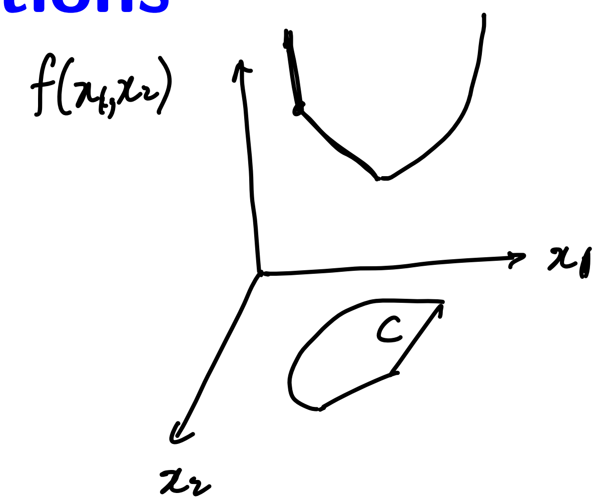
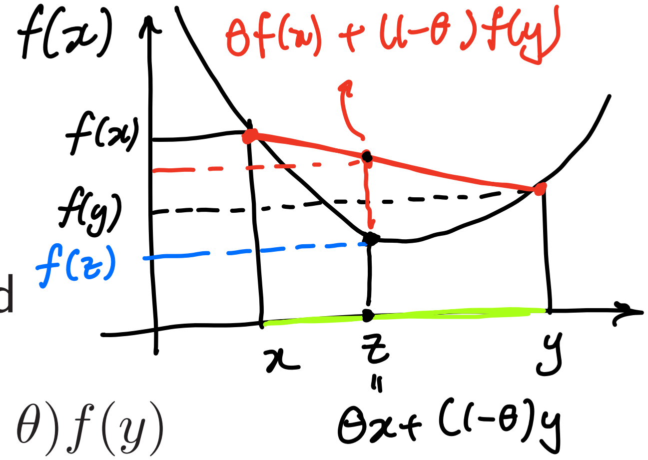
3. Convex functions



- 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

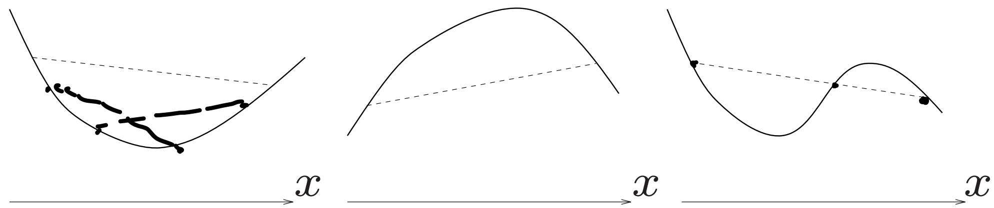
Definition



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

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

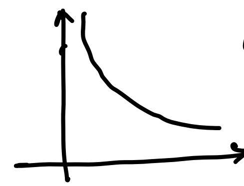
for all $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$



convex concave neither

$f(x) = 1/x$

$\mathrm{dom}\, f = \{x \mid x > 0\}$



convex

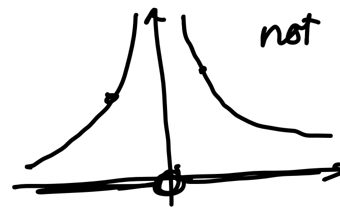
- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

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

$f(x) = 1/x$

$\mathrm{dom}\, f = \{x \mid x > 0 \text{ or } x < 0\}$

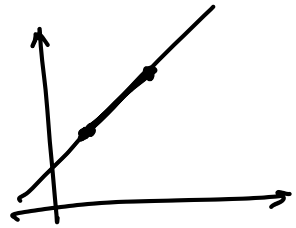


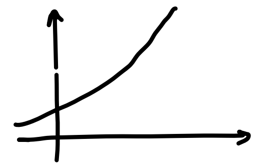
not convex

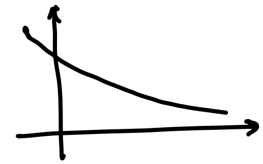
Examples on **R**

convex:

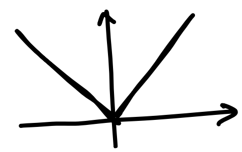
- affine: $ax + b$ on **R**, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$ ($x>0$), for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on **R**, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$







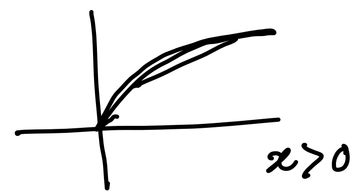
x^2 x^{-1}

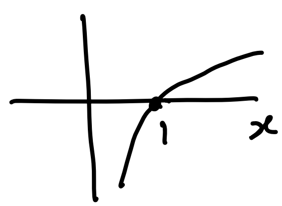


concave:

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

$x^{1/2}$





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

follows from 3 properties of norms

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$

- affine function $f(x) = a^T x + b = \langle a, x \rangle + b$ $f : \mathbb{R}^n \to \mathbb{R}$
- 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) $\langle A, X \rangle = \mathrm{Tr}\, A^T X = \mathrm{vec}(A)^T \mathrm{vec}(X) = \sum_{i,j} A_{ij} X_{ij}$

- affine function

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

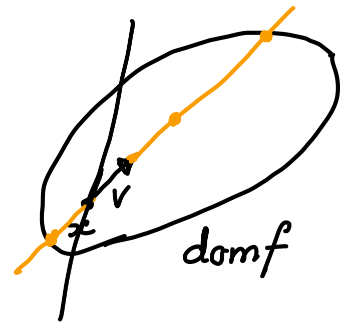
- spectral (maximum singular value) norm $\sigma_{max}(X)$ is a norm (need to verify 3 norm properties)

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

operator norm

$f(X) = \sqrt{\lambda_{max}(X^T X)} = \|X\|_2$

Restriction of a convex function to a line



thm: $f : \mathbf{R}^n \to \mathbf{R}$ 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 $f(X) = \log\det X$, $\mathbf{dom}\, X = \mathbf{S}^n_{++}$ $X \succ 0$

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

where λ_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

$f: S^n_{++} \to \mathbb{R}$

$f(X) = \log \det X$

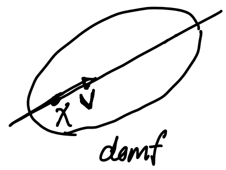
statistics: • max-likelihood estimation w/ Gaussians (Fisher information)

• experiment design (D-opt design) (also arises in online learning, multi-armed bandits)

• graphical models

• ellipsoidal fitting

$f(X) = \log\det X \qquad f: \underline{S^n_{++}} \to \mathbb{R}$



for all X, V $\quad X \succ 0, \; V \in S^n$, $\quad \{t \mid \underbrace{X}_{\succ 0} + t\underbrace{V}_{\in S^n} \succ 0\}$

$X = X^{1/2} X^{1/2}$

$\underline{g(t)} = f(\widetilde{X+tV}) = \log\det(X + tV)$

$= \log\det\left(X^{1/2}\left(I + t X^{-1/2} V X^{-1/2}\right) X^{1/2}\right)$

using - $\det(AB) = \det(BA)$

- $\det(AB) = \underline{\det(A)}\,\underline{\det(B)}$

$g(t) = \log\det(X^{1/2}X^{1/2}) + \log\det\left(I + t\, \underbrace{X^{-1/2} V X^{-1/2}}\right)$

$\lambda_i\left(X^{-1/2} V X^{-1/2}\right) = \lambda_i$

$= \log\det X + \log \prod_i^n \underbrace{\lambda_i\left(I + t X^{-1/2} V X^{-1/2}\right)}_{1 + t\lambda_i}$

$\lambda_i(I + tZ) = 1 + t\,\lambda_i(Z)$

$g(t) = \log\det X + \sum_{i=1}^n \log(1 + t\lambda_i)$

$g'(t) = \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i} \qquad g''(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1+t\lambda_i)^2} < 0 \quad \forall t \in \mathbb{R}$

$\Rightarrow g(t)$ is <u>concave</u> in t

convention:

Extended-value extension

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

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

often simplifies notation; for example, the condition

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

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

$x \in \text{dom} f$ $\quad$ $\tilde{f}(x) < \infty$

$y \in \text{dom} f$ $\quad$ $\tilde{f}(y) < \infty$ $\quad \Rightarrow \text{RHS} < \infty \Rightarrow \tilde{f}(\theta x + (1-\theta)y) < \infty$

$\Rightarrow \theta x + (1-\theta)y \in \text{dom} f, \ \forall\, 0 \le \theta \le 1$

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

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

First-order convexity condition

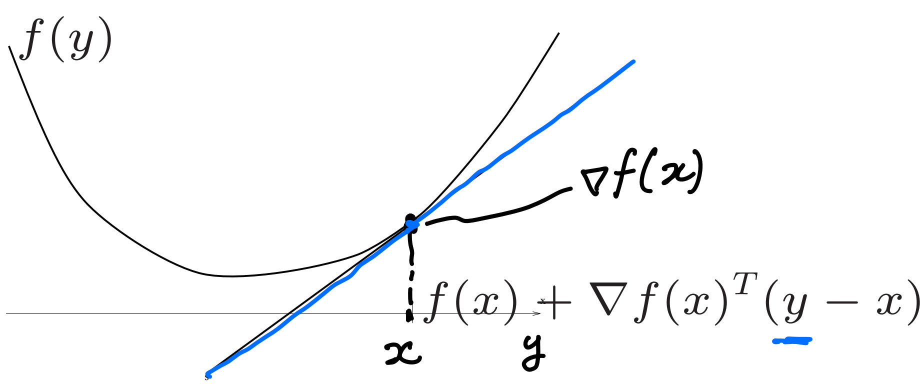
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

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

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

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



first-order approximation of f is global underestimator

Second-order convexity conditions

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

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

$$\begin{bmatrix} \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j} & \cdots \\ & & \ddots \end{bmatrix}$$

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

$\frac{\partial^2 f}{\partial x_i \partial x_i}$

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

$f''(x) \geq 0$

- f is convex if and only if

$$\begin{bmatrix} \square & \cdot & \cdot & \cdot & \cdot \\ \cdot & \square & & & \\ \cdot & & \ddots & & \\ \cdot & & & & \square \end{bmatrix} \succeq 0$$

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

$z^T \nabla^2 f(x) z \geq 0, \ \forall z \in \mathbb{R}^n$

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

converse doesn't hold: $f(x) = x^4$ $f'(x) = 4x^3$ $f''(x) = 12x^2$ $f''(0) = 0$

strictly convex

Examples

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

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

if $\lambda_{min}(P) = 0$

convex if $P \succeq 0$

$g(x) = x^T P x$ is flat for $x = v_{min}$

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

$= x^T A^T A x - 2b^T A x + b^T b$

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

$z^T(A^T A)z = \|Az\|_2^2 \geq 0, \forall z \in \mathbb{R}^n$

convex (for any A)

quadratic-over-linear: $f(x, y) = x^2/y$ $\qquad f: \mathbb{R}^2 \to \mathbb{R} \qquad \mathrm{dom} f = \{(x,y) \mid y \neq 0\}$

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

convex for $y > 0$

$$\begin{bmatrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{bmatrix}$$

$f(\underset{\in \mathbb{R}^n}{x}, y) = \dfrac{x^T x}{y} = \dfrac{\|x\|^2}{y}$

$A = I, \; b = 0$

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

$$\underbrace{\nabla^2 f(x)}_{\succeq 0\,?} = \overbrace{\underbrace{\frac{1}{\mathbf{1}^T z}}_{\succeq 0} \underbrace{\mathbf{diag}(z)}} - \underbrace{\frac{1}{(\mathbf{1}^T z)^2}} \underbrace{zz^T} \qquad (z_k = \exp x_k)$$

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

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

$$\begin{cases} x_k = v_k \sqrt{z_k} \\ y_k = \sqrt{z_k} \end{cases}$$

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)

$$v^T\left(\frac{1}{1^T z}\,\mathrm{diag}(z) - \frac{1}{(1^T z)^2}\, zz^T\right)v = \frac{\overbrace{(v^T \mathrm{diag}(z)\, v)}^{(\sum_k z_k v_k^2)}\overbrace{(1^T z)}^{(\sum z_k)} - \overbrace{v^T z z^T v}^{(v^T z)^2}}{(1^T z)^2}$$

$$= \frac{\sum x_k^2 \sum y_k^2 - (\sum x_k y_k)^2}{a} \geq 0 \text{ from CS ineq.}$$

log-sum-exp :

appendix A.4.2 (ex A.2, A.4) $\quad f(x) = \log \sum_{i=1}^{m} e^{a_i^T x + b_i} \quad A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$

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

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

$h(x) = g(f(x)) \quad \nabla h(x) = \underbrace{g'(f(x))}\nabla f(x)$

$$f(x) = \log \sum_{i=1}^{m} e^{a_i^T x + b_i} \qquad g(y) = \log\left(\sum_{i=1}^{m} e^{x_i}\right) \qquad \mathrm{diag}(z) = \begin{bmatrix} z_1 & & \\ & \ddots & \\ & & z_m \end{bmatrix}$$

$$\nabla g(y) = \frac{1}{\underbrace{\sum e^{y_i}}_{z_i}} \begin{bmatrix} e^{y_1} \\ \vdots \\ e^{y_m} \end{bmatrix}$$

$$\nabla f(x) = \frac{1}{1^T z} A^T z \qquad z_i = e^{a_i^T x + b_i}$$

$$\nabla^2 f(x) = A^T \left(\frac{1}{1^T z} \mathrm{diag}(z) - \frac{1}{1^T z} z z^T \right) A$$

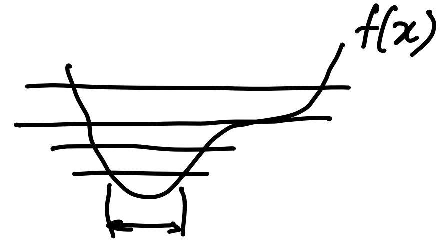
- Cauchy-Schwartz: $|x^T y|^2 \le \|x\|_2^2 \|y\|_2^2$

- $f(x) = \log \sum e^{x_k}$ $\quad$ ML & stats — softmax

$$\max\{x_1, \dots, x_m\} \le f(x) \le \max\{x_1, \dots, x_m\} + \log m$$

in engineering: power addition formula: $\underbrace{\log} \sum_{k=1}^{m} \underbrace{e^{\overbrace{x_k}^{\text{in dB}}}}_{\text{Watt}}$

Epigraph and sublevel set



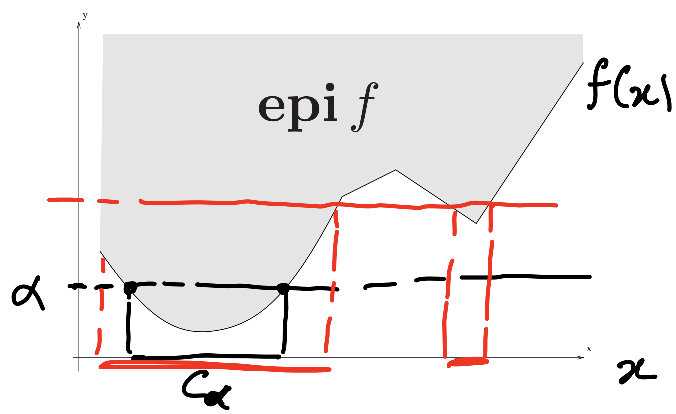
α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

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

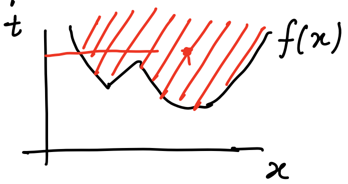
sublevel sets of convex functions are convex (converse is false)

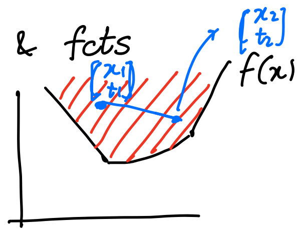
epigraph of $f : \mathbf{R}^n \to \mathbf{R}$:

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



thm: f is convex if and only if $\mathbf{epi}\, f$ is a convex set epigraph : link between convex sets & fcts





<u>thm:</u> f is a cvx fct <u>iff</u> epi$f$ is a cvx set

<u>proof:</u>

- suppose f is cvx & $\begin{bmatrix} x_1 \\ t_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ t_2 \end{bmatrix} \in$ epi f

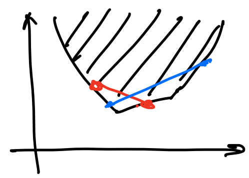
$$f(\underbrace{\theta x_1 + (1-\theta)x_2}_{x}) \le \theta \underbrace{f(x_1)}_{\le t_1} + (1-\theta)\underbrace{f(x_2)}_{\le t_2} \quad \forall\ 0 \le \theta \le 1 \quad \text{from convexity}$$

from def of epi$f$

$$\le \underbrace{\theta t_1 + (1-\theta)t_2}_{t}$$

$$\Longrightarrow \begin{bmatrix} \theta x_1 + (1-\theta)x_2 \\ \theta t_1 + (1-\theta)t_2 \end{bmatrix} \in \text{epi } f$$

- for the reverse: if epi$f$ is a convex set, line segments belong to set



Jensen's inequality

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

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

extension: if f is convex, then

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

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

for any random variable z

$$\mathbb{E} z = \theta x + (1-\theta) y$$

basic inequality is special case with discrete distribution

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

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 (partial)
- perspective

Positive weighted sum & composition with affine function

$g(x) = \int f(x,y)\,dy$ is convex in x, if $f(x,y)$ is convex in x for every y.

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

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

examples

- log barrier for linear inequalities

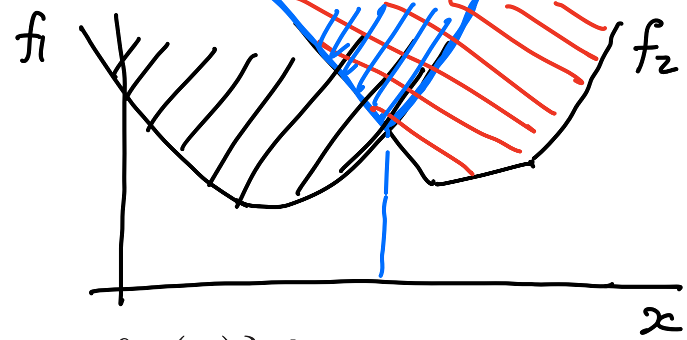
  $-\log x$ is convex

  slack: $b_i - a_i^T x > 0$

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

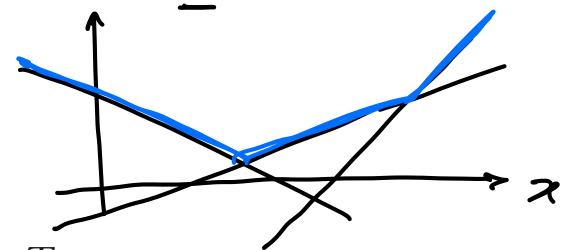
  polyhedron

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



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 $x \in \mathbf{R}^n$: $x_{[1]}$ largest, $x_{[r]}$: r^{th} largest

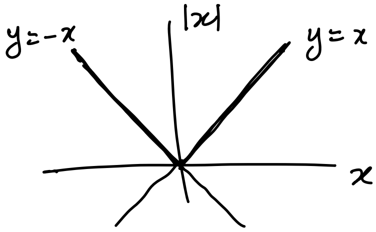
$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

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

proof:

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

$$\underbrace{[0\ 1\ 0\ 1 \cdots 0]}_{a_i}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_i^T x \qquad f(x) = \max_{i=1,\ldots,\binom{n}{r}}\{\underbrace{a_i^T x}_{\text{lin. in } x}\}$$



$$f(x) = |x| = \max\{x, -x\}$$

Pointwise supremum

"index" "index" set

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

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

any set !

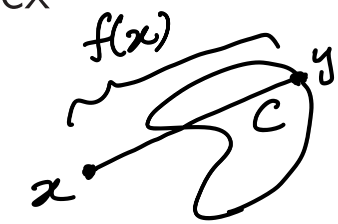
is convex

examples

f(x,y) is convex in x for each y

- 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:

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



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

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

Composition with scalar functions

recall:

$$\tilde{h}(u) = \begin{cases} h(u), & u \in \operatorname{dom} h \\ +\infty, & u \notin \operatorname{dom} \end{cases}$$

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

book, p.84

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

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

① ② ③

- proof (for $n = 1$, twice differentiable g, h) $h : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$

$$f''(x) = \underbrace{h''(g(x))}_{\text{from ②}, \ \geq 0} \underbrace{g'(x)^2}_{\geq 0} + \underbrace{h'(g(x))}_{\geq 0} \underbrace{g''(x)}_{\text{from ①}, \ \geq 0} \overset{?}{\geq} 0, \ \forall x$$

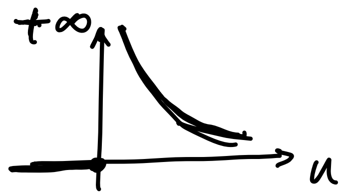
- note: monotonicity must hold for extended-value extension $\tilde{h}$

examples

- $\exp g(x)$ is convex if g is convex $h(u) = e^u$, $u \in \mathbb{R}$

- $1/g(x)$ is convex if g is concave and positive $h(u) = 1/u$, $\operatorname{dom} h = \{u \mid u > 0\}$

$$\tilde{h}(u) = \begin{cases} 1/u, & u > 0 \\ +\infty & u \leq 0 \end{cases}$$



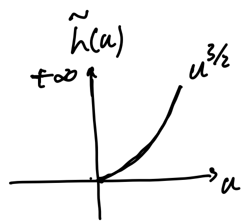
• example:

Is $f(x) = \|x-a\|^{3/2}$ convex ?

$g(x) = \|x-a\| \rightarrow$ convex in x

$\rightarrow h(u) = u^{3/2}$ $\quad$ $\mathrm{dom}\, h = \mathbb{R}_+$

$$\tilde{h}(u) = \begin{cases} u^{3/2}, & u \geq 0 \\ +\infty, & u < 0 \end{cases}$$



$\rightarrow$ not nondecreasing (non-monotonic)

$\Rightarrow$ cannot use comp. rule with the given $h(u)$!

$$\rightarrow \bar{h}(u) = \begin{cases} u^{3/2}, & u \geq 0 \\ \underline{\underline{0}}, & u < 0 \end{cases}$$

$\leftarrow$ nondecreasing

composition rule applies $f(x) = \bar{h}(g(x))$

$\Rightarrow$ $\underline{f(x) \text{ is convex}}$

Vector composition

$\begin{bmatrix} g_1(x) \\ \vdots \\ g_k(x) \end{bmatrix}$

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

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

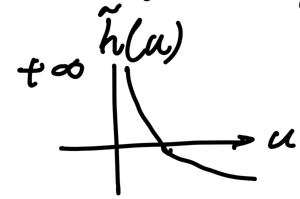
thm: f is convex if
- g_i convex ①, h convex ②, $\tilde{h}$ nondecreasing in each argument ③
- g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

proof (for $n = 1$, differentiable g, h) $x \in \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$ $g''(x) = \begin{bmatrix} g_1''(x) \\ \vdots \\ g_k''(x) \end{bmatrix}$

$$f''(x) = \underbrace{g'(x)^T \underbrace{\nabla^2 h(g(x))}_{\succeq 0 \text{ from } ②} g'(x)}_{\geq 0} + \underbrace{\underbrace{\nabla h(\tilde{g}(x))^T}_{\text{entrywise} \geq 0 \text{ (from ③)}} \underbrace{g''(x)}_{\text{entrywise} \geq 0 \text{ (from ①)}}}_{\geq 0} \geq 0, \ \forall x$$

examples

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



$h(u) = \log \sum_{i=1}^n e^{u_i}$ convex

$\tilde{h}(u) = h(u)$ h is nondecreasing in each u_i ✓

$\Rightarrow f(x)$ convex

- some earlier rules (that preserve convexity) can be obtained from comp. rules:

    - sum property can be obtained from comp. rule with

      $h(u) = \mathbf{1}^T u = u_1 + \cdots + u_n$

    - pointwise max " " " " " " "

      $\max\{\underbrace{g_1(x)}, \cdots, \underbrace{g_n(x)}\}$

      $h(u) = \max\limits_{i=1,\dots,n} u_i$

partial **Minimization**

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

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

is convex

$$f(x,y) = \begin{bmatrix} x^T & y^T \end{bmatrix} \underbrace{\begin{bmatrix} A & B \\ B^T & \boxed{C} \end{bmatrix}}_{\succeq 0 \rightarrow f \text{ is convex in } \begin{bmatrix} x \\ y \end{bmatrix}} \begin{bmatrix} x \\ y \end{bmatrix}$$

examples

- $f(x, y) = x^T A x + 2x^T B y + y^T C y$ with

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

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: $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

$f(x,y)$ is jointly convex in (x,y)

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

$g(x) = x^T A x - x^T B C^{-1} B^T x = x^T (A - BC^{-1}B^T) x \Rightarrow$ convex $\Rightarrow A - BC^{-1}B^T \succeq 0$ (Schur complement)

$t \in \mathbb{R}_{++}$

$\begin{bmatrix} x \\ t \end{bmatrix} \longrightarrow \begin{bmatrix} x/t \end{bmatrix}$

Perspective

the **perspective** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \underbrace{\mathbf{R}^n \times \mathbf{R}}_{\mathbb{R}^{n+1}} \to \mathbf{R}$,

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

thm: g is convex if f is convex

examples

quad-over-lin

$$g(x,t) = tf(x/t) = t\left(\frac{x}{t}\right)^T\left(\frac{x}{t}\right) = \frac{1}{t}x^Tx \qquad f(x,y) = \frac{x^2}{y}$$

- $f(x) = x^Tx$ $= \|x\|_2^2$ is convex; hence $g(x,t) = x^Tx/t$ is convex for $t > 0$
- negative logarithm $f(x) = -\log x$ is convex; hence relative entropy $g(x,t) = t\log t - t\log x$ is convex on $\mathbf{R}^2_{++}$
- if f is convex, then

$$x \longrightarrow \begin{bmatrix} Ax+b \\ c^Tx+d \end{bmatrix} = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$$

- compose f w/ affine fct

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

- apply perspective

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

The conjugate function

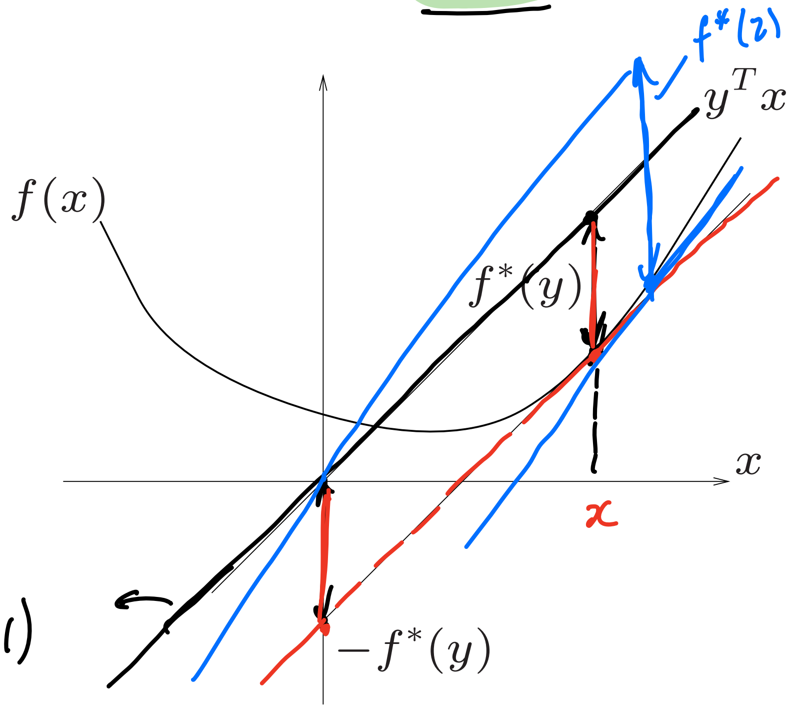
$f(x): \mathbb{R}^n \to \mathbb{R}$

affine in y for every (fixed) x

the **conjugate** of a function f is $\quad f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$

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

$y = 2$



$y=1$ (line w/ slope $=1$)

$f^*(y) = \sup_x (yx - f(x))$

$y=1: \quad f^*(1) = \sup_x (x - f(x))$

if f differentiable:

$1 - f'(x) = 0 \Rightarrow f'(x) = 1$

$\rightarrow y - \nabla_x f(x) = 0$

$\nabla_x f(x) = y \Rightarrow x^*$

$f^*(y)$

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

examples

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

$y + 1/x = 0 \Rightarrow x = -1/y$

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

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

$\mathbf{dom} f = \mathbb{R}^n$

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

$y - Qx = 0$

$x = Q^{-1}y$

$y^TQ^{-1}y - \frac{1}{2}y^T\cancel{Q^{-1}}\cancel{Q}Q^{-1}y$

$= \frac{1}{2}y^TQ^{-1}y$

$f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}x^Tx$

$f^*(y) = \frac{1}{2}\|y\|_2^2$

- exponential: $f(x) = e^x$

$$f^*(y) = \sup_x \left(yx - e^x \right) \qquad y - e^x = 0 \qquad e^x = y$$

- if $y > 0$: $x = \log y$ $\quad f^*(y) = y \log y - y$
- if $y < 0$: $f^*(y) = +\infty$
- if $y = 0$: $f^*(0) = \sup_x \left(-e^x \right) = 0$

- general norm: $f(x) = \| x \|$

Quasiconvex functions

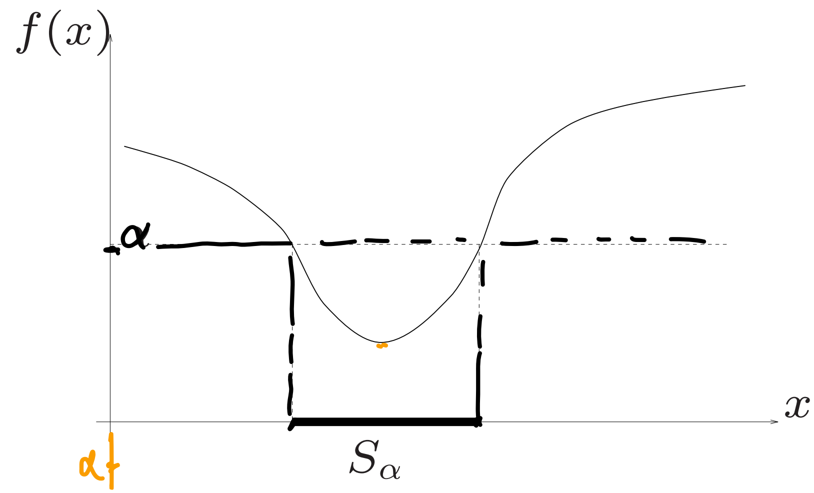
$f : \mathbf{R}^n \to \mathbf{R}$ is quasiconvex if $\mathbf{dom}\, f$ is convex and the sublevel sets

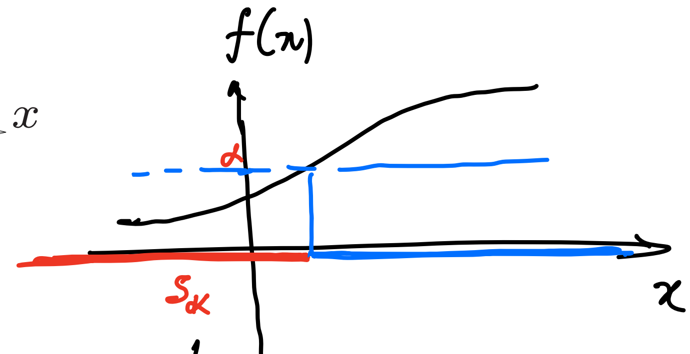
$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

are convex for all α

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

super-level set





- f is quasiconcave if $-f$ is quasiconvex
- f is quasilinear if it is quasiconvex and quasiconcave

Examples

- $\sqrt{|x|}$ is quasiconvex on **R**
- $\text{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$ is quasilinear
- $\log x$ is quasilinear on $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}^2_{++}$
- linear-fractional function

$$f(x) = \frac{a^T x + b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

  is quasilinear

- distance ratio

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \qquad \mathbf{dom}\, f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$$

  is quasiconvex

internal rate of return

- cash flow $x = (x_0, \ldots, x_n)$; x_i is payment in period i (to us if $x_i > 0$)
- we assume $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$
- present value of cash flow x, for interest rate r:

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$$

- internal rate of return is smallest interest rate for which $\mathrm{PV}(x, r) = 0$:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}$$

IRR is quasiconcave: superlevel set is intersection of halfspaces

$$\mathrm{IRR}(x) \geq R \quad \Longleftrightarrow \quad \sum_{i=0}^{n} (1+r)^{-i} x_i > 0 \text{ for } 0 \leq r < R$$

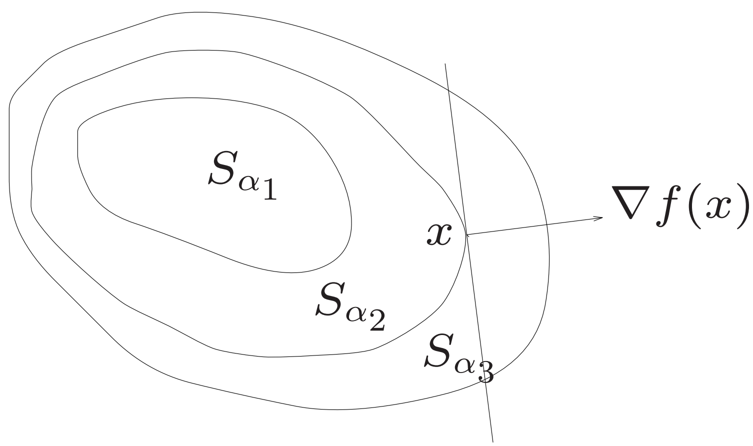
Properties

modified Jensen inequality: for quasiconvex f

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

first-order condition: differentiable f with cvx domain is quasiconvex iff

$$f(y) \leq f(x) \quad \Longrightarrow \quad \nabla f(x)^T (y - x) \leq 0$$



$\alpha_1 < \alpha_2 < \alpha_3$

sums of quasiconvex functions are not necessarily quasiconvex

Log-concave and log-convex functions

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

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

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

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

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

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

- product of log-concave functions is log-concave

- sum of log-concave functions is not always 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)

consequences of integration property

- convolution $f * g$ of log-concave functions f, g is log-concave

$$(f * g)(x) = \int f(x - y)g(y)dy$$

- if $C \subseteq \mathbf{R}^n$ convex and y is a random variable with log-concave pdf then

$$f(x) = \mathbf{prob}(x + y \in C)$$

  is log-concave

  proof: write $f(x)$ as integral of product of log-concave functions

$$f(x) = \int g(x + y)p(y)\, dy, \qquad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C, \end{cases}$$

  p is pdf of y

example: yield function

$$Y(x) = \mathbf{prob}(x + w \in S)$$

- $x \in \mathbf{R}^n$: nominal parameter values for product
- $w \in \mathbf{R}^n$: random variations of parameters in manufactured product
- S: set of acceptable values

if S is convex and w has a log-concave pdf, then

- Y is log-concave
- yield regions $\{x \mid Y(x) \geq \alpha\}$ are convex