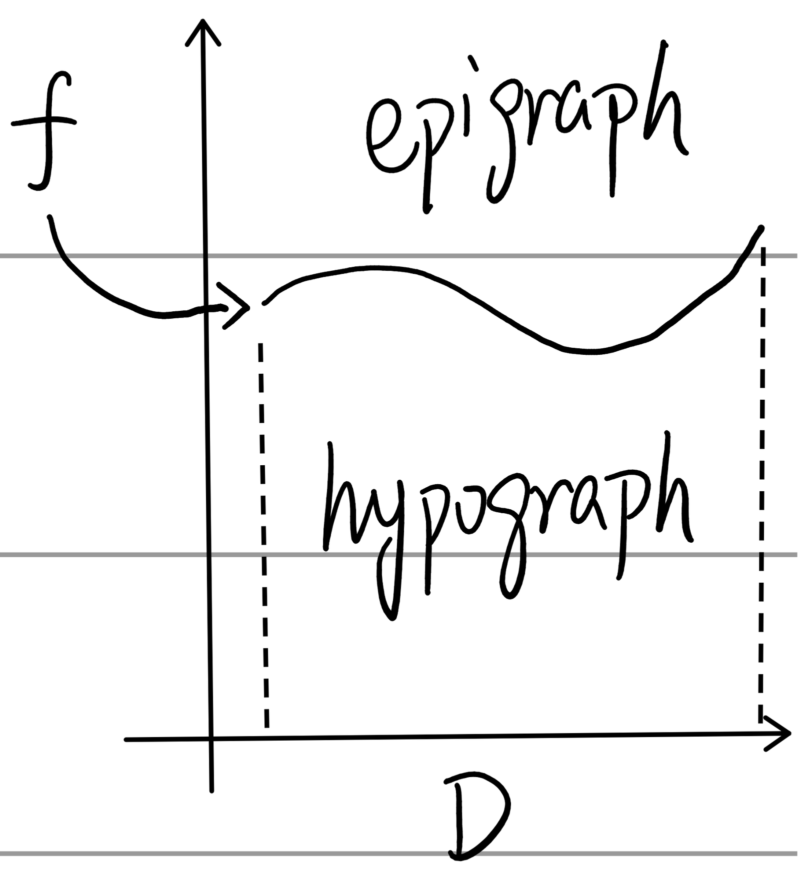
Lecture 6. Convex functions.

We now consider the convexity of functions. $f: D \subseteq \mathbb{R}^n \to \mathbb{R}$.

graph: $\{(\vec{x}, f(\vec{x})) \in \mathbb{R}^{n+1} : \vec{x} \in D\}$.

epigraph: $\operatorname{epi} f \triangleq \{(\vec{x}, y) \in \mathbb{R}^{n+1} : \vec{x} \in D,\ y \geq f(\vec{x})\}$

hypograph: $\operatorname{hyp} f \triangleq \{(\vec{x}, y) \in \mathbb{R}^{n+1} : \vec{x} \in D,\ y \leq f(\vec{x})\}$.

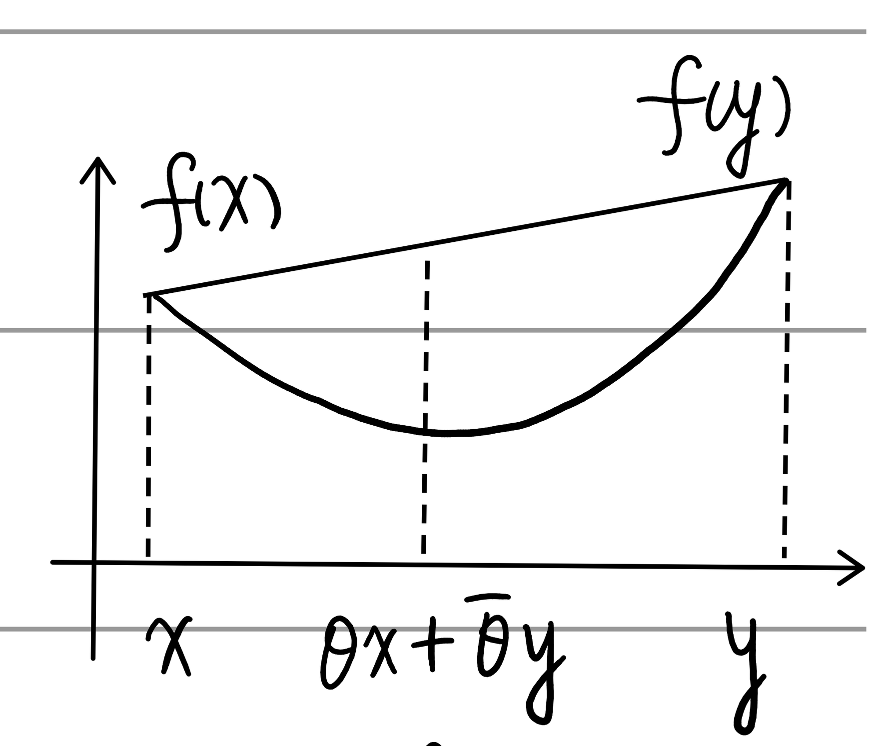


Definition: f is called a convex function iff $\operatorname{epi} f$ is a convex set.

Proposition 1. D is convex. D is a projection of a convex set $\operatorname{epi} f$.

Proposition 2. (Jensen's inequality). $\forall x, y \in D$.

$\forall \theta \in [0,1]$. $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$



$(x, f(x)), (y, f(y)) \in \operatorname{epi} f \Rightarrow (\theta x + \bar{\theta} y, f(\theta x + \bar{\theta} y)) \in \operatorname{epi} f$.

Theorem: f is convex iff Jensen's inequality holds for all $x, y \in D$.

Proof: "$\Leftarrow$". Given $(\vec{x}_1, x_2), (\vec{y}_1, y_2) \in \operatorname{epi} f$. let $z_i = \theta x_i + \bar{\theta} y_i$.

It suffices to show $(\vec{z}_1, z_2) \in \operatorname{epi} f$. Since $x_2 \geq f(\vec{x}_1)$, $y_2 \geq f(\vec{y}_1)$.

$z_2 = \theta x_2 + \bar{\theta} y_2 \geq \theta f(\vec{x}_1) + \bar{\theta} f(\vec{y}_1) \geq f(\theta \vec{x}_1 + \bar{\theta} \vec{y}_1) = f(\vec{z}_1)$ by

Jensen's inequality. So $\vec{z}_1 \in D$ and $z_2 \geq f(\vec{z}_1) \Rightarrow (\vec{z}_1, z_2) \in \operatorname{epi} f$. $\square$

Definition: f is strictly convex. if $\forall x \neq y \in D$. $\theta \in (0,1)$. strictly $<$.

Concavity: f is (strictly) concave, iff $-f$ is (strictly) convex.

Proposition: f is concave iff hyp f is convex.

Example: $f(x) = w^T x + b$ is convex and concave, but not strictly.

$f(x) = x^2$ is strictly convex. $f(x) = \frac{1}{x}$ is strictly convex on $\mathbb{R}_{>0}$.

Extended-value extension: for convenience, extend $\mathbb{R} \to \mathbb{R} \cup \{\infty\}$.

$\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} = \begin{cases} f(x), & x \in D \\ \infty & x \notin D. \end{cases}$ where $\begin{cases} x + \infty = \infty \\ 0 \cdot \infty = 0 \end{cases}$

Examples of convex functions: e^{ax}, $-\log x$, x^a if $a \geq 1$ or $a \leq 0$.

any norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is convex. (but not strictly, why?)

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

Verify convexity of $-\log x$: $\theta \log x + \bar{\theta} \log y \leq \log(\theta x + \bar{\theta} y) \Leftrightarrow$

$x^\theta y^{\bar{\theta}} \leq \theta x + \bar{\theta} y$. $\forall x, y > 0$. if $\theta = \frac{1}{2}$, easy. AM-GM inequality.

if $\theta \neq \frac{1}{2}$? Is it sufficient to verify only $\theta = \frac{1}{2}$?

Definition: f is midpoint convex if $\forall x, y \in D$, $f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}$.

Theorem (Jensen, 1905). If f is continuous and midpoint convex,

then f is convex. (o.w. $\exists$ counterexample, but not measurable).

Proof. By contradiction. Assume $\exists x, y, \theta \neq \frac{1}{2}$, $f(\theta x + \bar{\theta} y) > \theta f(x)$

$+\bar{\theta} f(y)$. Let $g(\alpha) = f(\alpha x + \bar{\alpha} y) - (\alpha f(x) + \bar{\alpha} f(y))$ $\alpha \in [0,1]$

$g(0) = g(1) = 0$. $g(\theta) > 0$. $\Rightarrow$ $M \triangleq \max\limits_{\alpha \in [0,1]} g(\alpha) > 0$. compactness.

Let $\alpha_0 = \inf \{\alpha : g(\alpha) = M\}$. Claim: $g(\alpha_0) = M$. by continuity. $\Rightarrow$

$\alpha_0 \neq 0, 1$. Choose δ sufficiently small s.t. $(\alpha_0 - \delta, \alpha_0 + \delta) \subseteq (0,1)$.

f is midpoint convex $\Rightarrow g(\alpha_0 - \delta) + g(\alpha_0 + \delta) \geq 2 g(\alpha_0)$. However.

$g(\alpha_0 - \delta) < M$. $g(\alpha_0) = M$. $g(\alpha_0 + \delta) \leq M$. contradiction. $\square$

Example: convexity of $f(X) = -\log \det(X)$. $X \in S_{++}^n$ (or $X \succ 0$).

Proof. Verify midpoint convexity: $\det\left(\frac{X+Y}{2}\right) \geq \det(X)^{1/2} \det(Y)^{1/2}$.

Since $X \succ 0$, X^{-1} exists and $\det(X^{-1}) > 0$. So our goal is equivalent

to $\det\left(\frac{I + X^{-1}Y}{2}\right) \geq \det(X^{-1}Y)^{1/2} \Longleftrightarrow \prod \left(\frac{1+\lambda_i(X^{-1}Y)}{2}\right) \geq \sqrt{\prod \lambda_i(X^{-1}Y)}$.

$\left(\det(\lambda I - X^{-1}Y) = 0 \Longleftrightarrow \det\left(\frac{\lambda I - X^{-1}Y}{2}\right) = 0 \Longleftrightarrow \det\left(\frac{1+\lambda}{2} I - \frac{I + X^{-1}Y}{2}\right) = 0\right)$

It suffices to show $\forall i, \lambda_i(X^{-1}Y) \geq 0$. $X \succ 0 \Rightarrow X = U \Lambda U^{-1} \Rightarrow$

$\exists X^{1/2} = U \Lambda^{1/2} U^{-1} \succ 0$, invertible. $\Rightarrow \lambda(X^{-1}Y) = \lambda(X^{1/2} X^{-1} Y X^{-1/2})$.

(if $X^{-1}Yv = \lambda v$. $X^{1/2} X^{-1} Y X^{-1/2} (X^{1/2} v) = X^{1/2}(X^{-1}Yv) = \lambda X^{1/2} v$).

$X^{1/2} X^{-1} Y X^{-1/2} = X^{-1/2} Y X^{-1/2}$ symmetric. $v^T X^{-1/2} Y X^{-1/2} v = u^T Y u > 0$.

where $u = X^{-1/2} v$, since $X^{-1/2}$ symmetric $\Rightarrow X^{-1/2} Y X^{-1/2} \succ 0$. $\square$

Criteria to verify convexity.

Theorem (Zeroth-order condition). (restriction to lines).

f is convex iff $\forall x \in \operatorname{dom} f$, $v \in \mathbb{R}^n$. $g(t) = f(x+tv)$ is convex.

Example: $f(x): \mathbb{R}^n \to \mathbb{R} = e^{\|x\|_1} = e^{x_1+x_2+\cdots+x_n}$ is convex.

$\forall u, v \in \mathbb{R}^n$. $g(t) = f(u+tv) = e^{u_1+\cdots+u_n} \cdot e^{(v_1+\cdots+v_n)t}$ is convex.

Proof. "$\Rightarrow$". assume f is convex. Fix $x \in \operatorname{dom} f$. $\forall v \in \mathbb{R}^n$. let t_1

$t_2 \in \operatorname{dom} g$. It suffices to show $\forall \theta \in [0,1]$, ① $\theta t_1 + \bar{\theta} t_2 \in \operatorname{dom} g$.

② $g(\theta t_1 + \bar{\theta} t_2) \le \theta g(t_1) + \bar{\theta} g(t_2)$. Let $x_i = x + t_i v$. $t_1, t_2 \in \operatorname{dom} g$

$\Rightarrow x_1, x_2 \in \operatorname{dom} f \Rightarrow \theta x_1 + \bar{\theta} x_2 = x + (\theta t_1 + \bar{\theta} t_2) v \in \operatorname{dom} f \Rightarrow$

$\theta t_1 + \bar{\theta} t_2 \in \operatorname{dom} g$. and $g(\theta t_1 + \bar{\theta} t_2) = f(\theta x_1 + \bar{\theta} x_2) \le \theta f(x_1) + \bar{\theta} f(x_2)$.

"$\Leftarrow$". assume g is always convex. Given $\forall x, y \in \operatorname{dom} f$. let $v = y - x$.

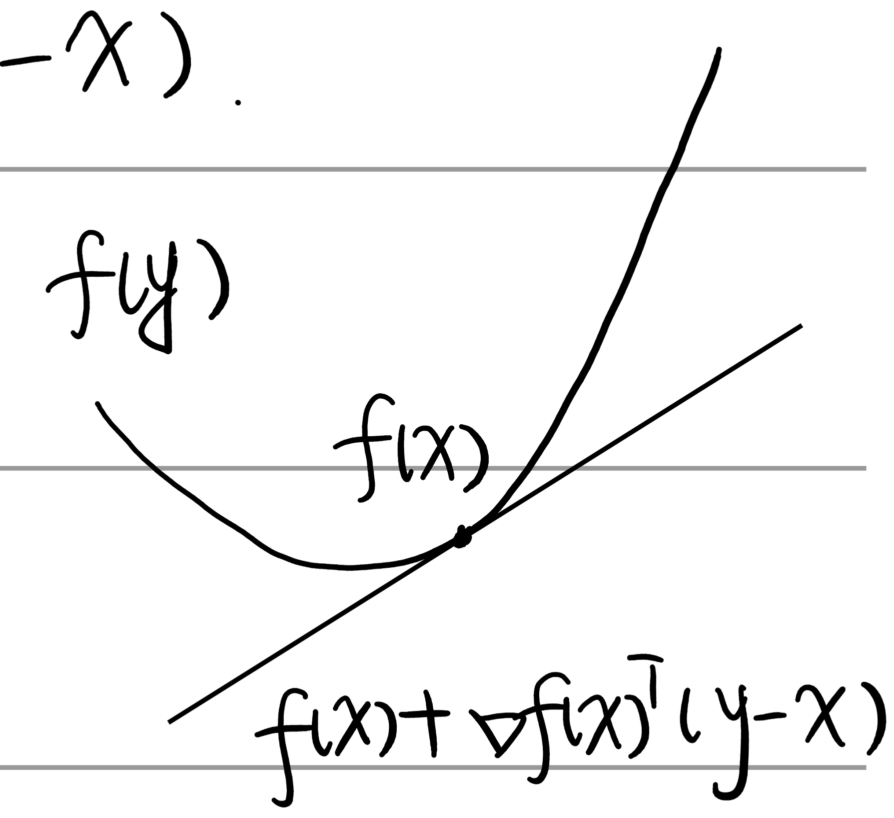
and $g(t) = f(x+tv)$. $x, y \in \operatorname{dom} f \Rightarrow 0, 1 \in \operatorname{dom} g \Rightarrow \forall \theta \in [0,1]$

$\theta \in \operatorname{dom} g \Rightarrow x + \theta(y-x) = \bar{\theta} x + \theta y \in \operatorname{dom} f$. In addition. $g(\theta)$

$= f(\bar{\theta} x + \theta y) \le \bar{\theta} g(0) + \theta g(1) = \bar{\theta} f(x) + \theta f(y) \Rightarrow f$ convex. $\square$

Theorem (First-order condition)

Suppose f is differentiable in an open convex set $\operatorname{dom} f$. Then f is convex iff $\forall x, y \in \operatorname{dom} f$, $f(y) \geq f(x) + \nabla f(x)^T (y-x)$. f is strictly convex iff $\forall x \neq y \in \operatorname{dom} f$, $f(y) > f(x) + \nabla f(x)^T (y-x)$.



Example. (Bernoulli's inequality).

$(1+x)^r \geq 1 + rx$ if $r \geq 1$, $x \geq -1$. $e^x \geq 1+x$.

Remark: The first order Taylor expansion is a global underestimator of a convex function, and vice versa. Local information $\Rightarrow$ global inequality. In particular, $\nabla f(x) = 0 \Rightarrow \forall y \in \operatorname{dom} f$, $f(y) \geq f(x)$.

$\{(\vec{x}, y) \in \mathbb{R}^{n+1} : y = f(\vec{x}_0) + \nabla f(\vec{x}_0)^T (\vec{x} - \vec{x}_0)\}$ supporting hyperplane of epi f.

Proof. "$\Rightarrow$". Fix any $x, y \in \operatorname{dom} f$. Let $v = y - x$. By Jensen's inequality, $\forall t \in [0,1]$, $f(x+tv) \leq (1-t) f(x) + t f(y)$. Rearranging we have $f(x+tv) - f(x) \leq t(f(y) - f(x))$. Recall $\nabla f(x)^T v = \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$

Taking $t \to 0$. $\nabla f(x)^T v \leq f(y) - f(x)$.

"$\Leftarrow$". Given x, y, θ. let $z = \theta x + \bar{\theta} y$. The first-order condition shows

$$\begin{cases} f(x) \geq f(z) + \nabla f(z)^T (x - z) \\ f(y) \geq f(z) + \nabla f(z)^T (y - z) \end{cases} \Rightarrow \theta f(x) + \bar{\theta} f(y) \geq f(z) = f(\theta x + \bar{\theta} y). \quad \square$$

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