CHAPTER 4

NONLINEAR PROGRAMMING: CONSTRAINED MINIMIZATION

We turn now to the study of minimization problems having constraints.

We begin by studying the necessary and sufficient conditions satisfied at solution points.

The general method used to derive necessary and sufficient conditions is a straightforward extension of that used before for unconstrained problems.

In this section we deal with general nonlinear programming problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m, \\ & g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, p, \\ & \mathbf{x} \in \Omega \subset \mathbb{R}^n, \end{aligned} \tag{1}$$

where $m \leq n$ and the functions f, h_i, g_j, $i = 1, \ldots, m, \ j = 1, \ldots, p$ are continuous, and usually assumed to possess continuous second partial derivatives.

For notational simplicity, we introduce the vector-valued functions

$$\mathbf{h} = (h_1, h_2, \dots, h_m)$$

and

$$\mathbf{g} = (g_1, g_2, \dots, g_p)$$

and rewrite (1) as

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{aligned} \tag{2}$$

The constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ are referred to as **functional constraints**, while the constraint $\mathbf{x} \in \Omega$ is a **set constraint**.

A point $\mathbf{x} \in \Omega$ that satisfies all the functional constraints is said to be **feasible**.

Assumption

We assume in most cases that either Ω is the whole space $\mathbb{R}^n$ or that the solution to (2) is in the interior of Ω.

Other variants are the **equality-constrained optimization problem**

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ & \mathbf{x} \in \Omega \end{aligned}$$

and the **inequality-constrained optimization problem**

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ & \mathbf{x} \in \Omega. \end{aligned}$$

An inequality constraint $g_j(\mathbf{x}) \leq 0$ is said to be **active** at a feasible point $\mathbf{x}$ if $g_j(\mathbf{x}) = 0$ and **inactive** at $\mathbf{x}$ if $g_j(\mathbf{x}) < 0$.

By convention we refer to any equality constraint $h_i(\mathbf{x}) = 0$ as active at any feasible point.

In studying the properties of a local minimum point, attention can be restricted to the active constraints.

Problems with Equality Constraints

A set of equality constraints on $\mathbb{R}^n$

$$h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m$$

defines a subset of $\mathbb{R}^n$ which is best viewed as a *hypersurface*.

If the functions h_i, $i = 1, 2, \ldots, m$, belong to C^1, the surface defined by them is said to be **smooth**.

Associated with a point on a smooth surface is the *tangent plane* at that point.

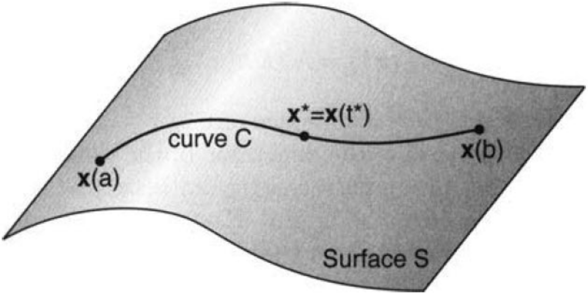
A **curve** on a surface S is a family of points $\mathbf{x}(t) \in S$ continuously parameterized by t for $a \leq t \leq b$.

The curve is **differentiable** if

$$\dot{\mathbf{x}}(t) := (d/dt)\mathbf{x}(t) = \left(\dot{x}_1(t), \dot{x}_2(t), \dots, \dot{x}_n(t)\right)$$

exists, and is **twice differentiable** if $\ddot{\mathbf{x}}(t)$ exists.

A curve $\mathbf{x}(t)$ is said to **pass** through the point $\mathbf{x}^*$ if $\mathbf{x}^* = \mathbf{x}(t^*)$ for some $t^* \in [a, b]$. The **derivative** of the curve at $\mathbf{x}^*$ is defined as $\dot{\mathbf{x}}(t^*) \in \mathbb{R}^n$.



Curve on a surface

Consider all differentiable curves on S passing through a point $\mathbf{x}^*$. The **tangent plane** at $\mathbf{x}^*$ is defined as the collection of the derivatives at $\mathbf{x}^*$ of all these differentiable curves.

The tangent plane is a subspace of $\mathbb{R}^n$.

We introduce the subspace

$$M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$$

and investigate under what conditions M is equal to the tangent plane at $\mathbf{x}^*$.

Definition 1.1

A point $\mathbf{x}^*$ satisfying the constraint $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ is said to be a **regular point** of the constraint if the gradient vectors

$$\nabla h_1(\mathbf{x}^*), \nabla h_2(\mathbf{x}^*), \ldots, \nabla h_m(\mathbf{x}^*)$$

are linearly independent.

In particular if the constraint has only one equation $h(\mathbf{x}) = 0$, then $\mathbf{x}^*$ is a regular point if and only if $\nabla h(\mathbf{x}^*) \neq \mathbf{0}$.

If $\mathbf{h}$ is affine, $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, regularity is equivalent to $\mathbf{A}$ having rank equal to m, and this condition is independent of $\mathbf{x}^*$.

Example 1.1 Consider an equality-constrained problem

(a) with the single constraint $h : \mathbb{R}^n \to \mathbb{R}$, $h(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - \alpha = 0$ at any point $\mathbf{x}$, where $\mathbf{a} \in \mathbb{R}^n$ is given;

(b) with the single constraint $g : \mathbb{R}^2 \to \mathbb{R}$, $g(\mathbf{x}) = \left(\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 1\right)^3 = 0$ at the feasible point $(1, 1)$;

(c) with the two constraints $h_1, h_2 : \mathbb{R}^3 \to \mathbb{R}$,

$$h_1(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 - 3 = 0$$
$$h_2(\mathbf{x}) = 2x_1 - 4x_2 + x_3^2 + 1 = 0$$

at the feasible point $(1, 1, 1)$.

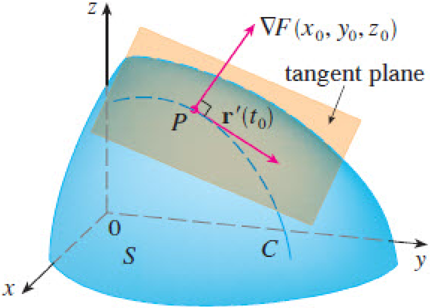
Assumption

It is assumed that $f, \mathbf{h} \in C^1$.

Theorem 1.1

At a regular point $\mathbf{x}^*$ *of the surface* S *defined by* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ *the tangent plane is equal to*

$$M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}.$$



$\nabla F(x_0, y_0, z_0)$ is a normal vector to surface $F(x, y, z) = 0$ at (x_0, y_0, z_0).

In this section we consider the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ & \mathbf{x} \in \Omega, \end{aligned}$$

where Ω is a nonempty open subset of $\mathbb{R}^n$.

Theorem 2.1

Let $\mathbf{x}^*$ *be a regular point of the constraints* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ *and a local extremum point of* f *subject to these constraints. Then all* $\mathbf{y} \in \mathbb{R}^n$ *satisfying*

$$\nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}$$

must also satisfy

$$\nabla f(\mathbf{x}^*)\mathbf{y} = 0.$$

Note The above theorem says that $\nabla f(\mathbf{x}^*)$ is orthogonal to the tangent plane.

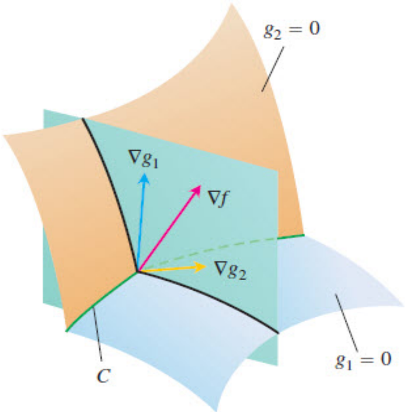
Theorem 2.2

Let $\mathbf{x}^*$ be a local extremum point of f subject to the equality constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Assume further that $\mathbf{x}^*$ is a regular point of these constraints. Then there is a $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

$\lambda_1, \lambda_2, \ldots, \lambda_m$ are called **Lagrange multipliers** and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is called then **Lagrange multiplier vector**.

4.2 FIRST-ORDER NECESSARY CONDITIONS (EQUALITY CONSTRAINTS)



For two constraints $g_1 = 0$, $g_2 = 0$, at an extremum point $\mathbf{x}^* = (x^*, y^*, z^*)$, ∇f is a linear combination of ∇g_1 and ∇g_2.

The first-order necessary conditions

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0} \tag{3}$$

is a system of n equations:

$$\frac{\partial f}{\partial x_j}(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \frac{\partial h_i}{\partial x_j}(\mathbf{x}^*) = 0, \quad j = 1, 2, \ldots, n.$$

Thus the first-order necessary conditions (3) together with the constraints

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

give a total of $n + m$ (generally nonlinear) equations in the $n + m$ variables comprising $\mathbf{x}^*$, $\boldsymbol{\lambda}$.

Definition 2.1

The function

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x})$$

is called the **Lagrangian** (or **Lagrangian function**) for problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{aligned}$$

The necessary conditions can then be expressed in the form

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \nabla_{\lambda} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{0}, \end{aligned}$$

the second of these being simply a restatement of the constraints.

Thus, in seeking an extremum of a function f whose variables are subject to the constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, one can write the Lagrange function with undetermined multipliers and look for its critical points.

Example 2.1 Consider the problem

$$\begin{array}{ll} \text{extremize} & x_1 + x_2 \\ \text{subject to} & x_1^2 + x_2^2 = 1. \end{array}$$

ANS. minimizer $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$;
maximizer $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

Example 2.2 Consider the problem

$$\begin{aligned} \text{minimize} \quad & x^2 + 4y^2 + 16z^2 \\ \text{subject to} \quad & xy = 1. \end{aligned}$$

ANS. Global minimizers are

$$(\sqrt{2}, 1/\sqrt{2}, 0) \quad \text{and} \quad (-\sqrt{2}, -1/\sqrt{2}, 0).$$

Example 2.3 (a) Solve the problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) = x_1 \cdot x_2 \cdots x_n \\ \text{subject to} \quad & x_1 + x_2 + \cdots + x_n = 1 \\ & x_1 \geq 0,\ x_2 \geq 0,\ \ldots, x_n \geq 0. \end{aligned}$$

(b) Prove the **Geometric-Arithmetic Mean Inequality**: for arbitrary nonnegative numbers $x_1, x_2, \ldots, x_n$ the inequality

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}$$

holds, with equality if and only if

$$x_1 = x_2 = \cdots = x_n.$$

Example 2.4 Consider a problem which is equivalent to the one given in Example 2.1:

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & (x_1^2 + x_2^2 - 1)^2 = 0. \end{aligned}$$

ANS. There is no Lagrange multiplier.

Example 2.5 Solve the optimization problem

$$\begin{array}{rl} \text{minimize} & 2x_1 + 3x_2 - x_3 \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1 \\ & x_1^2 + 2x_2^2 + 2x_3^2 = 2. \end{array}$$

ANS. Optimal solution: $\left(0, -\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$

Optimal value: $-\sqrt{10}$.

Example 2.6 Solve the quadratic problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

where $\mathbf{Q}$ is a symmetric and positive definite $n \times n$ matrix, $\mathbf{A}$ is an $m \times n$ matrix with $\operatorname{rank}(\mathbf{A}) = m \leq n$, and $\mathbf{b} \in \mathbb{R}^m$.

ANS. The unique global minimum point is

$$\mathbf{x}^* = \mathbf{Q}^{-1}\left[\mathbf{A}^T\left(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\right)^{-1}(\mathbf{b} + \mathbf{A}\mathbf{Q}^{-1}\mathbf{c}) - \mathbf{c}\right].$$

Example 2.7 Determine the minimal distance between the unit sphere

$$S = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = 1\}$$

and the hyperplane

$$L = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{y} = c\},$$

where $\mathbf{a} \in S$ and $c \in \mathbb{R}$, $c \geq 0$.

ANS. The minimal value equals $c - 1$ if $c > 1$ and 0 if $0 \leq c \leq 1$.

Assumption

Throughout this section it is assumed that

$$f, \mathbf{h} \in C^2.$$

Theorem 3.1 (Second-Order Necessary Conditions)

Suppose that $\mathbf{x}^*$ *is a local minimum of* f *subject to* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ *and that* $\mathbf{x}^*$ *is a regular point of these constraints. Then there is a* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

If we denote by M *the tangent plane* $M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$, *then the matrix*

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*)$$

is positive semidefinite on M, *that is,*

$$\mathbf{y}^T \mathbf{L}(\mathbf{x}^*)\mathbf{y} \geq 0 \quad \textit{for all} \quad \mathbf{y} \in M.$$

Theorem 3.2 (Second-Order Sufficiency Conditions)

Suppose there is a point $\mathbf{x}^*$ *satisfying* $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, *and* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) = \mathbf{0}.$$

Suppose also that the matrix

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*)$$

is positive definite on $M = \{\mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$, *that is, for* $\mathbf{y} \in M$, $\mathbf{y} \neq \mathbf{0}$ *there holds* $\mathbf{y}^T \mathbf{L}(\mathbf{x}^*)\mathbf{y} > 0$. *Then* $\mathbf{x}^*$ *is a strict local minimum of* f *subject to* $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. *If* $\mathbf{L}(\mathbf{x}^*)$ *is negative definite on* M, *then* $\mathbf{x}^*$ *is a strict local maximum.*

Example 3.1 Consider the problem

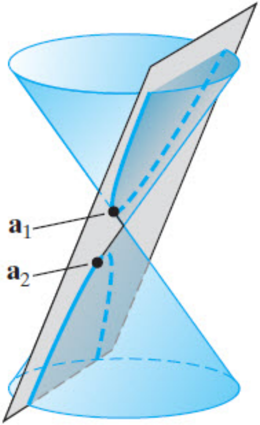
$$\begin{aligned} \text{maximize} \quad & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 3. \end{aligned}$$

ANS. Local maximum point $(1, 1, 1)$.

Example 3.2 Suppose the cone $z^2 = x^2 + y^2$ is sliced by the plane $z = x + y + 2$ so that a conic section C is created. Use Lagrange multipliers to find the points on C that are nearest to and farthest from the origin in $\mathbb{R}^3$.

ANS. Nearest point

$$\mathbf{a}_1 = \left(-2+\sqrt{2}, -2+\sqrt{2}, -2+2\sqrt{2}\right),$$
$$f(\mathbf{a}_1) = 24 - 16\sqrt{2}.$$


$\mathbf{a}_1$
$\mathbf{a}_2$

We consider now problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}. \end{aligned} \tag{4}$$

We assume that $f : \Omega \to \mathbb{R}$, $\mathbf{h} : \Omega \to \mathbb{R}^m$ (an m-dimensional function), and $\mathbf{g} : \Omega \to \mathbb{R}^p$ (a p-dimensional function).

First-Order Conditions

We assume that $f, \mathbf{h}, \mathbf{g} \in C^1$.

Definition 4.1

Let $\mathbf{x}^*$ be a point satisfying the constraints

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \tag{5}$$

and let J be the set of indices j for which $g_j(\mathbf{x}^*) = 0$,

$$J = \{j : g_j(\mathbf{x}^*) = 0\}.$$

Then $\mathbf{x}^*$ is said to be a **regular point** of the constraints (5) if the gradient vectors

$$\nabla h_i(\mathbf{x}^*),\ \nabla g_j(\mathbf{x}^*),\ 1 \leq i \leq m,\ j \in J$$

are linearly independent.

Theorem 4.1 (Karush-Kuhn-Tucker Conditions)

Let $\mathbf{x}^*$ *be a relative minimum point for the problem* (4) *and suppose* $\mathbf{x}^*$ *is a regular point for the constraints. Then there is a vector* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *and a vector* $\boldsymbol{\mu} \in \mathbb{R}^p$ *with* $\boldsymbol{\mu} \geq \mathbf{0}$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0} \quad (6)$$

$$\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = 0 \quad (7)$$

The relations (6)–(7) are called **Karush-Kuhn-Tucker conditions** (**KKT conditions**).

The conditions expressed in (7) are called **complementarity** conditions. It can be written as

$$\mu_j g_j(\mathbf{x}^*) = 0, \quad j = 1, 2, \ldots, p.$$

Definition 4.2

The function

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) \\ &= f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j g_j(\mathbf{x})\end{aligned}$$

is called the **Lagrangian** (or **Lagrangian function**) for Problem (4).

Remark We note that that the necessary condition (6) of Theorem 4.1 is equivalent to the n equations

$$\frac{\partial f}{\partial x_j}(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \frac{\partial h_i}{\partial x_j}(\mathbf{x}^*) + \sum_{i=1}^{p} \mu_i \frac{\partial g_i}{\partial x_j}(\mathbf{x}^*) = 0, \; j = 1, \ldots, n.$$

Condition (7) in component form gives the system of p equations

$$\mu_i g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, p.$$

Since $\mathbf{x}^*$ is a solution, it is feasible, and so we obtain an additional m equations

$$h_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, m.$$

Thus we have a system of $m + n + p$ equations

$$\begin{aligned}
\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) &= \mathbf{0} \\
\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) &= 0 \\
\mathbf{h}(\mathbf{x}^*) &= 0 \\
\mathbf{g}(\mathbf{x}^*) &\le 0 \\
\boldsymbol{\mu} &\ge 0
\end{aligned}$$

in the $m + n + p$ unknowns

$$\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*),\ \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m),\ \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p).$$

- **A geometric interpretation of the KKT conditions**

Example 4.1 We examine the problem

$$\begin{aligned} \text{maximize} \quad & \frac{1}{2}(x+1)^2 + \frac{1}{2}(y+1)^2 \\ \text{subject to} \quad & x^2 + y^2 \leq 2 \\ & y \leq 1. \end{aligned}$$

ANS. The global maximizer $(1,1)$.
The global minimizer $(-1,-1)$.

Theorem 4.2 (Sufficiency Conditions for Convex Problems)

Let $\mathbf{x}^*$ *be a feasible solution of Problem* (4). *Assume that* f *and* g_j *are continuously differentiable convex functions and* h_i *are affine functions. Suppose that there exist multipliers* $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$ *such that*

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0},$$

$$\boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) = \mathbf{0},$$

$$\boldsymbol{\mu} \geq \mathbf{0}.$$

Then $\mathbf{x}^*$ *is an optimal solution of* (4). *If in addition,* f *is strictly convex, then* $\mathbf{x}^*$ *is the only solution of the problem.*

Example 4.2 Consider the program

$$\begin{aligned} \text{minimize} \quad & f(x,y) = x^2 - 2x + y^2 + 1 \\ \text{subject to} \quad & x + y \le 0 \\ & x^2 - 4 \le 0. \end{aligned}$$

ANS. The global minimizer $(1/2, -1/2)$.

Second-Order Conditions

Theorem 4.3 (Second-Order Necessary Conditions)

Suppose the functions $f, \mathbf{g}, \mathbf{h} \in C^2$ *and that* $\mathbf{x}^*$ *is a regular point of the constraints* (5). *If* $\mathbf{x}^*$ *is a relative minimum point for problem* (4), *then there is a* $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\mu} \geq \mathbf{0}$ *such that* (6) *and* (7) *hold and such that*

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{G}(\mathbf{x}^*)$$

is positive semidefinite on the tangent subspace of the active constraints at $\mathbf{x}^*$.

Theorem 4.4 (Second-Order Sufficiency Conditions)

Let $f, \mathbf{g}, \mathbf{h} \in C^2$. Sufficient conditions that a point $\mathbf{x}^*$ satisfying (5) be a strict relative minimum point of problem (4) is that there exist $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$, such that

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}^*) &= \mathbf{0} \\ \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}^*) &= 0 \\ \boldsymbol{\mu} &\geq \mathbf{0}, \end{aligned}$$

Theorem 4.4 (cont'd)

and the Hessian matrix

$$\mathbf{L}(\mathbf{x}^*) = \mathbf{F}(\mathbf{x}^*) + \boldsymbol{\lambda}^T \mathbf{H}(\mathbf{x}^*) + \boldsymbol{\mu}^T \mathbf{G}(\mathbf{x}^*)$$

is positive definite on the subspace

$$M' = \left\{ \mathbf{y} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{y} = 0,\ \nabla g_j(\mathbf{x}^*)\mathbf{y} = 0 \text{ for all } j \in J' \right\},$$

where

$$J' = \left\{ j : g_j(\mathbf{x}^*) = 0,\ \mu_j > 0 \right\}.$$

Example 4.3 Consider the problem

$$\begin{array}{rl} \text{minimize} & f(x_1, x_2) = x_1 \\ \text{subject to} & (x_1+1)^2 + x_2^2 \geq 1 \\ & x_1^2 + x_2^2 \leq 2. \end{array}$$

Test whether the points $A = (0,0)$, $B = (-1,-1)$, and $C = (0, \sqrt{2}\,)$ are optimal.

ANS. $(0,0)$ is not a local minimizer; $(-1,-1)$ is a strict local minimizer; $(0, \sqrt{2}\,)$ does not satisfy the first-order necessary condition.

Example 4.4 Consider the problem

$$\begin{aligned} \text{minimize} \quad & (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{subject to} \quad & h(\mathbf{x}) = x_1 + x_2 - 2 = 0 \\ & g(\mathbf{x}) = -x_1^2 + x_2 \le 0. \end{aligned}$$

ANS. $\mathbf{x}_1 = (1, 1)$ and $\mathbf{x}_2 = (-2, 4)$ are local minimum points.

Penalization methods are procedures for approximating constrained optimization problems by unconstrained problems.

The hope is that in the limit, the solutions of the unconstrained problems will converge to the solution of the constrained problem.

The unconstrained problems involve an auxiliary function that incorporates the objective function together with penalty terms that measure violations of the constraints.

The general class of penalization methods includes two groups of methods:

- one group imposes a penalty for *violating a constraint*,
- the other imposes a penalty for *reaching the boundary* of an inequality constraint.

We refer to the first group as *penalty methods* and to the second group as *barrier methods*.

Penalty and barrier methods are of great interest to both the practitioner and the theorist.

To the practitioner they offer a simple straightforward method for handling constrained problems that can be implemented without sophisticated computer programming.

Suppose that our constrained problem is given in the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega, \end{aligned}$$

where Ω is the set of feasible points. Define

$$\sigma(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in \Omega \\ +\infty & \text{if } \mathbf{x} \notin \Omega. \end{cases}$$

The function σ can be considered as an "infinite penalty" for violating feasibility.

Then the constrained problem can be transformed into an equivalent unconstrained problem

$$\text{minimize} \quad f(\mathbf{x}) + \sigma(\mathbf{x}).$$

However, this is not a practical idea, since the objective function of the unconstrained minimization is not defined outside the feasible region.

Even if we were to replace the "∞" by a large number, the resulting unconstrained problem would be difficult to solve because of its discontinuities.

Penalization methods replace the "ideal" penalty σ by a continuous function that gradually approaches σ.

In barrier methods, this function (called a **barrier term**) approaches σ from the interior of the feasible region. It creates a barrier that prevents the iterates from reaching the boundary of the feasible region.

In penalty methods this function (called a **penalty term**) approaches σ from the exterior of the feasible region. It serves as a penalty for being infeasible.

Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega, \end{aligned} \tag{8}$$

where f is a continuous function on $\mathbb{R}^n$ and Ω is a constraint set in $\mathbb{R}^n$.

The idea of a penalty function method is to replace problem (8) by an unconstrained problem of the form

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x}) + cP(\mathbf{x}),$$

where c is a positive constant and P is a function on $\mathbb{R}^n$ satisfying:

(i) P is continuous;

(ii) $P(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$; and

(iii) $P(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in \Omega$.

The penalty method and the barrier method aim to construct a sequence of unconstrained problems such that the *minimum points of these problems converge to the minimum point of* (8).

Example 5.1 If

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x}) = 0,\ i = 1, 2, \dots, m\},$$

then the best-known such penalty is

$$P(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{m} \left(h_i(\mathbf{x})\right)^2.$$

Example 5.2 Suppose Ω is defined by a number of inequality constraints:

$$\Omega = \left\{\mathbf{x} : g_j(\mathbf{x}) \le 0, \quad j = 1, 2, \dots, p\right\}.$$

A very useful penalty function in this case is

$$P(\mathbf{x}) = \sum_{j=1}^{p} \left(g_j^+(\mathbf{x})\right)^2.$$

where g_j^+ is the **positive part** of the function g_j, defined by

$$g_j^+(\mathbf{x}) = \max\{0, g_j(\mathbf{x})\} = \begin{cases} g_j(\mathbf{x}) & \text{if } g_j(\mathbf{x}) \ge 0 \\ 0 & \text{if } g_j(\mathbf{x}) < 0. \end{cases}$$

Example 5.3 For the feasible region $\Omega = [a, b]$ with $a < b$ we obtain the constraints

$$g_1(x) = a - x \leq 0, \quad g_2(x) = x - b \leq 0.$$

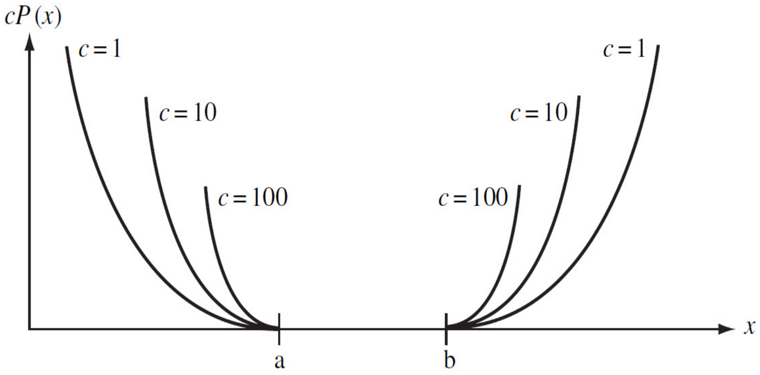
Therefore

$$g_1^+(x) = \begin{cases} a - x & \text{for } x \leq a \\ 0 & \text{for } x > a, \end{cases}$$

$$g_2^+(x) = \begin{cases} x - b & \text{for } x \geq b \\ 0 & \text{for } x < b, \end{cases}$$

and

$$P(x) = \left[g_1^+(x)\right]^2 + \left[g_2^+(x)\right]^2$$
$$= \begin{cases} (a-x)^2 & \text{for } x \le a \\ 0 & \text{for } a < x < b, \\ (b-x)^2 & \text{for } x \ge b. \end{cases}$$



Plot of $cP(x)$

Example 5.4 If

$$\Omega = \left\{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x}) = 0,\ 1 \leq i \leq m,\ g_j(\mathbf{x}) \leq 0,\ 1 \leq j \leq p\right\}$$

we can use

$$P(\mathbf{x}) = \sum_{i=1}^{m} |h_i(\mathbf{x})|^\alpha + \sum_{j=1}^{p} \left(g_j^+(\mathbf{x})\right)^\alpha, \quad \alpha \geq 1.$$

(In practice, α has usually the values 2 or 4.)

Then we get the following subsidiary problems

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x}) + cP(\mathbf{x}). \tag{9}$$

$f + cP$ is continuously differentiable, if f, h_i and g_j have this property. In this case, a great part of the methods for unrestricted minimization can be used to solve the problems of type (9).

Similarly, $f + cP$ is convex, if f, h_i and g_j are convex.

The Method

The procedure for solving problem (8) by the penalty function method is this:

Let $\{c_k\}$, $k = 1, 2, \ldots$, be a sequence tending to infinity such that for each k, $c_k \geq 0$, $c_{k+1} > c_k$.
Define the function

$$q(c, \mathbf{x}) = f(\mathbf{x}) + cP(\mathbf{x}).$$

For each k solve the problem

$$\text{minimize} \quad q(c_k, \mathbf{x}) := f(\mathbf{x}) + c_k P(\mathbf{x}), \tag{10}$$

obtaining a solution point $\mathbf{x}_k$.
We assume here that, for each k, problem (10) has a solution.

In practice, c_{k+1} can be defined by

$$c_{k+1} = \alpha c_k,$$

where $\alpha > 1$ may be arbitrary. However, to limit the number of necessary iterations, α should be sufficiently large.

Convergence

Theorem 5.2

Let $\{\mathbf{x}_k\}$ be a sequence generated by the penalty method. Then, any limit point of the sequence is a solution to (8).

Lemma 5.1

If $g(\mathbf{x})$ *has continuous first partial derivatives on* $\mathbb{R}^n$, *the same is true of* $\varphi(\mathbf{x}) = [g^+(\mathbf{x})]^2$. *Moreover,*

$$\frac{\partial \varphi}{\partial x_i}(\mathbf{x}) = 2g^+(\mathbf{x})\frac{\partial g}{\partial x_i}(\mathbf{x}), \quad i = 1, 2, \ldots, n,$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Example 5.5 Solve the problem

$$\begin{array}{ll} \text{minimize} & f(x_1, x_2) = x_1^2 + 4x_1x_2 + 5x_2^2 - 10x_1 - 20x_2 \\ \text{subject to} & h(x_1, x_2) = x_1 + x_2 - 2 = 0 \end{array}$$

using the penalty method.

ANS.

$$x_1^*(c) = \frac{5 + c}{1 + 2c}, \qquad x_2^*(c) = \frac{3c}{1 + 2c}.$$

Example 5.6 Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x_1, x_2) = (x_1+1)^2 + (x_2+2)^2 \\ \text{subject to} \quad & x_1 \geq 1, \; x_2 \geq 2. \end{aligned}$$

ANS.
$$\frac{\partial q}{\partial x_1} = \begin{cases} 2(x_1+1) + 2c(x_1-1) & \text{for } x_1 \leq 1, \\ 2(x_1+1) \quad \text{for } x_1 > 1, \end{cases}$$

$$\frac{\partial q}{\partial x_2} = \begin{cases} 2(x_2+2) + 2c(x_2-2) & \text{for } x_2 \leq 2, \\ 2(x_2+2) \quad \text{for } x_2 > 2. \end{cases}$$

$$\nabla q = \mathbf{0} \iff \begin{cases} 2(x_1+1) + 2c(x_1-1) & = 0 \\ 2(x_2+2) + 2c(x_2-2) & = 0 \end{cases}$$

$$\mathbf{x}^*(c) = \left(\frac{c-1}{c+1}, \frac{2c-2}{c+1}\right) \to \mathbf{x}^* = (1, 2).$$

Barrier methods are applicable to problems of the form

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega, \end{aligned} \tag{11}$$

where the constraint set Ω has a nonempty interior that is arbitrarily close to any boundary point of Ω.

This kind of set often arises in conjunction with inequality constraints, where Ω takes the form

$$\Omega = \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, p\}.$$

Barrier methods are also termed **interior methods** because in this method, we approach the optimum from the interior of the feasible region.

They work by establishing a barrier on the boundary of the feasible region that prevents a search procedure from leaving the region.

A **barrier function** is a function B defined on the interior of Ω such that:

(i) B is continuous,

(ii) $B(\mathbf{x}) \geq 0$;

(iii) $B(\mathbf{x}) \to \infty$ as $\mathbf{x}$ approaches the boundary of Ω.

Example 6.1 Let g_j, $j = 1, 2, \ldots, p$, be continuous functions on $\mathbb{R}^n$. Suppose

$$\Omega = \left\{\mathbf{x} : g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \ldots, p\right\}.$$

and suppose the interior of Ω is the set of $\mathbf{x}$'s where $g_j(\mathbf{x}) < 0$, $j = 1, 2, ..., p$. Then the function

$$B(\mathbf{x}) = -\sum_{j=1}^{p} \frac{1}{g_j(\mathbf{x})},$$

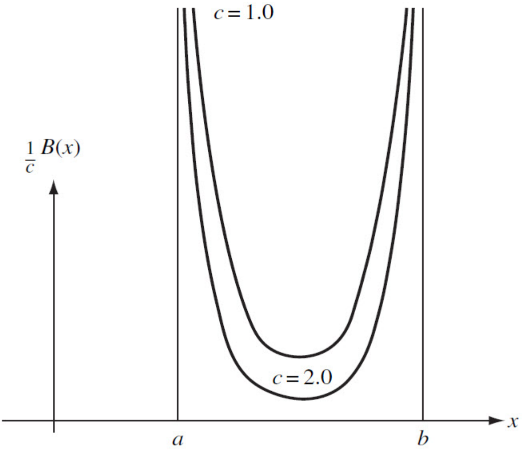
defined on the interior of Ω, is a barrier function.

Example 6.2 For the feasible region $\Omega = [a, b]$ with $a < b$, we get

$$g_1(x) = a - x, \quad g_2(x) = x - b.$$

Thus

$$B(x) = \frac{1}{x-a} + \frac{1}{b-x} \quad \text{for } a < x < b.$$



Barrier functions

Example 6.3 For the same situation as Example 6.1, we may use the logarithmic utility function

$$B(\mathbf{x}) = -\sum_{j=1}^{p} \log\big(-g_j(\mathbf{x})\big).$$

Corresponding to the problem (11), consider the approximate problem

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) + \dfrac{1}{c}B(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \operatorname{int}(\Omega), \end{array}$$

where c is a positive constant.

Alternatively, it is common to formulate the barrier method as

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) + \mu B(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \operatorname{int}(\Omega), \end{array}$$

When formulated with c we take c large (going to infinity); while when formulated with μ we take μ small (going to zero).

The Method

The barrier method is quite analogous to the penalty method.

Let $\{c_k\}$ be a sequence tending to infinity such that for each k, $k = 1, 2, \ldots$, $c_k \geq 0$, $c_{k+1} > c_k$. Define

$$r(c, \mathbf{x}) = f(\mathbf{x}) + \frac{1}{c} B(\mathbf{x}).$$

For each k solve the problem

$$\begin{array}{ll} \text{minimize} & r(c_k, \mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \operatorname{int}(\Omega), \end{array}$$

obtaining the point $\mathbf{x}_k$.

Convergence

Theorem 6.1

Any limit point of a sequence $\{\mathbf{x}_k\}$ generated by the barrier method is a solution to problem (11).

Example 6.4 Consider the problem with two variables

$$\begin{aligned} \text{minimize} \quad & f(x_1, x_2) = (x_1 - 3)^4 + (2x_1 - 3x_2)^2 \\ \text{subject to} \quad & x_1^2 - 2x_2 \leq 0. \end{aligned}$$

We get

$$g(\mathbf{x}) = x_1^2 - 2x_2, \quad B(\mathbf{x}) = -\frac{1}{g(\mathbf{x})} = \frac{1}{2x_2 - x_1^2}$$

$$r(c, \mathbf{x}) = (x_1 - 3)^4 + (2x_1 - 3x_2)^2 + \frac{1}{c(2x_2 - x_1^2)}.$$

Minimizing the function $r(c, \mathbf{x})$, we obtain the following results.

k	c_k	$x_1(c_k)$	$x_2(c_k)$
1	0.1	1.549353	1.801442
2	1	1.672955	1.631389
3	10	1.731693	1.580159
4	100	1.753966	1.564703
5	1000	1.761498	1.559926
6	10 000	1.763935	1.558429
7	100 000	1.764711	1.557955
8	1 000 000	1.764957	1.557806

The sequence of points generated by the barrier method converges to the minimum point

$$\mathbf{x}^* = (1.765071, 1.557737) \in \partial\Omega.$$

Example 6.5 Consider the nonlinear optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x_1, x_2) = x_1 - 2x_2 \\ \text{subject to} \quad & -x_1 + x_2^2 - 1 \leq 0 \\ & -x_2 \leq 0. \end{aligned}$$

Then the logarithmic barrier function gives the unconstrained problem

$$\text{minimize} \quad B_\mu(\mathbf{x}) := x_1 - 2x_2 - \mu \log(x_1 - x_2^2 + 1) - \mu \log x_2$$

for a sequence of decreasing barrier parameters.

The solution of this subsidiary problem is

$$\mathbf{x}^*(\mu) = \left(x_1^*(\mu), x_2^*(\mu)\right)$$

where

$$x_1^*(\mu) = \frac{\sqrt{1+2\mu} + 3\mu - 1}{2}, \quad x_2^*(\mu) = \frac{1+\sqrt{1+2\mu}}{2}.$$

It is obvious that $\mathbf{x}^*(\mu) \to \mathbf{x}^* = (0, 1)$ as $\mu \to 0^+$.

Barrier methods have several attractive features.

They converge under mild conditions. The barrier minimizers provide estimates of the Lagrange multipliers at the optimum.

However, barrier methods also have potential difficulties. The property for which barrier methods have drawn the most severe criticism is that the unconstrained problems become increasingly difficult to solve as the barrier parameter μ decreases.

Dual methods are based on the viewpoint that it is the Lagrange multipliers which are the fundamental unknowns associated with a constrained problem; once these multipliers are known determination of the solution point is simple.

Dual methods, therefore, do not attack the original constrained problem directly but instead attack an alternate problem, the dual problem, whose unknowns are the Lagrange multipliers of the first problem.

In this section we examine the following two questions:

1. How to define a new nonlinear optimization problem, where the unknown variables are the Lagrange multipliers?
2. Under what conditions will the solution to this new problem also provide us with a solution to the original problem?

Consider the following nonlinear programming which we call the **primal problem**.

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m, \\ & g_j(\mathbf{x}) \le 0, \quad j = 1, \ldots, p, \\ & \mathbf{x} \in \Omega \subset \mathbb{R}^n, \end{aligned} \tag{12}$$

where Ω is a convex subset of $\mathbb{R}^n$ and the functions f, h_i, and g_j are defined on $\mathbb{R}^n$.

The Lagrangian function for problem (12) is

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}),\\ \mathbf{x} \in \Omega,\ \ \boldsymbol{\lambda} \in \mathbb{R}^m,\ \ \boldsymbol{\mu} \in \mathbb{R}^p,\ \ \boldsymbol{\mu} \geq \mathbf{0}.\end{aligned}$$

The **(Lagrangian) dual problem** for (12) is

$$\begin{aligned}\text{maximize} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})\\ \text{subject to} \quad & \boldsymbol{\mu} \geq \mathbf{0},\end{aligned}$$

where

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) := \inf_{\mathbf{x} \in \Omega} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is called the **(Lagrangian) dual function**.

Theorem 7.1

The dual function is concave on the region where it is finite.

Example 7.1 Determine the dual function for the following problem

$$\begin{array}{rl} \text{minimize} & (x_1+3)^2 + x_2^2 \\ \text{subject to} & x_1^2 - x_2 \le 0. \end{array}$$

Find $\max \phi(\mu)$.

ANS. $\phi(\mu) = \frac{9\mu}{\mu+1} - \frac{\mu^2}{4}$. $\max_{\mu \ge 0} \phi(\mu) = \phi(2) = 5$.

For problem (12) we define

$$f^* = \inf\left\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0},\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{x} \in \Omega\right\} \text{ and}$$

$$\phi^* = \sup\{\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) :\ \boldsymbol{\lambda} \in \mathbb{R}^m,\ \boldsymbol{\mu} \in \mathbb{R}^p,\ \boldsymbol{\mu} \geq \mathbf{0}\}$$

where it is understood that the supremum is taken over the region where ϕ is finite.

Theorem 7.2 (Weak Duality)

$$\phi^* \leq f^*.$$

That is,

$$\sup_{\boldsymbol{\mu} \geq \mathbf{0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{\mathbf{x} \in \Omega} \left\{ f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0},\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \right\}.$$

If strict inequality holds true, a **duality gap** is said to exist.

A Geometric Interpretation

Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \le \mathbf{0} \\ & \mathbf{x} \in \Omega. \end{aligned} \tag{13}$$

The **primal function** associated with (13) is defined for $\mathbf{z} \in \mathbb{R}^p$ as

$$\omega(\mathbf{z}) = \inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \le \mathbf{z},\ \mathbf{x} \in \Omega\},$$

defined by letting the right hand side of inequality constraint take on arbitrary values.

If problem (13) has a solution $\mathbf{x}^*$ with value $f^* = f(\mathbf{x}^*)$, then $f^* = \omega(\mathbf{0})$ so it is the point on the vertical axis in $\mathbb{R}^{p+1}$ where the primal function passes through the axis.

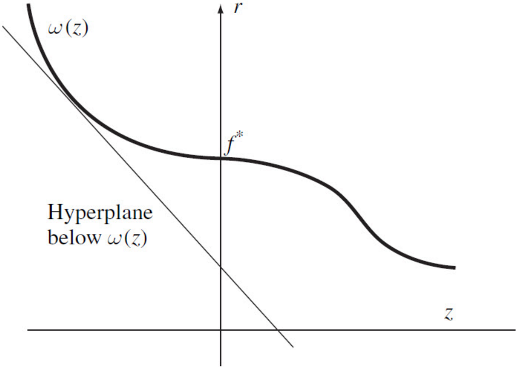
For a $(p+1)$-dimensional vector $(1, \boldsymbol{\mu}) \in \mathbb{R}^{p+1}$ with $\boldsymbol{\mu} \geq 0$ and a constant c, the set of vectors $(r, \mathbf{z})$ such that $(1, \boldsymbol{\mu})^T (r, \mathbf{z}) = r + \boldsymbol{\mu}^T \mathbf{z} = c$ defines a hyperplane in $\mathbb{R}^{p+1}$.

For a given $(1, \boldsymbol{\mu})$ we consider the lowest possible hyperplane of this form that just barely touches the region above the primal function of problem (13). Suppose $\mathbf{x}_1$ defines the touching point with values $r = f(\mathbf{x}_1)$ and $\mathbf{z} = \mathbf{g}(\mathbf{x}_1)$. Then

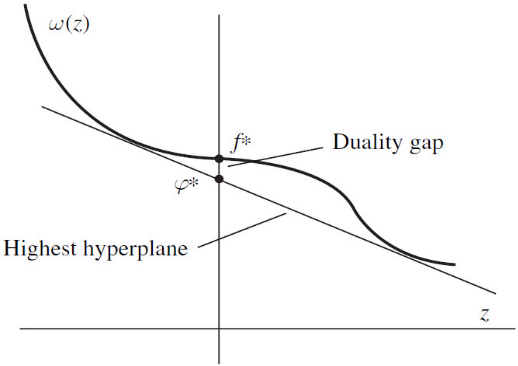
$$c = f(\mathbf{x}_1) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}_1) = \phi(\boldsymbol{\mu}).$$

The hyperplane intersects the vertical axis at a point of the form $(c, \mathbf{0})$ and $c = \phi(\boldsymbol{\mu})$.

Thus the dual function at $\boldsymbol{\mu}$ is equal to the intercept of the hyperplane defined by $(1, \boldsymbol{\mu})$ that just touches the epigraph of the primal function.



r
$\omega(z)$
f^*
Hyperplane
below $\omega(z)$
z



The highest hyperplane

Theorem 7.3 (Strong Duality)

Let Ω be convex, let f and $\mathbf{g}$ be convex, and let $\mathbf{h}$ be affine. Suppose that $\mathbf{h}$ is regular with respect to Ω and there is a point $\mathbf{x}_1 \in \Omega$ with $\mathbf{h}(\mathbf{x}_1) = \mathbf{0}$ and $\mathbf{g}(\mathbf{x}_1) < \mathbf{0}$.

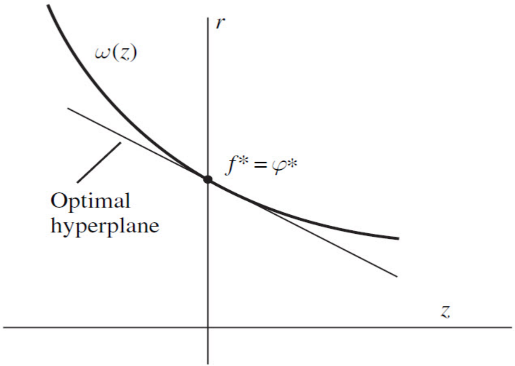
Suppose the problem has solution $\mathbf{x}^*$ with value $f(\mathbf{x}^*) = f^*$. Then for every $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{\lambda}$ there holds

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \phi^* \leq f^*.$$

Furthermore, there are $\boldsymbol{\mu}^* \geq \mathbf{0}$ and $\boldsymbol{\lambda}^*$ such that

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f^*.$$

and hence $\phi^* = f^*$. Moreover, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ are Lagrange multipliers for the original problem.



The strong duality theorem. There is no duality gap

Example 7.2 Given the problem

$$\begin{aligned} \text{minimize} \quad & f(x) = e^x \\ \text{subject to} \quad & x \in \Omega = [-1, 1]. \end{aligned}$$

Consider the dual function if Ω is written as the following

(a) $x^2 - 1 \leq 0$;

(b) $-1 \leq x \leq 1$.