OR4030 OPTIMIZATION Chapter 7

Penalty and Barrier Methods

7.1 A Brief Introduction

Main Idea of the Methods:

1. Solve a constrained optimization problem by solving a sequence of unconstrained optimization problems, and in the limit, the solutions of unconstrained problems will converge to the solution of the constrained problem.
2. Use an auxiliary function that incorporates the objective function together with "penalty" terms that measure violations of the constraints.

Two groups of classical methods:

- Barrier methods: impose a penalty for reaching the boundary of an inequality constraint from the interior area (prevent the iterative points from being out of the boundary).
- Penalty methods: impose a penalty for violating a constraint (force the iterative points to return to the feasible region gradually).

Common idea of the two groups of methods:

Consider the constrained problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & x \in S, \end{aligned} \tag{1}$$

where S is the feasible region of the problem.

Define

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

Function σ is an infinite penalty for violating feasibility.

Problem (1) can be transformed equivalently to

$$\min \; f(x) + \sigma(x). \tag{2}$$

But it is not practical to solve problem (2), because the objective function is not defined outside S, and discontinuous on the boundary.

Barrier and penalty methods solve a sequence of unconstrained sub-problems that gradually approximate problem (2) in which $\sigma(x)$ is replaced by a continuous function that gradually approaches $\sigma(x)$.

Barrier method generates a sequence of iterates that converge to a solution of the constrained problem (1) from the interior of the feasible region. – interior penalty method

Penalty method generates a sequence of iterates that converge to a solution of the constrained problem (1) from the exterior of the feasible region. – exterior penalty method

Consider the nonlinear inequality constrained problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g_i(x) \geq 0, i = 1, \ldots, m \end{aligned} \tag{3}$$

1. Assumption

1. f and g_i are twice continuously differentiable.
2. The feasible region has a nonempty interior

   $$S^0 = \{x | g_i(x) > 0, \quad i = 1, \ldots, m\},$$

   i.e., there exists a point $\bar{x}$ such that

   $$g_i(\bar{x}) > 0, \quad i = 1, \ldots, m.$$

3. Any point on the boundary can be approached by a sequence of interior points.

2. Barrier Terms and Barrier Functions

(1) We choose a continuous function ϕ defined on S^0 satisfying

$$\phi(x) \to \infty \quad \text{if any } g_i(x) \to 0_+.$$

Two examples of such a function ϕ are :

- Logarithmic function: $\phi(x) = -\sum_{i=1}^m \log(g_i(x))$.
- Inverse function: $\phi(x) = \sum_{i=1}^m \frac{1}{g_i(x)}$.

(2) Barrier terms:

For each feasible point x, the logarithmic function and the inverse function satisfy (as shown in Figure 7.1) that

$$\mu\phi(x) \text{ approaches } \sigma(x) \quad \text{as } \mu \to 0_+.$$

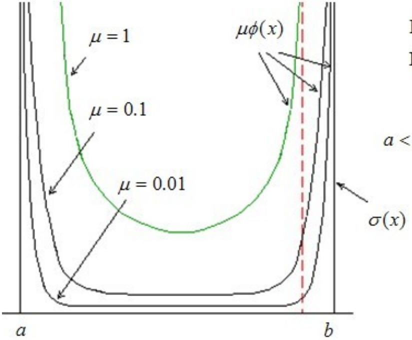

Figure 7.1
Effect of Barrier Term

We call

$\mu\phi(x)$: a barrier term;

μ : a barrier parameter.

Two frequently used barrier terms are

$$-\mu \sum_{i=1}^{m} \log(g_i(x)) \quad \text{and} \quad \mu \sum_{i=1}^{m} \frac{1}{g_i(x)}.$$

$\beta(x, \mu) = f(x) + \mu\phi(x)$: a barrier function.

- Logarithmic barrier function:

$$\beta(x,\mu) = f(x) - \mu \sum_{i=1}^{m} \log(g_i(x)).$$

- Inverse barrier function:

$$\beta(x,\mu) = f(x) + \mu \sum_{i=1}^{m} \frac{1}{g_i(x)}.$$

Barrier methods solve a sequence of "unconstrained" minimization problems

$$\min_{x \in S^0} \beta(x, \mu_k) \tag{4}$$

for a sequence of $\{\mu_k\}$ that decreases monotonically to zero.

The method transfers the constrained optimization problem (3) to problem (4). But (4) looks still a constrained optimization problem. What is the advantage to make this change?

Note that the region S^0 is an open set, that is, every point of S^0 is an interior point. So, if $\bar{x}$ is a minimum point of problem (4), it must be a unconstrained local minimum point of function $\beta(x, \mu_k)$, and hence

$$\nabla_x \beta(\bar{x}, \mu_k) = 0.$$

Therefore, we can use unconstrained minimization methods, such as the steepest descent method, or Newton's method, to solve this problem.

If we make a line search along a descent direction, as when the point is close to the boundary, the function value of $\beta(x, \mu_k)$ approaches $+\infty$, the minimum point along this direction would remain in S^0.

3. Why Don't Solve a Single Unconstrained Problem Using a Small Value of μ?

(1) No matter how small μ is, $\mu\phi(x)$ is different from $\sigma(x)$, and solving problem (4) cannot find a solution x_* of problem (1) if x_* is on the boundary.

(2) When μ is small, problem (4) is difficult to solve, especially if the initial point is far from the solution.

We need to start with an appropriate value of μ (not very small), and solve a sequence of problems (4) with decreasing μ,. The solution of problems (4) with $\mu = \mu_k$ is used as starting point for problems (4) with $\mu = \mu_{k+1}$ to facilitate computation.

Example 7.1 (Barrier Method) Consider the problem

$$\begin{aligned} \min \quad & f(x) = x_1 - 2x_2 \\ s.t. \quad & 1 + x_1 - x_2^2 \geq 0 \\ & x_2 \geq 0. \end{aligned}$$

The unconstrained problem by the logarithmic barrier function is:

$$\min_x \beta(x, \mu) = x_1 - 2x_2 - \mu \log(1 + x_1 - x_2^2) - \mu \log x_2.$$

For any fixed value $\mu > 0$, the first order necessary conditions for optimality are:

$$\begin{cases} 1 - \frac{\mu}{1+x_1-x_2^2} = 0, \\ -2 + \frac{2\mu x_2}{1+x_1-x_2^2} - \frac{\mu}{x_2} = 0. \end{cases}$$

$$
\begin{aligned}
\Longrightarrow \quad & -2 + 2x_2 - \frac{\mu}{x_2} = 0 \\
\Longrightarrow \quad & x_2^2 - x_2 - \frac{1}{2}\mu = 0 \\
\Longrightarrow \quad & x_2(\mu) = \frac{1 + \sqrt{1 + 2\mu}}{2} \\
& \text{(another solution } x_2 < 0 \text{ and hence is discarded),} \\
\Longrightarrow \quad & x_1(\mu) = \frac{\sqrt{1 + 2\mu} + 3\mu - 1}{2} \\
& \text{(note that } x_1 = x_2^2 - 1 + \mu\text{).}
\end{aligned}
$$

When $\mu \to 0$,

$$
\lim_{\mu \to 0} x_1(\mu) = 0, \quad \lim_{\mu \to 0} x_2(\mu) = 1.
$$

We may verify that the limit (0,1) is indeed the minimum point of the problem by the graph below.

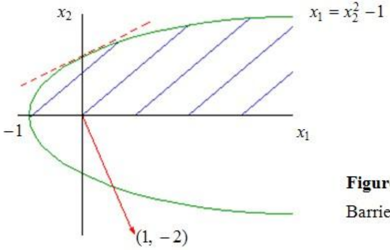

Figure (for Example 7.1)
Barrier Method

4 Features of Barrier Methods

(1) From the example, we see that when $\mu_k \to 0$, $x(\mu_k) \to x_*$ (the optimal solution). We shall prove this conclusion later.

(2) For different values of μ, the optimal solutions of problem (4) define a curve $x(\mu)$, called the barrier trajectory.

(3) If the logarithmic barrier function is used, the transformed equivalent problem is

$$\min_{x \in S^0} \beta(x, \mu) = f(x) - \mu \sum_{i=1}^{m} \log(g_i(x)).$$

The minimum point $x(\mu)$ satisfies

$$\nabla f(x) - \mu \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)} = 0.$$

If we define

$$\lambda_i(\mu) = \frac{\mu}{g_i(x)},$$

then $x = x(\mu)$ satisfies

$$\nabla f(x) - \sum_{i=1}^{m} \lambda_i(\mu) \nabla g_i(x) = 0.$$

So, all $x(\mu)$ and $\lambda(\mu)$ satisfy the following conditions:

$$g_i(x(\mu)) > 0, \quad i = 1, \ldots, m;$$

$$\nabla f(x(\mu)) - \sum_{i=1}^{m} \lambda_i(\mu) \nabla g_i(x(\mu)) = 0; \tag{5}$$

$$\lambda_i(\mu) g_i(x(\mu)) = \mu, \quad i = 1, \ldots, m; \tag{6}$$

$$\lambda_i(\mu) \geq 0, \quad i = 1, \ldots, m; \tag{7}$$

which resemble the first order necessary conditions for optimality, except that the RHS of (6) is μ, not 0.

When $\mu \to 0$, suppose $x(\mu) \to x_*$ and $\lambda(\mu) \to \lambda_*$
$= (\lambda_{*1}, \ldots, \lambda_{*m})^T$. Then from (5)-(7), we have

$$\begin{aligned}
&g_i(x_*) \geq 0, \quad i = 1, \ldots, m; \\
&\nabla f(x_*) - \sum_{i=1}^{m} \lambda_{*i} \nabla g_i(x_*) = 0; \\
&\lambda_{*i} g_i(x_*) = 0, \quad i = 1, \ldots, m; \\
&\lambda_{*i} \geq 0, \quad i = 1, \ldots, m,
\end{aligned}$$

i.e., $\lim_{\mu \to 0} \lambda(\mu)$ is the KKT multiplier of the problem (3).

Conclusion. The minimum points $x(\mu)$ of the barrier method provide estimates

$$\lambda_i(\mu) = \frac{\mu}{g_i(x(\mu))}$$

for the KKT multipliers λ_* at the optimal solution of problem (3). When $\mu \to 0$, the estimate $\lambda(\mu)$ approaches the exact λ_*.

Example 7.2 (KKT Multiplier Estimates) Consider the problem

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2 \\ s.t. \quad & g_1(x) = x_1 - 1 \geq 0 \\ & g_2(x) = x_2 + 1 \geq 0. \end{aligned}$$

It is easy to verify that the minimum point is $x_* = (1, 0)^T$.

KKT multipliers:

The second constraint is inactive $\Longrightarrow \lambda_{*2} = 0$. So,

$$\nabla f(x_*) - \lambda_{*1} \nabla g_1(x_*) = 0 \Longrightarrow \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \lambda_{*1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Longrightarrow \lambda_{*1} = 2,$$

i.e.,

$$\lambda_* = (2, 0)^T.$$

Now suppose the problem is solved by the logarithmic barrier method.

$$\min_{x \in S^0} \beta(x, \mu) = x_1^2 + x_2^2 - \mu \log(x_1 - 1) - \mu \log(x_2 + 1).$$

$$\frac{\partial \beta}{\partial x_1} = 0 \Longrightarrow 2x_1 - \frac{\mu}{x_1 - 1} = 0 \Longrightarrow 2x_1^2 - 2x_1 - \mu = 0,$$

$$\frac{\partial \beta}{\partial x_2} = 0 \Longrightarrow 2x_2 - \frac{\mu}{x_2 + 1} = 0 \Longrightarrow 2x_2^2 + 2x_2 - \mu = 0,$$

yielding

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

The KKT multiplier estimates are

$$\lambda_1(\mu) = \frac{\mu}{g_1(x(\mu))} = \frac{2\mu}{\sqrt{1+2\mu}-1} = \sqrt{1+2\mu}+1;$$

$$\lambda_2(\mu) = \frac{\mu}{g_2(x(\mu))} = \frac{2\mu}{\sqrt{1+2\mu}+1} = \sqrt{1+2\mu}-1.$$

When $\mu \to 0$,

$$x_1(\mu) \to 1 \quad x_2(\mu) \to 0;$$

$$\lambda_1(\mu) \to 2 \quad \lambda_2(\mu) \to 0.$$

We see that when $\mu \to 0$,

$$x(\mu) \to x_* \quad \text{and} \quad \lambda(\mu) \to \lambda_*.$$

7.3.1. Penalty Methods for Equality Constrained Problems

Consider the equality constrained problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & g(x) = (g_1(x), \ldots, g_m(x))^T = 0. \end{array} \tag{8}$$

Penalty terms and penalty function:

(1) A continuous function ψ with the following property:

$$\begin{cases} \psi(x) = 0, & \text{if } x \text{ is feasible;} \\ \psi(x) > 0, & \text{otherwise,} \end{cases} \tag{9}$$

can play a role of penalty for constraint violation.

For example, we can use the **quadratic-loss function**:

$$\psi(x) = \frac{1}{2}\sum_{i=1}^{m} g_i^2(x) \quad (= \frac{1}{2} g(x)^T g(x)),$$

or more generally,

$$\psi(x) = \frac{1}{\gamma} \sum_{i=1}^{m} |g_i(x)|^{\gamma} \quad (\gamma > 1).$$

(2) We need to introduce a parameter ρ (> 0) to control the weight of the penalty. ρ is called the penalty parameter.

- When $\rho \to \infty$, $\rho\psi(x) \to \sigma(x)$.
- As ρ increases,
  $\Longrightarrow$ the penalty is increased
  $\Longrightarrow$ the iterates are forced to move towards the feasible region.

(3) We call

- $\rho\psi(x)$: penalty term;
- $\pi(x, \rho) = f(x) + \rho\psi(x)$: **penalty function**.

How does the penalty method work?

For an increasing sequence $\{\rho_k\}$ of positive values tending to ∞, solve the unconstrained minimization problems

$$\min_{x \in R^n} \pi(x, \rho_k)$$

to obtain $\{x^k\}$. x^k shall approach the optimal solution x_* of problem (8).

How to approximate Lagrange multiplier by penalty method?

Suppose that we choose the quadratic-loss function as the penalty term, and let $x(\rho)$ be the minimum point of

$$\pi(x, \rho) = f(x) + \frac{1}{2}\rho \sum_{i=1}^{m} g_i(x)^2.$$

Then $x(\rho)$ satisfies

$$\nabla_x \pi(x(\rho), \rho) = \nabla f(x(\rho)) + \rho \sum_{i=1}^{m} g_i(x(\rho)) \nabla g_i(x(\rho)) = 0.$$

Define

$$\lambda_i(\rho) = \rho g_i(x(\rho)).$$

Then

$$\nabla f(x(\rho)) + \sum_{i=1}^{m} \lambda_i(\rho) \nabla g_i(x(\rho)) = 0.$$

When $\rho \to \infty$, suppose $x(\rho) \to x_*$ and $\lambda(\rho) \to \lambda_*$, then we see that x_* and λ_* satisfy

$$\nabla f(x_*) + \sum_{i=1}^{m} \lambda_{*i} \nabla g_i(x_*) = 0.$$

Hence the limit λ_* is the Lagrange multiplier vector of problem (8).

This mean that vector $\lambda(\rho)$ **can be used to estimate the Lagrange multiplier vector** λ_*.

For large values of ρ, function $\pi(x, \rho)$ is difficult to be minimized. Therefore, we need to

- minimize a sequence of functions $\pi(x, \rho_k)$ with increasing ρ_k which tend to ∞, and
- use the minimizer x^k of the function $\pi(x, \rho_k)$ as the initial point in minimizing function $\pi(x, \rho_{k+1})$.

Example 7.3 (Penalty Method). Consider the problem

$$\begin{aligned} \min \quad & f(x) = -x_1 x_2 \\ s.t. \quad & g(x) = x_1 + 2x_2 - 4 = 0. \end{aligned}$$

Use quadratic-loss penalty function,

$$\min_{x \in R^2} \pi(x, \rho) = -x_1 x_2 + \frac{1}{2}\rho(x_1 + 2x_2 - 4)^2.$$

$$\frac{\partial \pi}{\partial x_1} = 0 \Longrightarrow -x_2 + \rho(x_1 + 2x_2 - 4) = 0;$$

$$\frac{\partial \pi}{\partial x_2} = 0 \Longrightarrow -x_1 + 2\rho(x_1 + 2x_2 - 4) = 0.$$

For $\rho > \frac{1}{4}$, the above equations have the solution

$$x_1(\rho) = \frac{8\rho}{4\rho - 1}, \quad x_2(\rho) = \frac{4\rho}{4\rho - 1}.$$

Hence

$$\begin{aligned} g(x(\rho)) &= x_1(\rho) + 2x_2(\rho) - 4 \\ &= \frac{16\rho}{4\rho - 1} - 4 = \frac{4}{4\rho - 1} \quad (> 0). \end{aligned}$$

Note that $x(\rho)$ is not a feasible point of the original constrained problem. Let

$$\lambda(\rho) = \rho\, g(x(\rho)) = \frac{4\rho}{4\rho - 1}.$$

When $\rho \to \infty$,

$$x_1(\rho) \to 2, \quad x_2(\rho) \to 1, \quad \lambda(\rho) \to 1.$$

It can be verified easily that:

$x_* = (2, 1)^T$ is indeed the minimum point, and
$\lambda_* = 1$ is indeed the Lagrange multiplier
to the constrained minimization problem.

7.3.2. Penalty Methods for Inequality Constrained Problems

Penalty method is also available for inequality constrained problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) \geq 0, \quad i = 1, \ldots, m. \end{array} \tag{10}$$

Again, any continuous function ψ with the following property:

$$\begin{cases} \psi(x) = 0, & \text{if } x \text{ is feasible;} \\ \psi(x) > 0, & \text{otherwise,} \end{cases} \tag{11}$$

can play a role of penalty for constraint violation. In particular, the **quadratic-loss function** in this case is

$$\psi(x) = \frac{1}{2} \sum_{i=1}^{m} [\min(g_i(x), 0)]^2.$$

Note that

- When x is feasible, all $g_i(x) \geq 0$
  $\Longrightarrow \psi(x) = 0 \quad \sim$ no penalty.
- Otherwise, at least for one constraint function, say $g_j(x)$, the constraint is violated: $g_j(x) < 0 \Longrightarrow \psi(x) \geq \frac{1}{2} g_j^2(x) > 0 \quad \sim$ a penalty is imposed.

The penalty method is the same as before: for a sequence of ρ_k ($\rho_k \nearrow \infty$), solve unconstrained optimization problems

$$\min_{x \in R^n} \pi(x, \rho_k) = f(x) + \rho_k \psi(x).$$

When we minimize this penalty function, we want to know: what is $\nabla \psi(x)$?

It can be verified that, for the above quadratic-loss function,

$$\nabla\psi(x) = \sum_{i=1}^{m} \min(g_i(x), 0) \cdot \nabla g_i(x) \tag{12}$$

(see the textbook for proof).

Finally, if a problem contains both equality and inequality constraints, say

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & h_j(x) = 0, \quad j = 1, \ldots, n, && (13) \\ & g_i(x) \geq 0, \quad i = 1, \ldots, m. && (14) \end{aligned}$$

then the penalty function can be

$$\pi(x, \rho) = f(x) + \frac{\rho}{2}\{\sum_{j=1}^{n} h_j^2(x) + \sum_{i=1}^{m}[\min(g_i(x), 0)]^2\}.$$

7.4 Convergence of the Methods

Here we consider only the barrier methods (penalty methods can be analyzed in a similar way, see the textbook) applied to the following problem:

$$\begin{array}{ll} \min & f(x) \\ s.t. & g_i(x) \geq 0, \quad i = 1, \ldots, m. \end{array} \tag{15}$$

Let S and S^0 denote, respectively, the feasible region and its interior, i.e.,

$$S = \{x | g_i(x) \geq 0, \ i = 1, \ldots, m\},$$

$$S^0 = \{x | g_i(x) > 0, \ i = 1, \ldots, m\}.$$

Assumptions:

1. f and g_i $(i = 1, \ldots, m)$ are continuous, ϕ is a continuous function on S^0, and $\phi(x) \to +\infty$ when x approaches the boundary of S.
2. S^0 is nonempty.
3. Any y on the boundary of S can be approached by a sequence $\{x_k\}$ in S^0: $x_k \to y, \ x_k \in S^0$.

Theorem 7.1 (Convergence of the Barrier Method)
Let

$$\beta(x, \mu) = f(x) + \mu\phi(x); \quad \mu_1 \geq \mu_2 \geq \cdots ; \ \lim_{k\to\infty} \mu_k = 0,$$

and x_k be a global minimum point of problem $\min_{x\in S^0} \beta(x, \mu_k)$.
Then for $k = 1, \ldots,$

(a) $f(x_{k+1}) \leq f(x_k)$;

(b) $\phi(x_{k+1}) \geq \phi(x_k)$;

(c) if a subsequence $\{x_k | k \in K\}$ converges to $\hat{x}$, then $\hat{x}$ must be a global solution to problem (15).

Proof

(a) For each k,

$$x_k \text{ is the solution of } \min_{x \in S^0} \beta(x, \mu_k). \tag{16}$$

So,

$$\begin{aligned} & \beta(x_k, \mu_k) \le \beta(x_{k+1}, \mu_k) \\ \Longrightarrow\ & f(x_k) + \mu_k \phi(x_k) \le f(x_{k+1}) + \mu_k \phi(x_{k+1}). \end{aligned} \tag{17}$$

$$\begin{aligned} & \beta(x_{k+1}, \mu_{k+1}) \le \beta(x_k, \mu_{k+1}) \\ \Longrightarrow\ & f(x_{k+1}) + \mu_{k+1} \phi(x_{k+1}) \le f(x_k) + \mu_{k+1} \phi(x_k). \end{aligned} \tag{18}$$

$$
\begin{aligned}
& \mu_{k+1} \times (17) + \mu_k \times (18) \\
\Longrightarrow \quad & \mu_{k+1} f(x_k) + \mu_k f(x_{k+1}) \leq \mu_{k+1} f(x_{k+1}) + \mu_k f(x_k) \\
\Longrightarrow \quad & (\mu_k - \mu_{k+1}) f(x_{k+1}) \leq (\mu_k - \mu_{k+1}) f(x_k) \\
\Longrightarrow \quad & f(x_{k+1}) \leq f(x_k) \;\; (\text{as } \mu_k - \mu_{k+1} > 0). \qquad (19)
\end{aligned}
$$

(b) From (17) and (19)

$$
\begin{aligned}
\Longrightarrow \quad & \mu_k \phi(x_k) \leq \mu_k \phi(x_{k+1}) \\
\Longrightarrow \quad & \phi(x_k) \leq \phi(x_{k+1}).
\end{aligned}
$$

(c) Let $x_k \xrightarrow{K} \hat{x}$ (i.e., there is a subsequence of x_k which converges to $\hat{x}$). We need to prove that $\hat{x}$ must be an optimal solution of problem (15).

(i) First, since $x_k \in S^0$, $g_i(x_k) > 0$ for each k and i. Then, when $k \xrightarrow{K} \infty$,

$$g_i(\hat{x}) \geq 0, \ \forall i = 1, \dots, m.$$

$$\Longrightarrow \hat{x} \in S, \text{ i.e. } \hat{x} \text{ is a feasible solution of problem (15)}$$

(ii) Let x_* be a global minimum point of problem (15). We need to show that $f(\hat{x}) = f(x_*)$. By assumption (3) and the continuity of f, for **any** $\epsilon > 0$, there exists $x_\epsilon \in S^0$ such that

$$f(x_\epsilon) < f(x_*) + \epsilon.$$

Due to (16),

$$\beta(x_k, \mu_k) \le \beta(x_\epsilon, \mu_k),$$

i.e.,

$$f(x_k) + \mu_k \phi(x_k) \le f(x_\epsilon) + \mu_k \phi(x_\epsilon) < f(x_*) + \epsilon + \mu_k \phi(x_\epsilon). \quad (20)$$

We consider two cases:

Case A. $\hat{x} \in S^0$. Then $\phi(\hat{x})$ is a finite number, and

$$\phi(x_k) \xrightarrow{K} \phi(\hat{x}).$$

Let $k \xrightarrow{K} \infty$, from (20),

$$f(\hat{x}) + 0 \cdot \phi(\hat{x}) \leq f(x_*) + \epsilon + 0 \cdot \phi(x_\epsilon),$$

i.e.,

$$f(\hat{x}) \leq f(x_*) + \epsilon. \tag{21}$$

Case B. $\hat{x} \notin S^0$, i.e., $\hat{x}$ is on the boundary of S. Then $\phi(\hat{x}) = +\infty$. So, for large $k \in K$, $\phi(x_k) > 0$. We have from (20) that

$$f(x_k) \le f(x_k) + \mu_k \phi(x_k) < f(x_*) + \epsilon + \mu_k \phi(x_\epsilon), \quad \text{for large } k \in K.$$

Let $k \xrightarrow{K} \infty$, from the above inequalities we have

$$f(\hat{x}) \le f(x_*) + \epsilon. \tag{22}$$

(21) and (22) are the same result. As the two inequalities are true for any $\epsilon > 0$, they just mean that

$$\begin{aligned} f(\hat{x}) \leq f(x_*) &\implies f(\hat{x}) = f(x_*) \\ &\implies \hat{x} \text{ is a global minimum point of problem (15).} \end{aligned}$$

So, the proof of part (c) is completed.

7.5 Augmented Lagrangian Method

In the penalty method, the exact solution cannot be found unless the parameter $\rho \to \infty$. But when ρ is very big, the penalty function becomes ill conditioned. Are there some methods that can obtain the optimal solution without requiring the parameter $\rho \to \infty$? Yes, exact penalty method and augmented Lagrangian method are two this kind of methods. Here we introduce only the augmented Lagrangian method.

Consider the problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0, \quad i = 1, \ldots, m \\ & x \in \; R^n. \end{array} \tag{23}$$

Let x^* be an optimal solution with associated multiplier λ^*. Obviously x^* is also an optimal solution to the problem

$$\begin{aligned} \min \quad & L(x, \lambda) = f(x) + \lambda^T h(x) \\ \text{s.t.} \quad & h(x) = 0 \end{aligned} \tag{24}$$

where $h(x) = (h_1(x), \ldots, h_m(x))^T$, because when $h(x) = 0$, the function $L(x, \lambda) = f(x)$. We now use the penalty function method to solve problem (24), that is, we consider the unconstrained optimization problem

$$\min_x \mathcal{A}(x, \lambda, \rho) = f(x) + \lambda^T h(x) + \frac{1}{2}\rho h(x)^T h(x).$$

We call function $\mathcal{A}(x, \lambda, \rho)$ **an augmented Lagrangian function** because it adds a penalty term to the Lagrangian function. The method to find x^* by using penalty method on the augmented Lagrangian function is called **augmented Lagrangian method**.

Algorithm (Augmented Lagrangian Method)

Step 0. Give an initial guess (x^0, λ^0), and choose an initial penalty parameter value $\rho_0 > 0$. Set $k = 0$.

Step 1. Optimal Test: if $\nabla L(x^k, \lambda^k) = 0$, then output $(x^*, \lambda^*) = (x^k, \lambda^k)$ and stop.

Step 2. Solve the unconstrained optimization problem

$$\min_x \mathcal{A}(x, \lambda^k, \rho_k) = f(x) + (\lambda^k)^T h(x) + \frac{1}{2}\rho_k h(x)^T h(x)$$

using any of unconstrained optimization methods, and let the optimal solution be x^{k+1}.

Step 3. Update λ^k by formula

$$\lambda^{k+1} = \lambda^k + \rho_k h(x^{k+1}) \tag{25}$$

and choose $\rho_{k+1} \geq \rho_k$.

Step 4. Set $k := k + 1$ and return to Step 1.

Note that in Step 1, $\nabla L = (\nabla_x L, \nabla_\lambda L)$ as we explained in Chapter 6, where

$$\nabla_x L(x, \lambda) = \nabla f(x) + \nabla h(x)\lambda,$$

and

$$\nabla_\lambda L(x, \lambda) = h(x).$$

The reason to use formula (25) to update the multiplier vector λ is as follows.

If x^{k+1} minimizes $\mathcal{A}(x, \lambda^k, \rho_k)$, then

$$\nabla_x \mathcal{A}(x^{k+1}, \lambda^k, \rho_k) = 0,$$

or

$$\nabla f(x^{k+1}) + \nabla h(x^{k+1})\lambda^k + \rho_k \nabla h(x^{k+1}) h(x^{k+1}) = 0.$$

This can be rearranged as

$$\nabla f(x^{k+1}) + \nabla h(x^{k+1})[\lambda^k + \rho_k h(x^{k+1})] = 0.$$

If we use formula (25) to obtain λ^{k+1}, then

$$\nabla_x L(x^{k+1}, \lambda^{k+1}) = \nabla f(x^{k+1}) + \nabla h(x^{k+1})\lambda^{k+1} = 0,$$

that is, **the first order necessary conditions for optimality are always partially satisfied.**

The optimality test in Step 1 in fact only need to check if

$$\nabla_\lambda L(x^{k+1}, \lambda^{k+1}) = h(x^{k+1}) = 0?$$

i.e., if x^{k+1} is a feasible point? If yes, then we can finish computation.

Note that in the algorithm, **the penalty parameter ρ is unnecessary to go to ∞.**

Function $\mathcal{A}$ is more complicated than the penalty function $f + \frac{1}{2}\rho h^T h$. Then **what is the advantage if we solve problem (24) instead of problem (23)?** The main reason is that we can prove, under certain conditions, that for the augmented Lagrangian method, there exists a constant $M > 0$ such that

$$\|\lambda^{k+1} - \lambda^*\| \leq \frac{M}{\rho_k}\|\lambda^k - \lambda^*\|, \tag{26}$$

$$\|x^{k+1} - x^*\| \leq \frac{M}{\rho_k}\|\lambda^k - \lambda^*\|. \tag{27}$$

Remember that here x^* is the optimal solution to the original constrained minimization problem (23), and λ^* is the associated Lagrange multiplier. Therefore, if $\rho_k > M$, for example let all $\rho_k = \hat{\rho}$ and the constant $\hat{\rho} > M$, then from (26) we know that $\{\lambda^k\}$ converges to λ^* linearly. Then from (27) we see that $x_k \to x^*$.

Therefore, when we use the augmented Lagrangian method, we may obtain the solution x^* and its corresponding multiplier vector λ^* without requiring the penalty parameter ρ_k to increase to infinity. This is the main advantage of the method. But here we only assure the existence of constant M, not its exact value. Hence we do not know what value of ρ_k is already large enough for the convergence.

We now give an example.

Example 7.6 (Augmented Lagrangian Method) Consider the problem

$$\begin{array}{rl} \textit{minimize} & f(x_1, x_2) = \dfrac{1}{2}x_1^2 + \dfrac{1}{6}x_2^2 \\ \textit{subject to} & h(x_1, x_2) = x_1 + x_2 - 1 = 0. \end{array}$$

Form the augmented Lagrange function

$$\mathcal{A}(x, \lambda_k, \rho) = \frac{1}{2}x_1^2 + \frac{1}{6}x_2^2 + \lambda_k(x_1 + x_2 - 1) + \frac{1}{2}\rho(x_1 + x_2 - 1)^2.$$

Let the minimum solution to $\mathcal{A}(x, \lambda_k, \rho)$ be $x^{k+1} = (x_1^{k+1}, x_2^{k+1})$. To obtain this minimum solution, we solve the equations $\nabla_x \mathcal{A} = 0$, i.e.,

$$\begin{cases} x_1 + \lambda_k + \rho(x_1 + x_2 - 1) = 0, \\ \frac{1}{3}x_2 + \lambda_k + \rho(x_1 + x_2 - 1) = 0. \end{cases}$$

From the above equations, it is easy to see that $x_2 = 3x_1$. Substituting it into the first equation, we obtain that

$$
\begin{aligned}
x_1^{k+1} &= \frac{\rho - \lambda_k}{1 + 4\rho}, && (28) \\
x_2^{k+1} &= \frac{3(\rho - \lambda_k)}{1 + 4\rho}. && (29)
\end{aligned}
$$

The updated multiplier is

$$
\begin{aligned}
\lambda_{k+1} &= \lambda_k + \rho h(x^{k+1}) \\
&= \lambda_k + \rho(x_1^{k+1} + x_2^{k+1} - 1) \\
&= \lambda_k + \rho(\frac{4(\rho - \lambda_k)}{1 + 4\rho} - 1) \\
&= \frac{\lambda_k - \rho}{1 + 4\rho}. && (30)
\end{aligned}
$$

We now show that the sequence $\{\lambda_k\}$ has a limit $-\frac{1}{4}$.

It is easy to see that for any $\rho > 0$, the sequence $\{\lambda_k\}$ is decreasing because

$$\lambda_{k+1} < \frac{\lambda_k}{1+4\rho} < \lambda_k.$$

Also, if we choose the initial $\lambda_0 \geq -\frac{1}{4}$, then all $\lambda_k \geq -\frac{1}{4}$ because when $\lambda_k \geq -\frac{1}{4}$, by (30),

$$\lambda_{k+1} \geq \frac{-\frac{1}{4} - \rho}{1+4\rho} = -\frac{1}{4}.$$

So, the sequence $\{\lambda_k\}$ has a limit λ^*. Taking limits on both sides of (30), we obtain

$$\lambda^* = \frac{\lambda^* - \rho}{1+4\rho},$$

from which we see that

$$\lambda^* = -\frac{1}{4}.$$

Now taking limits in (28) and (29), we obtain

$$x_1^{k+1} \to \frac{\rho + \frac{1}{4}}{1 + 4\rho} = \frac{1}{4} = x_1^*;$$

$$x_2^{k+1} = 3x_1^{k+1} \to \frac{3}{4} = x_2^*.$$

It is easy to verify that $x^* = (\frac{1}{4}, \frac{3}{4})$ is indeed the optimal solution of the problem with a multiplier $\lambda^* = -\frac{1}{4}$. Note that in order to obtain the optimal solution, the parameter ρ can be **any positive number, and we do not require ρ to approach ∞**. This is the main advantage of the augmented Lagrangian method against the penalty method.