Lecture 6. Convex Functions (cont'd)

6.1 Convexity-preserving operations

For the convex sets, we have known that if C and D are both convex sets, then $C + D$, $C - D$, $C \times D$ and $C \cap D$ are all convex sets. We wonder if there exist similar convexity-preserving operations for functions.

Nonnegative Sums

Theorem

If $f_1, f_2, \ldots, f_n$ are convex, and $w_1, w_2, \ldots, w_n \geq 0$, then $f = w_1 f_1 + \ldots + w_m f_m$ is convex.
Furthermore, if a two-dimension function $f(x, y)$ is convex for any fixed y, and there exist a series of cofficients $w(y) \geq 0, y \in \Omega$, then

$$g(x) \triangleq \int_\Omega w(y) f(x, y) \, \mathrm{d}y$$

is also a convex function.

We can verify the convexity of objective functions via Jensen inequality.

Pointwise Maximum

Theorem

If $f_1, f_2, \ldots, f_m$ are convex, then

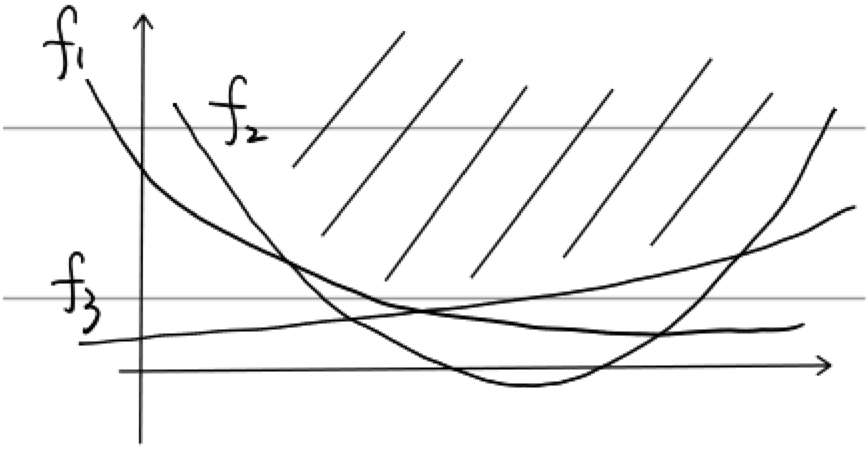
$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

is also convex.

Futhermore, if $f(x, y)$ is convex for any fixed y, then

$$g(x) = \sup_{y \in \Omega} f(x, y)$$

is also a convex function.



The proof is immediate by noting that the epigraph of the pointwise maximum function is the intersection of the epigraphs of all f's.

Affine mapping

Theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex/concave, and $\mathbf{A} \in \mathbb{R}^{n \times m}, \mathbf{b} \in \mathbb{R}^n$, then

$$g : \mathbb{R}^m \to \mathbb{R} \triangleq f(Ax + b)$$

is also convex or concave (the same as f).

Example

$f(x) = \|Ax + b\|$ is convex for any norm $\|\cdot\|$ function.

By the triangle inequality, it holds that

$$\|\theta x + \bar{\theta} u\| \leq \|\theta x\| + \|\bar{\theta} y\| = \theta \|x\| + \bar{\theta} \|y\| \, .$$

Therefore, $\|\cdot\|$ is convex and its affine transformation $f(x)$ is also convex due to the above theorem.

Scalar composition

Given two convex and differentiable functions $g : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}$, consider their scalar composition $h(x) = g(f(x))$. When will $h(x)$ also be convex?
When $n = 1$, we first compute the second order derivative of h:

$$h'(x) = g'(f(x)) \cdot f'(x) h''(x) = g''(f(x)) \cdot f'(x)^2 + g'(f(x)) \cdot f''(x) \,.$$

With the help of the direct computation results, we have the following theorem:

Theorem

- h is convex if g is convex and one of the following proposition is true:
  - g is increasing and f is convex;
  - g is decreasing and f is concave.
- h is convex if g is concave and one of the following proposition is true:
  - g is increasing and f is concave;
  - g is decreasing and f is convex.

Proof

We just show the proof of case 2, the other three cases can reuse this proof. If f is concave and g is decreasing, then

$$f(\theta x + \bar{\theta} y) \geq \theta f(x) + \bar{\theta} f(y) \,.$$

Let $z = \theta x + \bar{\theta} y$. We have

$$\begin{aligned} h(z) = g(f(z)) &\leq g\left(\theta f(x) + \bar{\theta} f(y)\right) \\ &\leq \theta g(f(x) + \bar{\theta} g(f(y)) = \theta h(x) + \bar{\theta} h(y) \,. \end{aligned}$$

Note that for other four cases, we can not determine whether h is convex just by the monotonic and convexity.

Example

- $h(x) = e^{x^\top Q x}$ is convex if $Q \succeq 0$ (but if $Q \not\succeq 0$, we have no idea whether $h(x)$ is convex or not).

- If $g(x) = e^{-x}$ and $f(x) = x^2$, then $h(x) = e^{-x^2}$ is neither convex nor concave.
- If $g(x) = -\log x$ and $f(x) = e^x + 1$, then $h(x) = -\log(e^x + 1)$ is concave.
- (*log-sum-exp*) If $g(x) = \log x$ and $f(x) = e^{x_1} + e^{x_2} + \ldots + e^{x_n}$, then

  $$h(x) = \log(e^{x_1} + \ldots + e^{x_n})$$

  is convex.

We should explain more on the log-sum-exp function. This function is very useful for approximately computing the maximum of $x_1, \ldots, x_n$ ($h(x) \approx \max\{x_1, x_2, \ldots, x_n\}$).
We usually hope the objective function of the optimization problem is *differentiable*. However $\max$ is not. So log-sum-exp gives a good approximation of $\max$ (log-sum-exp is a smooth function).
Given a series of points $x_1, \ldots x_n$, the *softmax* function $\frac{e^{x_i}}{\sum_j e^{x_j}}$ returns a probability distribution. Moreover, the distribution is equal to the gradient of the log-sum-exp.

Vector composition

Suppose $g : \mathbb{R}^\ell \to \mathbb{R}$, $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^\ell$ or $\boldsymbol{f} = (f_1, f_2, \ldots, f_\ell)$, $f_i : \mathbb{R}^n \to \mathbb{R}, \forall\, 1 \le i \le \ell$.
Let

$$h(\boldsymbol{x}) = g(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_\ell(\boldsymbol{x}))\,.$$

Then we have the following theorem by defining "increasing" as: $g(\boldsymbol{x}) \ge g(\boldsymbol{y})$ if $x_i \ge y_i$ for all $1 \le i \le \ell$.

Theorem

- h is convex if g is convex and one of the following proposition is true:
  - g is increasing and f_i is convex for all i;
  - g is decreasing and f_i is concave for all i.
- h is convex if g is concave and one of the following proposition is true:
  - g is increasing and f_i is concave for all i;
  - g is decreasing and f_i is convex for all i.

Minimization over convex sets

Theorem

Suppose $f(x, y)$ is convex, and $C \neq \emptyset$ is convex, then $g(x) \triangleq \inf_{y \in C} f(x, y)$ is convex. (e.g., $f(x) = \operatorname{dist}(x, C) = \inf_{y \in C} \|x - y\|$.)

Proof

We prove this theorem by verifying the Jensen's inequality. In other words, we want to show that

$$\forall\, x_1, x_2 \in \operatorname{dom} g, \qquad g(\theta x_1 + \bar{\theta} x_2) \leq \theta g(x_1) + \bar{\theta} g(x_2)\,.$$

By the definition of g, for any $\epsilon > 0$, there exist $y_1, y_2 \in C$ such that $f(x_i, y_i) < g(x_i) + \epsilon$. Therefore,

$$\begin{aligned} \theta f(x_1, y_1) + \bar{\theta} f(x_2, y_2) &< \theta\left(g(x_1) + \epsilon\right) + \bar{\theta}\left(g(x_2) + \epsilon\right) \\ &= \theta g(x_1) + \bar{\theta} g(x_2) + \epsilon\,. \end{aligned}$$

Since f is convex and C is convex, we have

$$\theta f(x_1, y_1) + \bar{\theta} f(x_2, y_2) \geq f\left(\theta x_1 + \bar{\theta} x_2, \theta y_1 + \bar{\theta} y_2\right) \geq g(\theta x_1 + \bar{\theta} x_2)\,.$$

Therefore, for any $\epsilon > 0$,

$$g(\theta x_1 + \bar{\theta} x_2) < \theta g(x_1) + \bar{\theta} g(x_2) + \epsilon\,.$$

Taking the limit $\epsilon \to 0$, it certifies the Jensen's inequality.

6.2 Applications of convexity

Now we consider the problem of the triangle inequality for general L^p-norms, which we omitted before. Let us first prove $\|u + v\|_2 \leq \|u\|_2 + \|v\|_2$ as warm-up.

$$\begin{aligned} \|u + v\|_2^2 &= \langle u + v, u + v \rangle \\ &= \|u\|^2 + \|v\|^2 + 2\langle u, v \rangle \\ &\leq \|u\|^2 + \|v\|^2 + 2\|u\| \cdot \|v\| \qquad \text{by Cauchy-Schwarz inequality} \\ &= (\|u\| + \|v\|)^2 \end{aligned}$$

To verify the triangle inequality for all p, we need a generalized version of Cauchy-Schwarz. We first introduce the monotonicity for L^p norms.

Proposition

Let $\boldsymbol{x} \in \mathbb{R}^n$. Then $\|\boldsymbol{x}\|_{p_1} \geq \|\boldsymbol{x}\|_{p_2}$ if $1 \leq p_2 \leq p_2$.

Proof

- If $p_2 = \infty$, $\|\boldsymbol{x}\|_{p_2} = \max\{|x_1|, \ldots, |x_n|\} \leq \|\boldsymbol{x}\|_{p_1}$.
- If $p_1 < \infty$. First we normalize $\boldsymbol{x}$. Let $q = \|\boldsymbol{x}\|_{p_1}$ and $\tilde{\boldsymbol{x}} = \boldsymbol{x}/q$ (namely, $\tilde{x}_i = x_i/q$), then we have $\|\tilde{\boldsymbol{x}}\|_{p_1} = 1$. Thus,

$$\begin{aligned}
\|\tilde{\boldsymbol{x}}\|_{p_2} &= \left(\sum |\tilde{x}_i|^{p_2}\right)^{\frac{1}{p_2}} \\
&= \left(\sum (|\tilde{x}_i|^{p_1})^{p_2/p_1}\right)^{\frac{1}{p_2}} \\
&\leq \left(\sum |\tilde{x}_i|^{p_1}\right)^{\frac{1}{p_2}} = 1 = \|\tilde{\boldsymbol{x}}\|_{p_1}.
\end{aligned}$$

Recall the detail of the proof of L^2-norm, we can notice that the key point is to apply the *Cauchy-Schwarz inequality*: for any two vectors $x, y \in \mathbb{R}^n$,
$\|x\|_2 \cdot \|y\|_2 \geq \langle x, y \rangle$.
When considering general L^p-norm, we wonder if there also exists an inequality in form of $\|x\|_\square \cdot \|y\|_\square \geq \langle x, y \rangle$. In fact, there exists an important inequality called *Hölder's inequality*.

Theorem (*Hölder' inequality*)

Let p and q be two conjugate exponents, i.e., $1/p + 1/q = 1$. Then for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, the following inequality holds:

$$\|\boldsymbol{u}\|_p \cdot \|\boldsymbol{v}\|_q \geq \langle \boldsymbol{u}, \boldsymbol{v} \rangle .$$

Proof

Without loss of generality, we assume that $u_i, v_i \geq 0$ for any i. If $\sum_i u_i v_i = 0$, the inequality obviously holds. Otherwise, we first normalize these two vectors. Let

$$\tilde{\boldsymbol{u}} = \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_p}, \quad \tilde{\boldsymbol{v}} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_q} .$$

Then $\|\tilde{\boldsymbol{u}}\|_p = \|\tilde{\boldsymbol{v}}\|_q = 1$. So our goal is to prove that $\langle \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \rangle = \sum_i \tilde{u}_i \tilde{v}_i \leq 1$.
We first claim that

$$\forall\, x, y > 0, \quad x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}.$$

Taking the logarithm on both sides, it is equivalent to the following inequality:

$$\forall\, x, y > 0, \quad \frac{1}{p}\log(x) + \frac{1}{q}\log(y) \leq \log\left(\frac{x}{p} + \frac{y}{q}\right).$$

By the Jensen's inequality, the above inequality holds since $f(x) = \log(x)$ is concave.
Next, applying this claim to $x = \tilde{u}_i^p$ and $y = \tilde{v}_i^q$, we have $\tilde{u}_i \tilde{v}_i \leq \frac{1}{p}\tilde{u}_i^p + \frac{1}{q}\tilde{v}_i^q$.
Summing them up we have

$$\sum_i \tilde{u}_i \tilde{v}_i \leq \sum_i \left(\frac{1}{p}\tilde{u}_i{}^p + \frac{1}{q}\tilde{v}_i{}^q\right) = \frac{1}{p}\left(\sum_i \tilde{u}_i^p\right) + \frac{1}{q}\left(\sum_i \tilde{v}_i^q\right) = \frac{1}{p} + \frac{1}{q} = 1\,.$$

Now we are going to show the triangle inequality for L^p norms, which is also called the *Minkowski inequality*.

Theorem (*Minkowski inequality*)

For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, and any p such that $1 \leq p \leq \infty$, the following inequality holds:

$$\|\boldsymbol{u}\|_p + \|\boldsymbol{v}\|_p \geq \|\boldsymbol{u} + \boldsymbol{v}\|_p\,.$$

Proof

Assuming $1 < p < \infty$, we have

$$\begin{aligned} \|\boldsymbol{u} + \boldsymbol{v}\|_p^p &= \sum_i |u_i + v_i|^p = \sum_i |u_i + v_i| \cdot |u_i + v_i|^{p-1} \\ &\leq \sum_i |u_i| \cdot |u_i + v_i|^{p-1} + |v_i| \cdot |u_i + v_i|^{p-1}\,. \end{aligned}$$

By the Hölder's inequality,

$$
\begin{aligned}
&\sum_i |u_i| \cdot |u_i + v_i|^{p-1} + |v_i| \cdot |u_i + v_i|^{p-1} \\
&\leq \|\boldsymbol{u}\|_p \left(\sum_i^n \left(|u_i + v_i|^{p-1} \right)^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} + \|\boldsymbol{v}\|_p \left(\sum_i^n \left(|u_i + v_i|^{p-1} \right)^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \\
&= \|\boldsymbol{u}\|_p \left(\sum_i^n |u_i + v_i|^p \right)^{\frac{p-1}{p}} + \|\boldsymbol{v}\|_p \left(\sum_i^n |u_i + v_i|^p \right)^{\frac{p-1}{p}} \\
&= \|\boldsymbol{u}\|_p \|\boldsymbol{u} + \boldsymbol{v}\|_q^{p-1} + \|\boldsymbol{v}\|_p \|\boldsymbol{u} + \boldsymbol{v}\|_q^{p-1} .
\end{aligned}
$$

Therefore, we conclude that $\|\boldsymbol{u} + \boldsymbol{v}\|_p \leq \|\boldsymbol{u}\|_p + \|\boldsymbol{v}\|_p$

6.3 Convex optimization problems

After defining and discussing properties of convex sets and convex functions, we now introduce what type of optimization problems we should consider in this course.

Recall that, in general, an optimization problem is to find the minimum value of $f(x)$ where x satisfies $g(x) = 0$ and $h(x) \leq 0$. Namely, it can be written as the following forms.

Definition (*Optimization Problem*)

The following problem P is the standard form of an optimization problem.

$$
\begin{array}{lll}
\text{minimize} & f(x) & \text{objective function} \\
\text{subject to} & g_i(x) = 0, 1 \leq i \leq k ; & \text{constraint functions} \\
 & h_j(x) \leq 0, 1 \leq j \leq \ell . & \text{constraint functions}
\end{array}
$$

- The *domain* of P is given by

$$D = \operatorname{dom} f \cap \left(\cap \operatorname{dom} g_i\right) \cap \left(\cap \operatorname{dom} h_j\right) .$$

- The *feasible set* of P is given by

$$\Omega = \{x \in D \mid \forall\, i \,, g_i(x) = 0 \text{ and } \forall\, j \,, h_j(x) \leq 0\} .$$

- The *optimal value* of P is

$$f^* = \inf_{x \in \Omega} f(x) \qquad \text{or } f^* = \min_{x \in \Omega} f(x) \text{ if exists} .$$

- The *optimal solution* of P (if exists) is

$$x^* = \arg\min_{x\in\Omega} f(x)\,.$$

For convenience, we usually allow f^* to take the extended value $\pm\infty$.
Conventionally,

- $f^* = \infty$ if P is infeasible (i.e., $\Omega = \emptyset$);
- $f^* = -\infty$ if $f(x)$ is unbounded below over Ω;
- x^* is an optimal solution iff $x^* \in \Omega$ and $f(x^*) = f^*$
- x^* is a locally optimal point if there exists $\delta > 0$ such that $f(x) \geq f(x^*)$ holds for all $x \in D \cap \mathcal{B}(x^*, \delta)$.

In particular, in this course, we mainly consider the *convex optimization*.

Definition (*Convex optimization problem*)

Given an optimization problem P, it is called *a convex optimization*, if the objective function f is convex, every equality constraint g_i is affine, and every inequality constraint h_j is convex.

Clearly, the domain of P is convex, since all domains of f, g_i and h_j are convex and the domain of P is their intersection.
We also note that the feasible set Ω is a convex set, since the solution sets $\{x \mid g_i(x) = 0\}$ is affine, the 0-sublevel sets $\{x \mid h_j(x) \leq 0\}$ are all convex, and Ω is their intersection.

Proposition

1. For a convex optimization problem, any local minimum is also a global minimum.
2. The set of optimal solutions $\Omega_{\text{opt}} = \{x^* \mid \forall\, x \neq x^*, f(x) \geq f(x^*)\}$ is also convex.
3. In particular, if f is strictly convex, there is at most one optimal solution.

Proof of item 2

For any two optimal solutions $x, y \in \Omega_{\text{opt}}$, for all $\theta \in [0, 1]$, $\theta x + (1-\theta)y \in \Omega$ since Ω is convex. By the Jensen's inequality, we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) = f^* .$$

Obviously, $f(\theta x + (1-\theta)y) \geq f^*$. Therefore, $\theta x + (1-\theta)y \in \Omega_{\text{opt}}$.

In fact, we can show that the *c-sublevel set* of a function, defined by

$$\{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq c\} ,$$

is convex if f is a convex function. (Exercise!)
Similarly, we can also define the *c-level set* of f as $\{\boldsymbol{x} \mid f(\boldsymbol{x}) = c\}$ and define the *c-superlevel set* of f as $\{\boldsymbol{x} \mid f(\boldsymbol{x}) \geq c\}$. The c-superlevel set of a concave function is convex.
Thus, another proof of item 2 is to note that Ω_{opt} is the intersection of two convex sets: Ω and the f^*-sublevel set of f.

Now we can say that a convex optimization problem is to compute the minimum value of a convex function over a convex set. However, the converse statement is not true. Calculating the minimum of a convex function on a convex set is not always a convex optimization problem. Consider the following example:

Example

The following optimization problem has convex objective function and convex feasible set. But it is **not** a convex optimization problem.

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 + x_2^2 \\ \text{subject to} \quad & g(x) = (x_1 + x_2)^2 = 0 \\ & h(x) = \frac{x_1}{x_2^2 + 1} \leq 0 . \end{aligned}$$

The feasible set Ω is just $\{(x_1, x_2) \mid x_1 + x_2 = 0, x_1 \leq 0\}$, which is a convex set. However, $g(x) = (x_1 + x_2)^2$ is not affine and $h(x) = \frac{x_1}{x_2^2+1}$ is not convex. Hence this problem is not convex.

Here are some canonical types of convex optimization problems.

Linear programming

A *linear programming* is a convex optimization where the objective function and constraint functions are all affine (linear).

Example (*Linear programming*)

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s. t.} \quad & \mathbf{A}_1\boldsymbol{x} = \boldsymbol{b}_1,\ \mathbf{A}_1 \in \mathbb{R}^{m\times n},\ \boldsymbol{b}_1 \in \mathbb{R}^m\,; \\ & \mathbf{A}_2\boldsymbol{x} \le \boldsymbol{b}_2,\ \mathbf{A}_2 \in \mathbb{R}^{\ell\times n},\ \boldsymbol{b}_2 \in \mathbb{R}^\ell\,. \end{aligned}$$

Quadratic programming

A *quadratic programming* is a convex optimization where the objective function is quadratic and constraint functions are all affine.

Example (*Quadratic programming*)

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} \quad (\text{convex iff } Q \succeq 0) \\ \text{s. t.} \quad & \mathbf{A}_1\boldsymbol{x} = \boldsymbol{b}_1,\ \mathbf{A}_1 \in \mathbb{R}^{m\times n},\ \boldsymbol{b}_1 \in \mathbb{R}^m\,; \\ & \mathbf{A}_2\boldsymbol{x} \le \boldsymbol{b}_2,\ \mathbf{A}_2 \in \mathbb{R}^{\ell\times n},\ \boldsymbol{b}_2 \in \mathbb{R}^\ell\,. \end{aligned}$$

Quadratically constrained quadratic programming

A *quadratically constrained quadratic programming* is a convex program where the objective function and inequality-constraint functions are all quadratic functions.

Example (*Quadratically constrained quadratic programming*)

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s. t.} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_i\boldsymbol{x} + \boldsymbol{w}_i^\top \boldsymbol{x} + d_i \le 0 \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b},\ \mathbf{A} \in \mathbb{R}^{m\times n},\ \boldsymbol{b} \in \mathbb{R}^n \end{aligned}$$

Note that it is convex iff $\boldsymbol{Q} \succeq 0$ and $\boldsymbol{Q}_i \succeq 0$ for all i.

The *linear least square regression* is a typical QP or QCQP. Given $y \in \mathbb{R}^n, x \in \mathbb{R}^{n\times p}$, our goal is to find $w \in \mathbb{R}^p$ to minimize $\|y - Xw\|$. By the direct calculation, we have

$$w^* = (X^\top X)^{-1} X^\top y$$

since

$$\nabla\|y - Xw\|^2 = \nabla(y - Xw)^\top(y - Xw) = 2X^\top Xw - 2X^\top y$$

and the optimal solution w^* satisfies $\nabla\|y - Xw^*\|^2 = 0$.