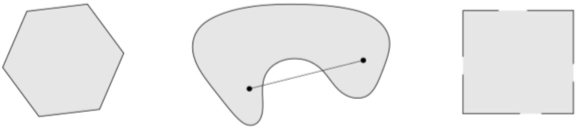
ECE 602 – Section 2
Convex sets and convex functions

- Convex optimization problems
- Convex sets and their examples
- Separating and supporting hyperplanes
- Projections on convex sets
- Convex functions, conjugate functions

Convex sets

- A set $C \subseteq \mathbf{R}^n$ is *convex* if for any $x_1, x_2 \in C$ and $0 \leq \theta \leq 1$, we have $\theta x_1 + (1-\theta)x_2 \in C$.



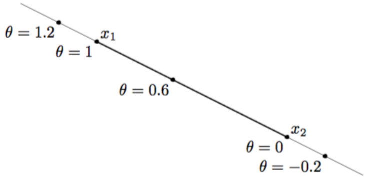
- We call a point of the form $\theta_1 x_1 + \ldots \theta_k x_k$, where $\sum_{i=1}^k \theta_i = 1$ and $\theta_i \geq 0$, a *convex combination* of the points $x_1, \ldots, x_k$.

- A convex combination of points $\{x_i\}_{i=1}^k$ can be thought of as a *mixture* or *weighted average* of the points, with θ_i being the fraction of x_i in the mixture.

Affine sets

- Suppose $x_1, x_2 \in \mathbf{R}^n$ and $x_1 \neq x_2$. The *line* through x_1 and x_2 is defined as

$$y = \theta x_1 + (1-\theta)x_2 = x_2 + \theta(x_1 - x_2), \quad \theta \in \mathbf{R}.$$



- A set $C \subseteq \mathbf{R}^n$ is *affine* if for any $x_1, x_2 \in C$ and $\theta \in \mathbf{R}$, we have that $\theta x_1 + (1-\theta)x_2 \in C$.

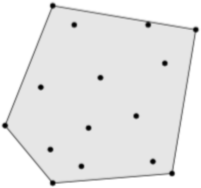
- We refer to a point of the form $\theta_1 x_1 + \ldots \theta_k x_k$, where $\sum_{i=1}^{k} \theta_i = 1$, as an *affine combination* of the points $x_1, \ldots, x_k$.

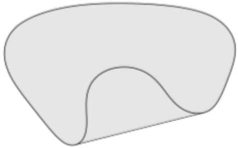
- E.g., the solution set of a system of linear equations, $C = \{x \mid Ax = b\}$, is an affine set. (The converse is true as well.)

Convex hull

- The *convex hull* **conv** C of a set C is the set of all convex combinations of points in C:

$$\mathbf{conv}\, C = \left\{ \theta_1 x_1 + \ldots \theta_k x_k \;\middle|\; x_i \in C,\, \theta_i \geq 0, \forall i,\, \&\, \sum_{i=1}^{k} \theta_i = 1 \right\}$$



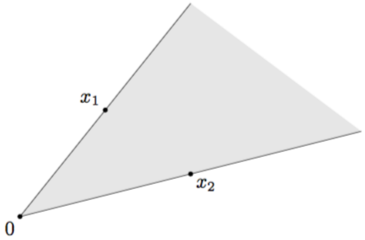


- The idea of a convex combination can be generalized to include infinite sums, integrals, and, more generally, probability distributions.

Cones

- A set C is called a *cone*, if for every $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$.

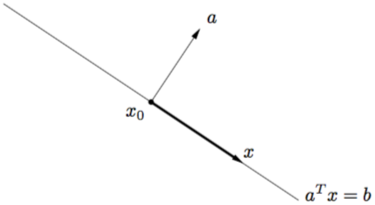
- A set C is a *convex cone* if it is convex and a cone, which means that for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have $\theta_1 x_1 + \theta_2 x_2 \in C$.



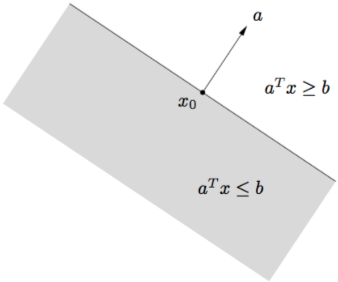
- A point of the form $\theta_1 x_1 + \ldots \theta_k x_k$, where $\theta_1, \ldots, \theta_k \geq 0$, is called a *conic combination* of the points $x_1, \ldots, x_k$.

Hyperplanes and half-spaces

- A *hyperplane* is a set of the form $\{x \mid a^T x = b\}$, where $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$. It is both affine and convex.



- A *closed half-space* is a set of the form $\{x \mid a^T x \leq b\}$. It is only convex.



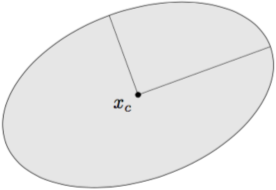
- A *Euclidean ball* in $\mathbf{R}^n$ is defined as

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\} = \left\{x \mid (x - x_c)^T (x - x_c) \le r^2\right\} = \\ = \{x_c + ru \mid \|u\|_2 \le 1\}$$

- A *Euclidean ellipsoid* in $\mathbf{R}^n$ has the form

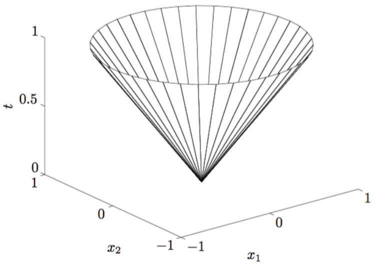
$$\mathcal{E}(x_c, P) = \left\{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\right\} = \left\{x_c + P^{1/2} u \mid \|u\|_2 \le 1\right\},$$

where $P \in \mathbf{S}^n_{++}$.



- Let $\|\cdot\|$ be some norm on $\mathbf{R}^n$.
- A *norm ball* of radius r centred at x_c is a convex set which is defined as $\{x \mid \|x - x_c\| \leq r\}$.
- The *norm cone* associated with the norm $\|\cdot\|$ is the set

$$C = \{(x,t) \mid \|x\| \leq t\} \in \mathbf{R}^{n+1}$$



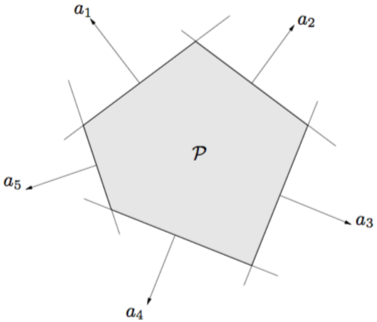
- The *second-order cone* is a norm cone defined with $\|\cdot\|_2$.

Polyhedra

- A *polyhedron* is defined as the solution set of a finite number of linear equalities and inequalities:

$$\mathcal{P} = \{x \mid Ax \preceq b, Cx = d\},$$

where $A \in \mathbf{R}^{m\times n}$, $C \in \mathbf{R}^{p\times n}$, $b \in \mathbf{R}^m$, and $d \in \mathbf{R}^p$.



- A polyhedron is the intersection of a finite number of halfspaces and hyperplanes (e.g., affine sets, subspaces, hyperplanes, lines, rays, line segments, and halfspaces).

The positive semidefinite cone

- As before, the set of *symmetric matrices* is denoted as

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n\times n} \mid X = X^T\}$$

- The set of *symmetric positive semidefinite matrices* is denoted by

$$\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$$

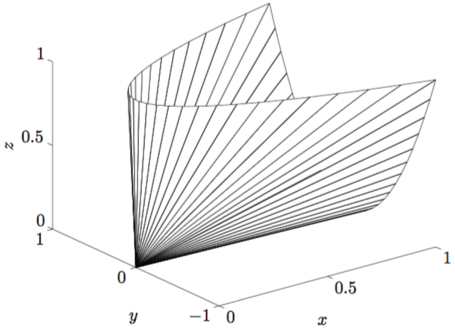
- The set of *symmetric positive definite matrices* is denoted by

$$\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$$

- The set $\mathbf{S}^n_+$ is a convex cone. (Why?)

- **Example:** *Positive semidefinite cone in* $\mathbf{S}^2$. We have

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \Longleftrightarrow x \geq 0,\ z \geq 0,\ xz \geq y^2$$



General characterization of convex sets

- Every closed convex set is a intersection of either finite or infinite number of halfspaces.

- In fact, a closed convex set S is the intersection of all halfspaces which contain it:

$$S = \cap \{\mathcal{H} \mid \mathcal{H} \text{ half-space }, S \subseteq \mathcal{H}\}$$

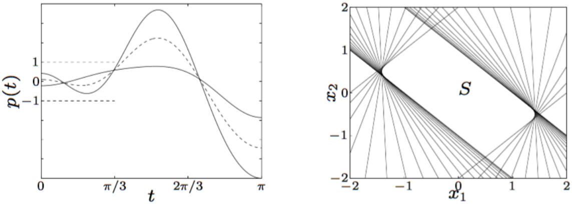
Example

- Consider the following set

$$S = \left\{ x \in \mathbf{R}^m \mid \left| \sum_{k=1}^{m} x_k \cos kt \right| \leq 1 \text{ for } |t| \leq \pi/3 \right\}$$

S is convex, since it can be expressed as an infinite intersection of *slabs*, $S = \cap_{|t| \leq \pi/3} S_t$, where

$$S_t = \left\{ x \mid -1 \leq (\cos t, \cos 2t \, \ldots, \cos kt)^T x \leq 1 \right\}$$

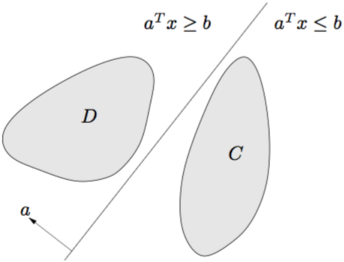


- Practical methods for establishing convexity of a given set C consist of either
  1. applying the definition, or
  2. showing that C can be obtained from a simpler convex by operations that preserve convexity.

- Operations that preserve convexity include
  1. intersections
  2. Cartesian products (e.g., $\{x \in \mathbf{R}^2 \mid \|x\|_\infty \leq 1\} = [-1, 1] \times [-1, 1]$)
  3. affine functions ($f(x) = A\,x + b$, with $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$). E.g., $\mathcal{E}(x_c, P)$ is the image of $\mathcal{B}(0,1)$ under $f(x) = P^{1/2}\,x + x_c$.
  4. perspective functions ($f(x,t) = x/t$, with $\mathbf{dom}\, f = \mathbf{R}^n \cup \mathbf{R}_{++}$)
  5. linear-fractional functions ($f(x) = (Ax+b)/(c^T x + d)$, with $\mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$)

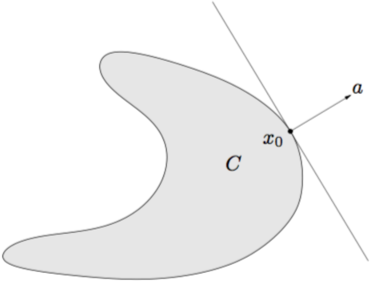
- If C and D are nonempty disjoint convex sets, there exist $a \neq 0$ and b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



- We say that the hyperplane $\{x \mid a^T x = b\}$ separates C and D.
- Strict separation requires additional assumptions.

- Let $C \subseteq \mathbf{R}^n$ and $x_0 \in \mathbf{bd}\, C$.
- If $a \neq 0$ satisfies $a^T x \leq a^T x_0$, then $\{x \mid a^T x = a^T x_0\}$ is called the *supporting hyperplane* to C at the point x_0.



- The *supporting hyperplane theorem* states that for any convex set $C \neq \emptyset$ and any $x_0 \in \mathbf{bd}\, C$, there exists a supporting hyperplane to C at x_0.

- Let C be a *closed* convex set. Then, there exists a unique *projection* $P_C(x)$ of x onto C given by

$$P_C(x) = \inf_{z \in C} \|x - z\|_2$$

- In other words, $P_C(x)$ solves the problem of minimizing (over z) of $\|x - z\|_2$ *subject to* $z \in C$.

- Fortunately, for many interesting practical cases, such projections can be computed in a closed form.

- **Hyperplane**: $C = \{x \in \mathbf{R} \mid a^T x = b\}$ (with $a \neq 0$)

$$P_C(x) = x + \frac{b - a^T x}{\|a\|_2^2} a$$

- **Affine set**: $C = \{x \in \mathbf{R} \mid Ax = b\}$ (with $A \in \mathbf{R}^{m \times n}$ and $\mathbf{rank}\, A = m$)

$$P_C(x) = x + A^T (AA^T)^{-1} (b - Ax)$$

(inexpensive if $m \ll n$ or $AA^T = I$)

- **Halfspace**: $C = \{x \in \mathbf{R} \mid a^T x \leq b\}$ (with $a \neq 0$)

$$P_C(x) = \begin{cases} x + \frac{b - a^T x}{\|a\|_2^2} a, & a^T x > x \\ x, & a^T x \leq b \end{cases}$$

- **Rectangle**: $C = \{x \in \mathbf{R} \mid l \preceq x \preceq u\}$

$$P_C(x)_k = \begin{cases} l_k, & x_k \leq l_k \\ x_k, & l_k \leq x_k \leq u_k \\ u_k, & x_k \geq u_k \end{cases}$$

- **Nonnegative orthant**: $C = \mathbf{R}^n_+$

$$P_C(x) = \max\{x, 0\} = x_+$$

- **Probability simplex**: $C = \{x \in \mathbf{R}^n_+ \mid \mathbf{1}^T x = 1\}$

$$P_C(x) = (x - \mathbf{1}\lambda)_+,$$

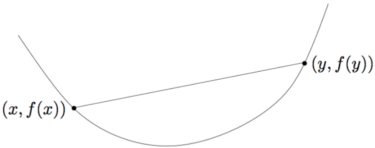
  where λ is the solution of

$$\mathbf{1}^T (x - \mathbf{1}\lambda)_+ = \sum_{i=1}^{n} \max\{0, x_k - \lambda\} = 1$$

Convex functions

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is *convex* if $\mathbf{dom}\, f$ is a convex set and if for all $x, y \in \mathbf{dom}\, f$, and θ with $0 \leq \theta \leq 1$, we have

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



- A function f is *strictly convex* if strict inequality holds.
- We say fš is *concave* if $-f$ is convex, and *strictly concave* if $-f$ is strictly convex.
- A function is convex if and only if it is convex when restricted to any line that intersects its domain. (Very useful for verifying convexity.)

First order conditions

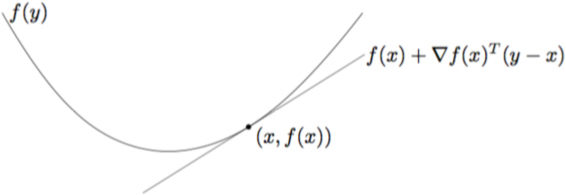
- Suppose f is differentiable. Then f is convex if and only if **dom** f is convex and, for all $x, y \in \mathbf{dom}\, f$,

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

while for *strict* convexity we require

$$f(y) > f(x) + \nabla f(x)^T (y - x), \quad x \neq y$$

Thus, for a convex f, the first-order Taylor approximation is a *global underestimator* of f. (The converse is also true.)



- **Important:** If $\nabla f(x) = 0$, then for all $y \in \mathbf{dom}\, f$, $f(y) \geq f(x)$, i.e., x is a global minimizer of f.

Second order conditions

- Suppose f is twice differentiable (i.e., $f \in \mathcal{C}^2(\mathbf{R})$). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and, for all $x \in \mathbf{dom}\, f$,

  $$\nabla^2 f(x) \succeq 0$$

  Thus, the graph of the function have positive (upward) curvature at x.

- If $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then the function f is strictly convex. (The converse, however, is not true; example: $f(x) = x^4$.)

- The requirement that $\mathbf{dom}\, f$ is convex cannot be dropped (example: $f(x) = 1/x^2$, with $\mathbf{dom}\, f = \mathbf{R} \backslash \{0\}$).

Examples

The following functions are convex.

- $f(x) = (1/2)x^T Px + q^T x + r$, with $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$.

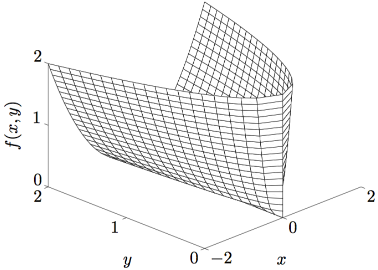
  Since $\nabla^2 f(x) = P$, f is convex iff $P \succeq 0$ and strictly convex iff $P \succ 0$.

- $f(x) = Ax + b$, with $x \in \mathbf{R}^n$.
- $f(x) = e^{ax}$, with $x, a \in \mathbb{R}$.
- $f(x) = x^a$, with $x \in \mathbf{R}_{++}$, $a \notin (0, 1)$.
- $f(x) = -\log x$, with $x \in \mathbf{R}_{++}$.
- $f(x) = x \log x$, with $x \in \mathbf{R}_+$. (Note that $0 \log 0 = 0$.)

  Indeed, we have $f'(x) = 1 + \log x$ and, therefore, $f''(x) = 1/x$, which is positive on $\mathbf{R}_+$. Thus, the negative entropy function is strictly convex.

More examples

- Every norm on $\mathbf{R}^n$ is convex. (Why ?)
- The *max function* $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbf{R}^n$.
- The *quadratic-over-linear function* $f(x, y) = x^2/y$, with $\mathbf{dom}\, f = \mathbf{R} \times \mathbf{R}_{++}$, is convex.



- The *log-sum-exp function* $f(x) = \log(e^{x_1} + \ldots + e^{x_n})$ is convex on $\mathbf{R}^n$.

More examples (cont.)

- The *geometric mean* $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{R}_{++}$.
- The *log-det function* $f(X) = \log\det X$ is concave on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

  Indeed, for arbitrary $Z, V \in \mathbf{S}^n$, define $g(t) = f(Z + tV)$ and restrict it to the interval $\{t \mid Z + tV \succ 0\}$ (that is assumed to include 0).

  $$g(t) = \log\det(Z + tV) = \log\det(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}) =$$
  $$= \sum_{i=1}^n \log(1 + t\lambda_i) + \log\det Z$$

  Therefore, we have

  $$g'(t) = \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i}, \quad g''(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

  Hence, $f(X)$ is concave.

Epigraph

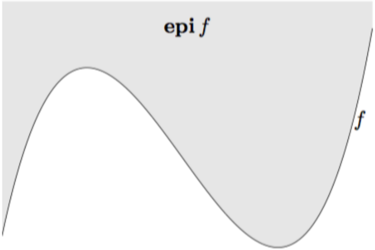
- The *graph* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\{(x, f(x)) \mid x \in \mathbf{dom}\, f\},$$

which is a subset of $\mathbf{R}^{n+1}$.

- The *epigraph* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

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



- A function is convex if and only if its epigraph is a convex set.

Operations that preserve convexity

- A *nonnegative weighted sum* of convex functions $\sum_{i=1}^{m} \omega_i f_i$ is convex.
- If $f : \mathbf{R}^n \to \mathbf{R}$ is convex, then $g(x) = f(Ax + b)$ is convex.
- If $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.
- The pointwise maximum of affine functions

  $$f(x) = \max\{a_1^T x + b_1, \ldots, a_m^T x + b_m\}$$

  is convex. (This function is, in fact, piecewise linear.)
- The sum of r largest components of $x \in \mathbf{R}^n$ is convex.
- If $f(x, y)$ is convex for each $y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex.

Operations that preserve convexity (cont.)

- Let $C \subseteq \mathbf{R}^n$, $C \neq \emptyset$. The *support function* of C,

$$S_C(x) = \sup\{x^T y \mid y \in C\},$$

is convex.

- Also, the distance to the farthest point of C,

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

is convex.

- The *maximum eigenvalue* λ_{max} of $X \in \mathbf{S}^n$ is convex. Note that, one can express λ_{max} as

$$\lambda_{max} = \sup\{y^T X y \mid y \in \mathbf{R}^n, \|y\|_2 = 1\},$$

which is a pointwise supremum of linear functions of X.

Composition

Let $h : \mathbf{R}^p \to \mathbf{R}$, $g : \mathbf{R}^n \to \mathbf{R}^p$, and $f = h \circ g : \mathbf{R}^n \to \mathbf{R}$, meaning

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

Then we have the following composition rules:

- f is convex if h is convex, h is nondecreasing in each argument, and g_i are convex,
- f is convex if h is convex, h is nonincreasing in each argument, and g_i are concave,
- f is concave if h is concave, h is nondecreasing in each argument, and g_i are concave.

Some 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.

- If f is convex in (x, y), and $C \in \mathbf{R}^n$ is a convex nonempty set, then

$$g(x) = \inf_{y \in \mathbf{R}^n} f(x, y)$$

is convex in x, provided $g(x) > -\infty$ for all x. (Note that we define $\mathbf{dom}\, g = \{x \mid (x, y) \in \mathbf{dom}\, f \text{ for some } y \in C\}$.)

- For example, define the *distance* (w.t.r. some $\|\cdot\|$) of x to $S \subseteq \mathbf{R}^n$ as

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

As $\|x - y\|$ is convex in (x, y), $\mathbf{dist}(x, S)$ is convex in x (if S is convex).

Perspective of a function

- If $f : \mathbf{R}^n \rightarrow R$, then the *perspective* of f is $g : \mathbf{R}^{n+1} \rightarrow \mathbf{R}$ given by

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

- If f is a convex function, then so is its perspective function g.

- For example, if $f(x) = x^T x = \|x\|_2^2$, then its perspective is

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

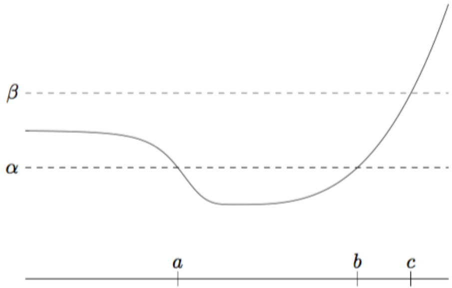
which is convex for $t > 0$.

Quasiconvex functions

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is called *quasiconvex* if $\mathbf{dom}\, f$ is convex and all its *sublevel sets*

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

  are convex for all $\alpha \in \mathbf{R}$.
- A function is *quasiconcave* if $-f$ is quasiconvex.
- Quasiconvex + quasiconcave = *quasilinear*.



Examples

- $f(x) = \sqrt{|x|}$ is quasiconvex on $\mathbf{R}$.
- $f(x) = \text{ceil}(x) = \inf\{z \in Z \mid z \geq x\}$ is quasilinear.
- $f(x) = \log x$ is quasilinear on $\mathbf{R}_{++}$.
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}^2_{++}$.
- The *linear-fractional function* function $f(x) = (a^T x + b)/(d^T x + c)$ is quasilinear on $\mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$.
- The *distance ratio function* $f(x) = \|x - a\|_2 / \|x - b\|_2$ is quasiconvex on $\mathbf{dom}\, f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$.

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is *log-concave* if $f(x) > 0$ for all $x \in \mathbf{dom}\, f$ and $\log f$ is concave.
- It is said to be *log-convex* if $\log f$ is convex.
- Thus, f is log-convex if and only if $1/f$ is log-concave.
- Some important examples include:
  - $f(x) = x^a$, on $\mathbf{R}_{++}$, is log-convex for $a \leq 0$, and log-concave for $a \geq 0$.
  - $f(x) = e^{ax}$ is log-convex and log-concave.
  - $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} \exp(-u^2/2) du$ is log-concave.
  - $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ is log-convex for $x \geq 1$.
  - $f(X) = \det X$ is log-concave on $\mathbf{S}^n_{++}$.
  - $f(X) = \det X / \mathbf{tr}\, X$ is log-concave on $\mathbf{S}^n_{++}$.