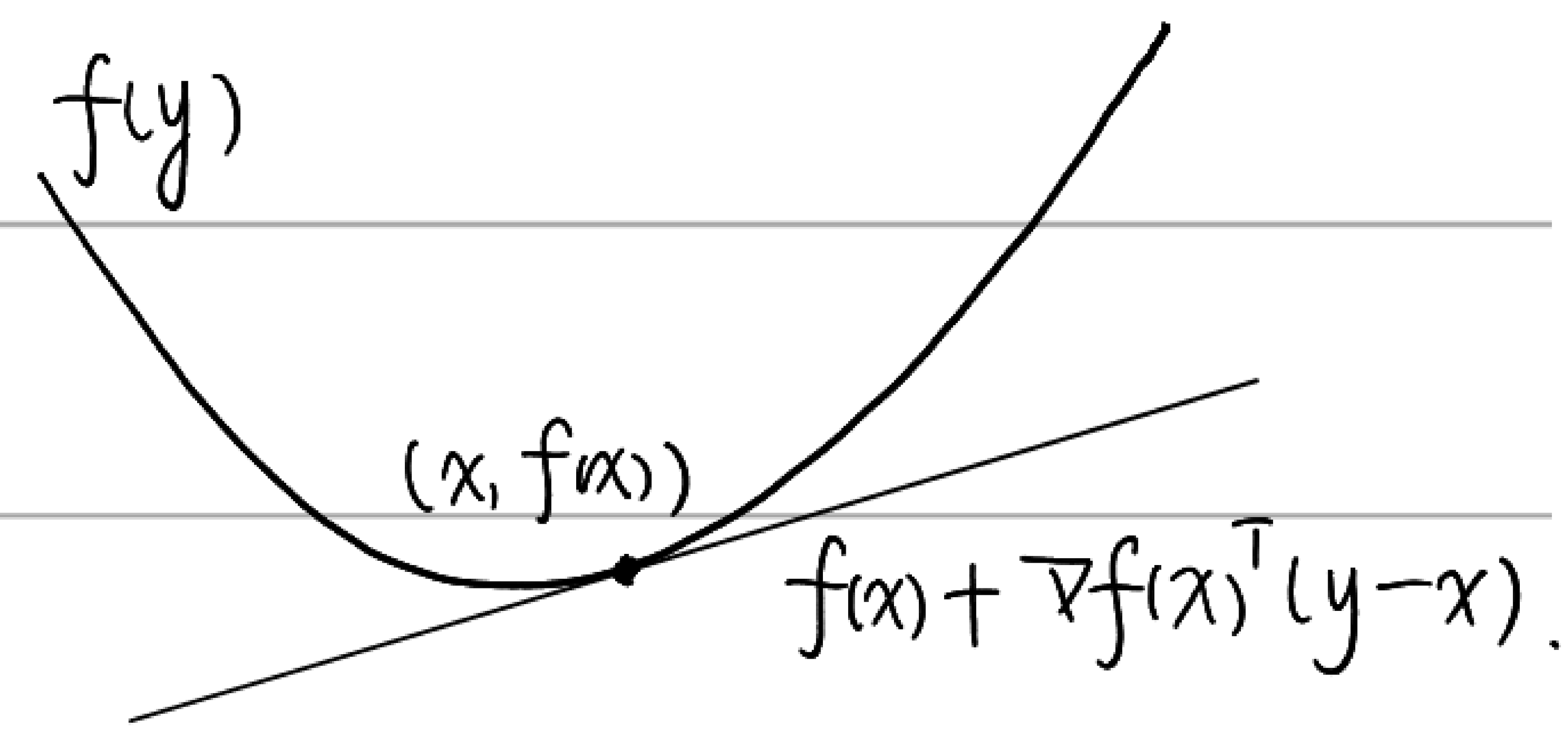
Lecture 7. Properties and characterization of convex functions

Zeroth-order condition:

f is convex iff f restricted to any direction is convex.

i.e. $\forall d \in \mathbb{R}^n$. $g(t) = f(x_0 + td)$ is convex.

First-order condition:



Suppose f is differentiable in an open. convex set $\mathrm{dom} f$. Then.

f is convex iff $f(y) \geq f(x) + \nabla f(x)^T (y-x) \quad \forall x. y \in \mathrm{dom} f$.

Example (Bernoulli's inequality): $(1+x)^r \geq 1 + rx$ if $r \geq 1.\ x \geq -1$.

in particular, $e^{rx} > (1+1/k)^{krx} \, (\forall k. \forall r > 0) \overset{\text{let } k = 1/x}{>} (1+x)^r$.

Remark: The first-order Taylor approximation. is a global. under-estimator of a convex function. and vice versa.

local information (value. gradient) $\Rightarrow$ global inequality.

In particular. $\nabla f(x) = 0 \Rightarrow f(y) \geq f(x) \quad \forall y \in \mathrm{dom} f$.

strictly convex: iff $f(y) > f(x) + \nabla f(x)^T (y-x) \quad \forall x \neq y$.

Note that $\mathrm{epi}(f)$ convex $\Rightarrow f(x) + \nabla f(x)^T (y-x)$ supporting hyperplane.

Proof: "$\Rightarrow$". Fix $x, y \in \text{dom} f$. and let $d = y - x$.

By Jensen's inequality. $f(x+td) \leq (1-t) f(x) + t f(y)$. $\forall t \in (0,1)$

$f(x+td) - f(x) \leq t(f(y) - f(x))$.

Taking the limit $t \to 0$. $\nabla f(x)^T \cdot d \leq f(y) - f(x)$

"$\Leftarrow$". Given x, y, θ. let $z = \theta x + \bar{\theta} y$.

the first-order condition $\Rightarrow \begin{cases} f(x) \geq f(z) + \nabla f(z)^T (x-z) \\ f(y) \geq f(z) + \nabla f(z)^T (y-z). \end{cases}$

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

Exercise. strictly convex. "$\Leftarrow$" trivial. but "$\Rightarrow$"?

Second-order condition:

Suppose f is twice differentiable in an open convex set $\text{dom} f$.

Then f is convex iff $\nabla^2 f(x) \succeq 0$ at $\forall x \in \text{dom} f$.

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

x_0 is a minima $\Rightarrow \nabla^2 g(x_0) = \nabla^2 f(x_0) \succeq 0$.

$\frac{1}{2} t^2 d^T \nabla^2 f(x_0) d + o(t^2 \|d\|^2) \geq 0$. $t \to 0 \Rightarrow \nabla^2 f(x_0) \succeq 0$.

"$\Leftarrow$". two problems: $\nabla^2 f(x_0) \succeq 0$ not sufficient; local $\to$ global.

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

Lagrange remainder: $R_n = \frac{1}{n!} f^{(n)}(x_0 + \theta(x - x_0))(x - x_0)^n$ for some $\theta \in (0,1)$.

Given $x, y \in \operatorname{dom} f$. let $d = y - x$.

$$f(y) = f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x + td)\, d \quad \text{for some } t \in (0,1)$$

$\operatorname{dom} f$ convex $\Rightarrow$ f defined over segment $[x, y]$.

$\nabla^2 f(x+td) \succeq 0 \Rightarrow f(y) \geq f(x) + \nabla f(x)^T d$. $\square$.

strictly convex: iff? "$\Leftarrow$" trivial. but "$\Rightarrow$"? $f(x) = x^4$

Exercise: give a proof or a counterexample. $f: \mathbb{R}^n \to \mathbb{R}$. $(n \geq 3)$

Example: negative entropy $f(x) = x \log x$. $f' = \log x + 1$. $f'' = \frac{1}{x}$.

Quadratic functions: $f(x) = \frac{1}{2} x^T Q x + w^T x + b$. with symmetric Q.

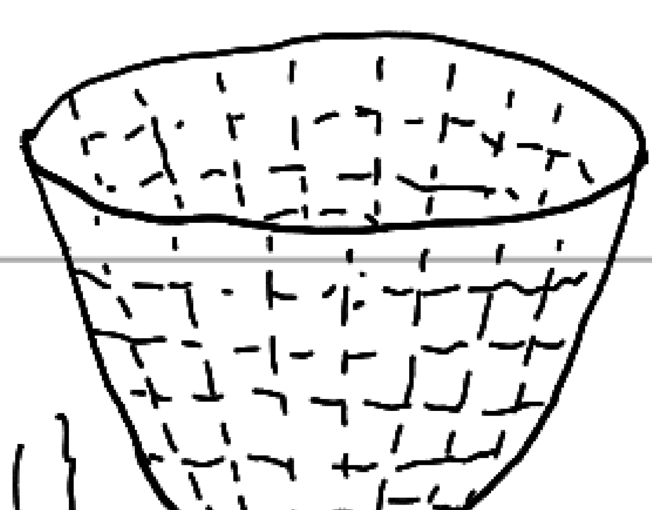
$f(x)$ is convex iff $Q \succeq 0$. is strictly convex iff $Q \succ 0$.

Remark: $\nabla^2 f(x) \succ 0$ not necessary in general. but necessary for quadratic.

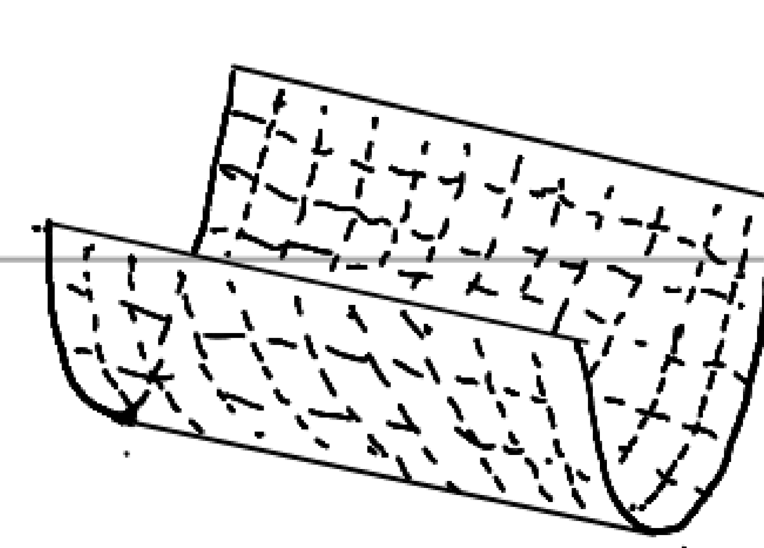
Proof: "$\Rightarrow$" part of strictly convex: note $\nabla f(x) = Qx + w$.

$$f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T Q d. \quad \forall d \neq 0.$$

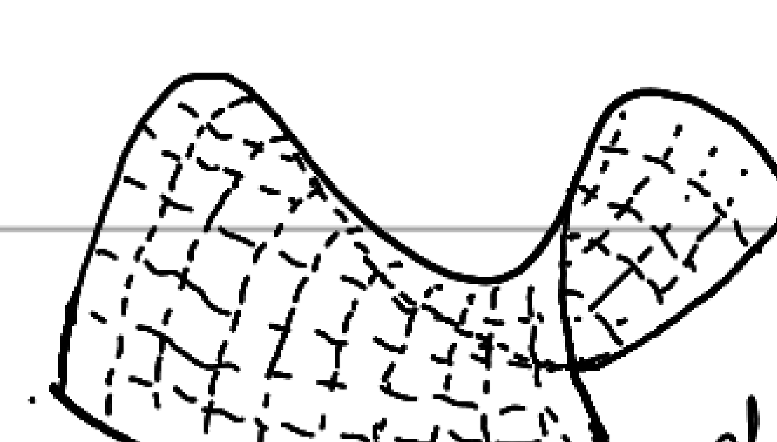
By first-order condition. $\frac{1}{2} d^T Q d > 0 \Rightarrow Q \succ 0$. $\square$



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



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



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

Convexity-preserving operations

Recall that C, D convex $\Rightarrow$ $C+D$. $C-D$. $C\cap D$. $AC+b$

How about $f+g$. $f-g$. $\max\{f,g\}$. $f(Ax+b)$ $f(g(x))$?

1. nonnegative weighted sums / conic combination.

let $f_1, f_2, \cdots, f_m$ are convex. $w_1, w_2, \cdots, w_m \geqslant 0$

$\Rightarrow f = w_1 f_1 + w_2 f_2 + \cdots + w_m f_m$ is also convex.

furthermore $g(x) = \int_\Omega w(y) f(x,y)\, dy$ is convex. if

$f(x,y)$ convex for any fixed $y \in \Omega$. $w \geqslant 0$. integral exists.

2. pointwise maximum and supremum.

$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}$ is convex.

$g(x) = \sup\limits_{y\in\Omega} f(x,y)$ is convex if $f(x,y)$ convex in x for $\forall y$.

3. composition: affine mapping / scalar / vector.

3.1. affine mapping: suppose $f: \mathbb{R}^n \to \mathbb{R}$. $A \in \mathbb{R}^{n\times m}$. $b \in \mathbb{R}^n$.

$g: \mathbb{R}^m \to \mathbb{R}$ $g(x) \triangleq f(Ax+b)$ convexity same as f.

Example. $f(x) = \|Ax - y\|$ is convex.

$f(x) = \log\left(\sum_{i=1}^n e^{w_i^T x + b_i}\right)$ $\begin{cases} g(y) = \log\left(\sum_{i=1}^n e^{y_i}\right). \\ y = (w_1, \cdots w_n)^T x + b. \end{cases}$

log-sum-exp.

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

$(n=1)$ $f'(x) = h'(g(x))\, g'(x) \Rightarrow f''(x) = h''(g(x))\, g'(x)^2 + h'(g(x))\, g''(x)$

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

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

Proof of case. 2: $g(\theta x + \bar\theta y) \geq \theta g(x) + \bar\theta g(y)$.

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

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

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

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

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

3.3. vector composition. $h: \mathbb{R}^k \to \mathbb{R}$. $g_i: \mathbb{R}^n \to \mathbb{R}$. $f(x) = h(g_1(x), \cdots g_k(x))$.

$(n=1)$. $f'(x) = \nabla h(g(x), g(x), \cdots, g(x))^T \cdot \mathbf{1} \cdot g'(x) \Rightarrow$

$f''(x) = \mathbf{1}^T \nabla^2 h(g(x), g(x), \cdots, g(x)) \cdot \mathbf{1} \cdot g'(x)^2 + \nabla h \cdot \mathbf{1} \cdot g''(x)$.

define $h: \mathbb{R}^k \to \mathbb{R}$ is increasing if $h(x) \geq h(y)$ $\forall x, y$ s.t. $\forall x_i \geq y_i$

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

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

4. minimization. f convex in (x, y). $C \neq \emptyset$ convex.

then $g(x) \triangleq \inf_{y \in C} f(x, y)$ is convex provided $g > -\infty$.

dom $g = \{x : \exists y \in C.\ \text{s.t. } (x, y) \in \operatorname{dom} f\}$. projection.

Proof: by verifying Jensen's inequality for $x_1, x_2 \in \operatorname{dom} g$.

Fix $\varepsilon > 0$. Then $\exists y_1 y_2$, s.t. $f(x_i, y_i) < g(x_i) + \varepsilon$. $\forall i$.

$$
\begin{aligned}
g(\theta x_1 + \bar\theta x_2) &= \inf_y f(\theta x_1 + \bar\theta x_2, y). \\
&\le f(\theta x_1 + \bar\theta x_2, \theta y_1 + \bar\theta y_2) \\
&\le \theta f(x_1, y_1) + \bar\theta f(x_2, y_2) \\
&\le \theta g(x_1) + \bar\theta g(x_2) + \varepsilon. \quad \forall \varepsilon > 0.
\end{aligned}
$$

$\Rightarrow g(\theta x_1 + \bar\theta x_2) \le \theta g(x_1) + \bar\theta g(x_2)$ $\square$.

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

epi(g) is a projection of another convex set on a convex set.

Example. distance to convex set. $\operatorname{dist}(x, C) = \inf_{y \in C} \|x - y\|$.

Example. geometric means. $(\prod x_i)^{1/n}$ concave if $x_i \ge 0$.

$$\frac{\partial^2 f}{\partial x_k^2} = -\frac{n-1}{n^2 x_k^2} (\prod x_i)^{1/n}. \qquad \frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{1}{n^2 x_k x_l} (\prod x_i)^{1/n}.$$

$\nabla^2 f = -\frac{f}{n^2} (n \operatorname{diag}\{q_1^2, \cdots, q_n^2\}) - qq^T$ where $q_i = 1/x_i$.

$v^T \nabla^2 f(x) v = -\frac{f}{n^2} (n \sum v_i^2/x_i^2 - (\sum v_i/x_i)^2) \le 0$ by Cauchy-Schwarz.