Convex Optimization (I)

Lecture 6, Nonlinear Programming

National Taiwan University

October 25, 2016

Table of contents

Conjugate functions

Conjugate functions

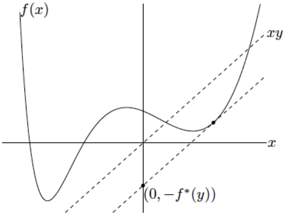
Let $f : \mathbf{R}^n \to \mathbf{R}$. The function $f^* : \mathbf{R}^n \to \mathbf{R}$, defined as

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

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



Example – Revenue and Profit Functions

- Let $r = (r_1, \ldots, r_n)$ denote the vector of resource quantities consumed, $S(r)$ denote the sales revenue derived from the product produced, $p = (p_1, \ldots, p_n)$ 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

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

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

Quasiconcave and quasilinear functions

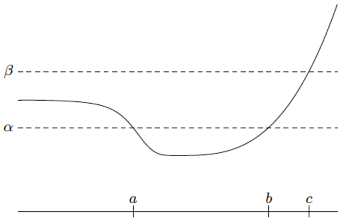
Quasiconcave and quasilinear functions

- 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 $\textbf{dom } f = \mathbf{R}^2_+$ 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

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

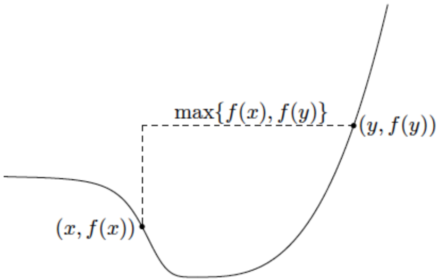
are convex sets for all α.

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 $x, y \in$ **dom** f and $0 \leq \theta \leq 1$,

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

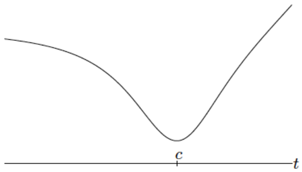


Quasiconvex functions – Basic Properties

Continuous quasiconvex functions on 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 $c \in \mathbf{dom}\ f$ such that for $t \leq c$ (and $t \in \mathbf{dom}\ f$), f is nonincreasing, and for $t \geq c$ (and $t \in \mathbf{dom}\ f$), 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 **dom** f is convex and for all $x, y \in$ **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

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

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 $x \in \mathbf{dom}\ f$ 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 $x \in \mathbf{dom}\ f$, is **log-concave** if and only if $\forall x, y \in \mathbf{dom}\ f$ 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

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

- 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

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

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

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

- 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

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

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

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

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

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

- 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

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

  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**

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

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

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

  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., $k = n - \text{rank } A$), 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

$$
\begin{array}{ll}
\text{minimize} & f_0(A_0 x + b_0) \\
\text{subject to} & f_i(A_i x + b_i) \leq 0, i = 1, ..., m \\
 & h_i(x) = 0, i = 1, ..., p,
\end{array}
$$

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 $A_i x + b_i$.

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 $i = 0, \dots, m$, and form the equivalent problem

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

- This problem has $k_0 + \dots + k_m$ new variables, $y_0 \in \mathbf{R}^{k_0}, \dots, y_m \in \mathbf{R}^{k_m}$, and $k_0 + \dots + k_m$ new equality constraints, $y_0 = A_0 x + b_0, \dots, y_m = A_m x + b_m$.
- 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

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

  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

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

  Then the problem is equivalent to

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

Epigraph problem form (1/2)

- The epigraph form of the standard problem is the problem

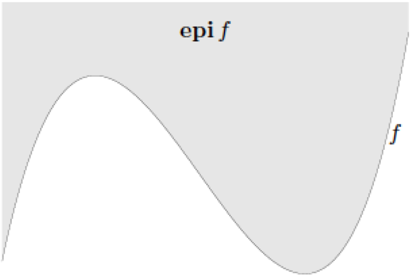
$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, i = 1, ..., m \\ & h_i(x) = 0, i = 1, ..., p, \end{array}$$

  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):



Convex optimization problems in standard form

A convex optimization problem is one of the form

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

where $f_0, \ldots, f_m$ are convex functions. Compared with the general standard form problem, the convex problem has three additional requirements:

- the objective function must be convex,
- the inequality constraint functions must be convex,
- the equality constraint functions $h_i(x) = a_i^T x - b_i$ must be affine.

Convex optimization problems in standard form

- The feasible set of a convex optimization problem is convex, since it is the intersection of
  - the domain of the problem

$$D = \bigcap_{i=0}^{m} \textbf{dom } f_i,$$

    (which is a convex set),
  - m (convex) sublevel sets $\{x \mid f_i(x) \leq 0\}$, and
  - p hyperplanes $\left\{x \mid a_i^T x = b_i\right\}$.
    - W.l.o.g., we assume that $a_i \neq 0$.
- In a convex optimization problem, we minimize a convex objective function over a convex set.

Quasiconvex Optimization Problems

- If f_0 is quasiconvex instead of convex, the problem

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, ..., m \\ & a_i^T x = b_i, i = 1, ..., p, \end{aligned}$$

  is called a (standard form) quasiconvex optimization problem.
- Since the sublevel sets of a convex or quasiconvex function are convex, we conclude that for a convex or quasiconvex optimization problem the ϵ-suboptimal sets are convex.
- In particular, the optimal set is convex.

Concave maximization problems

- We also refer to

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, ..., m \\ & a_i^T x = b_i, i = 1, ..., p, \end{array}$$

  as a convex optimization problem if the objective function f_0 is concave, and the inequality constraint functions $f_1, ..., f_m$ are convex.
- This concave maximization problem is readily solved by minimizing the convex objective function $-f_0$.
  - All of the results, conclusions, and algorithms that we describe for the minimization problem are easily transposed to the maximization case.
- In a similar way the above maximization problem is called quasiconvex if f_0 is quasiconcave.

Definition of Convex Optimization Problem

A closer look

- Consider the example with $x \in \mathbf{R}^2$,

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0, \end{array}$$

  which is in the standard form.
- This problem is not a convex optimization problem in standard form since the equality constraint function h_1 is not affine, and the inequality constraint function f_1 is not convex.
- Nevertheless the feasible set, which is $\{x \mid x_1 \leq 0, x_1 + x_2 = 0\}$, is convex.
- The problem, although not in a form of convex optimization problem, can be easily transformed to, and be shown to be equivalent to, a convex optimization problem.

Local and global optima (1/2)

- As an important property of convex optimization problems, any locally optimal point is also (globally) optimal.
- To see this, suppose that x is locally optimal for a convex optimization problem, i.e., x is feasible and

$$f_0(x) = \inf \left\{ f_0(z) \;\middle|\; z \text{ feasible}, ||z - x||_2 \leq R \right\},$$

  for some $R > 0$.
- Now suppose that x is not globally optimal, i.e., there is a feasible y such that $f_0(y) < f_0(x)$. Evidently $||y - x||_2 > R$, since otherwise $f_0(x) \leq f_0(y)$.

Local and global optima (2/2)

- Consider the point z given by

$$z = (1-\theta)x + \theta y, \theta = \frac{R}{2||y-x||_2}.$$

  Then we have $||z-x||_2 = R/2 < R$, and by convexity of the feasible set, z is feasible.

- By convexity of f_0 we have

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y) < f_0(x),$$

  which leads to a contradiction. So, x is globally optimal.

- It is not true that locally optimal points of quasiconvex optimization problems are globally optimal (to be shown later).

An optimality criterion for differentiable f_0

- Suppose that the objective f_0 in a convex optimization problem is differentiable, so that for all $x, y \in \textbf{dom } f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (y - x).$$

- Let X denote the feasible set, i.e.,

$$X = \{x \mid f_i(x) \leq 0, i = 1, ..., m, h_i(x) = 0, i = 1, ..., p\}.$$

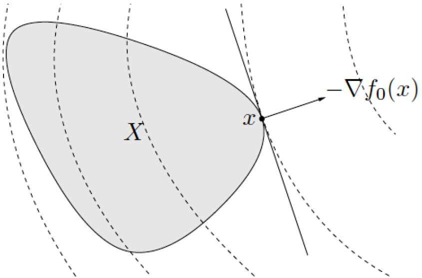
Then x is optimal if and only if $x \in X$ and

$$\nabla f_0(x)^T (y - x) \geq 0$$

for all $y \in X$.

An optimality criterion for differentiable f_0

- The optimality criterion can be understood geometrically: If $\nabla f_0(x) \neq 0$, it means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x.



Proof of optimality condition

- The "if" part is obvious.
- For the "only if" part, suppose x is optimal, but the optimality condition $\nabla f_0(x)^T(y-x) \geq 0$ does not hold, i.e., for some $y \in X$ we have
$$\nabla f_0(x)^T(y-x) < 0.$$
- Consider the point $z(t) = ty + (1-t)x$, where $t \in [0,1]$ is a parameter. Since $z(t)$ is on the line segment between x and y, and the feasible set is convex, $z(t)$ is feasible. Note that
$$\left[\frac{d}{dt} f_0(z(t))\right]\Bigg|_{t=0} = \nabla f_0(x)^T(y-x) < 0,$$
so for small positive t, we have $f_0(z(t)) < f_0(x)$, which proves that x is not optimal.

Unconstrained problems

- For an unconstrained problem (i.e., $m = p = 0$), the optimality condition

$$\nabla f_0(x)^T(y - x) \geq 0$$

reduces to the well known necessary and sufficient condition

$$\nabla f_0(x) = 0$$

for x to be optimal.

Unconstrained problems

- To see this, suppose x is optimal, which means here that $x \in \textbf{dom } f_0$, and for all feasible y we have $\nabla f_0(x)^T(y-x) \geq 0$. Since f_0 is differentiable, its domain is (by definition) open, so all y sufficiently close to x are feasible.
- Let us take $y = x - t\nabla f_0(x)$. Then for t small and positive, y is feasible, and so

$$\nabla f_0(x)^T(y-x) = -t||\nabla f_0(x)||_2^2 \geq 0,$$

  from which we conclude $\nabla f_0(x) = 0$.
- If $\nabla f_0(x) = 0$ has no solutions, then there are no optimal points, possibly
  - the problem is unbounded below, or
  - the optimal value is finite, but not attained.
- On the other hand, $\nabla f_0(x) = 0$ can have multiple solutions.
  - In this case, each such solution is a minimizer of f_0.

Example – Unconstrained quadratic optimization.

- Consider the problem of minimizing the quadratic function
$$f_0(x) = (1/2)x^T Px + q^T x + r,$$
where $P \in \mathbf{S}_+^n$ (which makes f_0 convex).
- The necessary and sufficient condition for x to be a minimizer of f_0 is
$$\nabla f_0(x) = Px + q = 0.$$
- Several cases can occur, depending on whether this (linear) equation has no solutions, one solution, or many solutions.
  - If $q \notin \mathcal{R}(P)$, then there is no solution. In this case f_0 is unbounded below.
  - If $P \succ 0$ (which is the condition for f_0 to be strictly convex), then there is a unique minimizer, $x^* = -P^{-1}q$.
  - If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is the (affine) set $X_{opt} = -P^\dagger q + \mathcal{N}(P)$, where $P^\dagger$ denotes the pseudo-inverse of P.

Problems with equality constraints only (1/2)

- Consider the case where there are equality constraints but no inequality constraints, i.e.,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b. \end{array}$$

  Here the feasible set is affine. We assume that it is nonempty.
- The optimality condition for a feasible x is that

$$\nabla f_0(x)^T (y - x) \geq 0$$

  must hold for all y satisfying $Ay = b$.
- Since x is feasible, every feasible y has the form $y = x + v$ for some $v \in \mathcal{N}(A)$. The optimality condition can therefore be expressed as: $\nabla f_0(x)^T v \geq 0$ for all $v \in \mathcal{N}(A)$.

Problems with equality constraints only (2/2)

- If a linear function is nonnegative on a subspace, then it must be zero on the subspace, so it follows that $\nabla f_0(x)^T v = 0$ for all $v \in \mathcal{N}(A)$. In other words, $\nabla f_0(x) \perp \mathcal{N}(A)$.
- Using the fact that $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$, this optimality condition can be expressed as $\nabla f_0(x) \in \mathcal{R}(A^T)$, i.e., there exists a $\nu \in \mathbf{R}^p$ such that

$$\nabla f_0(x) + A^T \nu = 0.$$

  Together with the requirement $Ax = b$ (i.e., that x is feasible), this is the classical Lagrange multiplier optimality condition.

Minimization over the nonnegative orthant (1/2)

- We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \succeq 0, \end{array}$$

where the only inequality constraints are nonnegativity constraints on the variables. The optimality condition is then

$$x \succeq 0, \quad \nabla f_0(x)^T (y - x) \geq 0 \text{ for all } y \succeq 0.$$

- The term $\nabla f_0(x)^T y$, which is a linear function of y, is unbounded below on $y \succeq 0$, unless we have $\nabla f_0(x) \succeq 0$.

Minimization over the nonnegative orthant (2/2)

- The condition then reduces to $-\nabla f_0(x)^T x \geq 0$. But $x \succeq 0$ and $\nabla f_0(x) \succeq 0$, so we must have $\nabla f_0(x)^T x = 0$, i.e.,

$$\sum_{i=1}^{n} [\nabla f_0(x)]_i x_i = 0.$$

- Therefore, $[\nabla f_0(x)]_i x_i = 0$ for $i = 1, \ldots, n$. The optimality condition can therefore be expressed as

$$x \succeq 0, \quad \nabla f_0(x) \succeq 0, \quad x_i [\nabla f_0(x)]_i = 0, \quad i = 1, \ldots, n.$$

- The last condition is called **complementarity**, since it means that the set of indices corresponding to nonzero components of the vectors x and $\nabla f_0(x)$ are complementary (i.e., have empty intersection).

Quasiconvex optimization

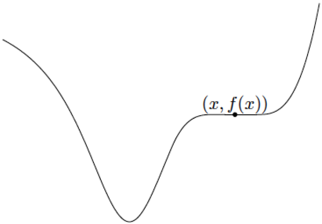
- Recall that a quasiconvex optimization problem has the standard form

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

  where the inequality constraint functions $f_1, ..., f_m$ are convex, and the objective f_0 is quasiconvex (instead of convex).
- Some basic differences between convex and quasiconvex optimization problems will be studied.
    - It would be shown that how solving a quasiconvex optimization problem can be reduced to solving a sequence of convex optimization problems.

Locally optimal solutions and optimality conditions

- The most important difference between convex and quasiconvex optimization is that a quasiconvex optimization problem can have locally optimal solutions that are not (globally) optimal.
- This phenomenon can be seen even in the simple case of unconstrained minimization of a quasiconvex function on $\mathbf{R}$.



Locally optimal solutions and optimality conditions

- Nevertheless, a variation of the optimality condition for convex problems ($\nabla f_0(x)^T(y-x) \geq 0$ for all $y \in X$) does hold for quasiconvex optimization problems with differentiable objective function.
- Let X denote the feasible set for the quasiconvex optimization problem described in a previous page.
- We first recognize that

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

for any quasiconvex differentiable function f.
- It then follows that x is optimal if

$$x \in X, \quad \nabla f_0(x)^T(y-x) > 0 \text{ for all } y \in X \setminus \{x\}.$$

Quasiconvex optimization via convex feasibility problems

- One general approach to quasiconvex optimization relies on the representation of the sublevel sets of a quasiconvex function via a family of convex inequalities.
- Let $\phi_t : \mathbf{R}^n \to \mathbf{R}$, $t \in \mathbf{R}$, be a family of convex functions that satisfy

$$f_0(x) \le t \iff \phi_t(x) \le 0,$$

  and also, for each x, $\phi_t(x)$ is a nonincreasing function of t, i.e., $\phi_s(x) \le \phi_t(x)$ whenever $s \ge t$.
- Let p^* denote the optimal value of the quasiconvex optimization problem. If the feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & \phi_t(x) \le 0 \\ & f_i(x) \le 0, i = 1, \ldots, m \\ & Ax = b, \end{array}$$

  is feasible, then we have $p^* \le t$. Otherwise, we have $p^* \ge t$.

Bisection for Quasiconvex Optimization (1/2)

Algorithm 4.1 Bisection method for quasiconvex optimization.

- given $l \leq p^*$, $u \geq p^*$, tolerance $\epsilon > 0$.
  repeat
  1. $t := (l + u)/2$.
  2. Solve the convex feasibility problem

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

  3. If the previous problem is feasible, $u := t$; else $l := t$.

  until $u - l \leq \epsilon$.

Bisection for Quasiconvex Optimization (2/2)

- The interval $[l, u]$ is guaranteed to contain p^*, i.e., we have $l \leq p^* \leq u$ at each step.
- In each iteration the interval is divided in two, i.e., bisected, so the length of the interval after k iterations is $2^{-k}(u - l)$, where $u - l$ is the length of the initial interval.
- It follows that exactly $\lceil \log_2((u - l)/\epsilon) \rceil$ iterations are required before the algorithm terminates.
- Each step involves solving the convex feasibility problem

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

Quasiconvex Optimization Problem – An Example

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & ||Ax - b|| \leq \epsilon, \end{array}$$

  where $f_0(x) = \text{length}(x) = \min\{k \mid x_i = 0 \text{ for } i > k\}$. The problem variable is $x \in \mathbf{R}^n$; the problem parameters are $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\epsilon > 0$.
- This is to find the minimum number of columns of A, taken in order, that can approximate the vector b within ϵ.
- It can be shown to be a quasiconvex optimization problem.
- The bisection algorithm can be applied by finding an appropriate family of functions $\phi_t(x)$ that satisfies

$$f_0(x) \leq t \Longleftrightarrow \phi_t(x) \leq 0.$$