm{dom} f$, we have

$$f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x) + \nabla f(\tilde{t}y + (1-\tilde{t})x)^T (y - x)(t - \tilde{t}), \tag{9}$$

i.e., $g(t) \geq g(t) + g'(\tilde{t})(t - \tilde{t})$. We have seen that this implies that g is convex.

Definition 1: Norm functions Given a vector space X, a norm on X is a real-valued function $p: X \to R$ with the following properties, where $|s|$ denotes the usual absolute value of a scalar s:

- Subadditivity/Triangle inequality: $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$;
- Absolute homogeneity: $p(sx) = |s|\, p(x)$ for all $x \in X$ and all scalars s.
- Positive definiteness/Point-separating: for all $x \in X$, if $p(x) = 0$, then $x = 0$.

Theorem 2: Norm functions are convex.

Proof: If $f: R^n \to R$ is an norm, and $0 \leq \theta \leq 1$,then $f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y) = \theta f(x) + (1-\theta) f(y)$. The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

Problem 2: The maximum function $f(x) = \max\{x_1, x_2, ..., x_n\}$, $x \in R^m$ is convex.

Solution: The function $f(x) = \max_i x_i$ satisfies, for $0 \leq \theta \leq 1$,

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i (\theta x_i + (1-\theta) y_i) \\ &\leq \theta \max_i x_i + (1-\theta) \max_i y_i \\ &= \theta f(x) + (1-\theta) f(y) \end{aligned} \tag{10}$$

Problem 3: Determine whether $f(x) = x^{-2}, x \neq 0$ is convex function.

Solution: The domain of this function is not a convex set.

Problem 4: L_0 norm: $||x||_0$ is the number of non-zero elements in x. Explain that whether $L0$ norm is a norm function, whether it is a convex function.

Solution: Let $x = (0, 1)$. There is $||x||_0 = 1$ and $||2x||_0 = 1$. However, $2||x||_0 = 2$, which violates the homogeneity. Therefore, L_0 norm is not a norm function.

Let $x = (0, 1), y = (1, 0)$. Let $\theta = 0.5$,

$$f(\theta x + (1-\theta)y) = ||(0.5, 0.5)||_0 = 2, \tag{11}$$

$$\theta f(x) + (1-\theta) f(y) = 0.5 + 0.5 = 1, \tag{12}$$

Therefore, L_0 norm is not convex.

Problem 5: $f(x, y) = \frac{x^2}{y}$, $y > 0$ is a convex function.

Solution: The Hessian matrix of $f(x, y)$ is (for $y > 0$)

$$\nabla^2 f = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0. \tag{13}$$

Problem 6: Compute the conjugate of the functions:
(1) $f(x) = ax + b, \mathrm{dom} f = \mathbb{R}$.
(2) $f(x) = -\log x, \mathrm{dom} f = \mathbb{R}_{++}$

Solution: (1). As a function of x, $yx - ax - b$ is bounded if and only if $y = a$, in which case it is constant. Therefore the domain of the conjugate function f^* is the singleton $\{a\}$, and $f^*(a) = -b$.

(2) The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathrm{dom} f^* = \{y | y < 0\} = -\mathbb{R}_{++}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.

Problem 7: Determine whether the following functions are convex

- $f(x) = e^x - 1, x \in \mathbb{R}$
- $f(x_1, x_2) = x_1 x_2, x_1, x_2 \in \mathbb{R}$
- $h(z) = \log(\sum_{i=1}^k e^{x_i}), x_i \in \mathbb{R}$

Solution: (1) The function is convex as $\nabla^2 f = e^x > 0$.

(2) The function is not convex as $\nabla^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is not positive semi-definite.

(3) The Hessian of the log-sum-exp function is

$$\nabla^2 f(x) = \frac{1}{(1^T z)^2}((1^T z)\mathrm{diag}(z) - zz^T),$$

where $z = (e^{x_1}, ..., e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that for all $v, v^T \nabla^2 f(x) v \geq 0$, i.e.,

$$v^T \nabla^2 f(x) v = \frac{1}{(1^T z)^2}((\sum_{i=1}^n z_i)(\sum_{i=1}^n v_i^2 z_i) - (\sum_{i=1}^n v_i z_i)^2) \geq 0$$

. But this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to the vectors with components $a_i = v_i \sqrt{z_i}, b_i = \sqrt{z_i}$.

Problem 8: The KL-divergence is convex $D_{KL}(u, v) = \sum_{i=1}^n (u_i \log \frac{u_i}{v_i} - u_i + v_i)$

Solution: First, we show that $g(x, t) = -t \log(x/t)$ is convex on R^2_{++}: Let $\mathrm{dom} f = C$. The domain $\mathrm{dom} g = R^2_{++}$ is a convex set.

The function $g(\cdot)$ is convex because the function is defined as a scaled and shifted version of convex function $f(x)$.

Therefore, the relative entropy of two vectors $u, v \in R^n_{++}$, defined as

$$\sum_{i=1}^n u_i \log(u_i/v_i)$$

is convex in (u, v), since it is a sum of relative entropies of u_i, v_i.

The KL-divergence is convex as it is the relative entropy plus a linear function of (u, v).

Problem 9: Convex-concave functions and saddle-points. We say the function $f : R^n \times R^m \to R$ is convex-concave if $f(x, z)$ is a concave function of z, for each fixed x, and a convex function of x, for each fixed z. We also require its domain to have the product form $\mathrm{dom} f = A \times B$, where $A \subseteq R^n$ and $B \subseteq R^m$ are convex.

(1) Give a second-order condition for a twice differentiable function $f : R^n \times R^m \to R$ to be convex-concave, in terms of its Hessian $\nabla^2 f(x, z)$.

(2) Suppose that $f : R^n \times R^m \to R$ is convex-concave and differentiable, with $\nabla f(\tilde{x}, \tilde{z}) = 0$. Show that the saddle-point property holds: for all x, z, we have

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

. Show that this implies that f satisfies the *strong max-min property*:

$$\sup_z \inf_x f(x, z) = \inf_x \sup_z f(x, z)$$

(and their common value is $f(\tilde{x}, \tilde{z})$.

(3) Now suppose that $f : R^n \times R^m \to R$ is differentiable, but not necessarily convex-concave, and the saddle-point property holds at $\tilde{x}, \tilde{z}$:

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

for all x, z. Show that $\nabla f(\tilde{x}, \tilde{z}) = 0$

Solution: (1) The Hessian matrix of f is

$$\nabla^2 f = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where $A_{11} \in R^{n\times n}$, $A_{12} \in R^{n\times m}$, $A_{21} \in R^{m\times n}$, $A_{22} \in R^{m\times m}$. And there is $\nabla^2 f = A_{11}$, $\nabla_z^2 f = A_{22}$. As f is convex when z is fixed, then A_{11} is positive semi-definite; As f is concave when x is fixed, then A_{22} is negative semi-definite.

(2) For the first inequality, as $f(x, z)$ is convex of x, there is

$$f(x, \tilde{z}) \geq f(\tilde{x}, \tilde{z}) + \nabla_x f(\tilde{x}, \tilde{z})(x - \tilde{x}) = f(\tilde{x}, \tilde{z})$$

The same is true for the left part of the inequality.

Next we prove $\sup_z \inf_x f(x, z) = f(\tilde{x}, \tilde{z})$. First

$$\inf_x f(x, z) \leq f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}),$$

As

$$\sup_z \inf_x f(x, z) \leq f(\tilde{x}, \tilde{z}),$$

and

$$\sup_z \inf_x f(x, z) \geq \inf_x f(x, \tilde{z}) \geq f(\tilde{x}, \tilde{z}),$$

Therefore, there is $\sup_z \inf_x f(x, z) = f(\tilde{x}, \tilde{z})$. The same is true for $\inf_x \sup_z f(x, z) = f(\tilde{x}, \tilde{z})$, which proves the equality.

(3) We need to prove $\nabla_x f(\tilde{x}, \tilde{z}) = 0$ and $\nabla_z f(\tilde{x}, \tilde{z}) = 0$. We prove by contradiction. Assume $\nabla_x f(\tilde{x}, \tilde{z}) \neq 0$, let $v = (\nabla_x f(\tilde{x}, \tilde{z})^T, 0)^T$. Get the first order of $f(\tilde{x} + tv, \tilde{z})$ at $(\tilde{x}, \tilde{z})$ $(t \neq 0)$,

$$f(\tilde{x} + tv, \tilde{z}) = f(\tilde{x}, \tilde{z}) + t||\nabla_x f(\tilde{x}, \tilde{z})||^2 + O(t^2).$$

Take $t < 0$ with small enough absolute value, there is $f(\tilde{x} + tv, \tilde{z}) < f(\tilde{x}, \tilde{z})$, which contradicts with the assumption. Therefore, there is $\nabla_x f(\tilde{x}, \tilde{z}) = 0$ and $\nabla_z f(\tilde{x}, \tilde{z}) = 0$.

Problem 10: Show that $f(Ax + b)$ is convex if $f(x)$ is a convex function.

Solution: The domain of $f(Ax + b)$ is the same with $f(x)$, which is a convex set. For any two points $x, y \in \text{dom} f$, $0 \leq \theta \leq 1$, there is,

$$f(A(\theta x + (1-\theta)y) + b) = f(\theta(Ax + b) + (1-\theta)(Ay + b)) \leq \theta f(Ax + b) + (1-\theta) f(Ay + b). \quad (14)$$

Therefore, $f(Ax + b)$ is convex.