Convex Functions (III)

Lecture 5, Nonlinear Programming

National Taiwan University

October 18, 2016

Table of contents

Basic Operations that Preserve Convexity

- If f is convex and $\alpha \geq 0$, then αf is also convex.
- If both f_1 and f_2 are convex, then $f_1 + f_2$ is also convex.
- More generally, if $f_1, \ldots, f_n$ are convex functions, then any of their "conic combinations",

$$f = w_1 f_1 + \cdots + w_n f_n,$$

  is also convex (with $w_1, \ldots, w_n \geq 0$). This is also called the **nonnegative weighted sum**.
- Extension: if $f(x, y)$ is convex in x for any $y \in \mathcal{A}$, and $w(y) \geq 0$ for any $y \in \mathcal{A}$, then the function

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y) dy$$

  is convex in x.

Basic Operations that Preserve Convexity

- Suppose $f : \mathbf{R}^n \to \mathbf{R}$, $A \in \mathbf{R}^{n \times m}$, and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by
$$g(x) = f(Ax + b)$$
with $\mathbf{dom}\ g = \left\{ x \;\middle|\; Ax + b \in \mathbf{dom}\ f \right\}$.
- If f is convex, then g is also convex.
- If f is concave, so is g.

Pointwise maximum

- If f_1 and f_2 are convex functions then their **pointwise maximum** f, defined as

$$f(x) = \max\{f_1(x), f_2(x)\},$$

  with $\mathbf{dom}\ f = \mathbf{dom}\ f_1 \cap \mathbf{dom}\ f_2$, is also convex.
- Proof:

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \\ &\leq \\ &\leq \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

- It can be easily extended: if $f_1, \ldots, f_m$ are convex, then their pointwise maximum

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\},$$

  is also convex.

Pointwise maximum – Examples

Piecewise-linear functions

A **piecewise-linear** function $f(x) = \max\left\{a_1^T x + b_1, \ldots, a_L^T x + b_L\right\}$ is convex, since the affine functions $a_i^T x + b_i$ are all convex.

Sum of r largest components

For $x \in \mathbf{R}^n$, we denote by $x_{[i]}$ the ith largest component of x, i.e.,

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$$

are the components of x sorted in nonincreasing order $\left(\left\{x_{[1]}, \ldots, x_{[n]}\right\} = \{x_1, \ldots, x_n\}\right)$. Then the function $f(x) = \sum_{i=1}^{r} x_{[i]}$ is convex.

- Note that, as a generalization, the function $f(x) = \sum_{i=1}^{r} w_i x_{[i]}$ is also convex as long as $w_1 \geq w_2 \geq \ldots \geq w_r \geq 0$.

Pointwise supremum

- If for each $y \in \mathcal{A}$, $f(x, y)$ is convex in x, then the function g, defined as

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

  is convex in x. Here

$$\textbf{dom } g = \left\{ x \;\middle|\; (x, y) \in \textbf{dom } f, \quad \forall y \in \mathcal{A}, \quad \sup_{y \in \mathcal{A}} f(x, y) < \infty \right\}.$$

- Similarly, the pointwise infimum of a set of concave functions is a concave function.

Note: the **supremum** and **infimum** of a set is defined as

$$\sup \mathcal{A} = \min \{ y \mid y \geq x, \forall x \in \mathcal{A} \}, \text{(i.e., the minimum upper bound of } \mathcal{A})$$

$$\inf \mathcal{A} = \max \{ y \mid y \leq x, \forall x \in \mathcal{A} \}, \text{(i.e., the maximum lower bound of } \mathcal{A})$$

respectively.

Pointwise supremum

- In terms of epigraphs, the pointwise supremum of functions corresponds to the intersection of epigraphs: if

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

then we have

$$\textbf{epi } g = \bigcap_{y \in \mathcal{A}} \textbf{epi } f(\cdot, y).$$

- Thus, the result follows from the fact that the intersection of a family of convex sets is convex.

Pointwise supremum – Examples

Support function of a set

Let $C \subseteq \mathbf{R}^n$ with $C \neq \phi$. The **support function** S_C associated with the set C, defined as

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

with **dom** $S_C = \left\{ x \mid \sup_{y \in C} x^T y < \infty \right\}$, is convex.

Distance to farthest point of a set

Let $C \subseteq \mathbf{R}^n$. The distance (in any norm) to the farthest point of C,

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

is convex.

Pointwise supremum – Examples

Maximum eigenvalue of a symmetric matrix

The function $f(X) = \lambda_{max}(X)$, with **dom** $f = \mathbf{S}^m$, is convex.

Proof:

$$f(X) = \sup \left\{ y^T X y \mid ||y||_2 = 1 \right\}.$$

Norm of a matrix

The function $f(X) = ||X||_2$ with **dom** $f = \mathbf{R}^{p \times q}$, where $|| \cdot ||_2$ denotes the spectral norm or maximum singular value, is convex.

Proof:

$$f(X) = \sup \left\{ u^T X v \mid ||u||_2 = 1, ||v||_2 = 1 \right\},$$

is the pointwise supremum of a family of linear functions of X.

Convexity of composition of functions

Convexity of composition of functions

Let $h : \mathbf{R} \to \mathbf{R}$, and $g : \mathbf{R} \to \mathbf{R}$ and $f = h \circ g : \mathbf{R} \to \mathbf{R}$, $f(x) = h(g(x))$. Let $\mathbf{dom}\ f = \mathbf{dom}\ g = \mathbf{dom}\ h = \mathbf{R}$ and f, g, h be differentible. Then,

- f is convex if h is convex and nondecreasing, and g is convex,
- f is convex if h is convex and nonincreasing, and g is concave,
- f is concave if h is concave and nondecreasing, and g is concave,
- f is concave if h is concave and nonincreasing, and g is convex.

Proof (for the case where h and g are both twice differentiable):

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

Examples – Convexity of composition of functions

- If g is convex then $\exp g(x)$ is convex.
- If g is concave and positive, then $\log g(x)$ is concave.
- If g is concave and positive, then $1/g(x)$ is convex.
- If g is convex and nonnegative and $p \geq 1$, then $g(x)^p$ is convex.
- If g is convex then $-\log(-g(x))$ is convex on $\{x \mid g(x) < 0\}$.

A generalization

Convexity of composition of functions

Let $h : \mathbf{R} \to \mathbf{R}$, and $g : \mathbf{R}^n \to \mathbf{R}$ and $f = h \circ g : \mathbf{R}^n \to \mathbf{R}$, $f(x) = h(g(x))$. Let $\mathbf{dom}\ f = \mathbf{dom}\ g = \mathbf{R}^n, \mathbf{dom}\ h = \mathbf{R}$, and f, g, h be differentible. Then,

- f is convex if h is convex and nondecreasing, and g is convex,
- f is convex if h is convex and nonincreasing, and g is concave,
- f is concave if h is concave and nondecreasing, and g is concave,
- f is concave if h is concave and nonincreasing, and g is convex.

Proof idea: convexity is determined by the behavior of a function on arbitrary lines that intersect its domain.

Vector composition – A further generalization

Vector Composition

Suppose $f(x) = h(g(x)) = h(g_1(x), ..., g_k(x))$, with $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, ..., k$. Then,

- 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.

Proof: W.l.o.g., we can assume $n = 1$.

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x),$$

Vector composition examples

- Let $h(z) = z_{[1]} + \ldots + z_{[r]}$, the sum of the r largest components of $z \in \mathbf{R}^k$. Then h is convex and nondecreasing in each argument.
- Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbf{R}^n$. Then the composition function $f = h \circ g$, i.e., the pointwise sum of the r largest g_i's, is convex.
- The function $h(z) = \log(\sum_{i=1}^k e^{z_i})$ is convex and nondecreasing in each argument, so $\log(\sum_{i=1}^k e^{g_i})$ is convex whenever g_i are.
- For $0 < p \leq 1$, the function $h(z) = (\sum_{i=1}^k z_i^p)^{1/p}$ on $\mathbf{R}_+^k$ is concave, and its extension (which has the value $-\infty$ for $z \not\succeq 0$) is nondecreasing in each component. So if g_i are concave and nonnegative, we conclude that $f(x) = (\sum_{i=1}^k g_i(x)^p)^{1/p}$ is concave.

Vector composition examples

- Suppose $p \geq 1$, and $g_1, ..., g_k$ are convex and nonnegative. Then the function $(\sum_{i=1}^{k} g_i(x)^p)^{1/p}$ is convex.
    - Proof idea: The l_p-norm is convex, and is nondecreasing in each argument if the considered domain is $\textbf{dom}\ ||\cdot||_p = \mathbf{R}_+$.
- The geometric mean $h(z) = (\prod_{i=1}^{k} z_i)^{1/k}$ on $\mathbf{R}_+^k$ is concave and its extension is nondecreasing in each argument. It follows that if $g_1, ..., g_k$ are nonnegative concave functions, then so is their geometric mean,

$$\left(\prod_{i=1}^{k} g_i\right)^{1/k}.$$

Minimization

Minimization and convexity

If f is convex in (x, y), and C is a convex nonempty set, then

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

is convex in x, provided $g(x) > -\infty$ for some x (which implies $g(x) > -\infty$ for all x), with

$$\mathbf{dom}\ g = \{x \mid (x, y) \in \mathbf{dom}\ f,\ \exists y \in C\}.$$

- Proof: For $x_1, x_2 \in \mathbf{dom}\ g$. Let $\epsilon > 0$. Then $\exists y_1, y_2 \in C$ such that $f(x_i, y_i) \leq g(x_i) + \epsilon$ for $i = 1, 2$. For any $\theta, 0 \leq \theta \leq 1$, we have

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\leq \theta g(x_1) + (1-\theta) g(x_2) + \epsilon \end{aligned}$$

Minimization

Minimization and convexity

If f is convex in (x, y), and C is a convex nonempty set, then

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

is convex in x, provided $g(x) > -\infty$ for some x (which implies $g(x) > -\infty$ for all x), with

$$\textbf{dom}\ g = \{x \mid (x, y) \in \textbf{dom}\ f,\ \exists y \in C\}\,.$$

- Alternative proof (based on epigraph): Since

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

we have

$$\textbf{epi}\ g = \{(x, t) \mid (x, y, t) \in \textbf{epi}\ f,\ \ \exists y \in C\}\,.$$

Example – Distance to a set

- The distance of a point x to a set $S \subseteq \mathbf{R}^n$, in the norm $||\cdot||$, is defined as
$$\textbf{dist}\ (x, S) = \inf_{y \in S} ||x - y||.$$
- The function $||x - y||$ is convex in (x, y), so if the set S is convex, the distance function $\textbf{dist}\ (x, S)$ is a convex function of x.

Example

- Suppose h is convex. Then the function g defined as

$$g(x) = \inf \{h(y) \mid Ay = x\}$$

  is convex.

- Proof: we define f by

$$f(x, y) = \begin{cases} h(y) & \text{if } Ay = x \\ \infty & \text{otherwise} \end{cases}$$

  which is convex in (x, y). Then g is the minimum of f over y, and hence is convex. (It is not hard to show directly that g is convex.)

Conjugate functions

Conjugate functions

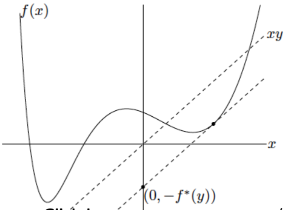
Let $f : \mathbf{R}^n \to \mathbf{R}$. The function $f^* : \mathbf{R}^n \to \mathbf{R}$, defined as

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

is called the **conjugate** of the function f. The domain of f^* is

$$\mathbf{dom}\ f^* = \left\{ y \in \mathbf{R}^n \mid \exists z \in \mathbf{R} \text{ s.t. } \forall x \in \mathbf{dom}\ f,\ y^T x - f(x) < z \right\}$$

Example:
$f : \mathbf{R}^1 \to \mathbf{R}, f^* : \mathbf{R}^1 \to \mathbf{R}$



Example – Revenue and Profit Functions

- Let $r = (r_1, \dots, r_n)$ denote the vector of resource quantities consumed, $S(r)$ denote the sales revenue derived from the product produced, $p = (p_1, \dots, p_n)$ denote the vector of unit prices of resources.
- Then the profit is

$$S(r) - p^T r.$$

- Given the price vector p, the maximum profit is given by

$$M(p) = \sup_r \left(S(r) - p^T r \right),$$

or

$$M(p) = (-S)^*(-p).$$

Conjugate functions

- Conjugate functions

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

are convex.

  - $\because$ it is the pointwise supremum of a family of convex (indeed, affine) functions of y.
  - This is true whether or not f is convex.
  - Note that when f is convex, the subscript $x \in \mathbf{dom}\ f$ is not necessary since $y^T x - f(x) = -\infty$ for $x \notin \mathbf{dom}\ f$.

Conjugate Functions – Examples for $f : \mathbf{R} \to \mathbf{R}$

- Affine function $f(x) = ax + b$. The function, $yx - ax - b$ is bounded if and only if $y = a$. Therefore $\mathbf{dom}\ f^* = \{a\}$, and $f^*(a) = -b$.
- Negative logarithm. $f(x) = -\log x$, with $\mathbf{dom}\ f = \mathbf{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathbf{dom}\ f^* = \{y \mid y < 0\} = -\mathbf{R}_{++}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.
- Exponential. $f(x) = e^x$. $xy - e^x$ is unbounded if $y < 0$. It can be shown that $\mathbf{dom}\ f^* = \mathbf{R}_+$ and

$$f^*(y) = \begin{cases} y \log y - y, & y > 0 \\ 0, & y = 0 \end{cases}.$$

Conjugate Functions – Examples for $f : \mathbf{R} \to \mathbf{R}$

- Negative entropy. $f(x) = x \log x$, with $\mathbf{dom}\ f = \mathbf{R}_+$ (and $f(0) = 0$). The function $xy - x \log x$ is bounded above on $\mathbf{R}_+$ for all y, hence $\mathbf{dom}\ f^* = \mathbf{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^*(y) = e^{y-1}$.
- Inverse. $f(x) = 1/x$ on $\mathbf{R}_{++}$. For $y > 0$, $yx - 1/x$ is unbounded above. For $y = 0$ this function has supremum 0; for $y < 0$ the supremum is attained at $x = (-y)^{-1/2}$. Therefore we have $f^*(y) = -2(-y)^{1/2}$, with $\mathbf{dom}\ f^* = -\mathbf{R}_+$.

Conjugate Functions – Examples for $f : \mathbf{R}^n \to \mathbf{R}$

- Strictly convex quadratic function. Consider $f(x) = \frac{1}{2}x^T Qx$, with $Q \in \mathbf{S}^n_{++}$. The function $y^T x - \frac{1}{2}x^T Qx$ is bounded above as a function of x for all y. It attains its maximum at $x = Q^{-1}y$, so
$$f^*(y) = \frac{1}{2}y^T Q^{-1}y.$$
- Log-sum-exp function. Consider
$$f(x) = \log\left(\sum_{i=1}^n e^{x_i}\right).$$
Then, **dom** $f^* = \{y \mid \mathbf{1}^T y = 1, y \succeq 0\}$ and
$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i, & \text{if } y \succeq 0 \text{ and } \mathbf{1}^T y = 1 \\ \infty & \text{otherwise} \end{cases}.$$

Conjugate Functions – Examples for $f : \mathbf{S}^n_{++} \to \mathbf{R}$

- Log-determinant. We consider $f(X) = \log\det X^{-1}$ on $\mathbf{S}^n_{++}$. The conjugate function is defined as

$$f^*(Y) = \sup_{X \succ 0} \left(\mathbf{tr}\,(YX) + \log\det X\right),$$

since $\mathbf{tr}\,(YX)$ is the standard inner product on $\mathbf{S}^n$. It can be shown that $\mathbf{dom}\ f^* = -\mathbf{S}^n_{++}$ and

$$f^*(Y) = \log\det(-Y)^{-1} - n.$$

Quasiconvex functions

Quasiconvex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is called quasiconvex if its domain and all its sublevel sets

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

for $\alpha \in \mathbf{R}$, are convex sets.

Quasiconcave and quasilinear functions

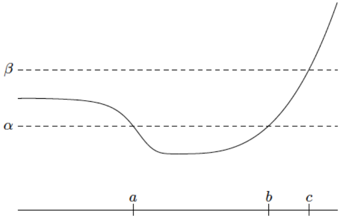
Quasiconcave and quasilinear functions

- A function is **quasiconcave** if $-f$ is quasiconvex, i.e., every superlevel set $\{x|f(x) \geq \alpha\}$ is convex.
- A function that is both quasiconvex and quasiconcave is called **quasilinear**.

- If a function f is quasilinear, then its domain, and every **level set** $\{x \mid f(x) = \alpha\}$ is convex.

Convex functions are quasiconvex functions

- For a function on $\mathbf{R}$, quasiconvexity requires that each sublevel set be an interval (including an infinite interval).
- Convex functions have convex sublevel sets, and so are quasiconvex. But the converse is not true.



Quasiconvex functions – Examples

Some examples on $\mathbf{R}$:

- Logarithm. $\log x$ on $\mathbf{R}_{++}$ is quasiconvex (and quasiconcave, hence quasilinear).
- Ceiling function. $\operatorname{ceil}(x) = \inf \{z \in Z | z \geq x\}$ is quasiconvex (and quasiconcave).

An example on $\mathbf{R}^n$:

- The **length** of $x \in \mathbf{R}^n$, defined as the largest index of a nonzero component, i.e.,

$$f(x) = \begin{cases} \max \{i \mid x_i \neq 0\} & x \neq 0 \\ 0 & x = 0 \end{cases},$$

  is quasiconvex.

Quasiconvex functions – Examples

- Consider $f : \mathbf{R}^2 \to \mathbf{R}$, with $\mathbf{dom}\ f = \mathbf{R}^2_+$ and $f(x_1, x_2) = x_1 x_2$. Then, f is neither convex nor concave since

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

  has eigenvalues ± 1 (not definite).

- But f is quasiconcave on $\mathbf{R}^2_+$, since the superlevel sets

$$\left\{x \in \mathbf{R}^2_+ \mid x_1 x_2 \geq \alpha\right\}$$

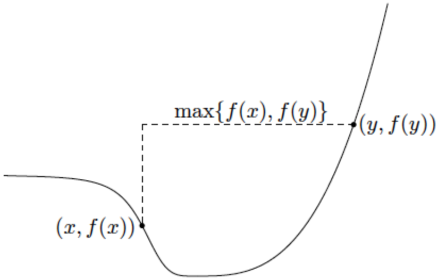
  are convex sets for all α.

Quasiconvex functions – Basic Properties

Jensen's inequality for quasiconvex functions

A function f is quasiconvex if and only if **dom** f is convex and for any $x, y \in \mathbf{dom}\ f$ and $0 \leq \theta \leq 1$,

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

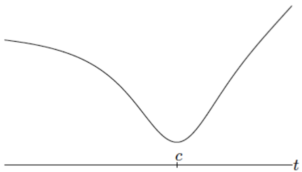


Quasiconvex functions – Basic Properties

Continuous quasiconvex functions on $\mathbf{R}$

A continuous function $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex if and only if at least one of the following conditions holds:

- f is nondecreasing.
- f is nonincreasing.
- there is a point $c \in \textbf{dom } f$ such that for $t \leq c$ (and $t \in \textbf{dom } f$), f is nonincreasing, and for $t \geq c$ (and $t \in \textbf{dom } f$), f is nondecreasing.



Differentiable quasiconvex functions

First-Order Conditions

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable. Then f is quasiconvex if and only if $\mathbf{dom}\ f$ is convex and for all $x, y \in \mathbf{dom}\ f$

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

Proof Idea: It suffices to prove the result for a function on $\mathbf{R}$; the general result follows by restriction to an arbitrary line.

Representation via family of convex functions

Representation via family of convex functions

We can always find a family of convex functions $\phi_t : \mathbf{R}^n \to \mathbf{R}$, indexed by $t \in \mathbf{R}$, with

$$f(x) \leq t \iff \phi_t(x) \leq 0,$$

i.e., the t-sublevel set of the quasiconvex function f is the 0-sublevel set of the convex function ϕ_t.

- Evidently ϕ_t must satisfy the property that for all $x \in \mathbf{R}^n$, $\phi_t(x) \leq 0 \Rightarrow \phi_s(x) \leq 0$ for $s \geq t$. This is satisfied if for each x, $\phi_t(x)$ is a nonincreasing function of t, i.e., $\phi_s(x) \leq \phi_t(x)$ whenever $s \geq t$.
- One (straightforwards) example:

$$\phi_t(x) = \begin{cases} 0 & f(x) \leq t \\ \infty & \text{otherwise,} \end{cases}$$

- Another example: if the sublevel sets of f are closed, we can take

$$\phi_t(x) = \mathbf{dist}\ (x, \{z | f(z) \leq t\}).$$

We are usually interested in a family ϕ_t with nice properties, such as differentiability.

Log-convex and log-concave functions

Log-convex and log-concave functions

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is **logarithmically concave** or **log-concave** if $f(x) > 0$ for all $x \in \textbf{dom } f$ and $\log f$ is concave.
- It is said to be **logarithmically convex** or **log-convex** if $\log f$ is convex.

- f is log-convex if and only if $1/f$ is log-concave.

Log-concavity

A function $f : \mathbf{R}^n \to \mathbf{R}$, with convex domain and $f(x) > 0$ for all $x \in \textbf{dom } f$, is **log-concave** if and only if $\forall x, y \in \textbf{dom } f$ and $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta}$$

- The value of a log-concave function at the average of two points is at least the geometric mean of the values at the two points.

Log-convex and log-concave functions – Some Properties

- A log-convex function is convex (since e^h is convex if h is convex).
- A nonnegative concave function is log-concave.
- A log-convex function is quasiconvex; a log-concave function is quasiconcave (since the logarithm is monotone increasing).

Optimization Problems

- The notation

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p \end{array}$$

  is used to describe an optimization problem of finding an x that minimizes $f_0(x)$ among all x that satisfy the conditions $f_i(x) \leq 0, i = 1, ..., m$ and $h_i(x) = 0, i = 1, ..., p$.
- $x \in \mathbf{R}^n$: the **optimization variables**.
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: the **objective function**.
- $f_i : \mathbf{R}^n \to \mathbf{R}$: the **inequality constraint functions**.
  - $f_i(x) \leq 0$: the **inequality constraints**.
- $h_i : \mathbf{R}^n \to \mathbf{R}$: the **equality constraint functions**.
  - $h_i(x) = 0$: the **equality constraints**.

Optimization Problems

Optimization Problems

Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array}$$

- The set
$$\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$$
is called the **domain** of the problem.
- A point $x \in \mathcal{D}$ is **feasible** if $f_i(x) \leq 0$ for all $i = 1, \ldots, m$ and $h_i(x) = 0$ for all $i = 1, \ldots, p$.
- The problem is called **feasible** if there exists $x \in \mathcal{D}$ that is feasible; the problem is called **infeasible** if there is no feasible point in $\mathcal{D}$.
- The set of all feasible points is called the **feasible set**.
- If there are no constraints (i.e., $m = p = 0$), then the feasible set equals $\mathcal{D} = \textbf{dom } f_0$, and the problem is called **unconstrained**.

Optimization Problems – Optimal Values

Optimal Values

In the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p, \end{array}$$

- the **optimal value** p^* is defined as

$$p^* = \inf \{ f_0(x) \mid f_i(x) \leq 0, i = 1, ..., m, h_i(x) = 0, i = 1, ..., p \}.$$

- If the problem is infeasible, we have $p^* = \infty$.
- If there are feasible points x_k with $f_0(x_k) \to -\infty$ as $k \to \infty$, then $p^* = -\infty$, and the problem is said to be **unbounded below**.

Optimization Problems – Optimal Points

Optimal Point

Suppose the optimal value of the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

is p^*. Then we say x^* is an **optimal point** if

- x^* is feasible, and
- $f_0(x^*) = p^*$.

- The set of all optimal points is the **optimal set**, denoted

$$X_{opt} = \{x \mid f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p, f_0(x) = p^*\}.$$

Optimization Problems – Optimal Points

- If there exists an optimal point for an optimal problem, we say the optimal value is **attained** or **achieved**, and the problem is **solvable**.
- If X_{opt} is empty, we say the optimal value is not attained or not achieved.
  - e.g., this always occurs when the problem is unbounded below.
- A feasible point x with $f_0(x) \leq p^* + \epsilon$ (where $\epsilon > 0$) is called **ϵ-suboptimal**.
  - The set of all ϵ-suboptimal points is called the **ϵ-suboptimal set** for the optimization problem.

Optimization Problem

- We say a feasible point x is **locally optimal** if there exists an $R > 0$ such that

$$f_0(x) = \inf \{ f_0(z) \mid f_i(z) \leq 0, i = 1, ..., m, h_i(z) = 0, i = 1, ..., p, ||z - x||_2 \leq R \} .$$

  - This means x minimizes f_0 over nearby points in the feasible set.

- If x is feasible and $f_i(x) = 0$, we say the ith inequality constraint $f_i(x) \leq 0$ is **active** at x.
- If $f_i(x) < 0$, we say the constraint $f_i(x) \leq 0$ is **inactive**.
- We say that a constraint is **redundant** if deleting it does not change the feasible set.

Optimization Problems – Examples

We consider the following unconstrained problems as examples, with $f_0 : \mathbf{R} \to \mathbf{R}$ and $\mathbf{dom}\ f_0 = \mathbf{R}_{++}$. Recall that

$$p^* = \inf \{ f_0(x) \mid x \text{ is feasible} \}.$$

- $f_0(x) = 1/x$: $p^* = 0$, but the optimal value is not achieved.
- $f_0(x) = -\log x$: $p^* = -\infty$, so this problem is unbounded below.
- $f_0(x) = x \log x$: $p^* = -1/e$, achieved at the (unique) optimal point $x^* = 1/e$.

Feasibility problems

- If the objective function is identically zero, the optimal value is either
  - 0, if the feasible set is nonempty, or
  - ∞, if the feasible set is empty.
- We call this the **feasibility problem**, and will sometimes write it as

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \leq 0, i = 1, ..., m \\ & h_i(x) = 0, i = 1, ..., p. \end{array}$$

- The feasibility problem is thus to determine whether the constraints are consistent, and if so, find a point that satisfies them.

Expressing Problems in Standard Forms

- An optimization problem in the form of

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p, \end{array}$$

  is called in the **standard form**, i.e., the righthand side of the inequality and equality constraints are zeros.
- An equality constraint in a non-standard form $g_i(x) = \tilde{g}_i(x)$ can be reformulated as $h_i(x) = 0$ where $h_i(x) = g_i(x) - \tilde{g}_i(x)$.
- An inequality constraint of the form $f_i(x) \geq 0$ can be rewritten as $-f_i(x) \leq 0$.

Expressing Problems in Standard Forms – Examples

The optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i \leq x_i \leq u_i, \quad i = 1, \ldots, n \end{array}$$

can be expressed in standard form as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i - x_i \leq 0 \quad i = 1, \ldots, n \\ & x_i - u_i \leq 0 \quad i = 1, \ldots, n \end{array}$$

There are $2n$ inequality constraint functions:

$$f_i(x) = l_i - x_i \quad i = 1, \ldots, n,$$

and

$$f_i(x) = x_{i-n} - u_{i-n} \quad i = n+1, \ldots, 2n.$$

Expressing Problems in Standard Forms – Examples

The **maximization problem**

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

can be solved by minimizing the function $-f_0(x)$ subject to the same constraints.

Equivalent Problems

We call two problems **equivalent** (informally) if from a solution of one, a solution of the other is readily found, and vice versa.

Example

$$\begin{array}{ll} \text{minimize} & \tilde{f}(x) = \alpha_0 f_0(x) \\ \text{subject to} & \tilde{f}_i(x) = \alpha_i f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(x) = \beta_i h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

(where $\alpha_i > 0, i = 0, \ldots, m,\ \beta_i \neq 0, i = 1, \ldots, p$) and

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

are equivalent problems.

Change of Variables

- Suppose $\phi : \mathbf{R}^n \to \mathbf{R}^n$ is one-to-one, with image covering the problem domain $\mathcal{D}$, i.e., $\mathcal{D} \subseteq \phi(\textbf{dom } \phi)$.
- Now consider the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(z) \\ \text{subject to} & \tilde{f}_i(z) \leq 0, i = 1, \ldots, m \\ & \tilde{h}_i(z) = 0, i = 1, \ldots, p, \end{array}$$

  with variable z, where we define functions $\tilde{f}_i$ and $\tilde{h}_i$ as $\tilde{f}_i(z) = f_i(\phi(z)), i = 0, \ldots, m$, $\tilde{h}_i(z) = h_i(\phi(z)), i = 1, \ldots, p$.
- Then, we say that the problem and the standard form problem are equivalent and related by the change of variable or substitution of variable $x = \phi(z)$.

Transformation of objective and constraint functions

- Suppose that
  - $\phi_0 : \mathbf{R} \to \mathbf{R}$ is monotone increasing,
  - $\phi_1, \ldots, \phi_m : \mathbf{R} \to \mathbf{R}$ satisfy $\phi_i(u) \leq 0$ if and only if $u \leq 0$, and
  - $\phi_{m+1}, \ldots, \phi_{m+p} : \mathbf{R} \to \mathbf{R}$ satisfy $\phi_i(u) = 0$ if and only if $u = 0$.
- We define functions $\tilde{f}_i$ and $\tilde{h}_i$ as the compositions
  - $\tilde{f}_i(x) = \phi_i(f_i(x)), i = 0, \ldots, m,$
  - $\tilde{h}_i(x) = \phi_{m+i}(h_i(x)), i = 1, \ldots, p.$
- Then, the associated problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x) \\ \text{subject to} & \tilde{f}_i(x) \leq 0, i = 1, \ldots, m \\ & \tilde{h}_i(x) = 0, i = 1, \ldots, p \end{array}$$

and the standard form problem are equivalent.

Slack variables

- Observation: $f_i(x) \leq 0$ if and only if there is an $s_i \geq 0$ that satisfies $f_i(x) + s_i = 0$.
- Based on the observation we obtain the transformed problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & s_i \geq 0, i = 1, \ldots, m \\ & f_i(x) + s_i = 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p, \end{array}$$

  where the variables are $x \in \mathbf{R}^n$ and $s \in \mathbf{R}^m$.
- This problem has $n + m$ variables, m inequality constraints (the nonnegativity constraints on s_i), and $m + p$ equality constraints.
- The new variable s_i is called the **slack variable** associated with the original inequality constraint $f_i(x) \leq 0$.

Eliminating equality constraints

- Suppose the function $\phi : \mathbf{R}^k \to \mathbf{R}^n$ is such that x satisfies $h_i(x) = 0, i = 1, ..., p$ if and only if there is some $z \in \mathbf{R}^k$ such that $x = \phi(z)$.
- Then, the optimization problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(z) = f_0(\phi(z)) \\ \text{subject to} & \tilde{f}_i(z) = f_i(\phi(z)) \leq 0, i = 1, ..., m \end{array}$$

  is then equivalent to the original standard form problem.
- This transformed problem has variable $z \in \mathbf{R}^k$, m inequality constraints, and no equality constraints.
- If z is optimal for the transformed problem, then $x = \phi(z)$ is optimal for the original problem.
- Conversely, if x is optimal for the original problem, then any z that satisfies $x = \phi(z)$ is optimal for the transformed problem.

Eliminating linear equality constraints

- Consider the standard form problem with linear equality constraints

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b. \end{array}$$

- Suppose $Ax = b$ is consistent. Then the solution set of $Ax = b$ can be parametrized as $\left\{Fz + x_0 \mid z \in \mathbf{R}^k\right\}$ where $F \in \mathbf{R}^{n \times k}$ is chosen to be any full rank matrix with $\mathcal{R}(F) = \mathcal{N}(A)$ (i.e., $k = n - \operatorname{rank} A$), and x_0 is any particular solution of $Ax = b$.
- Then we can eliminate these linear constraints and create an equivalent problem, as in

$$\begin{array}{ll} \text{minimize} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, i = 1, \ldots, m, \end{array}$$

where we introduced new variables $z \in \mathbf{R}^k$.

Introducing equality constraints (1/2)

- We can also introduce equality constraints and new variables into a problem.
- As a typical example, consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(A_0x + b_0) \\ \text{subject to} & f_i(A_ix + b_i) \leq 0, i = 1, ..., m \\ & h_i(x) = 0, i = 1, ..., p, \end{array}$$

where $x \in \mathbf{R}^n, A_i \in \mathbf{R}^{k_i \times n}$, and $f_i : \mathbf{R}^{k_i} \to \mathbf{R}$. In this problem the objective and constraint functions are given as compositions of the functions f_i with affine transformations defined by $A_ix + b_i$.

Introducing equality constraints (2/2)

- We introduce new variables $y_i \in \mathbf{R}^{k_i}$, as well as new equality constraints $y_i = A_i x + b_i$, for $i = 0, \ldots, m$, and form the equivalent problem

$$\begin{array}{ll} \text{minimize} & f_0(y_0) \\ \text{subject to} & f_i(y_i) \leq 0, i = 1, \ldots, m \\ & y_i = A_i x + b_i, i = 0, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p. \end{array}$$

- This problem has $k_0 + \ldots + k_m$ new variables, $y_0 \in \mathbf{R}^{k_0}, \ldots, y_m \in \mathbf{R}^{k_m}$, and $k_0 + \ldots + k_m$ new equality constraints, $y_0 = A_0 x + b_0, \ldots, y_m = A_m x + b_m$.
- The objective and inequality constraints in this problem are independent, i.e., involve different optimization variables.

Optimizing over some variables (1/2)

- Note that we always have

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

  where $\tilde{f}(x) = \inf_y f(x,y)$.
- Therefore, we can always minimize a function by first minimizing over some of the variables, and then minimizing over the remaining ones.

Optimizing over some variables (2/2)

- Suppose the variable $x \in \mathbf{R}^n$ is partitioned as $x = (x_1, x_2)$, with $x_1 \in \mathbf{R}^{n_1}$, $x_2 \in \mathbf{R}^{n_2}$, and $n_1 + n_2 = n$. Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \leq 0, i = 1, \ldots, m_1 \\ & \tilde{f}_i(x_2) \leq 0, i = 1, \ldots, m_2, \end{array}$$

  in which the constraints are independent, in the sense that each constraint function depends on x_1 or x_2.
- We first minimize over x_2. Define the function $\tilde{f}_0$ of x_1 by

$$\tilde{f}_0(x_1) = \inf \left\{ f_0(x_1, z) \mid \tilde{f}_i(z) \leq 0, i = 1, \ldots, m_2 \right\}.$$

  Then the problem is equivalent to

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \leq 0, i = 1, \ldots, m_1. \end{array}$$

Epigraph problem form (1/2)

- The epigraph form of the standard problem is the problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, i = 1, ..., m \\ & h_i(x) = 0, i = 1, ..., p, \end{array}$$

  with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$.
- It is equivalent to the original problem: (x, t) is optimal for the epigraph form problem if and only if x is optimal for the original problem and $t = f_0(x)$.

Epigraph problem form (2/2)

- Note that the objective function of the epigraph form problem is a linear function of the variables x, t.
- The epigraph form problem can be interpreted geometrically as an optimization problem in the 'graph space' (x, t):