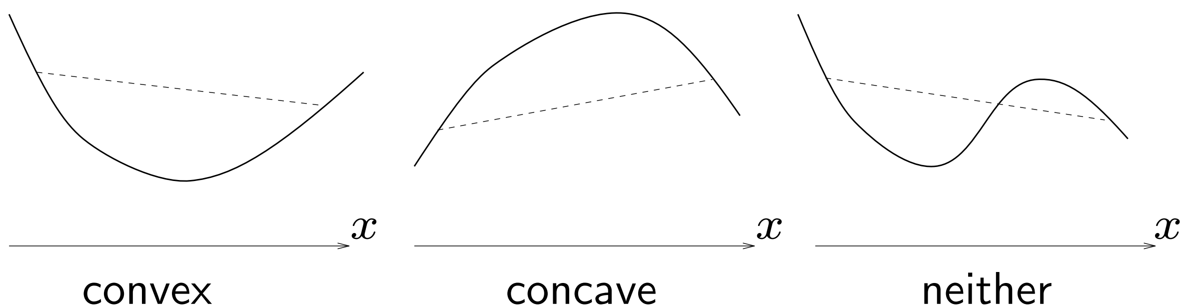
Lecture 2: Convex functions

$f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if $\mathbf{dom}\, f$ is convex and for all $x, y \in \mathbf{dom}\, f$, $\theta \in [0, 1]$

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

f is concave if $-f$ is convex



examples (on $\mathbf{R}$)

- $f(x) = x^2$ is convex
- $f(x) = \log x$ is concave ($\mathbf{dom}\, f = \mathbf{R}_{++}$)
- $f(x) = 1/x$ is convex ($\mathbf{dom}\, f = \mathbf{R}_{++}$)

Extended-valued extensions

for f convex, it's convenient to define the *extension*

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \textbf{dom}\, f \\ +\infty & x \notin \textbf{dom}\, f \end{cases}$$

inequality

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

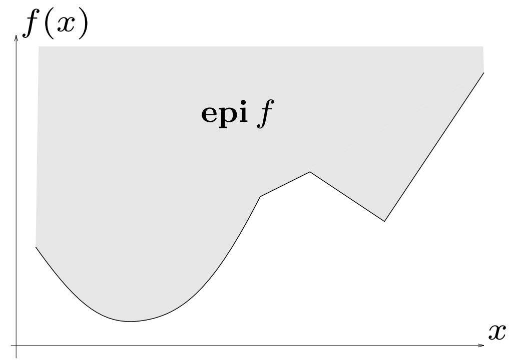
holds for all $x, y \in \mathbf{R}^n$, $0 \leq \theta \leq 1$
(as an inequality in $\mathbf{R} \cup \{+\infty\}$)

we'll use same symbol for f and its extension, $i.e.$, we'll implicitly assume convex functions are extended

Epigraph & sublevel sets

epigraph of a function f is

$$\mathbf{epi}\, f = \{(x,t) \mid x \in \mathbf{dom}\, f, \quad f(x) \le t\}$$



f convex function $\Leftrightarrow$ $\mathbf{epi}\, f$ convex set

the (α-)**sublevel set** of f is

$$C(\alpha) \triangleq \{x \in \mathbf{dom}\, f \mid f(x) \le \alpha\}$$

f convex $\Rightarrow$ sublevel sets are convex (converse false)

Differentiable convex functions

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

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T \quad \text{(evaluated at } x\text{)}$$

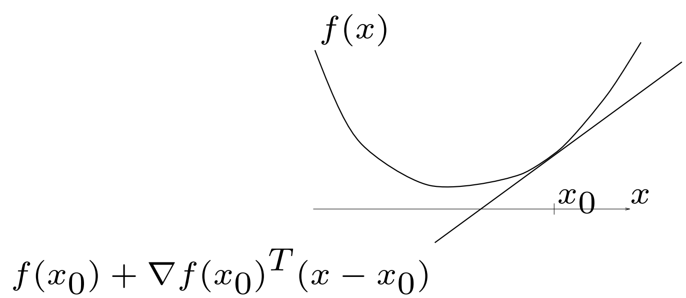
first order Taylor approximation at x_0:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

first-order condition: for f differentiable,

f is convex $\Longleftrightarrow$ for all $x, x_0 \in \mathbf{dom}\, f$,

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$



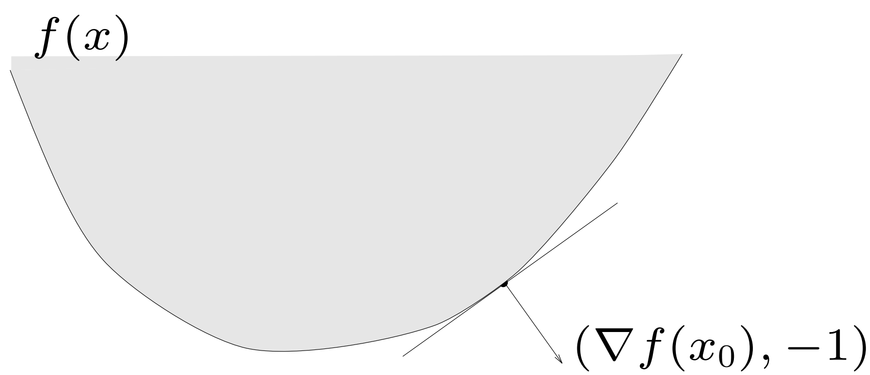
$i.e.$, 1st order approx. is a *global underestimator*

epigraph interpretation

for all $(x, t) \in \mathbf{epi}\, f$,

$$\begin{bmatrix} \nabla f(x_0) \\ -1 \end{bmatrix}^T \begin{bmatrix} x - x_0 \\ t - f(x_0) \end{bmatrix} \leq 0,$$

$i.e.$, $(\nabla f(x_0), -1)$ defines supporting hyperplane to $\mathbf{epi}\, f$ at $(x_0, f(x_0))$



Hessian of a twice differentiable function:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

(evaluated at x)

2nd order Taylor series expansion around x_0:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

second order condition: for f twice differentiable,
f is convex $\Longleftrightarrow$ for all $x \in \mathbf{dom}\, f$, $\nabla^2 f(x) \succeq 0$

Simple examples

- linear and affine functions are convex and concave
- quadratic function $f(x) = x^T Px + 2q^T x + r$
  convex $\Longleftrightarrow P \succeq 0$; concave $\Longleftrightarrow P \preceq 0$
  $(P = P^T)$
- any norm is convex

examples on R:

- x^α is convex on $\mathbf{R}_{++}$ for $\alpha \geq 1$, $\alpha \leq 0$; concave for $0 \leq \alpha \leq 1$
- $\log x$ is concave on $\mathbf{R}_{++}$, $x \log x$ is convex on $\mathbf{R}_{+}$
- $e^{\alpha x}$ is convex
- $|x|$, $\max(0, x)$, $\max(0, -x)$ are convex
- $\log \int_{-\infty}^{x} e^{-t^2} dt$ is concave

Elementary properties

- a function is convex iff it is convex on all lines:

$$f \text{ convex } \iff f(x_0 + th) \text{ convex in } t \text{ for all } x_0, h$$

- positive multiple of convex function is convex:

$$f \text{ convex}, \alpha \geq 0 \Longrightarrow \alpha f \text{ convex}$$

- sum of convex functions is convex:

$$f_1, f_2 \text{ convex } \Longrightarrow f_1 + f_2 \text{ convex}$$

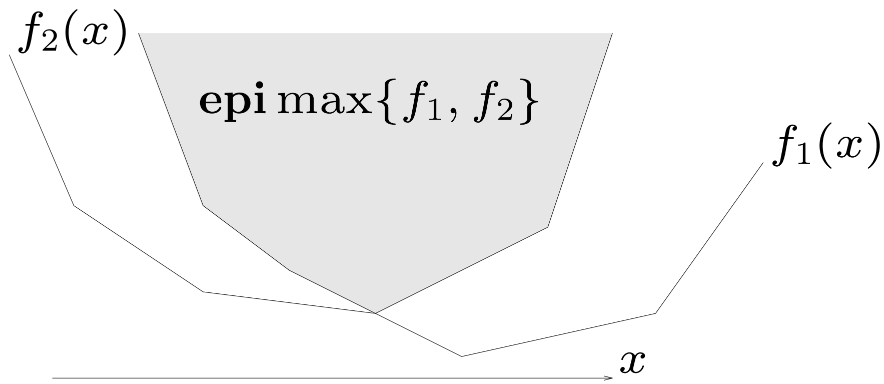
- extends to infinite sums, integrals:

$$g(x, y) \text{ convex in } x \Longrightarrow \int g(x, y) dy \text{ convex}$$

- pointwise maximum:

$$f_1, f_2 \text{ convex } \Longrightarrow \max\{f_1(x), f_2(x)\} \text{ convex}$$

(corresponds to intersection of epigraphs)



- pointwise supremum:

$$f_\alpha \text{ convex } \Longrightarrow \sup_{\alpha\in\mathcal{A}} f_\alpha \text{ convex}$$

- affine transformation of domain

$$f \text{ convex } \Longrightarrow f(Ax+b) \text{ convex}$$

More examples

- piecewise-linear functions: $f(x) = \max_i \{a_i^T x + b_i\}$ is convex in x ($\mathbf{epi}\, f$ is polyhedron)
- max distance to any set, $\sup_{s \in S} \|x - s\|$, is convex in x
- $f(x) = x_{[1]} + x_{[2]} + x_{[3]}$ is convex on $\mathbf{R}^n$ ($x_{[i]}$ is the ith largest x_j)
- $f(x) = \left(\prod_i x_i\right)^{1/n}$ is concave on $\mathbf{R}^n_+$
- $f(x) = \sum_{i=1}^m \log(b_i - a_i^T x)^{-1}$ is convex ($\mathbf{dom}\, f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$)
- least-squares cost as functions of weights,

$$f(w) = \inf_x \sum_i w_i (a_i^T x - b_i)^2,$$

is concave in w

Convex functions of matrices

- $\mathbf{Tr}\, A^T X = \sum_{i,j} A_{ij} X_{ij}$ is linear in X on $\mathbf{R}^{n\times n}$
- $\log\det X^{-1}$ is convex on $\{X \in \mathbf{S}^n \mid X \succ 0\}$
  proof: let λ_i be the eigenvalues of $X_0^{-1/2} H X_0^{-1/2}$

$$\begin{aligned} f(t) &\triangleq \log\det(X_0 + tH)^{-1} \\ &= \log\det X_0^{-1} + \log\det(I + tX_0^{-1/2} H X_0^{-1/2})^{-1} \\ &= \log\det X_0^{-1} - \sum_i \log(1 + t\lambda_i) \end{aligned}$$

  is a convex function of t
- $(\det X)^{1/n}$ is concave on $\{X \in \mathbf{S}^n \mid X \succ 0\}$
- $\lambda_{\max}(X)$ is convex on $\mathbf{S}^n$. **proof:** $\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y$
- $\|X\|_2 = \sigma_1(X) = (\lambda_{\max}(X^T X))^{1/2}$ is convex on $\mathbf{R}^{m\times n}$

  proof: $\|X\|_2 = \sup_{\|u\|_2=1} \|Xu\|_2$

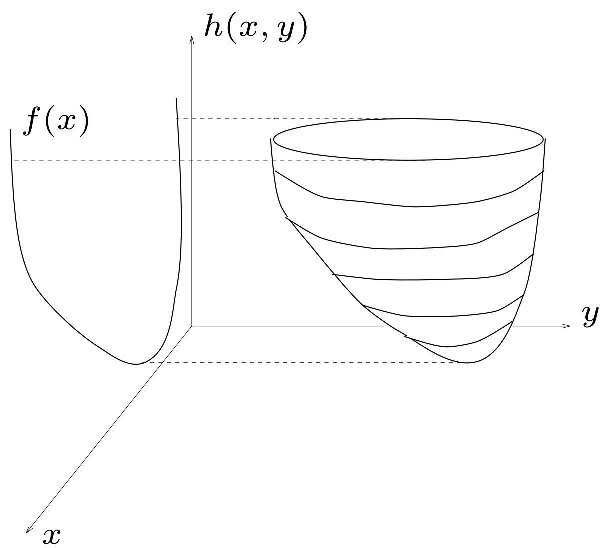
Minimizing over some variables

if $h(x, y)$ is convex in x and y, then

$$f(x) = \inf_{y} h(x, y)$$

is convex in x

corresponds to projection of epigraph, $(x, y, t) \to (x, t)$



examples

- if $S \subseteq \mathbf{R}^n$ is convex then (min) distance to S,

$$\mathbf{dist}(x, S) = \inf_{s \in S} \|x - s\|$$

  is convex in x

- if g is convex, then

$$f(y) = \inf\{g(x) \mid Ax = y\}$$

  is convex in y

  proof: (assume $A \in \mathbf{R}^{m \times n}$ has rank m)
  find B s.t. $\mathcal{R}(B) = \mathcal{N}(A)$; then $Ax = y$ iff

$$x = A^T(AA^T)^{-1}y + Bz$$

  for some z, and hence

$$f(y) = \inf_z g(A^T(AA^T)^{-1}y + Bz)$$

Composition — one-dimensional case

$f(x) = h(g(x))$ ($g : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R} \to \mathbf{R}$) is convex if

- g convex; h convex, nondecreasing
- g concave; h convex, nonincreasing

proof: (differentiable functions, $x \in \mathbf{R}$)

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

examples

- $f(x) = \exp g(x)$ is convex if g is convex
- $f(x) = 1/g(x)$ is convex if g is concave, positive
- $f(x) = g(x)^p$, $p \geq 1$, is convex if $g(x)$ convex, positive
- $f(x) = -\sum_i \log(-f_i(x))$ is convex on $\{x \mid f_i(x) < 0\}$ if f_i are convex

Composition — k-dimensional case

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

with $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$ is convex if

- h convex, nondecreasing in each arg.; g_i convex
- h convex, nonincreasing in each arg.; g_i concave
- etc.

proof: (differentiable functions, $n = 1$)

$$f'' = \nabla h^T \begin{bmatrix} g_1'' \\ \vdots \\ g_k'' \end{bmatrix} + \begin{bmatrix} g_1' \\ \vdots \\ g_k' \end{bmatrix}^T \nabla^2 h \begin{bmatrix} g_1' \\ \vdots \\ g_k' \end{bmatrix}$$

examples

- $f(x) = \max_i g_i(x)$ is convex if each g_i is
- $f(x) = \log \sum_i \exp g_i(x)$ is convex if each g_i is

Jensen's inequality

$f : \mathbf{R}^n \to \mathbf{R}$ convex

- two points: $\theta_1 + \theta_2 = 1$, $\theta_i \geq 0 \implies f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2)$
- more than two points: $\sum_i \theta_i = 1$, $\theta_i \geq 0 \implies f(\sum_i \theta_i x_i) \leq \sum_i \theta_i f(x_i)$
- continuous version: $p(x) \geq 0, \ \int p(x)\ dx = 1 \implies$

$$f(\int xp(x)\ dx) \leq \int f(x)p(x)\ dx$$

- most general form: for any prob. distr. on x,

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

these are all called *Jensen's inequality*

interpretation of Jensen's inequality:
(zero mean) randomization, dithering increases average value of a convex function

many (some people claim most) inequalities can be derived from Jensen's inequality

example: arithmetic-geometric mean inequality

$$a, b \geq 0 \Rightarrow \sqrt{ab} \leq (a+b)/2$$

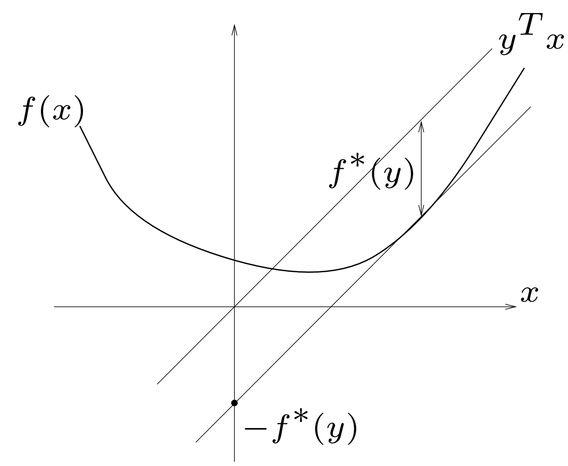
proof: $f(x) = \log x$ is concave on $\{x | x > 0\}$, so for $a, b > 0$,

$$\frac{1}{2}(\log a + \log b) \leq \log\left(\frac{a+b}{2}\right)$$

Conjugate functions

the **conjugate** function of $f : \mathbf{R}^n \to \mathbf{R}$ is

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



- f^* is convex (even if f isn't)

Examples

$f(x) = -\log x$ ($\mathbf{dom}\, f = \{x \mid x > 0\}$):

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

$f(x) = x^T P x$ ($P \succ 0$):

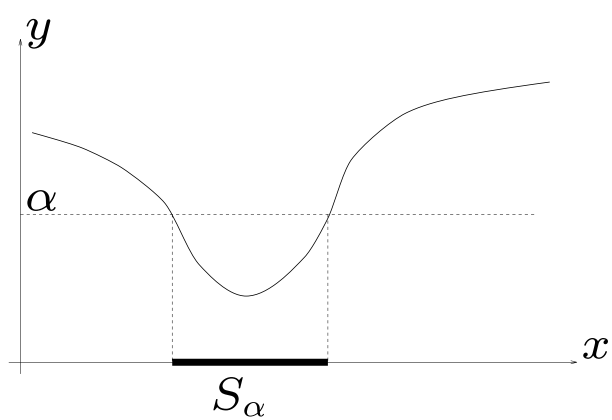
$$f^*(y) = \sup_x (y^T x - x^T P x) = \frac{1}{4} y^T P^{-1} y$$

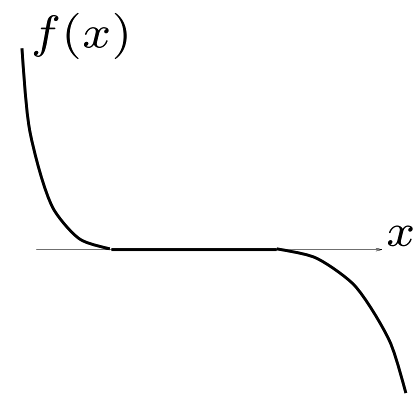
Quasiconvex functions

$f : \mathbf{R}^n \to \mathbf{R}$ is *quasiconvex* if every sublevel set

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

is convex





- can have 'locally flat' regions
- f is *quasiconcave* if $-f$ is quasiconvex, *i.e.*, superlevel sets $\{x \mid f(x) \geq \alpha\}$ are convex
- a function which is both quasiconvex and quasiconcave is called *quasilinear*
- f convex (concave) $\Rightarrow$ f quasiconvex (quasiconcave)

Examples

- $f(x) = \sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
- $f(x) = \log x$ is quasilinear on $\mathbf{R}_+$
- linear fractional function,

$$f(x) = \frac{a^T x + b}{c^T x + d}$$

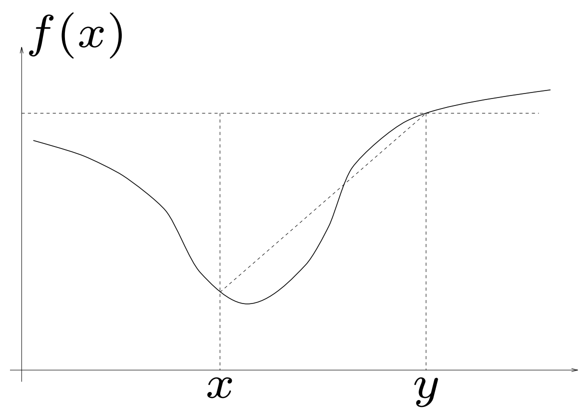
  is quasilinear on the halfspace $c^T x + d > 0$

- $f(x) = \frac{\|x-a\|_2}{\|x-b\|_2}$ is quasiconvex on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$
- $f(a) = \text{degree}(a_0 + a_1 t + \cdots + a_k t^k)$ on $\mathbf{R}^{k+1}$

Properties

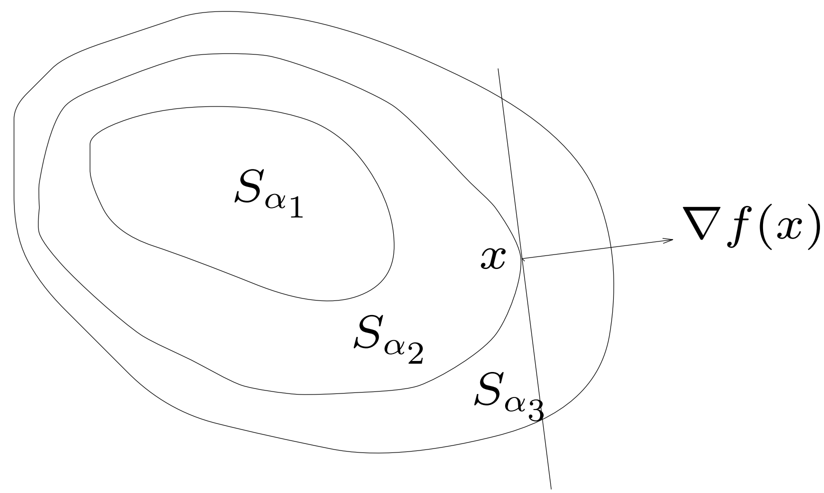
- f is quasiconvex if and only if it is quasiconvex on lines, $i.e.$, $f(x_0 + th)$ quasiconvex in t for all x_0, h
- modified Jensen's inequality: f is quasiconvex iff for all $x, y \in \mathbf{dom}\, f$, $\theta \in [0,1]$,

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



- for f differentiable, f quasiconvex $\iff$ for all $x, y \in \mathbf{dom}\, f$

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



$$\alpha_1 < \alpha_2 < \alpha_3$$

- positive multiples

$$f \text{ quasiconvex}, \alpha \geq 0 \Longrightarrow \alpha f \text{ quasiconvex}$$

- pointwise maximum

$$f_1, f_2 \text{ quasiconvex } \Longrightarrow \max\{f_1, f_2\} \text{ quasiconvex}$$

(extends to supremum over arbitrary set)

- affine transformation of domain

$$f \text{ quasiconvex} \Longrightarrow f(Ax+b) \text{ quasiconvex}$$

- linear-fractional transformation of domain

$$f \text{ quasiconvex} \Longrightarrow f\left(\frac{Ax+b}{c^Tx+d}\right) \text{ quasiconvex}$$

on $c^Tx + d > 0$

- composition with monotone increasing function

$$f \text{ quasiconvex, } g \text{ monotone increasing} \Longrightarrow g(f(x)) \text{ quasiconvex}$$

- sums of quasiconvex functions are **not** quasiconvex in general
- f quasiconvex in x, $y \Longrightarrow g(x) = \inf_y f(x,y)$ quasiconvex in x

Nested sets characterization

f quasiconvex $\Rightarrow$ sublevel sets S_α are convex, nested, $i.e.$,

$$\alpha_1 \leq \alpha_2 \Rightarrow S_{\alpha_1} \subseteq S_{\alpha_2}$$

converse: if T_α is a nested family of convex sets, then

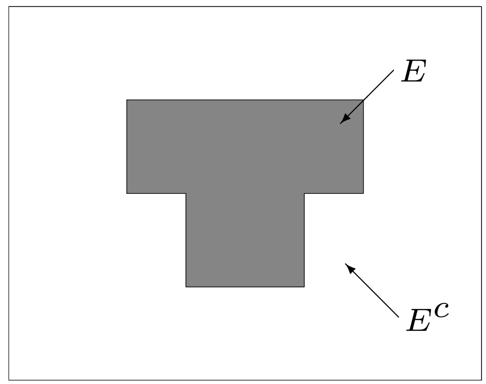
$$f(x) = \inf\{\alpha \mid x \in T_\alpha\}$$

is quasiconvex.

engineering interpretation: T_α are specs, tighter for smaller α

Example of Quasiconvex Functions via Nested Sets: Electron-beam Lithography

$E \subseteq [0,1] \times [0,1]$: desired exposure region
$E^c = [0,1] \times [0,1] \backslash E$: desired non-exposure region



$I(p)$: e-beam intensity at position $p \in [0,1] \times [0,1]$

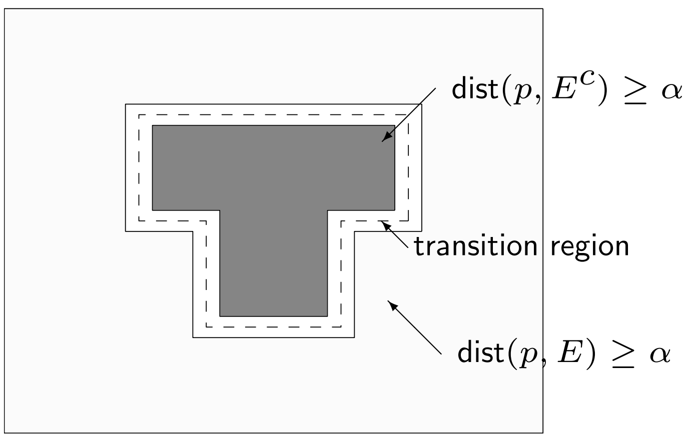
$$I(p) = \sum_i x_i g(p - p_i), \quad i = 1, \ldots, N$$

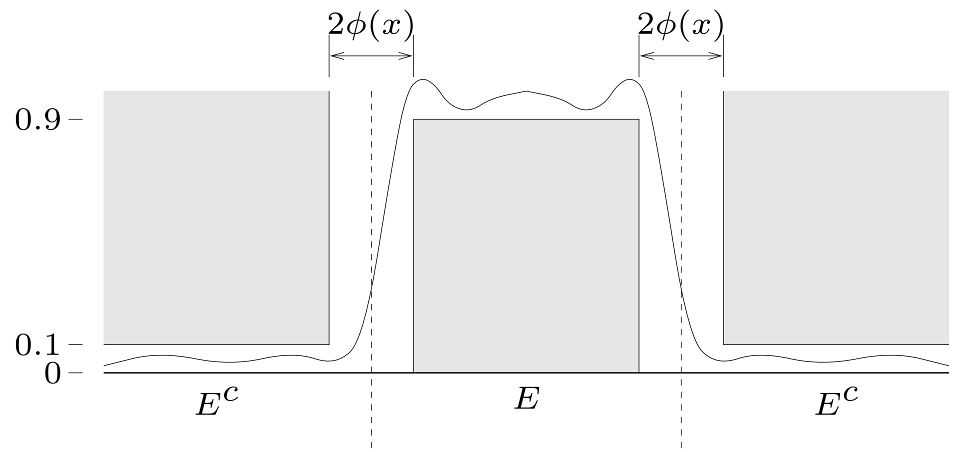
x_i: intensity of electron beam directed at pixel i
$g(p)$: given (point-spread) function

pattern transition width

define $\phi(x)$ as minimum α s.t.

$$
\begin{array}{ll}
I(p) \geq 0.9 & \text{for } \mathbf{dist}(p, E^c) \geq \alpha \\
I(p) \leq 0.1 & \text{for } \mathbf{dist}(p, E) \geq \alpha
\end{array}
$$





$\phi(x)$ is quasiconvex

Log-concave functions

$f : \mathbf{R}^n \to \mathbf{R}_+$ is log-concave (log-convex) if $\log f$ is concave (convex)

log-convex $\Rightarrow$ convex; concave $\Rightarrow$ log-concave

examples

- normal density, $f(x) = e^{-(1/2)(x-x_0)^T\Sigma^{-1}(x-x_0)}$
- erfc, $f(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$
- indicator function of convex set C:

$$I_C(x) = \begin{cases} 1 & x \in C \\ 0 & x \notin C \end{cases}$$

Properties

- sum of log-concave functions not always log-concave
  (but sum of log-convex functions **is** log-convex)
- products

$$f, g \text{ log-concave} \implies fg \text{ log-concave}$$

  (immediate)
- integrals

$$f(x, y) \text{ log-concave in } x, y \implies \int f(x, y)dy \text{ log-concave}$$

  (not easy to show!)
- convolutions

$$f, g \text{ log-concave} \implies \int f(x-y)g(y)dy \text{ log-concave}$$

  (immediate from the properties above)

Log-concave probability densities

many common probability density functions are log-concave

- normal ($\Sigma \succ 0$)

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

- exponential ($\lambda_i > 0$)

$$f(x) = \left(\prod_{i=1}^{n} \lambda_i \right) e^{-(\lambda_1 x_1 + \cdots + \lambda_n x_n)}, \quad x \in \mathbf{R}_+^n$$

- uniform distribution on convex (bounded) set C

$$f(x) = \begin{cases} 1/\alpha & x \in C \\ 0 & x \notin C \end{cases}$$

where α is Lebesgue measure of C ($i.e.$, length, area, volume . . .)

Example: manufacturing yield

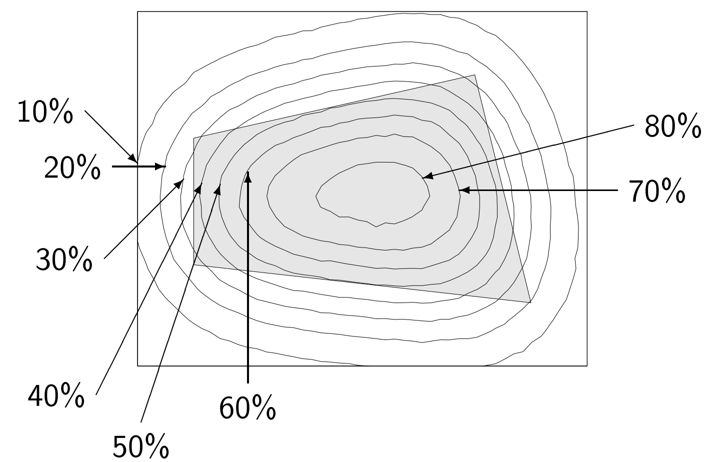
$$x_{\text{manu}} = x + v$$

- $x \in \mathbf{R}^n$: nominal value of design parameters
- $v \in \mathbf{R}^n$: manufacturing errors; zero mean random variable
- $S \subseteq \mathbf{R}^n$: specs, *i.e.*, acceptable values of x_{manu}

the *yield* $Y(x) = \mathbf{Prob}(x + v \in S)$ is log-concave

if

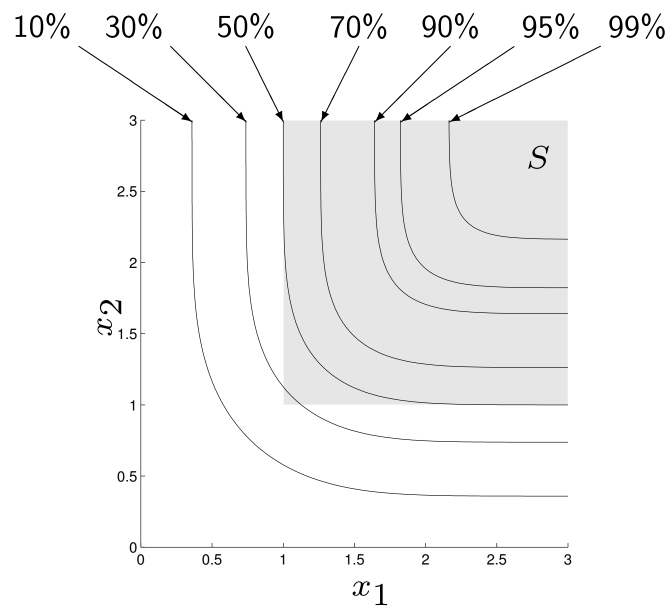
- S is a convex set
- the probability density of v is log-concave



example

- $S = \{y \in \mathbf{R}^2 \mid y_1 \geq 1, y_2 \geq 1\}$
- v_1, v_2: independent, normal with $\sigma = 1$

$$\text{yield}(x) \quad = \quad \mathbf{Prob}(x + v \in S) = \frac{1}{2\pi}\left(\int_{1-x_1}^{\infty} e^{-t^2/2} dt\right)\left(\int_{1-x_2}^{\infty} e^{-t^2/2} dt\right)$$



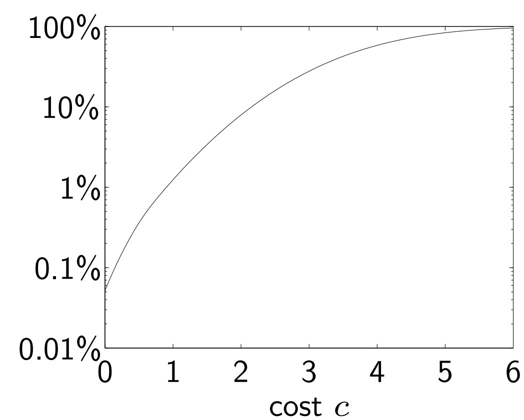
example (continued): max yield vs. cost

manufacturing cost $c = x_1 + 2x_2$; max yield for given cost is

$$Y^{\mathrm{opt}}(c) = \sup_{\substack{x_1 + 2x_2 = c \\ x_1, x_2 \geq 0}} Y(x)$$

$Y(x)$ is log-concave

$$-\log Y^{\mathrm{opt}}(c) = \inf_{\substack{x_1 + 2x_2 = c \\ x_1, x_2 \geq 0}} -\log Y(x_1, x_2)$$



K-convexity

cvx. cone $K \subseteq \mathbf{R}^m$ induces generalized inequality $\preceq_K$

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if $0 \leq \theta \leq 1 \quad \Longrightarrow$

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

example. K is PSD cone (called *matrix convexity*). $f(X) = X^2$ is K-convex on $\mathbf{S}^m$

let's show that for $\theta \in [0,1]$,

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

for any $u \in \mathbf{R}^m$, $u^T X^2 u = \|Xu\|_2^2$ is a (quadratic) convex fct of X, so

$$u^T(\theta X + (1-\theta)Y)^2 u \leq \theta u^T X^2 u + (1-\theta)u^T Y^2 u$$

which implies (1)