# 3. Convex functions

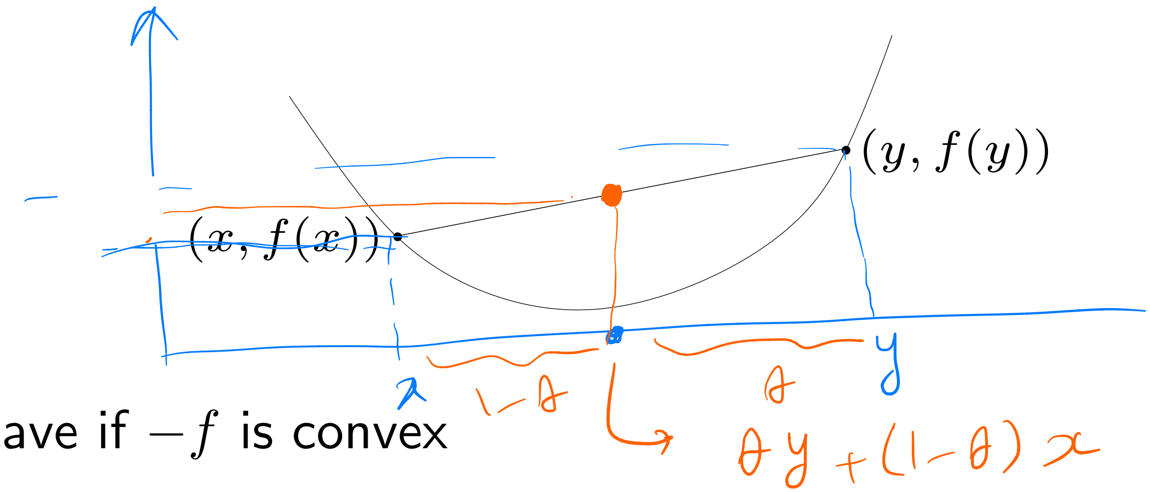
- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions
- log-concave and log-convex functions
- convexity with respect to generalized inequalities

# Definition

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

for all $x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$




- $f$ is concave if $-f$ is convex
- $f$ is strictly convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

for $x, y \in \mathbf{dom}\, f$, $x \ne y$, $0 < \theta < 1$

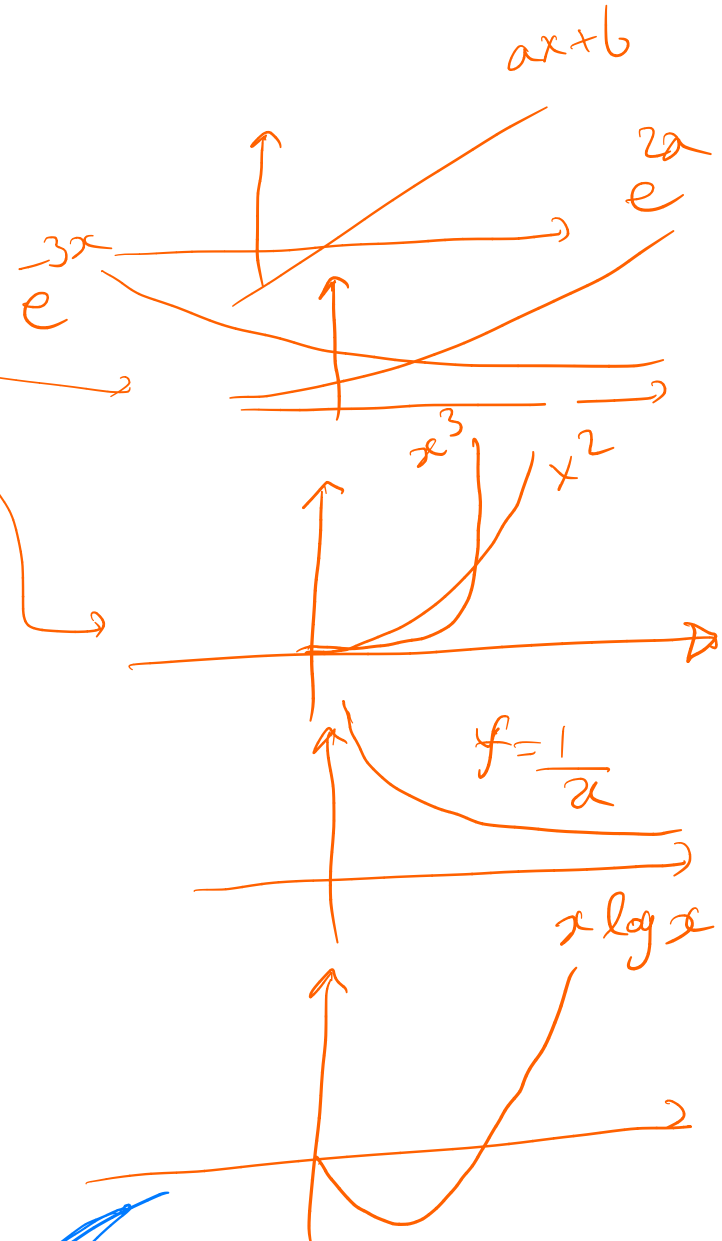
$\mathbf{R}_+ \{x \geq 0\}$

$\mathbf{R}_{++} \{x > 0\}$

# Examples on R

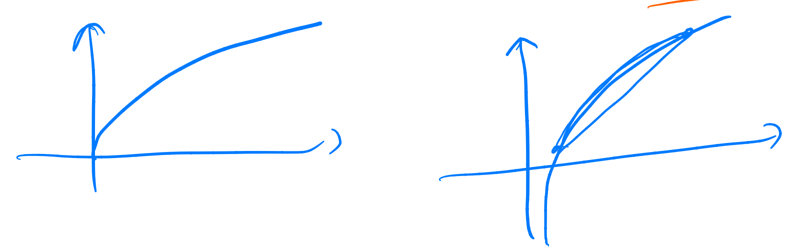
convex:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: $e^{ax}$, for any $a \in \mathbf{R}$
- powers: $x^\alpha$ on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

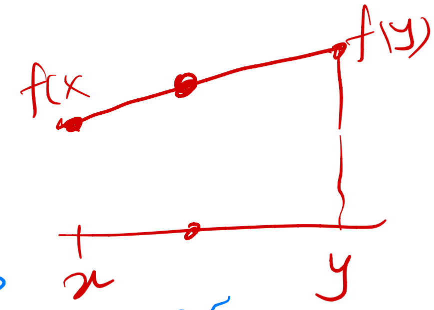



concave:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- powers: $x^\alpha$ on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$



# Examples on $\mathbf{R}^n$ and $\mathbf{R}^{m\times n}$




affine functions are convex and concave; all norms are convex

**examples on $\mathbf{R}^n$**

$\sum a_i x_i + b$

- affine function $f(x) = a^T x + b$

$f : \mathbb{R}^n \to \mathbb{R}$

- norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

**examples on $\mathbf{R}^{m\times n}$** ($m \times n$ matrices)

- affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

تابع آفین

$f(ax + (1-a)\,y)$ $a \in \mathbb{R}$

$= a f(x)$;

$+ (1-a) f(y)$

- spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

norm : $\mathbb{R}^n \longrightarrow \mathbb{R}_+$

$$\|cx\| = |c|\,\|x\|$$

$$\|x+y\| \le \|x\| + \|y\|$$

$$\|\quad\| : \mathbb{R}^n \longrightarrow \mathbb{R}$$

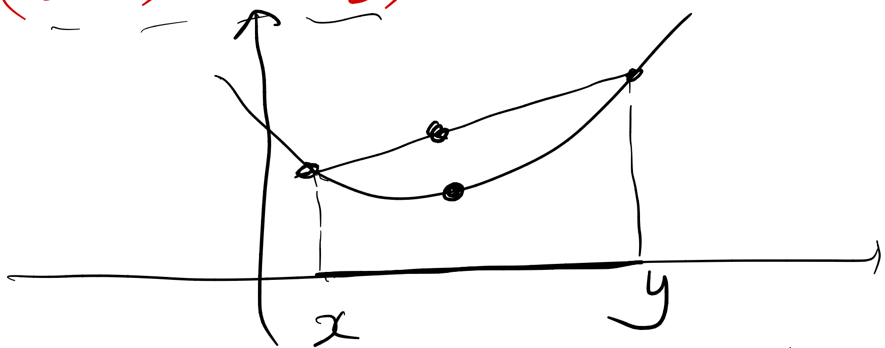
6 در 6 $\longrightarrow \mathbb{R}$

$$\sigma_{max}(Z) = \max_{u,v \,:\, \|u\| = \|v\| = 1} u^T Z v$$

$$\sigma_{max}(X+Y) = \max_{u,v} \underbrace{u^T (X+Y) v}_{u^T X v \,+\, u^T Y v}$$

$$\le \max_{u,v} u^T X v + \max_{u,v} u^T Y v$$

$$f((1-t)x+ty) \le (1-t)f(x)+tf(y)$$

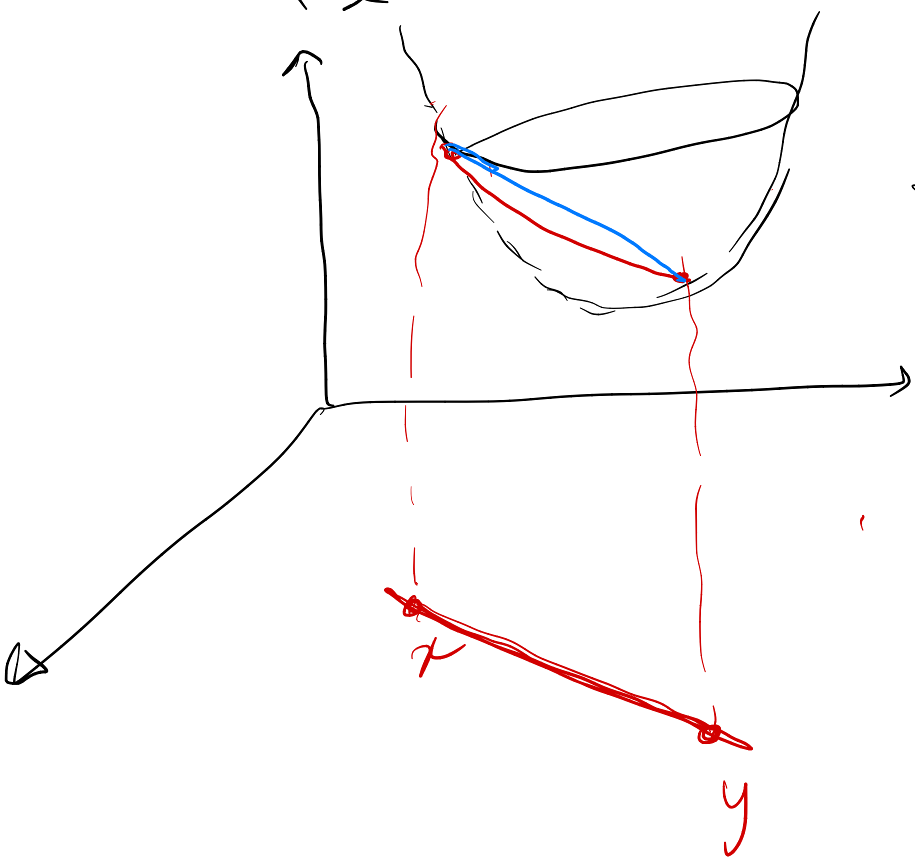



• توابع محدب •

تابعی محدب است اگر و فقط اگر

تحدید آن تابع به هر پاره خط

یک تابع محدب از یک بعد به یک بعد باشد




$$x(t) = (1-t)x + ty$$

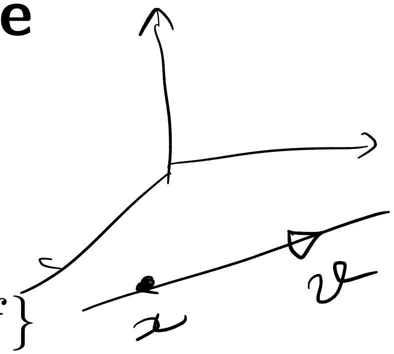
( نقاط روی پاره خط )

$$g(t) = f(x(t))$$

$$g : [0,1] \longrightarrow \mathbb{R}$$

$$\begin{aligned} g(t) &= f(x(t)) \\ &\le (1-t)f(x) + t f(y) \\ &= (1-t)\,g(0) + t\,g(1) \end{aligned}$$

# Restriction of a convex function to a line




$f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function $g : \mathbf{R} \to \mathbf{R}$,

$$g(t) = f(x + tv), \qquad \mathbf{dom}\, g = \{t \mid x + tv \in \mathbf{dom}\, f\}$$

is convex (in $t$) for any $x \in \mathbf{dom}\, f$, $v \in \mathbf{R}^n$

can check convexity of $f$ by checking convexity of functions of one variable

**example.** $f : \mathbf{S}^n_{++} \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\, f = \mathbf{S}^n_{++}$

ماتریس متقارن

$$\begin{aligned} g(t) = \log\det\underbrace{(X + tV)}_{X^{+1/2}(I+\ \ )X^{+1/2}} &= \log\det X + \log\det(I + tX^{-1/2}VX^{-1/2}) \\ &= \log\det X + \sum_{i=1}^{n} \log(1 + t\lambda_i) \end{aligned}$$

$\lambda_i(X^{-1/2}VX^{-1/2})$

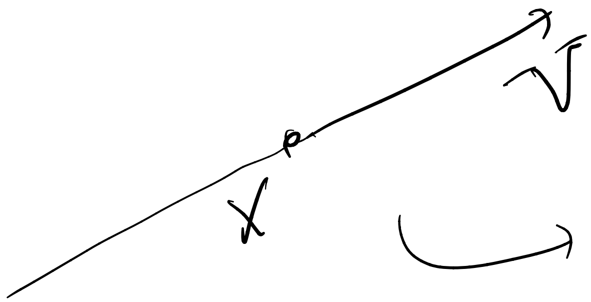
$\boxed{\det(AB) = \det(A)\det(B)}$

where $\lambda_i$ are the eigenvalues of $X^{-1/2}VX^{-1/2}$

$g$ is concave in $t$ (for any choice of $X \succ 0$, $V$); hence $f$ is concave

$1 + 2t > 0 \to t > -\frac{1}{2}$

$1 - 3t > 0 \to t < \frac{1}{3}$



$X \in S_{++}$

$V \in S$

$g(t) = f(X + tV)$

نسبت به t ~~تعریف~~ است !!

$X + tV$

قطعا [illegible] +

$u^T X u > 0$

$u^T X u + t u^T V u$

---

جذر ماتریس :

$A \in S_+$: $A = \sum \lambda_i u_i u_i^T = U^T \Lambda U$

$$A^2 = \sum \lambda_i^2 u_i u_i^T = (U^T \Lambda \underbrace{U)(U^T}_{I} \Lambda U)$$

$$= U^T \Lambda^2 U$$

$$\rightarrow A^{1/2} = \sum \sqrt{\lambda_i}\, u_i u_i^T = U^T \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} U$$

$$\log\det A = \sum_{i=1}^{n} \log \lambda_i(A)$$

---

$$I + X$$

$X$ [illegible]

$$\lambda_i(I+X) = 1 + \lambda_i(X)$$

$A, B$

$AB$

---

$$X + tV \longleftrightarrow I + t X^{-1/2} V X^{-1/2}$$

$$u^T(X+tV)\,u = \underbrace{u^T X^{1/2}}_{w^T} \underbrace{(I + \cdots)}\, \underbrace{X^{1/2} u}_{w}$$

$$u^T (I + \cdots)\, u = u^T X^{-1/2} (X+tV) X^{-1/2} u$$

---

$A, B \in S_{++}$

$AB \notin S_+$

$A^{1/2} B A^{1/2} \in S_+$

$f(x)$

# Extended-value extension

extended-value extension $\tilde{f}$ of $f$ is

$$\tilde{f}(x) = f(x), \quad x \in \mathbf{dom}\, f, \qquad \tilde{f}(x) = \infty, \quad x \notin \mathbf{dom}\, f$$

often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad \tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$$

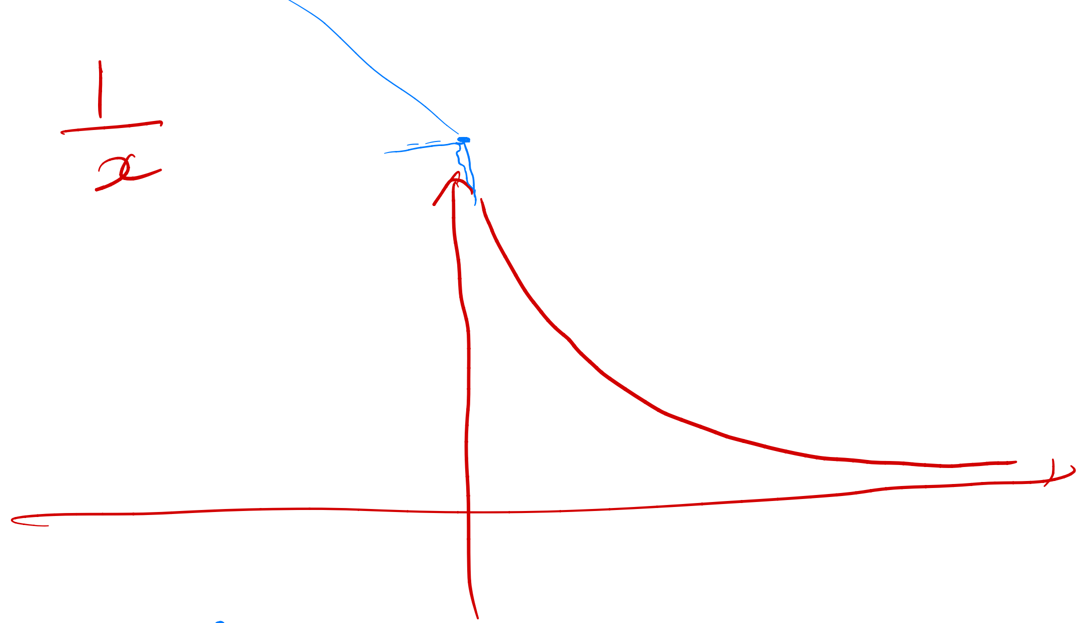
(as an inequality in $\mathbf{R} \cup \{\infty\}$), means the same as the two conditions

- $\mathbf{dom}\, f$ is convex
- for $x, y \in \mathbf{dom}\, f$,

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

$$f(x) = \frac{1}{x}$$

$$f : \mathbb{R}_+ \longrightarrow \mathbb{R}$$



$$\tilde{f} : \mathbb{R} \longrightarrow \mathbb{R} \cup \{\infty\}$$

$\tilde{f}$ روی $\mathbb{R}$ تعریف باشد و

$$\tilde{f}\big|_{\mathbb{R}_+} = f$$

$$\tilde{f}(x) = \infty \qquad x \leq 0$$

# First-order condition

$f$ is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$$

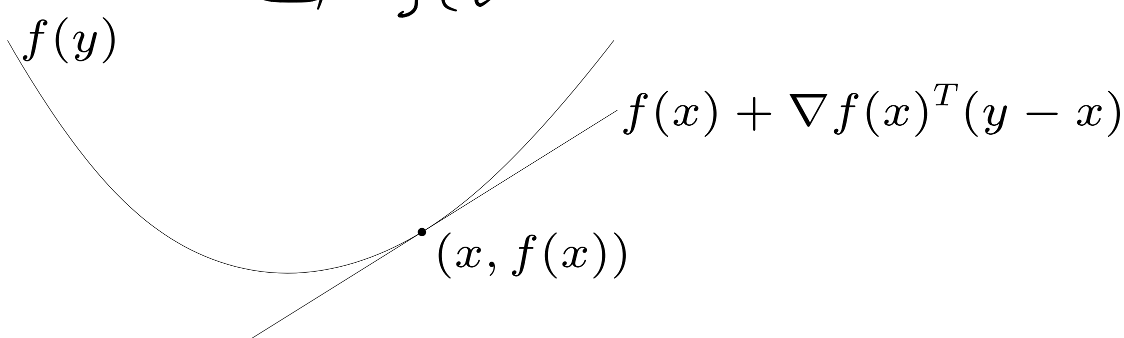
exists at each $x \in \mathbf{dom}\, f$

**1st-order condition:** differentiable $f$ with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$

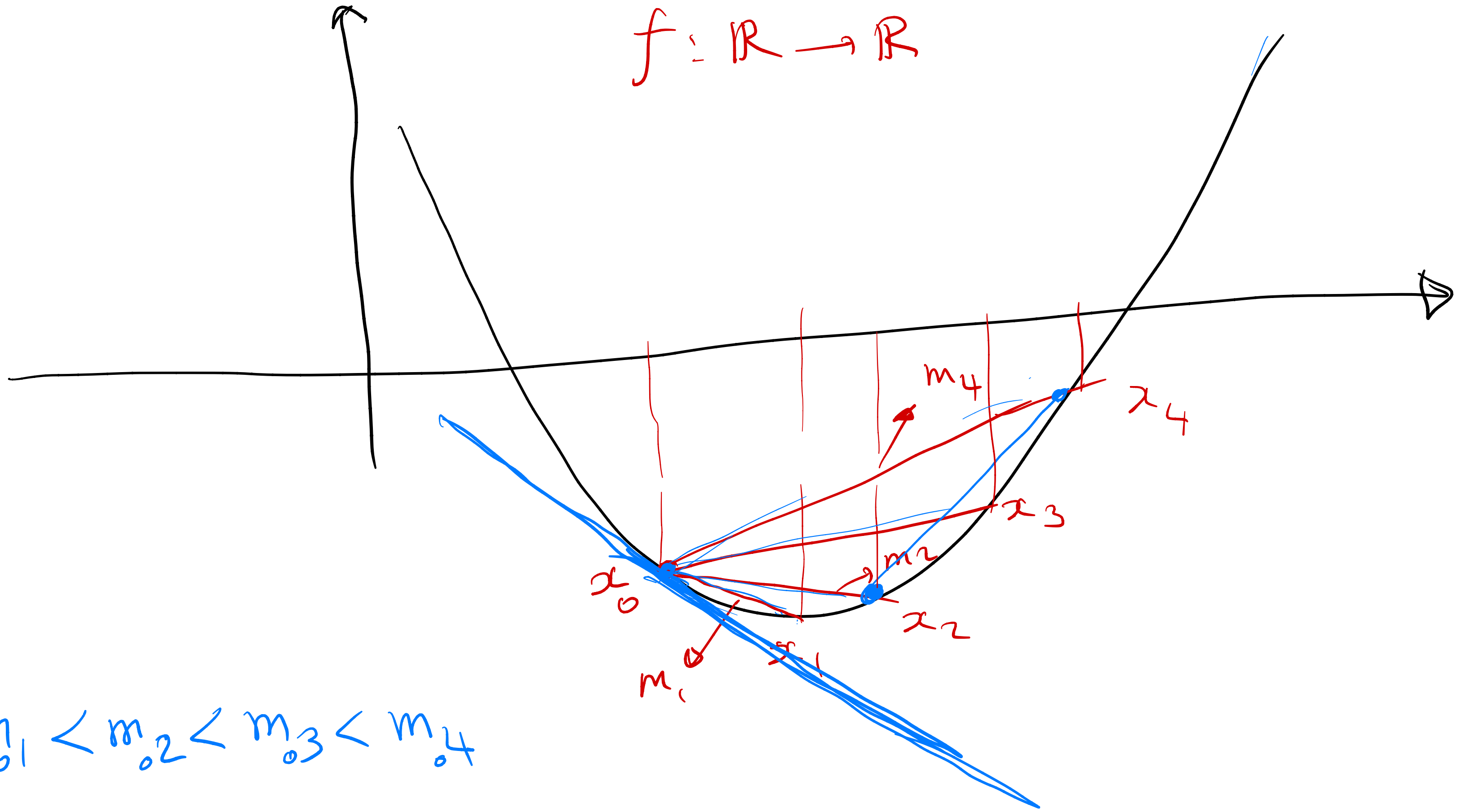
$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

$\hookrightarrow$ gll




first-order approximation of $f$ is global underestimator

$f : \mathbb{R} \to \mathbb{R}$




$m_{01} < m_{02} < m_{03} < m_{04}$

$m_{02} < m_{24}$

$\forall \quad x_0 < x_1 < x_2$

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq \frac{f(x_2) - f(x_0)}{x_2 - x_0} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

$$\frac{f(x_1)-f(x_0)}{x_1-x_0} \le \frac{f(x_2)-f(x_0)}{x_2-x_0}$$

$$\xrightarrow{x_1 \to x_0} \quad f'(x_0) \le \frac{f(x_2)-f(x_0)}{x_2-x_0}$$

$$f'(x_0)(x_2-x_0)+f(x_0) \le f(x_2)$$

---

$$g(t) = f(\vec{x}+t(\vec{y}-\vec{x})) \; ; \; t \in \mathbb{R}$$

خط گذرنده از x و y ؛ تحدید تابع f روی خط

$$g'(t) = \left(\nabla f(x+t(y-x))\right)^T \cdot (y-x)$$

$$= \sum \left(\frac{\partial f}{\partial x_i}\right)(y_i-x_i)$$

---

چون g محدب است ،

$$\underbrace{g(1)}_{f(y)} \ge \underbrace{g(0)}_{f(x)} + \underbrace{g'(0)\times 1}_{\nabla f(x)^T(y-x)\cdot 1}$$

# Second-order conditions

$f$ is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n,$$

$$\begin{array}{cc} & j \\ i & \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_i \partial x_j} \end{pmatrix} \end{array}$$

exists at each $x \in \mathbf{dom}\, f$

**2nd-order conditions:** for twice differentiable $f$ with convex domain

- $f$ is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then $f$ is strictly convex

$$g(t) = f(x + t(y-x))$$

$$g''(t) \geq 0 \quad ?$$

$$g'(t) = \left(\nabla f(x + t\,\underset{(y-x)}{\cancel{v}})\right)^T \underbrace{(y-x)}_{v} = \langle \nabla f(x+tv), v \rangle$$

$$g''(0) = \lim_{\varepsilon \downarrow 0} \frac{\langle \nabla f(x+\varepsilon v), v\rangle - \langle \nabla f(x), v\rangle}{\varepsilon}$$

$$= \Big\langle \underbrace{\frac{d}{d\varepsilon} \nabla f(x+\varepsilon v)}_{\underbrace{\nabla^2 f(x)}_{n\times n \text{ ماتریس}} \cdot \, v}, v \Big\rangle \Big|_{\varepsilon=0}$$

$$= \underbrace{\langle \nabla^2 f(x)\, v, v \rangle}_{v^T \nabla^2 f\, v} \geq 0$$

ماتریس هسین مثبت نیمه معین $\Rightarrow$ محدب

اگر مثبت معین ، اکیدا محدب .

# Examples

**quadratic function:** $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

**least-squares objective**: $f(x) = \frac{1}{2}\|Ax - b\|_2^2 \;=\; \langle Ax-b, Ax-b \rangle$

$$\nabla f(x) = \cancel{2}A^T(Ax - b), \qquad \nabla^2 f(x) = \cancel{2}A^T A$$

convex (for any $A$)

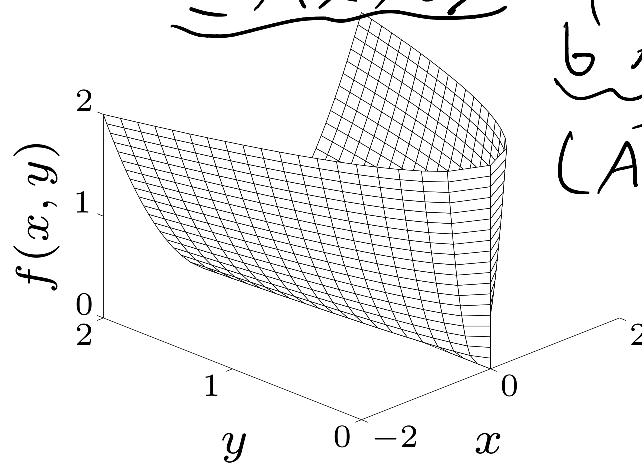
$$= \frac{1}{2}x^T A^T A x + (b^T b)/2 - \underbrace{\langle Ax, b \rangle}$$

$$\underbrace{b^T A}_{(A^T b)} x$$

**quadratic-over-linear:** $f(x, y) = x^2/y$

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

convex for $y > 0$



$$x^T P x = \sum_{i,j} x_i P_{ij} x_j$$

$$\rightarrow \frac{\partial^2 x^T P x}{\partial x_i \partial x_j} = 2 P_{ij} \longrightarrow \nabla^2 x^T P x = 2P$$

$$\frac{\partial x^T P x}{\partial x_i} = 2 \sum_j P_{ij} x_j = 2 \left( \text{[illegible]} \; P \right) \vec{x}$$

$$\rightarrow \left( \quad \right) = 2 \left( P \; \text{[illegible]} \right) \vec{x}$$

$x_1, \ldots, x_n$

**log-sum-exp**: $f(x) = \log \sum_{k=1}^n \exp x_k$ is convex

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z} \mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \qquad (z_k = \exp x_k)$$

to show $\nabla^2 f(x) \succeq 0$, we must verify that $v^T \nabla^2 f(x) v \geq 0$ for all $v$:

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$ (from Cauchy-Schwarz inequality)

**geometric mean**: $f(x) = (\prod_{k=1}^n x_k)^{1/n}$ on $\mathbf{R}_{++}^n$ is concave

(similar proof as for log-sum-exp)

$$f(x) = \log \sum_{k=1}^{n} e^{x_k}$$

$$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{\sum e^{x_k}}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = -\frac{e^{x_i} e^{x_j}}{(\sum \quad)^2} \qquad j \neq i$$

$$\frac{\partial^2 f}{\partial x_i^2} = -\frac{e^{2x_i}}{(\sum \quad)^2} + \frac{e^{x_i}}{\sum} \qquad j = i$$

$$P_i = \frac{e^{x_i}}{\sum e^{x_k}} \quad ; \quad \begin{cases} \dfrac{\partial^2 f}{\partial x_i \partial x_j} = -P_i P_j \\[2ex] \dfrac{\partial^2 f}{\partial x_i^2} = -P_i^2 + P_i \end{cases}$$

$$\nabla^2 f = -\begin{pmatrix} p_1^2 & p_1 p_2 & \cdots & p_1 p_n \\ p_1 p_2 & p_2^2 & \cdots & p_2 p_n \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} + \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$$

$$= -\underbrace{\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}}_{p} \underbrace{(p_1 \cdots p_n)}_{p^T} + \underbrace{\begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}}_{\mathrm{diag}(p_1, \dots, p_n)}$$

---

$$v^T \nabla^2 f(x)\, v \overset{?}{\geq} 0$$

چک کردن شرط ؟

$$v^T \nabla^2 v = -\underbrace{v^T p\, p^T v}_{(\sum p_i v_i)(\sum p_i v_i)} + \sum p_i v_i^2$$

$$= \underbrace{\sum p_i v_i^2}_{E[V^2]} - \Big(\underbrace{\sum p_i v_i}_{E[V]}\Big)^2 \overset{?}{\geq} 0$$

$$\Big(\sum p_i v_i^2\Big)\Big(\sum p_i\Big) \geq_{C.S.} \cdots$$

$$f(x) = \log \sum e^{x_i}$$

• تابعی است که هموار . مشتق از هر مرتبه‌ای دارد

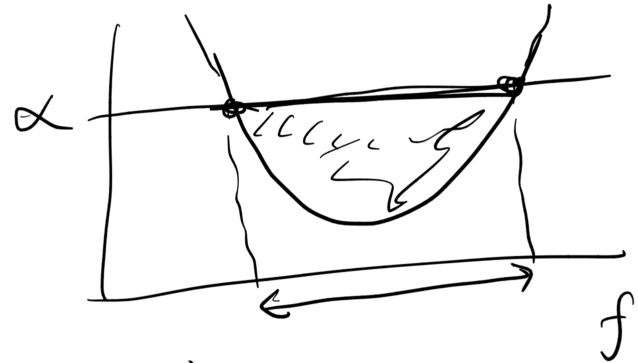
$$\max(x_1, \dots, x_n) \leq \log \underline{\sum e^{x_i}}$$

$$\leq n \max(e^{x_1}, \dots, e^{x_n})$$

$$= \log n + \underbrace{\log \max(e^{x_1}, \dots, e^{x_n})}_{\max(x_1, \dots, x_n)}$$

# Epigraph and sublevel set

$\alpha$-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

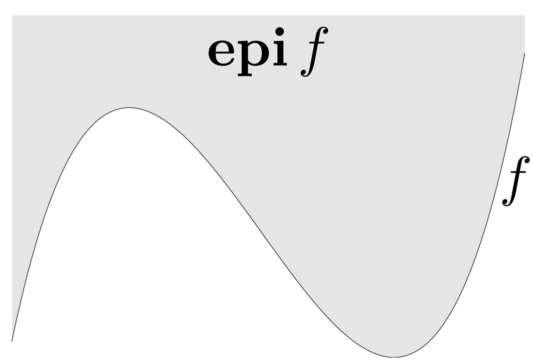



$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

sublevel sets of convex functions are convex (converse is false)

**epigraph** of $f : \mathbf{R}^n \to \mathbf{R}$:

$$\mathbf{epi}\, f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\, f,\ f(x) \leq t\}$$




$f$ is convex if and only if $\mathbf{epi}\, f$ is a convex set

- $f, g$ ; $h = \max(f, g)$ .

$$\operatorname{epi}(h) = \left\{ (x, t) : \quad t \geq h(x) \iff \begin{matrix} t > f(x) \\ t > g(x) \end{matrix} \right\}$$

$$= \operatorname{epi}(f) \cap \operatorname{epi}(g)$$

- اگر $f$ و $g$ محدب باشد آنگاه $\max(f, g)$ محدب است.

# Jensen's inequality

**basic inequality:** if $f$ is convex, then for $0 \le \theta \le 1$,

$$\underbrace{f(\theta x + (1-\theta)y)}_{E z} \le \underbrace{\theta f(x) + (1-\theta) f(y)}_{E f(z)}$$

$$z = \begin{cases} x & p = \theta \\ y & 1-\theta \text{ احتمال} \end{cases}$$

**extension:** if $f$ is convex, then

$$f(\mathbf{E}\, z) \le \mathbf{E}\, f(z)$$

for any random variable $z$

اگر f مشتق پذیر باشد

$$f(z) \ge f(x) + \nabla f(x)^T (z - x), \quad x \to E z$$

$$E f(z) \ge E(\text{کل عبارات}) = \cancel{E} f(E z) + E\left[\nabla f(E z)^T \cdot (z - E z)\right]$$

$$\cancel{E\left[\nabla f(E z)^T \cdot (z - E z)\right]} = 0$$

basic inequality is special case with discrete distribution

$$\mathbf{prob}(z = x) = \theta, \qquad \mathbf{prob}(z = y) = 1 - \theta$$

# Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

3. show that $f$ is obtained from simple convex functions by operations that preserve convexity

   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective

# Positive weighted sum & composition with affine function

**nonnegative multiple:** $\alpha f$ is convex if $f$ is convex, $\alpha \geq 0$ $\left\{ \alpha < 0 \rightarrow \text{مقعر} \right.$

**sum:** $f_1 + f_2$ convex if $f_1, f_2$ convex (extends to infinite sums, integrals)

**composition with affine function**: $\underbrace{f(Ax+b)}_{}$ $g(x) =$ $f(Ax + b)$ is convex if $f$ is convex

$E\,g(X) > g(EX)$ ?

**examples**

$E\,f(\underbrace{Ax+b}_{U}) > f(E\,U)$ محدب $f$

$= f(E[Ax+b]) = f(A\,EX + b)$

$= g(EX)$

- log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \qquad \mathbf{dom}\, f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$$

- (any) norm of affine function: $f(x) = \|Ax + b\|$

• مروری بر روش‌های بررسی تحدب

* (۱) استفاده از تعریف $\theta$ و $(1-\theta)$ یا معادلاً استفاده از $E$

• کاربرد: جهت بررسی قضایا (هیچ اطلاعی از خود تابع به جز تحدب)

* شرط‌های مشتق

• کاربرد: یک تابع خاص که مشتق آن محاسبه‌پذیر است.

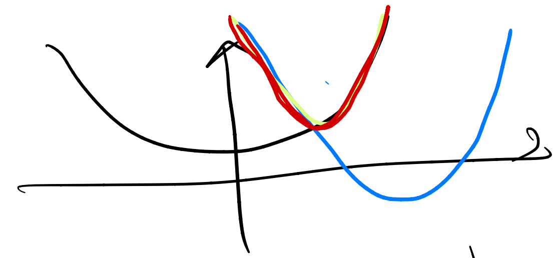
$f'' \succ 0$ ; $\nabla^2 f \succeq 0$ .

* تحدید تابع به روی پاره‌خط یا خط

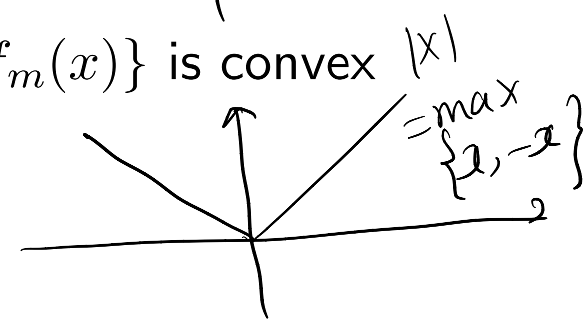
• کاربرد: بررسی تحدب توابعی از ماتریس‌ها (نیمه‌معین)

$g(t)=f(X+tV)$

# Pointwise maximum



if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex




## examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ is convex
- sum of $r$ largest components of $x \in \mathbf{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

  is convex ($x_{[i]}$ is $i$th largest component of $x$)

  proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n\}$$

# Pointwise supremum

$f_1, \ldots, f_n$ محدب

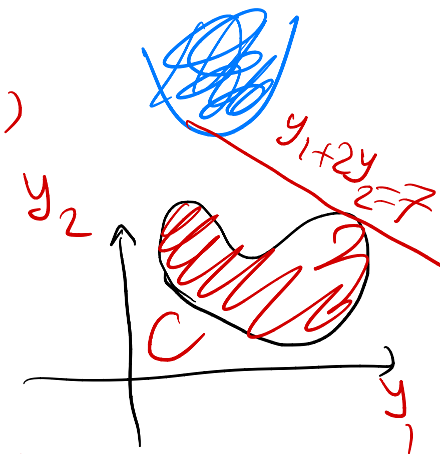
$\max\{\underbrace{f_1(x)}_{f(x,1)}, \ldots, \underbrace{f_n(x)}_{f(x,n)}\}$ محدب

if $f(x, y)$ is convex in $x$ for each $y \in \mathcal{A}$, then

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

is convex

$$\mathrm{Epi}_x(g) = \bigcap_y \mathrm{Epi}_x f(x, y)$$




**examples**

- support function of a set $C$: $S_C(x) = \sup_{y \in C} y^T x$ is convex

$x = (1, 2)$

$$\sup_{y \in C} y_1 + 2y_2$$

- distance to farthest point in a set $C$:


$$f(x) = \sup_{y \in C} \|x - y\|$$

- maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n$,

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} \underbrace{y^T X y}_{f(X, y)}$$

$$g(\theta x + (1-\theta)\tilde{x}) \overset{?}{\leq} \theta g(x) + (1-\theta) g(\tilde{x})$$

$$\sup_y f(\theta x + (1-\theta)\tilde{x}, y) \overset{?}{\leq} \theta \sup_y f(x,y) + (1-\theta) \sup_y f(\tilde{x}, y)$$

$$f(\theta x + (1-\theta)\tilde{x}, y) < \theta f(x,y) + (1-\theta) f(\tilde{x}, y) \quad \forall x, \tilde{x}, y$$

$$\sup_y f(\qquad) < \sup_y \left( \theta f(x,y) + (1-\theta) f(\tilde{x}, y) \right)$$

$$\boxed{\begin{aligned} &\sup_z f(z) + g(z) \\ &< \sup_z f(z) + \sup_z g(z) \end{aligned}}$$

$$< \theta \sup_y f(x,y) + (1-\theta) \sup_y f(\tilde{x}, y)$$

# Composition with scalar functions

composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$:

$$f(x) = h(g(x))$$

$f$ is convex if $\begin{array}{l} g \text{ convex, } h \text{ convex, } \tilde{h} \text{ nondecreasing} \\ g \text{ concave, } h \text{ convex, } \tilde{h} \text{ nonincreasing} \end{array}$

مثال نقض وقتی $h$ صعودی نباشد

$g(x) = \cos x$ on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

$h(x) = x^2$

$f = h(g) = \cos^2 x$

محدب نیست!

- proof (for $n = 1$, differentiable $g, h$)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

- note: monotonicity must hold for extended-value extension $\tilde{h}$

**examples**

- $\exp g(x)$ is convex if $g$ is convex

$h = e^x$ : محدب صعودی

و $g$

$h(g(x)) = e^{g(x)}$

- $1/g(x)$ is convex if $g$ is concave and positive

$h = \frac{1}{x}$ $\quad x > 0$

$$g(x_1,x_2) = -\log x_1 - \log x_2 \quad ; \; x_1,x_2 > 0$$

$$= -\log x_1 x_2 = \log \frac{1}{x_1 x_2}$$

$$e^{g(x_1,x_2)} = \frac{1}{x_1 x_2} \longrightarrow$$ تابعی محدب از $x_1,x_2$

$$g(x_1,\dots,x_n) = \frac{1}{x_1^{\alpha_1}\cdots\cdots x_n^{\alpha_n}} \quad ; \; \alpha_1,\dots,\alpha_n > 0$$

تابع محدبی از $X_1,\dots,x_n$ می‌شود.

---

$h(g(X))$ محدب ✓ : $E f(X) \overset{?}{\geq} f(EX) \longrightarrow$ ✓

$f(X) =$

$$E\, h(g(X)) \geq h(\underbrace{E\, g(X)}_{\geq g(EX)})$$ محدب $h$

صعودی بودن $h$ $\longrightarrow$ $E\, h(g(X)) \geq h(g(EX))$

# Vector composition

$g$

composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

$f$ is convex if $\begin{array}{l} g_i \text{ convex, } h \text{ convex, } \tilde{h} \text{ nondecreasing in each argument} \\ g_i \text{ concave, } h \text{ convex, } \tilde{h} \text{ nonincreasing in each argument} \end{array}$ $h(y_1, \ldots, y_k)$

proof (for $n = 1$, differentiable $g, h$)

$g : \mathbb{R} \to \mathbb{R}^k \; ; \; h : \mathbb{R}^k \to \mathbb{R}$

$g = (g_1, \ldots, g_k)$

$$f''(x) = \underbrace{g'(x)^T \nabla^2 h(g(x)) g'(x)}_{\geq 0} + \nabla h(g(x))^T g''(x)$$

$g' = \begin{pmatrix} g_1' \\ \vdots \\ g_k' \end{pmatrix}$

$\begin{pmatrix} \frac{\partial h}{\partial y_1}(g(x)) \\ \frac{\partial h}{\partial y_2}(g(x)) \end{pmatrix}$ $\begin{pmatrix} g_1'' \geq 0 \\ \vdots \\ g_n'' \geq 0 \end{pmatrix}$

**examples**

- $\sum_{i=1}^m \log g_i(x)$ is concave if $g_i$ are concave and positive
- $\log \sum_{i=1}^m \exp g_i(x)$ is convex if $g_i$ are convex

$f(x_1, \ldots, x_n) = \log \sum_{i=1}^n e^{x_i}$ محدب

$g_i(x)$

$$g_i(\vec{x}) = -\log x_i \longrightarrow$$ محدب

$$\log \sum e^{g_i(x)} = \log \sum_{i=1}^{n} \frac{1}{x_i}$$ محدب ✓

---

محاسبه مشتق و $f''$:

$$f(x+\varepsilon) = h\big(g_1(x+\varepsilon), \dots, g_k(x+\varepsilon)\big)$$

$$\approx f(x) + \varepsilon f'(x)$$

$$\approx h\big(\underbrace{g_1(x)}_{u_1} + \varepsilon \underbrace{g_1'(x)}_{v_1}, \dots, \underbrace{g_k(x)}_{u_k} + \varepsilon \underbrace{g_k'(x)}_{v_k}\big)$$

$$\boxed{\begin{aligned} &h(\vec{u} + \varepsilon \vec{v}) \\ &\approx h(u) \\ &+ \varepsilon v^T \nabla h(u) \end{aligned}}$$

$$\approx \underbrace{h(g_1(x), \dots, g_k(x))}_{f(x)} + \varepsilon \Big\langle \underbrace{\begin{pmatrix} g_1'(x) \\ \vdots \\ g_k'(x) \end{pmatrix}}_{g'(x)}, \nabla h(g(x)) \Big\rangle$$

$$\longrightarrow f'(x) = g'(x)^T \nabla h(g(x)) = \langle g'(x), \nabla h(g(x)) \rangle$$

$$f''(x) = \langle g''(x), \nabla h(g(x)) \rangle + \langle g'(x), (\nabla h(g(x)))' \rangle$$

$$(\nabla h(g(x)))' \overset{\text{قاعده زنجیره}}{=\!=} \nabla^2 h(g(x)) \cdot g'(x)$$

# Minimization

f روی x برای هر y محدب

$\sup_{y} f(x,y)$

if $f(x,y)$ is convex in $(x,y)$ and $C$ is a convex set, then

$$g(x) = \inf_{y \in C} f(x,y)$$

is convex

**examples**

- $f(x,y) = x^T A x + 2x^T B y + y^T C y$ with $\nabla_y f = 2Cy + 2B^T x = 0$ → $y = -C^{-1} B^T x$

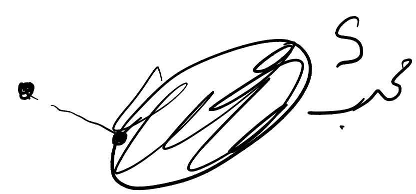
  A و C متقارن هستند

  $$D = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \qquad C \succ 0$$

  minimizing over $y$ gives $g(x) = \inf_y f(x,y) = x^T(A - BC^{-1}B^T)x$

  $g$ is convex, hence Schur complement $A - BC^{-1}B^T \succeq 0$

- distance to a set: $\mathbf{dist}(x,S) = \inf_{y \in S} \underbrace{\|x - y\|}_{f(x,y)}$ is convex if $S$ is convex



$$f(x,y) = \begin{bmatrix} \vec{x}^T & \vec{y}^T \end{bmatrix} D \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}$$

تابع محدب از $x,y$ است چون $D \succ 0$ .

---

$$g(x) = \min_{y \in \mathcal{C}} f(x,y)$$

$$g(\theta x + (1-\theta)\tilde{x}) \overset{?}{\leq} \theta g(x) + (1-\theta) g(\tilde{x})$$

$$g(x) = f(x, y_0) \;;$$ $y_0$ مینیمم کننده $f(x,y)$

$$g(\tilde{x}) = f(\tilde{x}, \tilde{y}_0)$$ $\tilde{y}_0$ — — $f(\tilde{x},y)$

$$\underline{\underline{\theta}}\, f(x,y_0) + \underline{\underline{\bar{\theta}}}\, f(\tilde{x},\tilde{y}_0) \geq f(\theta x + \bar{\theta}\tilde{x}, \theta y_0 + \bar{\theta}\tilde{y}_0)$$

$f$ نسبت به جفت $(x,y)$ محدب ✓

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

# Perspective

the **perspective** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$,

$++$

$$g(x, t) = t f(x/t), \qquad \mathbf{dom}\, g = \{(x, t) \mid x/t \in \mathbf{dom}\, f,\ t > 0\}$$

$g$ is convex if $f$ is convex

$g(x,t) = t \cdot \left(\frac{x}{t}\right)^T \left(\frac{x}{t}\right)$

**examples**

- $f(x) = x^T x$ is convex; hence $g(x, t) = x^T x/t$ is convex for $t > 0$

  $n = 1$

  $g(x,t) = \frac{x^2}{t}$

- negative logarithm $f(x) = -\log x$ is convex; hence relative entropy $g(x, t) = t \log t - t \log x$ is convex on $\mathbf{R}^2_{++}$

  $f(x) = -\log x \longrightarrow$

  $g(x,t) = t\left(-\log \frac{x}{t}\right)$

- if $f$ is convex, then

$$g(x) = (c^T x + d) f\left((Ax + b)/(c^T x + d)\right)$$

is convex on $\{x \mid c^T x + d > 0,\ (Ax + b)/(c^T x + d) \in \mathbf{dom}\, f\}$

یادآوری: نگاشت پرسپکتیو $\quad t>0;\ \vec{x}\in\mathbb{R}^n$

$$P(\vec{x},t)=\frac{\vec{x}}{t}$$

$$\underbrace{\mathcal{C}}_{\text{محدب}}\subseteq\mathbb{R}^n\times\mathbb{R}_+\xrightarrow{P}\underbrace{P(\mathcal{C})}_{\text{محدب}}$$

$$\underbrace{\mathcal{C}}_{\text{محدب}}\subseteq\mathbb{R}^n\xrightarrow{P^{-1}}\underbrace{P^{-1}(\mathcal{C})}_{\text{محدب}}=\left\{(\alpha\vec{x},\alpha):\ \begin{array}{l}\alpha>0\\ x\in\mathcal{C}\end{array}\right\}\subseteq\mathbb{R}^n\times\mathbb{R}_+$$

---

فرض: $f:\mathbb{R}^n\to\mathbb{R}$ محدب است $\Rightarrow$ $\underbrace{Epi(f)}_{\text{محدب}}=\{(\vec{x},\beta):\beta\geq f(x)\}\subseteq\mathbb{R}^n\times\mathbb{R}$

$$\Rightarrow P^{-1}(Epi(f))=\{(\alpha\vec{x},\alpha\beta,\alpha):\alpha>0;\ (\vec{x},\beta)\in Epi(f)\}$$

زیرمجموعه محدب از $\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}_+$ است.

$$\xrightarrow[\text{درایه ۲ و ۳}]{\text{جابه‌جا کردن}}\{(\alpha\vec{x},\alpha,\alpha\beta):\alpha>0;\ (\vec{x},\beta)\in Epi(f)\}$$

زیرمجموعه محدب از $\mathbb{R}^n\times\mathbb{R}_+\times\mathbb{R}$ است.

$$\begin{pmatrix} \vec{\tilde{x}} \\ t \\ \tilde{\beta} \end{pmatrix} = \begin{pmatrix} \alpha \vec{x} \\ \alpha \\ \alpha\beta \end{pmatrix}$$

قرار دهید

$$\Longrightarrow \left\{ (\vec{\tilde{x}}, t, \tilde{\beta}) \;:\; t > 0 \;;\; \left(\frac{1}{t}\tilde{x}, \frac{1}{t}\tilde{\beta}\right) \in Epi(f) \right\}$$

ولی

$$\left(\frac{1}{t}\tilde{x}, \frac{1}{t}\tilde{\beta}\right) \in Epi(f)$$

$$\Longleftrightarrow \frac{1}{t}\tilde{\beta} \geq f\left(\frac{1}{t}\vec{\tilde{x}}\right)$$

$$\Longleftrightarrow \tilde{\beta} \geq t\, f\left(\frac{\vec{\tilde{x}}}{t}\right) = g(\vec{\tilde{x}}, t)$$

$$\Longleftrightarrow (\vec{\tilde{x}}, t, \tilde{\beta}) \in Epi\left(\underbrace{t f\left(\frac{\vec{\tilde{x}}}{t}\right)}_{= g(\tilde{x}, t)}\right)$$

$$= Epi(g)$$

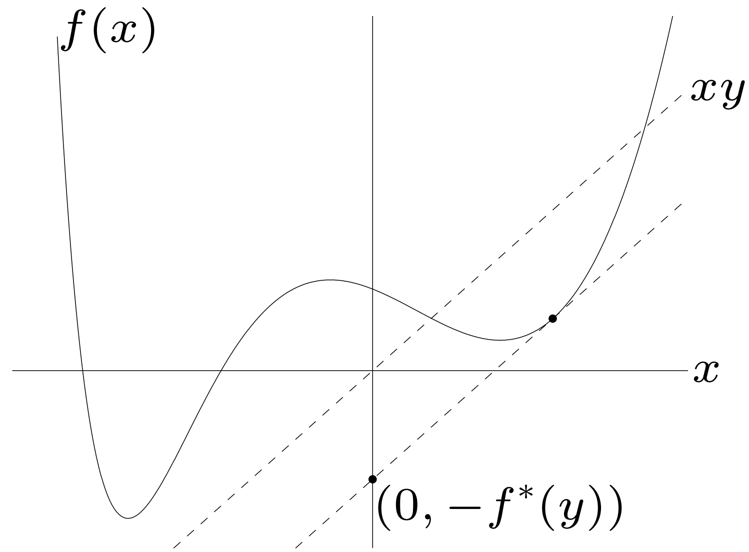
# The conjugate function

the **conjugate** of a function $f$ is

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$

$f^* : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$




- $f^*$ is convex (even if $f$ is not)
- will be useful in chapter 5

## examples

- negative logarithm $f(x) = -\log x$

$$
\begin{aligned}
f^*(y) &= \sup_{x>0}(xy + \log x) \\
&= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases}
\end{aligned}
$$

- strictly convex quadratic $f(x) = (1/2)x^TQx$ with $Q \in \mathbf{S}^n_{++}$

$$
\begin{aligned}
f^*(y) &= \sup_{x}(y^Tx - (1/2)x^TQx) \to y - Qx = 0 \\
& \qquad \to x = Q^{-1}y \\
&= \frac{1}{2}y^TQ^{-1}y
\end{aligned}
$$

$$\to (f^*)^*(z) = \frac{1}{2}z^TQz = f(z)$$

# Convexity with respect to generalized inequalities

$f : \mathbf{R}^n \to \mathbf{R}^m$ is $K$-convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta) f(y)$$

for $x$, $y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$

**example** $f : \mathbf{S}^m \to \mathbf{S}^m$, $f(X) = X^2$ is $\mathbf{S}^m_+$-convex

proof: for fixed $z \in \mathbf{R}^m$, $z^T X^2 z = \|Xz\|_2^2$ is convex in $X$, *i.e.*,

$$z^T(\theta X + (1-\theta)Y)^2 z \le \theta z^T X^2 z + (1-\theta) z^T Y^2 z$$

for $X, Y \in \mathbf{S}^m$, $0 \le \theta \le 1$

therefore $(\theta X + (1-\theta)Y)^2 \preceq \theta X^2 + (1-\theta)Y^2$