Lecture 7. Convex functions

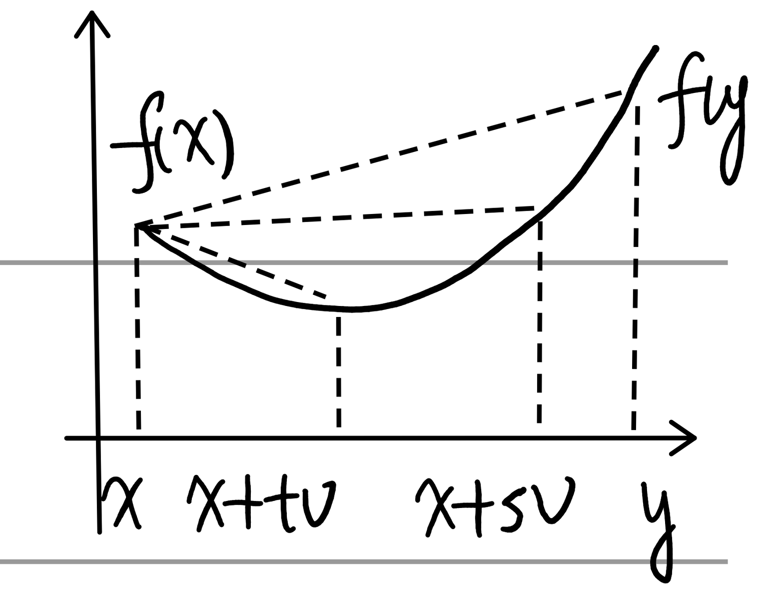
Recall the definition of convex functions: convexity of epigraphs.

Jensen's inequality; properties: midpoint convexity + continuity, zeroth condition / first-order condition: $f(y) \geq f(x) + \nabla f(x)^T(y-x)$.

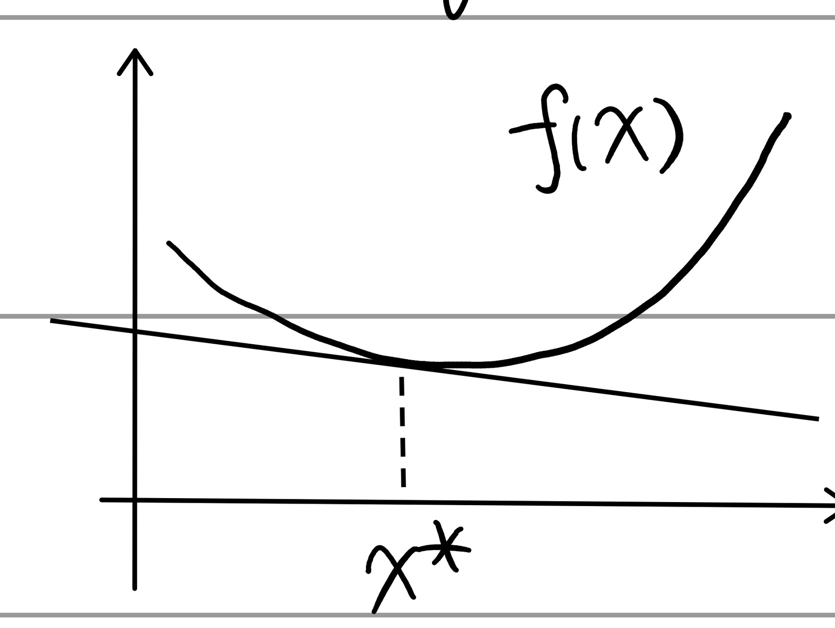
Remaining question: strict convexity $\Rightarrow f(y) > f(x) + \nabla f(x)^T(y-x)$?

Jensen's inequality $\Rightarrow f(x+tv) - f(x) < t(f(y)-f(x))$

Choose $0 < t < s < 1 \Rightarrow \dfrac{f(x+tv)-f(x)}{t} < \dfrac{f(x+sv)-f(x)}{s}$



Corollary: if f has a local minimum x^*, then x^* is a global one.



Second-order condition.

Suppose f is twice differentiable in an open convex set $\mathrm{dom}\, f$. Then f is convex iff $\forall x \in \mathrm{dom}\, f$, $\nabla^2 f(x) \succeq 0$.

Proof: "$\Rightarrow$". $\forall x_0$, let $g(x) = f(x) - (f(x_0) + \nabla f(x_0)^T(x-x_0)) \geq 0$.

$\Rightarrow x_0$ is a local minimum point of $g \Rightarrow \nabla^2 f(x_0) = \nabla^2 g(x_0) \succeq 0$.

"$\Leftarrow$": However, this direction is not easy. 2 problems $\begin{cases} \nabla^2 f \succeq 0 \to \min \\ \text{local} \to \text{global} \end{cases}$

Taylor expansion. $f(x) = f(x_0) + f'(x_0)(x-x_0) + \dfrac{f''(x_0)}{2}(x-x_0)^2 + \cdots + R_n$.

Lagrange reminder: $R_n = \dfrac{1}{n!} f^{(n)}(x_0 + \theta(x-x_0))(x-x_0)^n$ for some $\theta \in [0,1]$

For $f: \mathbb{R}^n \to \mathbb{R}$. given $x, y \in \mathrm{dom}\, f$. let $v = y - x$. $\mathrm{dom}\, f$ convex.

$\Rightarrow f$ defined over segment $[x, y]$. Then $\exists\, \theta \in (0,1)$. s.t. $f(y) =$

$f(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x + \theta v) v \geq f(x) + \nabla f(x)^T v$. $\square$.

Strict convexity: iff? "$\Leftarrow$" easy. but "$\Rightarrow$" not true. $f(x) = x^4$.

Example: entropy function $f(x) = -x \log x$ is concave, since $f''(x) = -\frac{1}{x}$.

In particular, for quadratic functions $f(x) = \frac{1}{2} x^T Q x + w^T x + b$. with

symmetric Q. f is convex iff $Q \succeq 0$, is strictly convex iff $Q \succ 0$.

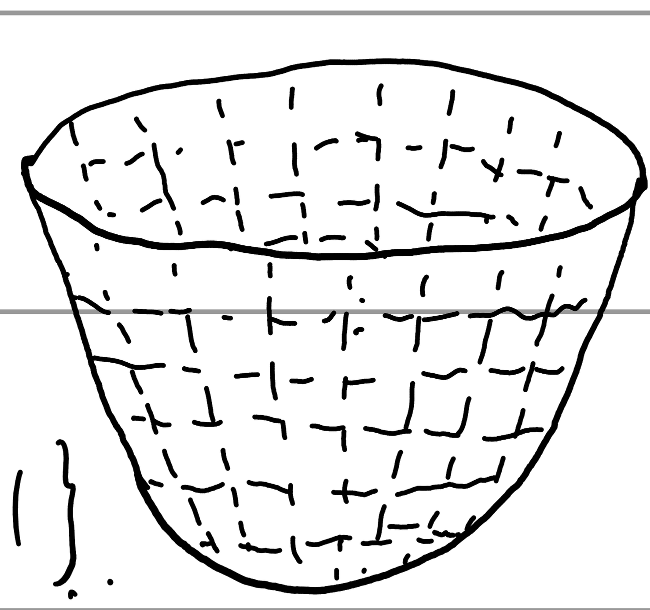
Remark: if Q is not symmetric. note that $x^T Q x = (x^T Q x)^T =$

$x^T Q^T x = x^T \frac{Q + Q^T}{2} x$. so it's no harm to assume Q is always symmetric.

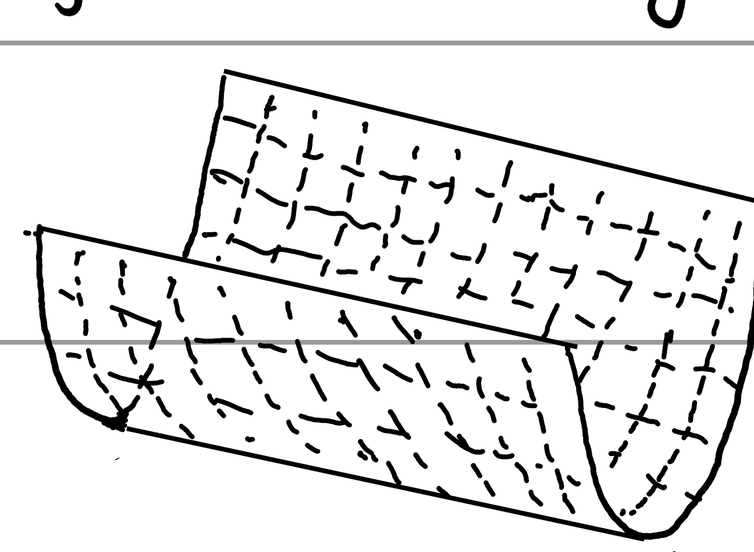
Proof: Only need to prove "$\Rightarrow$" direction for strict convexity.

Note that $\nabla f(x) = Qx + w$. $f(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$.

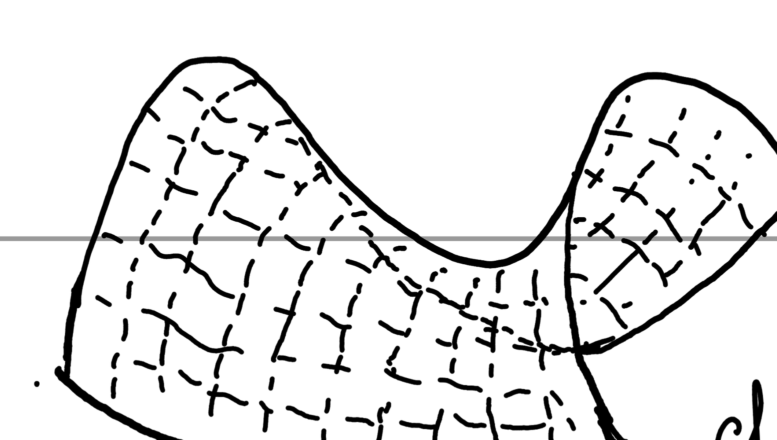
$\forall v \neq 0$. So $v^T Q v = v^T \nabla^2 f(x) v > 0$ by the first-order condition. $\square$.



$Q = \mathrm{diag}\{1, 1\}$.



$Q = \mathrm{diag}\{0, 1\}$



$Q = \mathrm{diag}\{1, -1\}$.

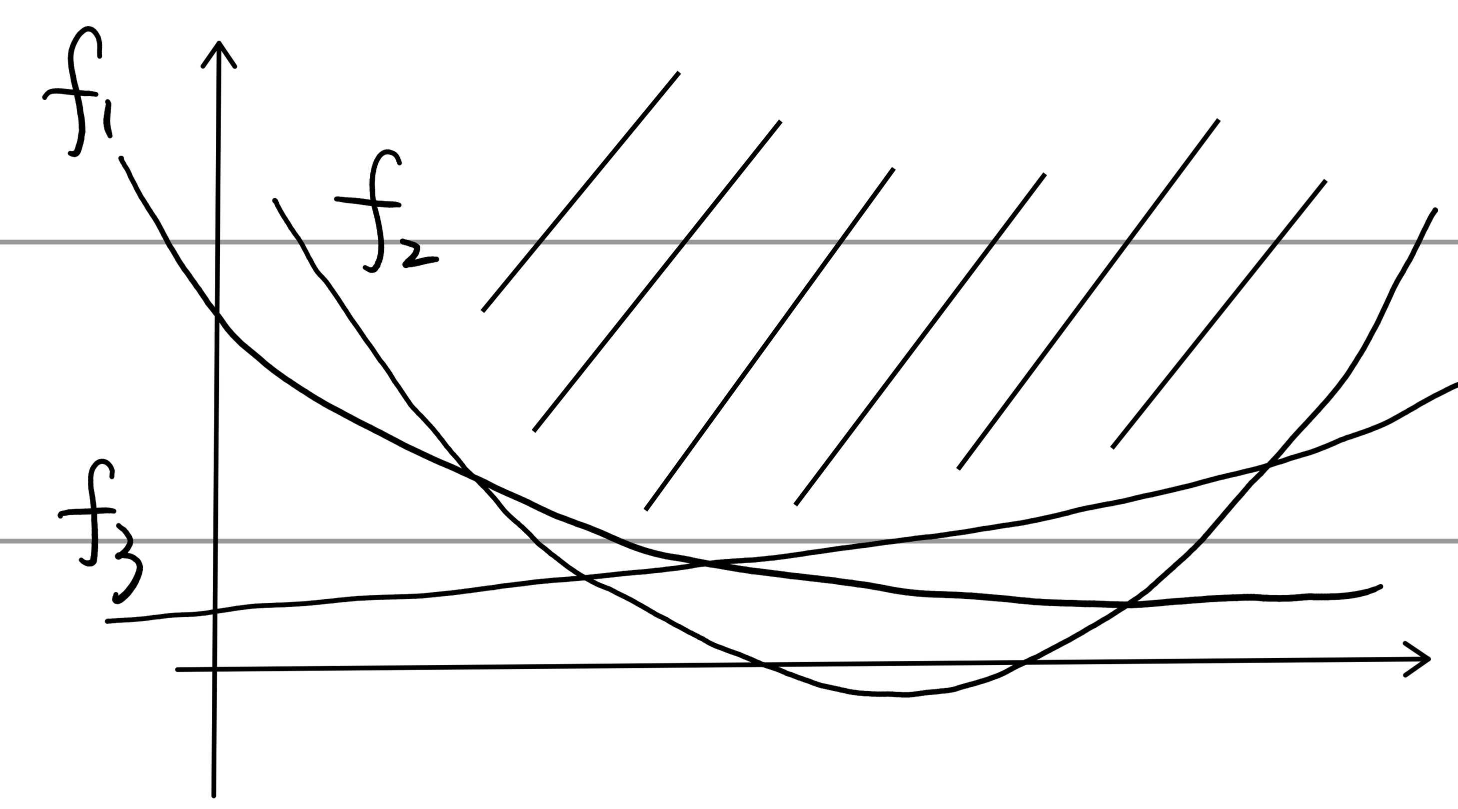
Convexity-preserving operations.

Recall that if C, D are convex, so is $C+D$. $C-D$. $C \cap D$. $C \times D$ ...

1. nonnegative weighted sums / conic combination.

Let $f_1, f_2, \ldots, f_m$ are convex, $\omega_1, \omega_2, \ldots, \omega_m \geq 0 \Rightarrow f = \sum \omega_i f_i$ $= \omega_1 f_1 + \omega_2 f_2 + \cdots + \omega_m f_m$ is also convex. Furthermore, if $f(x, y)$ convex for any y, and $\omega(y) \geq 0$ then $g(x) = \int_\Omega \omega(y) f(x, y)\, dy$ convex.

2. pointwise maximum and supremum.



If $f_1, f_2, \ldots, f_m$ are convex, so is $f(x) = \max\{f_1(x), f_2(x), \ldots, f_m(x)\}$.

If $f(x, y)$ convex for any y, then $g(x) = \sup\limits_{y \in \Omega} f(x, y)$ convex.

3. composition: affine mapping / scalar / vector

3.1. affine mapping: suppose $f: \mathbb{R}^n \to \mathbb{R}$ convex / concave. $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$. $g: \mathbb{R}^m \to \mathbb{R} \triangleq f(Ax + b)$ is convex / concave (same as f).

Example: $g(x) = \|Ax + b\|$ is convex since $f(x) = \|x\|$ convex.

3.2. scaler composition. suppose $g: \mathbb{R} \to \mathbb{R}$, $f: \mathbb{R}^n \to \mathbb{R}$, $h(x) = g(f(x))$

(inspired by the case $n = 1$: $h''(x) = g''(f(x)) f'(x)^2 + g'(f(x)) f''(x)$).

h is convex if g is convex, and $\begin{cases} g \text{ is increasing, } f \text{ is convex.} \\ g \text{ is decreasing, } f \text{ is concave.} \end{cases}$

h is concave if g is concave, and $\begin{cases} g \text{ is increasing, } f \text{ is concave} \\ g \text{ is decreasing, } f \text{ is convex.} \end{cases}$

Proof of Case 2. f is concave $\Rightarrow f(\theta x+\bar{\theta}y) \geq \theta f(x) + \bar{\theta} f(y)$.

$\Rightarrow h(z) = g(f(z)) \leq g(\theta f(x) + \bar{\theta} f(y)) \leq \theta g(f(x)) + \bar{\theta} g(f(y))$. $\square$

Example. $e^{x^T Q x}$ is convex if $Q \succeq 0$. concave if ? unknown.

Remark. If conditions fail. convexity is indetermined in general.

$g(x) = e^{-x}$. $f(x) = x^2$. $h(x) = e^{-x^2}$ neither convex nor concave.

$g(x) = -\log x$. $f(x) = e^x + 1$. $h(x) = -\log(e^x+1)$ is concave.

$g(x) = \log x$. $f(x) = e^{x_1} + e^{x_2} + \cdots + e^{x_n}$. $h(x) = \log(e^{x_1} + \cdots + e^{x_n})$.

$= \log(\Sigma e^{x_i})$ (log-sum-exp) is convex. More on log-sum-exp:

(true) softmax / smooth max: smooth function to approximate max.

(so-called) softmax: $\dfrac{e^{x_i}}{\Sigma e^{x_v}} = \nabla \mathrm{lse}(x)$. softargmax. lse = entropy.

3.3. vector composition: suppose $g: \mathbb{R}^l \to \mathbb{R}$. $f_i: \mathbb{R}^n \to \mathbb{R}$. or $f: \mathbb{R}^n \to \mathbb{R}^l$

$h(x) = g(f_1(x), f_2(x), \ldots, f_l(x))$. define increasing: $g(x) \geq g(y)$ if $\forall i, x_i \geq y_i$

h is convex if g is convex, and $\begin{cases} g \text{ is increasing. } f_i \text{ are convex.} \\ g \text{ is decreasing, } f_i \text{ are concave.} \end{cases}$

h is concave if g is concave, and $\begin{cases} g \text{ is increasing, } f_i \text{ are concave} \\ g \text{ is decreasing. } f_i \text{ are convex.} \end{cases}$

3.4. minimization over convex sets.

Suppose $f(x,y)$ is convex, $C \neq \phi$ is convex. So is $g(x) \triangleq \inf_{y \in C} f(x,y)$.

Remark. note that $\text{dom}\, g = \{x : \exists\, y \in C \text{ s.t. } (x,y) \in \text{dom}\, f\}$ is convex.

Proof: We verify Jensen's inequality for $\forall\, x_1, x_2 \in \text{dom}\, g$. Fix $\varepsilon > 0$.

$\exists\, y_1, y_2$ s.t. $f(x_i, y_i) < g(x_i) + \varepsilon$. $g(\theta x_1 + \bar\theta x_2) = \inf\limits_y f(\theta x_1 + \bar\theta x_2, y)$

$\le f(\theta x_1 + \bar\theta x_2, \theta y_1 + \bar\theta y_2) \le \theta f(x_1, y_1) + \bar\theta f(x_2, y_2)$ (since f is convex)

$< \theta g(x_1) + \bar\theta g(x_2) + \varepsilon$. $\forall \varepsilon > 0 \Rightarrow g(\theta x_1 + \bar\theta x_2) \le \theta g(x_1) + \bar\theta g(x_2)$. $\square$

Another proof: $\text{epi}(g) = \{(x,t) : \exists\, y \in C, \text{ s.t. } (x,y,t) \in \text{epi}(f)\}$

Example: distance to a convex set. $\text{dist}(x, C) = \inf\limits_{y \in C} \|x - y\|$ convex.

Jensen's inequality. $f(\theta_1 x_1 + \cdots + \theta_m x_m) \le \theta_1 f(x_1) + \cdots + \theta_m f(x_m)$. $\sum \theta_i = 1$.

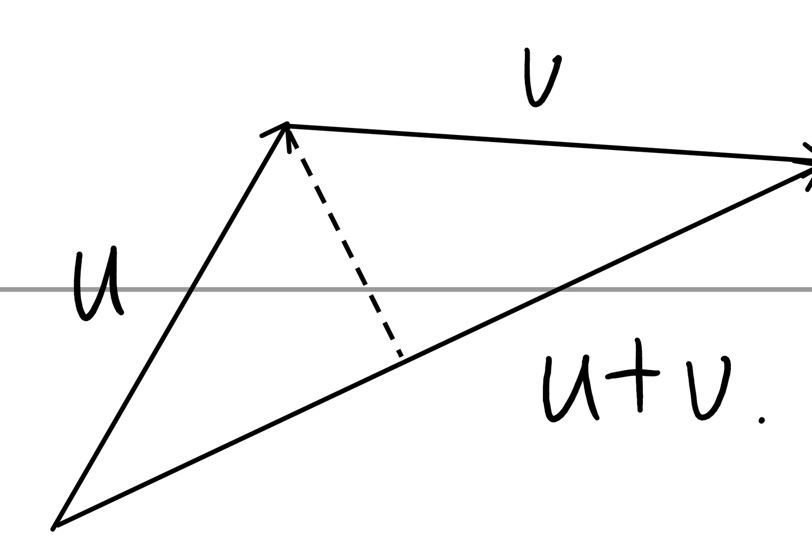
Integrals: if $p(x) \geq 0$ defined on $\Omega \subseteq \text{dom}\, f$. $\int_\Omega p(x)\,dx = 1$. then

$f\left(\int_\Omega p(x)\, x\, dx\right) \le \int_\Omega p(x) f(x)\, dx$. provided integrals exist.

in particular, if $p(x)$ is the distribution of X, then $f(\mathbb{E}[X]) \le \mathbb{E}[f(X)]$.

Verifying triangle inequality for l_p-norms: Minkowski's inequality.

Warnup: $\|u+v\|_2 \le \|u\|_2 + \|v\|_2$ in $\mathbb{R}^2$.



consider $\|u+v\|_2^2 = \langle u+v, u+v \rangle = \|u\|^2 + \|v\|^2$

$+ 2\langle u, v \rangle \le \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| = (\|u\| + \|v\|)^2$. $\square$

Rewrite: $\sum (u_i + v_i)^2 \le \sum \underbrace{u_i(u_i + v_i)}_{\langle u, u+v \rangle} + v_i(u_i + v_i)$

By Cauchy-Schwarz inequality. $\langle u, u+v\rangle \le \|u\|_2 \|u+v\|_2$.

How about $\sum u_i(u_i+v_i)^{p-1}$? Generalization of Cauchy-Schwarz.

Proposition: Given $1 \le p_1 < p_2 \le \infty$. then $\|x\|_{p_1} \ge \|x\|_{p_2}$ $\forall x \in \mathbb{R}^n$.

Proof. If $p_2 = \infty$. done. Assume $p_2 < \infty$. Let $q = \|x\|_{p_1}$. $\tilde{x}_i = x_i/q$.

Then $\|\tilde{x}\|_{p_1} = 1$. $\|\tilde{x}\|_{p_2} = \left(\sum (|\tilde{x}_i|^{p_1})^{p_2/p_1}\right)^{1/p_2} \le \left(\sum |\tilde{x}_i|^{p_1}\right)^{1/p_2} = 1$. $\square$

Hölder's inequality: Let $p, q \in (1, \infty)$ be conjugate exponents. i.e.

$\frac{1}{p} + \frac{1}{q} = 1$. Then $\langle u, v\rangle \le \|u\|_p \|v\|_q$. $\forall u, v \in \mathbb{R}^n$. namely.

$$\sum_{i=1}^{n} |u_i v_i| \le \left(\sum_{i=1}^{n} |u_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |v_i|^q\right)^{1/q}.$$