3. Convex functions

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

(domain)

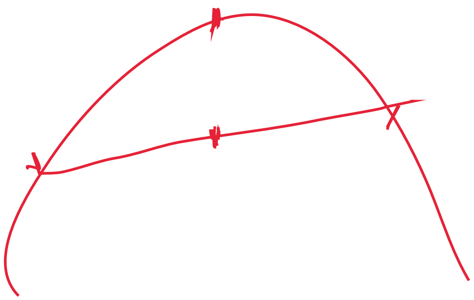
Definition

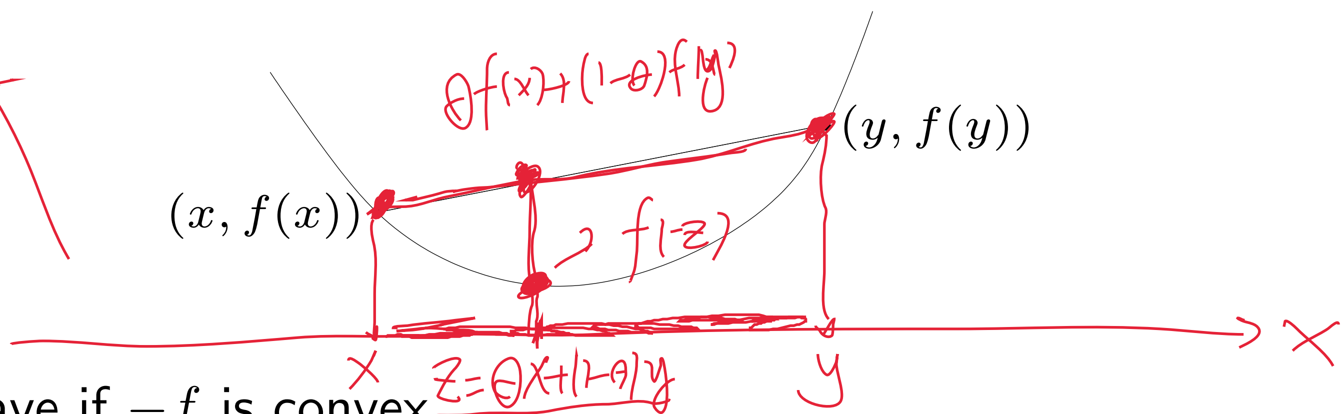
$x \in R^n$, $f(x) \in R$.

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

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

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

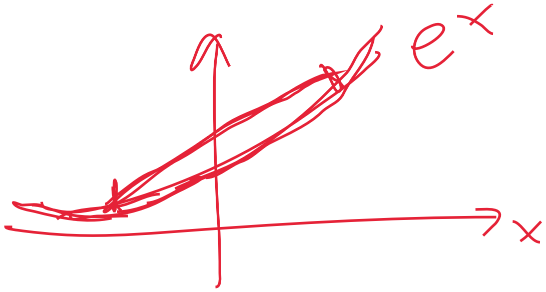




- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

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



$f: \mathbf{R} \to \mathbf{R}, \quad x \in \mathbf{R}$

Examples on R

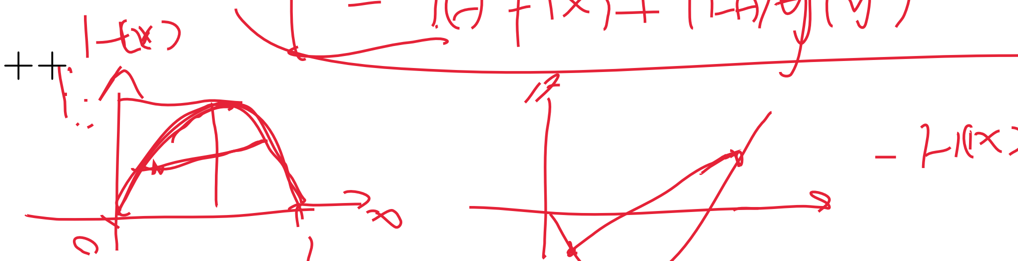
$$\left(\begin{array}{l} f(x) = ax + b \\ f(y) = ay + b. \end{array} \right)$$

$$\begin{aligned} f(\theta x + (1-\theta) y) &= a(\theta x + (1-\theta) y) + b \\ &= \theta(ax + b) + (1-\theta)(ay + b) \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

convex:

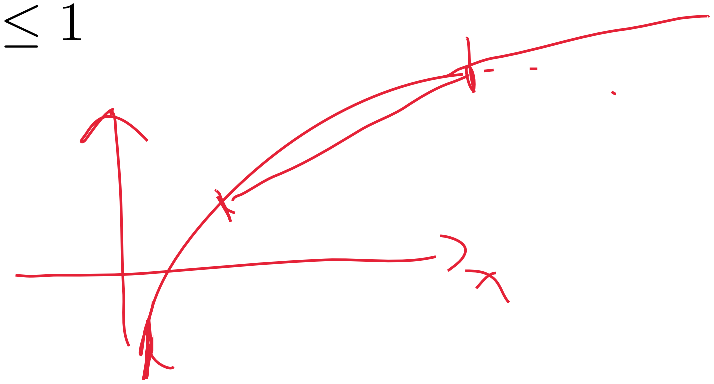
- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

$H(x) = x \log \frac{1}{x}$



concave:

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



$f: R^n \to R,$ or $R^{m\times n} \to R.$

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

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$ $f(x) = \langle a, x\rangle + b$

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

$\|x\|_2$

$\|X\|_F$

- affine function $f(X) = \langle A, X\rangle + B$

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

- spectral (maximum singular value) norm

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

$f: R^n \to R$

First-order condition

$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial \cdot}{\partial} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$

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

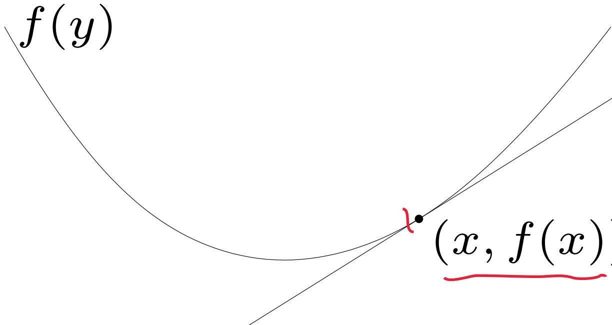
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T \in R^n$$

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

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

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

1-st order Taylor expansion



$= g(y)$

$(y \in \mathrm{dom} f)$

first-order approximation of f is global underestimator

$f: \mathbb{R}^n \to \mathbb{R}$. $\nabla f(x) \in \mathbb{R}^n$ | $\nabla^2 f = \frac{\partial \nabla f(x)}{\partial x} = \frac{\partial [\,]}{\partial [\,]} = \boxed{\cdot}$ Hessian.

Second-order conditions

$\nabla^2 f(x)$

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

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

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

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

$\|A\cancel{x} + b\|_2^2$

- f is convex if and only if

$\nabla^2 f(x) \in S^n_+$

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

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

Examples

$R \to R$

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

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

convex if $P \succeq 0$ $\quad$ PSD, $P \in S^n_+$ $\quad$ $\|a\|_2^2 = \langle a, a \rangle$

least-squares objective: $f(x) = \|Ax - b\|_2^2$ $= \langle Ax-b, Ax-b \rangle$

$= \langle Ax, Ax \rangle - 2\langle Ax, b \rangle + \langle b, b \rangle$

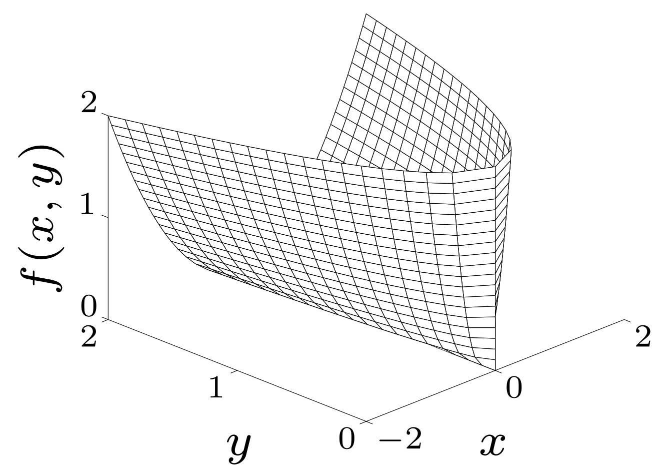
$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A \in S^n_+$$

convex (for any A) $\quad$ $f(x) = \|Ax-b\|_2^2 + \lambda \|x\|_2^2$.

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 $y > 0$



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)

$|\langle a, b \rangle| \leq \|a\|_2 \|b\|_2$

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)

Epigraph and sublevel set

Graph: $\{(x, f(x)) \mid x \in \text{Dom} f\}$

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

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

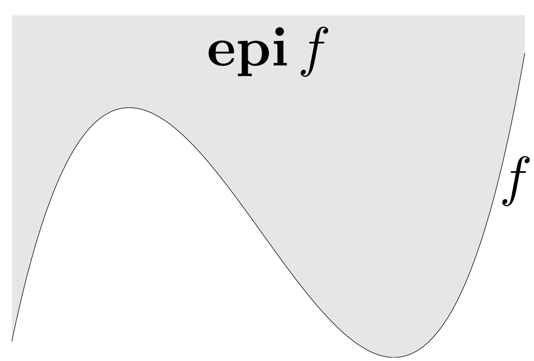
sublevel sets of convex functions are convex (converse is false)

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

epi: above.

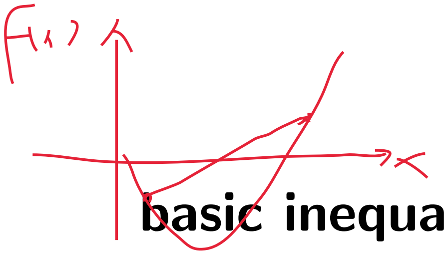
hypo: below

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

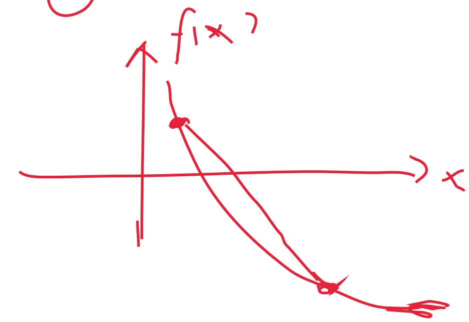


f is convex if and only if $\mathbf{epi}\, f$ is a convex set

$f(x) = x\log(x)$, $H(x) = x\log\frac{1}{x}$ convex $f(x) = -\log x$, $(x>0)$



Jensen's inequality



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

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

$-\log\left(\frac{1}{2}x+\frac{1}{2}y\right) \leq -\frac{1}{2}\log x - \frac{1}{2}\log y$.

$\boxed{\Rightarrow \sqrt{xy} \leq \frac{x+y}{2}}$

extension: if f is convex, then

$y = -f$: concave

$g(\mathbf{E}z) \geq \mathbf{E}g(z)$

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

$\left(x^{\theta} y^{1-\theta} \leq \theta x + (1-\theta) y \right)$

Hölder's inequality.

for any random variable z

$|\langle x, y\rangle| \leq \|x\|_p \|y\|_q$

$\left(\frac{1}{p}+\frac{1}{q}=1\right)$

basic inequality is special case with discrete distribution

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

$(p, q > 1)$.

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

affine, exp
negative entropy
negative log
...

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

Positive weighted sum & composition with affine function

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

$\|Ax-b\|_2^2 + \lambda \|x\|_2^2$

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

composition with affine function: $f(Ax + b)$ is convex if f is convex

$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$ $(p \geq 1)$

$\|x+y\| \leq \|x\| + \|y\|$

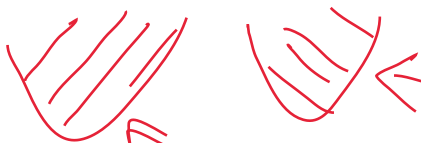
examples $\left(f(x) = -\log x \right)$

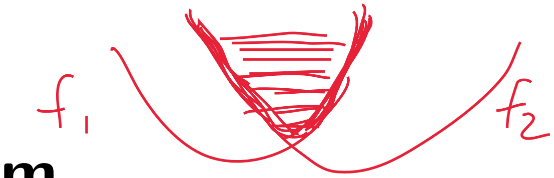
- log barrier for linear inequalities

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

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

$f(x) = \|x\|$





Pointwise maximum

f_1, f_2 $\quad$ epi f_1, epi f_2 convex set, $\quad f(x) = \max\{f_1, f_2\}$

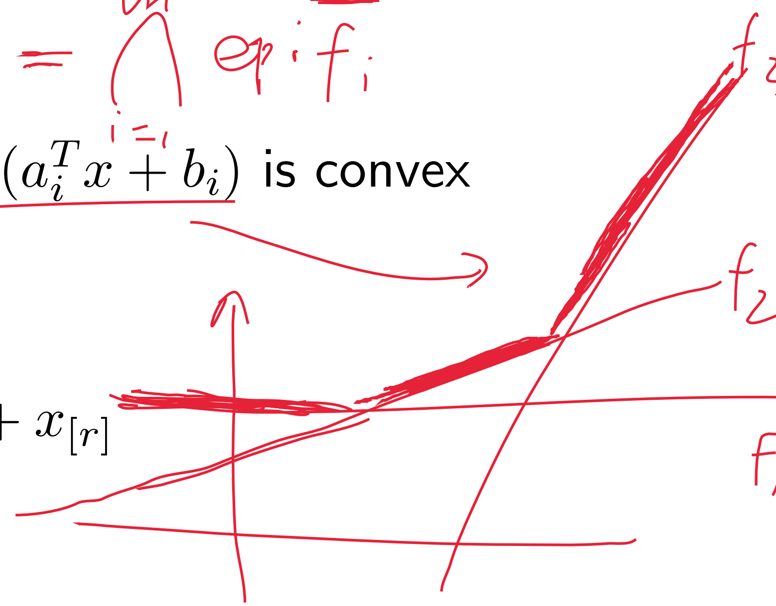
if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

$$\text{epi} f = \text{epi} f_1 \cap \text{epi} f_2 = \bigcap_{i=1}^{m} \text{epi} f_i$$

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 ith 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

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

(supreme ← maximum
infimum ← minimum)

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

is convex

(epi g = $\bigcap_{y \in A}$ epi f(·, y))

examples

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

f(x, y)

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

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

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y = \sum_{i,j} X_{ij} y_i \cdot y_j$$

f(x, y)

linear in X

Composition with scalar functions

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

$$f(x) = h(g(x)) = h \circ g\,(x)$$

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

$f'(x) = h'(g(x)) \cdot g'(x)$

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

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

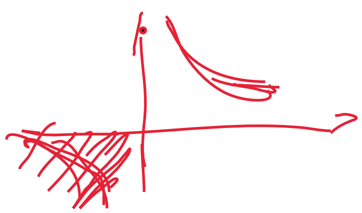
$$f''(x) = \underbrace{h''(g(x))}_{\geq 0}\underbrace{g'(x)^2}_{\geq 0} + \underbrace{h'(g(x))}_{\geq 0 \;/\; \leq 0}\underbrace{g''(x)}_{\geq 0 \;/\; \leq 0}$$

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

examples

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

$h(x) = \frac{1}{x}$, $g(x)$

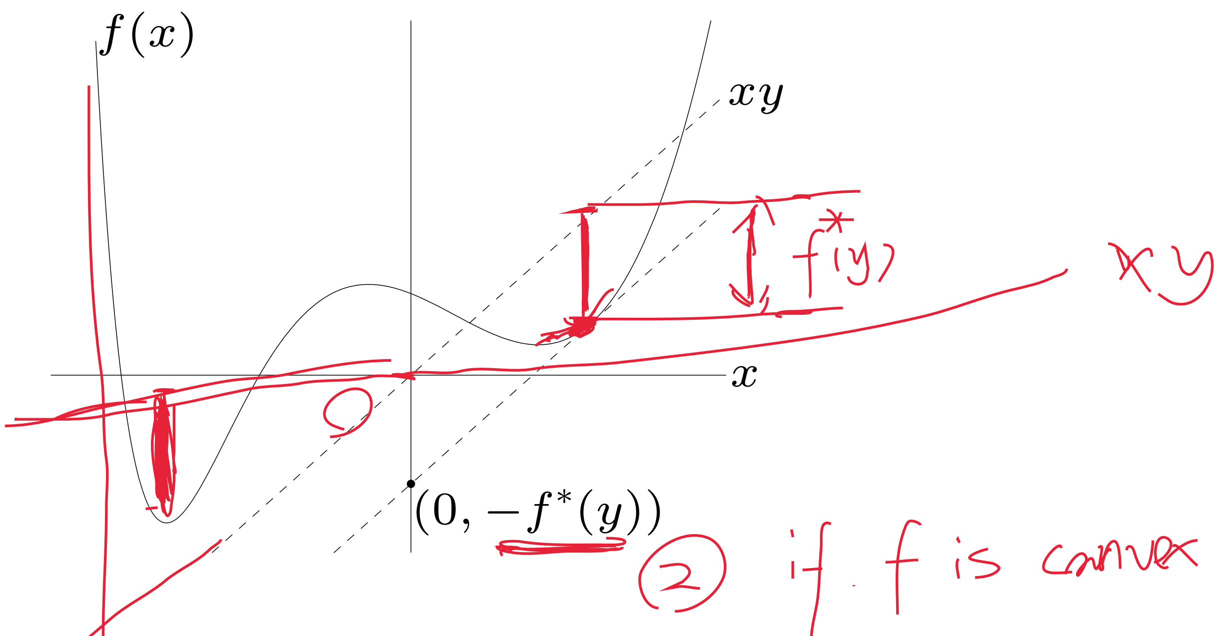


The conjugate function

the **conjugate** of a function f is

maximum gap between y^Tx and $f(x)$

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



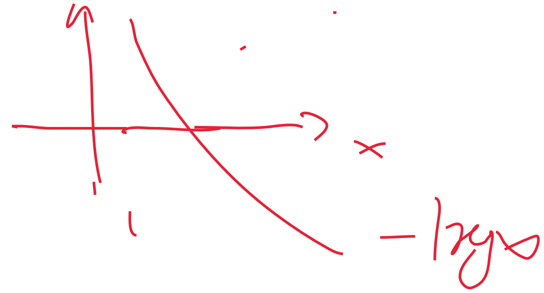
② if f is convex

$(f^{**} = f)$

① • f^* is convex (even if f is not)

• will be useful in chapter 5

examples



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

$f^*(y) = \sup(xy - f(x))$, $x \in \operatorname{dom} f$

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

concave in x

$y + \frac{1}{x} = 0$

$\Rightarrow x = -\frac{1}{y} > 0$

$f''(x) = Q \succeq 0$

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

$$
\begin{aligned}
f^*(y) &= \sup_x (y^Tx - (1/2)x^TQx) \\
&= \frac{1}{2}y^TQ^{-1}y
\end{aligned}
$$

concave in x

$y - Qx = 0$

Log-concave and log-convex functions

a positive function f is log-concave if $\log f$ is concave:

log f(x) is concave

$$\log f(\theta x + (1-\theta)y) \geq \log f(x)^\theta f(y)^{1-\theta} \quad \text{for } 0 \leq \theta \leq 1$$

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

f is log-convex if $\log f$ is convex

- powers: x^a on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
- many common probability densities are log-concave, *e.g.*, normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

- cumulative Gaussian distribution function Φ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

Properties of log-concave functions

$(\nabla^2 f(x) \succeq 0) \iff f(x) = \text{convex}$

- twice differentiable f with convex domain is log-concave if and only if

$$f(x)\nabla^2 f(x) \preceq \nabla f(x)\nabla f(x)^T$$

MLE

for all $x \in \mathbf{dom}\, f$

~~log~~ likelihood

log-concave

- product of log-concave functions is log-concave

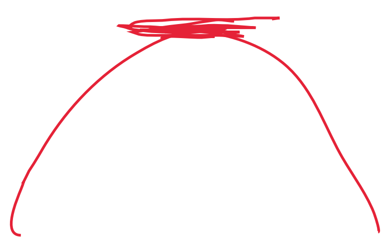
$$p(x) \overset{\text{i.i.d}}{=} \prod_{i=1}^{n} p(x_i)$$

- sum of log-concave functions is not always log-concave

log-concave

- integration: if $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)\, dy$$



is log-concave (not easy to show)

$f: R^n \to R$. $f(\theta x+(1-\theta)y) \le \theta f(x)+(1-\theta)f(y)$: f: convex

Convexity with respect to generalized inequalities

K: proper cone.

solid:
closed:
pointed

$f : \mathbf{R}^n \to \mathbf{R}^m$ 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 $x,\, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$

$K = R^n_+ \leftarrow$ vector
$= S^n_+ \leftarrow$ matrix

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