Optimization

Màster de Fonaments de Ciència de Dades

Gerard Gómez

Lecture VI. Penalty methods for constrained optimization problems

Penalty function methods. General idea

- Consider the following **constrained optimization problem**: seek a minimum of a real-valued function f on a proper subset $X \subset \mathbb{R}^n$.
- This is problem **can be transformed into an unconstrained optimization one** after some modification of the objective function f.
- Define
$$P(x) = \begin{cases} 0 & x \in X, \\ +\infty & x \notin X, \end{cases}$$
- Consider the unconstrained minimization of the **augmented objective function** F defined by
$$\min_{x \in \mathbb{R}^n} F(x) = \min_{x \in \mathbb{R}^n} (f(x) + P(x)),$$
where f is assumed to be defined on $\mathbb{R}^n$.
- The function P is called a **penalty function**, for it imposes an (infinite) penalty on points lying outside the feasible set X.
- Clearly, a point x^* minimizes F in $\mathbb{R}^n$ if and only if it also minimizes f over X.

Penalty function methods. General idea

- In practice, the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} (f(x) + P(x)),$$

  cannot be, in general, carried out because:
  - The **discontinuity of** F on the boundary of X, and
  - The **infinite values outside** X**.**
- Replacing $+\infty$ by some large finite penalty will not simplify the problem, since the numerical difficulties would still remain.
- The idea for solving these problems involves a **sequence of unconstrained minimization problems**.
- In each problem of the sequence **a penalty parameter** is adjusted from one minimization to the next one.
- The sequence of unconstrained minima converges to a feasible point of the constrained problem.

Exterior and interior penalty function methods

Consider the general minimization problem (P) defined by:

$$\begin{array}{llll} (P) & \min & f(x) & \\ & \text{subject to} & g_i(x) \geq 0, & i = 1, ..., m \\ & & h_j(x) = 0, & j = 1, ..., p, \end{array}$$

where $f, g_1, ..., g_m, h_1, ..., h_p$ are assumed to be continuous on $\mathbb{R}^n$.

- **Exterior methods**
  - Exterior penalty function methods usually solve the above problem by a sequence of unconstrained minimization problems **whose optimal solutions approach the solution of the problem from outside the feasible set.**
  - In the sequence, a penalty is imposed such that it is increased from problem to problem.
- **Interior methods**
  - Interior penalty function methods solve inequality constrained nonlinear problems through a sequence of unconstrained minimization problems **whose solutions are points that strictly satisfy the constraints –that is- they are in the interior of the feasible set.**
  - Staying in the interior is ensured by formulating a “barrier” function by which an infinitely large penalty is imposed for crossing the boundary of the feasible set from the inside.

Exterior penalty functions

Recall that we need to define a **penalty function** P and the associated **augmented objective function** F

$$F(x) = f(x) + P(x).$$

- For the definition of the penalty function, we will use the real-valued continuous functions ψ and ξ of the variable $\eta \in \mathbb{R}$ defined by

$$\psi(\eta) = |\min(0, \eta)|^{\alpha}, \quad \xi(\eta) = |\eta|^{\beta},$$

where $\alpha \geq 1$ and $\beta \geq 1$ are given constants, usually equal to 1 or 2.

- Let

$$s(x) = \sum_{i=1}^{m} \psi(g_i(x)) + \sum_{j=1}^{p} \xi(h_j(x)),$$

this is

$$s(x) = \sum_{i=1}^{m} |\min[0, g_i(x)]|^{\alpha} + \sum_{j=1}^{p} |(h_j(x)|^{\beta}.$$

The function s is continuousand is called a **loss function** for problem (P).

- Note that since if $x \in X$, then $g_i(x) \geq 0$ and $h_j(x) = 0$, then:

$$s(x) = 0 \;\text{ if }\; x \in X, \quad \text{and} \quad s(x) > 0 \;\text{ if }\; x \notin X.$$

Exterior penalty functions

- For any positive number $\rho > 0$, define the **augmented objective function** for the minimization problem as

$$F(x, \rho) = f(x) + \frac{1}{\rho} s(x).$$

- Observe that, since

$$s(x) = 0 \ \text{ if } \ x \in X, \quad \text{and} \quad s(x) > 0 \ \text{ if } \ x \notin X,$$

  then:
  - $F(x, \rho) = f(x)$ if and only if x is feasible,
  - otherwise $F(x, \rho) > f(x)$

- The **continuous penalty function** $s(x)/\rho$ **approximates the discontinuous penalty function** $P(x)$ as $\rho \to 0$

Exterior penalty function method

- The exterior penalty function method consists of solving a sequence of unconstrained optimizations for $k = 0, 1, 2, \ldots$ given by

$$(EP^k) \quad \min_{x \in \mathbb{R}^n} F(x, \rho^k) = \min_{x \in \mathbb{R}^n} \left(f(x) + \frac{1}{\rho^k} \left\{ \sum_{i=1}^{m} |\min[0, g_i(x)]|^\alpha + \sum_{j=1}^{p} |(h_j(x)|^\beta \right\} \right),$$

  using a **strictly decreasing sequence** of positive numbers $\{\rho^k\}$.

- Defining x^{k*} as the optimal solution of (EP^k), we construct a sequence $\{x^{k*}\}$ which under rather mild conditions **has a subsequence** converging to an optimum of the original minimization problem x^*.

- Of course, in any real problem the **unconstrained minimizations** of $F(x, \rho^k)$ must be done by some algorithm, as the ones that we have already seen.

Example of the exterior penalty function method

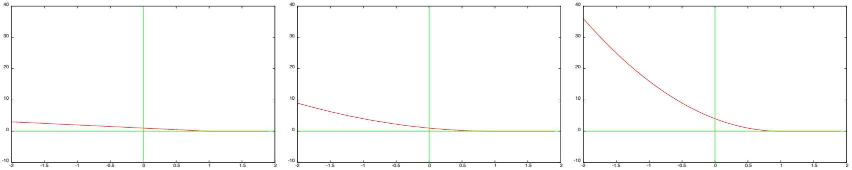
Example. We want to seek the minimum of

$$f(x) = x^2, \quad x \in \mathbb{R} \quad \text{subject to} \quad x \geq 1.$$

The optimal solution is $x^* = 1$. Note that $x \geq 1 \iff x - 1 \geq 0$.

Let us form the augmented objective function $F(x, \rho^k)$ with $\alpha = 2$. It gives the unconstrained optimization problem

$$\min_{x \in \mathbb{R}} F(x, \rho^k) = \min_{x \in \mathbb{R}} \left(x^2 + \frac{1}{\rho^k} [\min(0, x-1)]^2 \right)$$



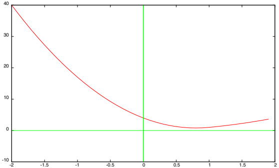
For $\rho = 0.5$, $k = 2$, plot of the functions: $|\min(0, x-1)|$, $[\min(0, x-1)]^2$ and $\frac{1}{\rho^k}[\min(0, x-1)]^2$

Example of the exterior penalty function method (cont.)

- For any given $\rho^k > 0$, the function

$$F(x, \rho^k) = x^2 + \frac{1}{\rho^k}[\min(0, x-1)]^2$$

is convex and its minimum is at the left of $x = 1$



For $\rho = 0.5$, $k = 2$, plot of the function $x^2 + \frac{1}{\rho^k}[\min(0, x-1)]^2$

- Since at the left of $x = 1$ the function is

$$F(x, \rho) = x^2 + \frac{1}{\rho^k}[x-1]^2 \quad \Rightarrow \quad F_x = 2x + \frac{2(x-1)}{\rho^k} = 0 \quad \Rightarrow \quad x(\rho^k + 1) = 1$$

- The minimum of $F(x, \rho^k)$ is achieved at the point

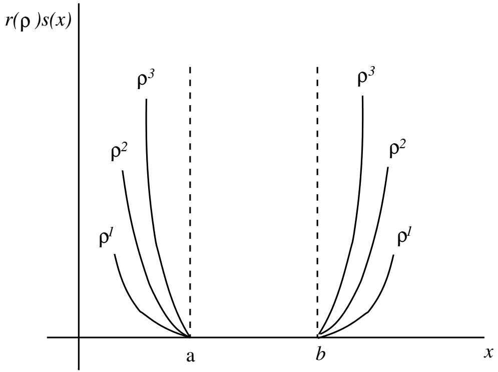
$$x^{k*} = \frac{1}{\rho^k + 1}.$$

- Note that, for every $\rho^k > 0$, **this point is infeasible** for the original problem.
- As $\rho^k \to 0$, the points x^{k*} approach $x^* = 1$ from outside the feasible set.

Exterior penalty function generalization

- The previous construction of the exterior penalty function $(1/\rho)\, s$, can be generalized.
- Let r be a continuous real-valued function of the variable $\rho \in \mathbb{R}$, such that
$$\rho^1 > \rho^2 > 0 \quad \Rightarrow \quad r(\rho^2) > r(\rho^1) > 0,$$
and if $\{\rho^k\}$ is a strictly decreasing sequence of positive numbers such that
$$\lim_{k\to\infty} \rho^k = 0 \quad \text{then} \quad \lim_{k\to\infty} r(\rho^k) = +\infty.$$
- Let s be any continuous function such that $s(x) = 0$ if $x \in X$, and $s(x) > 0$ if $x \notin X$, where X is the feasible set.
- Then
$$r(\rho)s(x)$$
is an **exterior penalty function** and
$$(EP^k) \quad \min_{x\in\mathbb{R}^n} F(x,\rho^k) = \min_x \left(f(x) + r(\rho^k)s(x) \right),$$
is the corresponding unconstrained optimization problem.

Penalty function example



Example of exterior penalty function $r(\rho^k)s(x)$ for a case in which the feasible set X is the closed interval $[a, b] \subset \mathbb{R}$.

Exterior penalty function method. Convergence

Lemma

Let F be given by

$$F(x, \rho^k) = f(x) + r(\rho^k)s(x)$$

and let

$$\rho^k \geq \rho^{k+1} > 0.$$

Assume that $F(x, \rho^k)$ and $F(x, \rho^{k+1})$ attain their minima on $\mathbb{R}^n$ at x^{k*} and x^{k+1*}, respectively, Then

$$\begin{aligned} F(x^{k+1*}, \rho^{k+1}) &\geq F(x^{k*}, \rho^k) && (1) \\ s(x^{k*}) &\geq s(x^{k+1*}) && (2) \\ f(x^{k+1*}) &\geq f(x^{k*}) && (3) \end{aligned}$$

Proof: Since $r(\rho)s(x) \geq 0$, and r is increasing as ρ is decreasing, we obtain

$$\begin{aligned} F(x^{k+1*}, \rho^{k+1}) &= f(x^{k+1*}) + r(\rho^{k+1})s(x^{k+1*}) \\ &\geq f(x^{k+1*}) + r(\rho^k)s(x^{k+1*}) \\ &\geq f(x^{k*}) + r(\rho^k)s(x^{k*}) = F(x^{k*}, \rho^k). \end{aligned}$$

where the last inequality follows from the fact that x^{k*} minimizes $F(x, \rho^k)$.

Exterior penalty function method. Convergence proof

From the above system of inequalities we get

$$f(x^{k*}) + r(\rho^k)s(x^{k*}) \leq f(x^{k+1*}) + r(\rho^k)s(x^{k+1*}),$$

and by the definition of x^{k+1*}

$$f(x^{k+1*}) + r(\rho^{k+1})s(x^{k+1*}) \leq f(x^{k*}) + r(\rho^{k+1})s(x^{k*}).$$

Adding the last two inequalities we get

$$r(\rho^k)[s(x^{k*}) - s(x^{k+1*})] \leq r(\rho^{k+1})[s(x^{k*}) - s(x^{k+1*})],$$

and since $r(\rho^k) \leq r(\rho^{k+1})$, the second inequality of the Lemma, $s(x^{k*}) \geq s(x^{k+1*})$, follows. The last inequality follows from the inequalities:

$$s(x^{k*}) \geq s(x^{k+1*}), \quad \text{and} \quad f(x^{k+1*}) - f(x^{k*}) \geq r(\rho^k)[s(x^{k*}) - s(x^{k+1*})].$$

□

Exterior penalty function method. Convergence theorem

Theorem

Suppose that the feasible set X in problem (P) is nonempty and there exists an $\epsilon > 0$ such that the set

$$X^{\epsilon} = \{x \mid x \in \mathbb{R}^n;\ g_i(x) \geq -\epsilon, i = 1, ..., m;\ |h_j(x)| \leq \epsilon, j = 1, ..., p\}$$

is compact. Also suppose that the $F(x, \rho^k)$ attain their unconstrained minima on $\mathbb{R}^n$ for all k. If $\{\rho^k\}$ is a strictly decreasing sequence of positive numbers converging to zero, then there exists a **convergence subsequence** $\{x^{k_i *}\}$ *of the optimal solutions to (EP^k), $\{x^{k*}\}$, and the limit of any such convergent subsequence is optimal for (P).*

Proof: Since X is compact and f is continuous, there exists at least one point $x^* \in X$ where f attains its minimum, that is: x^* is optimal for (P).

For $k = 0, 1, ...$

$$f(x^*) = f(x^*) + r(\rho^k)s(x^*) \geq F(x^{k*}, \rho^k)$$

so the sequence $\{F(x^{k*}, \rho^k)\}$ is bounded from above; by the preceeding Lemma, the sequence $\{F(x^{k*}, \rho^k)\}$ is an increasing sequence, so it converges to a limit F^0.

Similarly, the sequence $\{f(x^{k*})\}$ is increasing, and

$$f(x^{k*}) \leq f(x^{k*}) + r(\rho^k)s(x^{k*}) = F(x^{k*}, \rho^k).$$

Exterior penalty function method. Convergence theorem (cont.)

Combining the last inequalities we get

$$f(x^{k*}) \leq F(x^{k*}, \rho^k) \leq f(x^*) \quad \Rightarrow \quad f(x^{k*}) \leq f(x^*)$$

so $\{f(x^{k*})\}$ is bounded and, consequently, it converges to a limit f^0.

Furthermore, since $F = f + r\,s$, we get

$$\lim_{k\to\infty} \left(r(\rho^k) s(x^{k*}) \right) = \lim_{k\to\infty} \left(F(x^{k*}, \rho^k) - f(x^{k*}) \right) = F^0 - f^0.$$

Since $\lim_{k\to\infty} r(\rho^k) = +\infty$, it follows that

$$\lim_{k\to\infty} s(x^{k*}) = 0.$$

Using this limit and recalling the definition of s

$$s(x) = \sum_{i=1}^{m} |\min[0, g_i(x)]|^{\alpha} + \sum_{j=1}^{p} |(h_j(x)|^{\beta},$$

and that

$$s(x) = 0 \text{ if } x \in X, \quad \text{and} \quad s(x) > 0 \text{ if } x \notin X,$$

we conclude that for every $\delta > 0$, there exists a natural number $K(\delta)$ such that for $k \geq K(\delta)$

$$x^{k*} \in X^{\delta} = \{x \mid x \in \mathbb{R}^n;\ g_i(x) \geq -\delta, i = 1, \ldots, m;\ |h_j(x)| \leq \delta, j = 1, \ldots, p\}$$

Exterior penalty function method. Convergence theorem (cont.)

Then, for a sufficiently large $\hat{K}(\delta)$, the points x^{k*} will be in the compact set X^{δ} for all $k \geq \hat{K}(\delta)$.

Hence, there is a subsequence $\{x^{k_i*}\}$ that converges to a limit x^0 and, since $\lim_{k\to\infty} s(x^{k*}) = 0$, it follows that $s(x^0) = 0$, thus $x^0 \in X$.

From the optimality of x^*, we obtain $f(x^*) \leq f(x^0)$. Now, for all k_i in the convergent subsequence

$$f(x^{k_i*}) \leq f(x^{k_i*}) + r(\rho^{k_i})s(x^{k_i*}) \leq f(x^*) + r(\rho^{k_i})s(x^*) = f(x^*),$$

(since $s(x^*) = 0$), thus

$$\lim_{k_i\to\infty} f(x^{k_i*}) = f(x^0) \leq f(x^*),$$

with which we get

$$f(x^0) = f(x^*),$$

and x^0 must be a solution for (P). □

Interior penalty functions methods

- Using **interior penalty functions methods**, inequality constrained nonlinear problems are solved through a **sequence of unconstrained optimization problems whose minima are points that strictly satisfy the constraints** –that is, in the interior of the feasible set.
- Staying in the interior will be ensured by formulating a “barrier” function by which an infinitely large penalty is imposed for crossing the boundary of the feasible set from the inside.
- Since the methods require that the interior of the feasible set to be nonempty, **no equality constraints can be handled by the procedure that will be described**, although there are other interior-type penalty function methods that are capable of solving these problems.

Interior penalty functions. Regularity conditions

- Consider the problem

$$(PI) \quad \min \quad f(x)$$
$$s.t. \quad g_i(x) \geq 0, \quad i = 1, ..., m$$

where $f, g_1, ..., g_m$ are assumed to be continuous on $\mathbb{R}^n$.

- Let X be the feasible set

$$X = \{x \mid x \in \mathbb{R}^n;\ g_i(x) \geq 0, i = 1, ..., m\},$$

and S be the interior of X, then the following two **regularity conditions** are assumed:

1. The set X is closed, S is nonempty, and X is the closure of S.
2. There is a point $x^0 \in X$, with $f(x^0) = \alpha^0$, such that the intersection of the level set $\Gamma(f, \alpha^0)$ with X

$$\Gamma(f, \alpha^0) \cap X = \left\{x \mid x \in X,\ f(x) \leq \alpha^0\right\} \cap X$$

is compact.

Interior penalty functions

As in the case of exterior methods, in interior methods the penalty function is the product of two other functions $t(\rho)$ and $q(x)$ that we are going to define.

- Let q be a real-valued function on $\mathbb{R}^n$, such that is continuous at every point of the interior S of the feasible set X.
- If $\{x^k\}$ is any sequence of points in S that converges to some point $\hat{x}$ on the boundary of X, then we will assume that

$$\lim_{k\to\infty} q(x^k) = +\infty.$$

Remark: If $\hat{x}$ is in the boundary of X, then

$$I(\hat{x}) = \{i \mid g_i(\hat{x}) = 0\} \neq \emptyset$$

- Let t be a real-valued function on $\mathbb{R}$ such that

$$\rho^1 > \rho^2 > 0 \quad \Rightarrow \quad t(\rho^1) > t(\rho^2) > 0,$$

and

$$\lim_{k\to\infty} \rho^k = 0 \quad \Rightarrow \quad \lim_{k\to\infty} t(\rho^k) = 0.$$

The function $t(\rho^k)q(x)$ is called **interior penalty function** or **barrier function**.

The interior penalty method

The interior penalty method can be stated as follows:

- For $k = 0, 1, \ldots$ define

  $$G(x, \rho^k) = f(x) + t(\rho^k)q(x),$$

  to be the augmented objective function to be minimized in a sequence of unconstrained optimization problems given by

  $$(IP^k) \qquad \min_{x \in \mathbb{R}^n} G(x, \rho^k).$$

- Let $x^0 \in S$ be the starting point and assign a positive value to ρ^0.
- Solve (IP^0) by some unconstrained minimization technique starting at x^0, and let x^{0*} be a solution of (IP^0).
- Decrease ρ^0 to ρ^1 and solve (IP^1) starting at x^{0*}. Denote the optimal solution of (IP^1) by x^{1*}.
- Continue solving (IP^k) for a strictly decreasing sequence ρ^k starting always at x^{k-1*}.

The functions t and q of the interior penalty method

The most common choices for the function $t(\rho)$ are:

$$\mathbf{t}_1(\rho) = \rho, \quad t_2(\rho) = \rho^2,$$

and some common choices for the function $q(x)$ are

$$q_1(x) = -\sum_{i=1}^{m} \log g_i(x), \quad \mathbf{q}_2(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)},$$
$$q_3(x) = \sum_{i=1}^{m} \frac{1}{(g_i(x))^2}, \quad q_4(x) = \sum_{i=1}^{m} \frac{1}{\max(0, g_i(x))}.$$

Remark: Interior penalty methods are based on the idea proposed by C.W. Carroll in 1961 of transforming a constrained nonlinear problem into a sequence of constrained problems, by using the above $\mathbf{t}_1$ and $\mathbf{q}_2$ function.

The interior penalty method. Example

Example. Consider the following problem in one variable

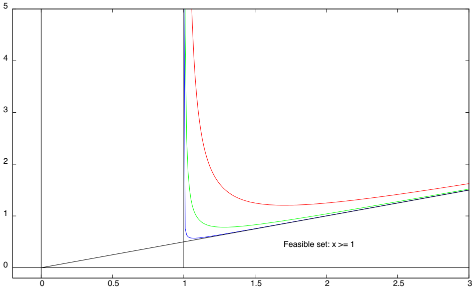
$$\min f(x) = \min \frac{1}{2}x,$$

subject to $g(x) = x - 1 \geq 0$.

The optimal solution is at $x^* = 1$ and $f(x^*) = 1/2$.

Suppose that we use the above t_2 and q_2 functions to define the barrier function

$$G(x, \rho^k) = f(x) + t(\rho^k)q(x) = \frac{1}{2}x + (\rho^k)^2 \frac{1}{x-1}.$$



Graph of $G(x, \rho^k)$ for $\rho = 0.5$ (red), 0.2 (green) and 0.05 (blue)

The interior penalty method. Example (cont.)

Since

$$G_x(x, \rho^k) = \frac{1}{2} - (\rho^k)^2 \frac{1}{(x-1)^2},$$

the unconstrained minimum of $G(x, \rho^k)$ is

$$x^{k*} = 1 + \sqrt{2}\rho^k > 1 \quad \text{if} \quad \rho^k > 0,$$

and

$$f(x^{k*}) = \frac{1}{2} + \frac{\rho^k}{\sqrt{2}}.$$

Thus the optimal unconstrained minima are all in S and converge to x^* as the values of ρ^k are successively reduced.

The interior penalty method. Convergence

In order to present a simple convergence proof of this method, we will assume that the q functions are positive for $x \in S$.

Remark: This assumption may not hold for the penalty

$$q_1(x) = -\sum_{i=1}^{m} \log g_i(x),$$

although convergence of the interior penalty method using this type of penalty can also be proved.

Theorem

- *Assume that the two regularity conditions already stated are satisfied, and that the q function is positive for $x \in S$.*
- *Suppose that the $G(x, \rho^k)$ attain their unconstrained minima in S for all k.*

If $\{\rho^k\}$ is a strictly decreasing sequence of positive numbers converging to zero, then there exists a convergence subsequence $\{x^{k_i *}\}$ of the optimal solutions to (IP^k), and the limit of any such convergent subsequence is optimal for (IP).

The interior penalty method. Convergence proof

Proof:

- By the second regularity condition, the intersection, $\Gamma(f, \alpha^0) \cap X$, of the level set $\Gamma(f, \alpha^0)$ with X is nonempty and compact.
- As a consequence, the continuous function f attains its minimum value $f(x^*)$ on X at some point $x^* \in X$.
- Let $G(x^{k*}, \rho^k)$ be the minimum value of the augmented objective function in problem (IP^k). Since

$$t(\rho^0) > t(\rho^1) > \dots \quad \Rightarrow \quad G(x, \rho^0) > G(x, \rho^1) > \dots$$

  and $G(x, \rho) \geq f(x)$, since both t and q are positive, we have that

$$G(x^{0*}, \rho^0) \geq G(x^{1*}, \rho^1) \geq \dots \geq f(x^*).$$

- Since $\{G(x^{k*}, \rho^k)\}$ is a strictly decreasing sequence bounded from below, it converges to a limit $\hat{G} \geq f(x^*)$.
- If we suppose that $\hat{G} > f(x^*)$ we will reach a contradiction.
- From the first regularity condition (X closed, S nonempty, X is the closure of S) and the continuity of f, we conclude that there exists $\delta > 0$ and $N_\delta(x^*)$ such that

$$S \cap N_\delta(x^*) \neq \emptyset, \quad \text{and} \quad f(x) \leq \frac{1}{2}(\hat{G} + f(x^*)) = \hat{G} - \frac{1}{2}(\hat{G} - f(x^*)) \quad \forall x \in N_\delta(x^*)$$

The interior penalty method. Convergence proof (cont.)

- Take any point $\overline{x} \in S \cap N_\delta(x^*)$.
- From the above inequality and the properties of the function t, it follows that there exists a natural number K such that for every $k \geq K$

$$t(\rho^k)q(\overline{x}) < \frac{1}{4}(\hat{G} - f(x^*)).$$

  Thus

$$G(x^{k*}, \rho^k) \leq f(\overline{x}) + t(\rho^k)q(\overline{x}) < \hat{G} - \frac{1}{4}(\hat{G} - f(x^*))$$

  for $k \geq K$, contradicting that $\{G(x^{k*}, \rho^k)\}$ monotonically converges to $\hat{G}$. Hence $\hat{G} = f(x^*)$.

- From the second regularity condition, there exists a $\hat{K}$ such that for all $k \geq \hat{K}$ the points x^{k*} are in a compact set, and so there exists a subsequence $\{x^{k_i*}\}$ that converges to a limit $\hat{x} \in X$.
- Suppose that $\hat{x}$ is not optimal for (PI), then $f(\hat{x}) > f(x^*)$, and the sequence

$$\{f(x^{k_i*}) + t(\rho^{k_i})q(x^{k_i*}) - f(x^*)\}$$

  does not converge to zero, thereby contradicting

$$\lim_{k\to\infty} G(x^{k*}, \rho^k) = f(x^*).$$

  Hence we must have $f(\hat{x}) = f(x^*)$, and $\hat{x}$ is optimal for (PI). $\square$

The interior penalty method. Strongly consistent problems

The **most important assumption in the above theorem** is that the $G(x, \rho^k)$ attain their minima in S or, equivalently, that the problems (IP^k) have their solutions in S.

Definition
Problem (PI) is said to be **strongly consistent** is the first regularity condition:

"the set X is closed, S is nonempty, and X is the closure of S"

holds, and the interior of X is nonempty :

$$S = \{x \mid x \in \mathbb{R}^n,\ g_i(x) > 0,\ i = 1, ..., m\} \neq \emptyset.$$

The next Lemma will give a sufficient condition that ensures the existence of these solutions.

Lemma
Assume that $X \subset \mathbb{R}^n$ is compact and that problem (PI) is strongly consistent. Then, the $G(x, \rho^k)$ attain their unconstrained minima in S.

The interior penalty method. Proof of the Lemma

Proof:

- Let
$$\inf_{x\in S} G(x,\rho) = \alpha.$$
- According to the definition of infimum, there exists a sequence of points $x^i \in S$ such that $\lim_{i\to\infty} G(x^i,\rho) = \alpha$.
- Since $\{x^i\}$ is contained in the compact set X, it has a convergent subsequence $\{x^{i_j}\}$ such that
$$\lim_{i_j\to\infty} x^{i_j} = \hat{x} \in X.$$
- Assume that $\hat{x} \in S$, then, by continuity of G, and since any subsequence of $\{G(x^i,\rho)\}$ must converge to α, we have
$$\lim_{i_j\to\infty} G(x^{i_j},\rho) = \lim_{i_j\to\infty} f(x^{i_j}) + \lim_{i_j\to\infty} t(\rho)q(x^{i_j}) = f(\hat{x}) + t(\rho)q(\hat{x}) = \alpha,$$
hence
$$G(\hat{x},\rho) = \lim_{i_j\to\infty} G(x^{i_j},\rho) = \min_{x\in S} G(x,\rho).$$

The interior penalty method. Proof of the Lemma (cont.)

- Suppose now that $\hat{x} \notin S$, then $\hat{x}$ is on the boundary of X.
- Using that q satisfies $\lim_{k\to\infty} q(x^k) = +\infty$, the positivity of $t(\rho)$ and the above chain of equalities, we get

$$\inf_{x\in S} G(x,\rho) = f(\hat{x}) + \lim_{i_j\to\infty} t(\rho)q(x^{i_j}) = +\infty,$$

which is a **contradiction**, so $\hat{x} \in S$. $\square$

Parameter-free penalty methods. The selection of ρ^k

An open question on the implementation of penalty methods concerns the choice of the parameters ρ.

- It is necessary to decide on the **initial value** of the parameter ρ^0, and on the **rule to modify** the value of ρ in order to obtain a monotonically decreasing sequence that converges to zero.

- These questions can be avoided by modifying penalty function methods so that **the parameters are automatically chosen** or, equivalently, the **methods** are modified so that they **become parameter free**.

- Both, exterior and interior penalty methods can be modified with this purpose.

Parameter-free exterior penalty methods

Parameter-free exterior penalty methods

Suppose that we have a **lower estimate** w^0 **of the value of the objective function** f at its global minimum x^* over a feasible set X of problem (P), that is

$$w^0 \leq f(x^*).$$

Assume also that we solve the unconstrained optimization

$$(EPF^0) \qquad \min_{x\in\mathbb{R}^n} \hat{F}(x, w^0) = \psi(w^0 - f(x)) + \sum_{i=1}^{m} \psi(g_i(x)) + \sum_{j=1}^{p} \xi(h_j(x))$$
$$= |\min(0, w^0 - f(x))|^\alpha + \sum_{i=1}^{m} |\min(0, g_i(x))|^\alpha + \sum_{j=1}^{p} |h_j(x)|^\beta$$

for certain $\alpha \geq 1$ and $\beta \geq 1$.

Let x^{0*} be the optimal solution of (EPF^0). If the optimal solution of (P), x^*, happens to be the unconstrained minimum of f over $\mathbb{R}^n$, then $x^{0*} = x^*$ and the algorithm terminates.

Otherwise

$$w^0 \leq f(x^{0*}) \leq f(x^*),$$

provided the regularity conditions are satisfied. Then we proceed as follows:

Parameter-free exterior penalty methods (cont.)

Let

$$(EPF^k) \qquad \min_{x \in \mathbb{R}^n} \hat{F}(x, w^k) = \psi(w^k - f(x)) + \sum_{i=1}^{m} \psi(g_i(x)) + \sum_{j=1}^{p} \xi(h_j(x))$$

$$= |\min(0, w^k - f(x))|^\alpha + \sum_{i=1}^{m} |\min(0, g_i(x))|^\alpha + \sum_{j=1}^{p} |h_j(x)|^\beta,$$

where $\{w^k\}$ is a strictly increasing sequence whose elements are computed from w^{k-1} and the optimal solution of (EPF^{k-1}).

It can be proved that solvig (EPF^k) for k =1,2,... we obtain a sequence of points $\{x^{k*}\}$ that contains a subsequence that converges to the optimal solution of (P)

Parameter-free interior penalty methods

Consider the problem

$$(PI) \quad \min \quad f(x) \qquad s.t. \quad g_i(x) \geq 0, \quad i = 1, ..., m$$

with the regularity conditions stated for the family of interior penalty methods.

Parameter-free interior penalty methods are based on solving a sequence of unconstrained optimization problems such as

$$(IPF^k) \quad \min_{x \in \mathbb{R}^n} \hat{G}(x, x^{k-1*}) = \frac{1}{f(x^{k-1*}) - f(x)} + \sum_{i=1}^{m} \frac{1}{g_i(x)},$$

for k =1,2,... where x^{0*} is an arbitrary point in X^0 and x^{k-1^*} is the optimal solution of (IPF^{k-1}). This method is called **SUMT without parameters**.

Using the q_1 function instead of q_2, the method obtained is called **method of centres**, which is based on solving the sequence of problems

$$\min_{x \in \mathbb{R}^n} \hat{G}_1(x, x^{k-1*}) = -\log(f(x^{k-1*}) - f(x)) - \sum_{i=1}^{m} \log\, g_i(x),$$

which is equivalent to

$$\max_{x \in \mathbb{R}^n} (f(x^{k-1*}) - f(x)) \prod_{i=1}^{m} g_i(x).$$

Parameter-free interior penalty methods

Again, under conditions similar to those stated for the general interior penalty methods, it can be shown that parameter-free interior penalty methods converge to an optimal solution of (PI).

Example. Consider again the problem

$$\begin{array}{ll} \text{minimize} & f(x) = \frac{1}{2}x, \\ \text{subject to} & g(x) = x - 1 \geq 0, \end{array}$$

and solve it using SUMT without parameters.

At iteration k, we solve the unconstrained optimization problem

$$\min_{x\in\mathbb{R}} \left(\frac{1}{f(x^{k-1*}) - f(x)} + \frac{1}{x-1} \right) = \min_{x\in\mathbb{R}} \left(\frac{2}{x^{k-1*} - x} + \frac{1}{x-1} \right).$$

The optimal solution of this problem is

$$\frac{2}{(x^{k-1*} - x)^2} - \frac{1}{(x-1)^2} = 0 \quad \Rightarrow \quad x^{k*} = \frac{\sqrt{2} + x^{k-1*}}{\sqrt{2}+1}.$$

Since $x^* = 1$, we obtain

$$\frac{x^{k*} - x^*}{x^{k-1*} - x^*} = \frac{1}{\sqrt{2}+1},$$

which shows that the rate of convergence is linear.

The logarithmic barrier function

- The **logarithmic barrier functions** (q_1)
$$B(x) = -\sum_{j=1}^{r} \log(-g_j(x))$$
have been central in the development of interior point methods that have extensively been applied to linear and quadratic problems.
- Consider the optimization problem written as
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X \quad \text{and} \quad g_j(x) \leq 0, \quad j = 1, \ldots, r, \end{array}$$
where f and g_i are continuous real-valued functions and X is a closed set.
- Note that $B(x)$ is convex if all the constraint functions g_j are convex.
- The iterior of the feasible set X is
$$S = \{x \in X \mid g_j(x) < 0, \ j = 1, \ldots, r\}.$$
- Wewill assume that S is nonempty and that any feasible point that is not in S can be approached arbitrarily closely by a point from S (it can be seen that this property holds if the constraint functions g_j are convex.
- Note also that the barrier function is defined only on the interior set S (so, the successive iterates of any interior point method must be interior points).

The logarithmic barrier function

Example. Consider the 2-dimensional problem

$$\begin{array}{ll} \text{minimize} & f(x) = \min \frac{1}{2}(x^2 + y^2), \\ \text{subject to} & 2 \leq x, \quad (2 - x \leq 0) \end{array}$$

with optimal solution $x^* = (2, 0)$.

For the case of the logarithmic barrier function $B(x) = -\log(x - 2)$, we have

$$x^k = \min_{x>2} \left(\frac{1}{2}(x^2 + y^2) - \epsilon^k \log(x - 2) \right),$$

$$F(x, y, \epsilon) = \frac{1}{2}(x^2 + y^2) - \epsilon^k \log(x - 2),$$

$$F_x(x, y, \epsilon) = x - \frac{\epsilon^k}{x - 2} = 0 \quad \Rightarrow \quad x = 1 \pm \sqrt{1 + \epsilon^k},$$

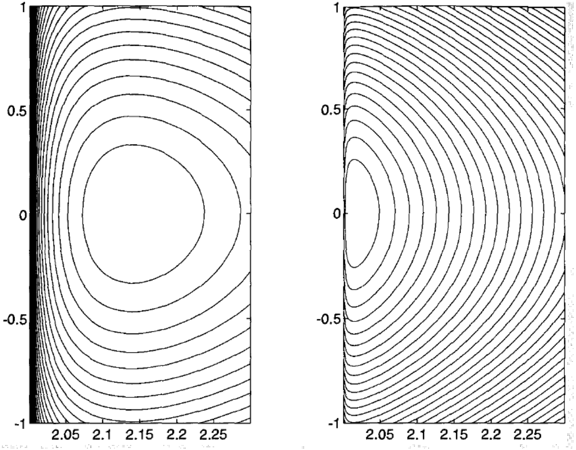
$$F_y(x, y, \epsilon) = y = 0 \quad \Rightarrow \quad y = 0,$$

so, since $x > 2$,

$$x^k = \left(1 + \sqrt{1 + \epsilon^k}, 0\right),$$

and, as ϵ^k is decreased the unconstrained minimum x^k approaches the constrained minimum x^*.

The logarithmic barrier function



Equal cost surfaces of $f(x) + \epsilon B(x)$ for $\epsilon = 0.3$ (left) and 0.03 (right).

Linear programming and the logarithmic barrier

Consider the linear programming problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array}$$

with $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and A an $m \times n$ matrix of rank m, and where $x \geq 0$ means that all the coordinates of x are positive or zero. Assume that the problem has at least on optimal solution.

A note on duality: *It can be shown that the dual problem*

$$\begin{array}{ll} \textit{maximize} & b^T y \\ \textit{subject to} & A^T y \leq c, \end{array}$$

has also an optimal solution. Furthermore, the optimal values of the primal and the dual problems are equal.

The logarithmic barrier method involves finding for several $\epsilon > 0$

$$x(\epsilon) = \min_{x \in S} F(x, \epsilon) = \min_{x \in S} \left(c^T x - \epsilon \sum_{i=1}^{n} \log x_i \right),$$

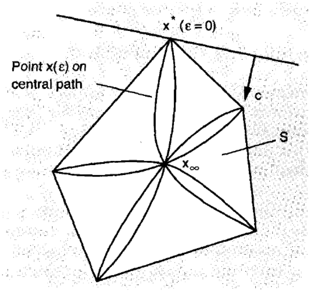
where S is the open set

$$S = \{x \mid Ax = b, \ x > 0\}$$

Linear programming and the logarithmic barrier

- We assume that S is nonempty and bounded.
- Since $\log x_i$ grows to $+\infty$ as $x_i \to 0$, there axists at least one global minimum of $F(x, \epsilon)$ over S (due to Weierstrass theorem).
- The minimum of the objective function $c^T x$ must be unique because the function $F(x, \epsilon)$ can be seen to be strictly convex.

- For given A, b and c, as ϵ goes to 0, $x(\epsilon)$ follows a trajectory that is known as the **central path**.



Central paths associated to 10 different values of the cost vector c.

Linear programming and the logarithmic barrier

Note that:

- In the above figure, all the paths start at the same point x_∞, the **analytic center** which corresponds to $\epsilon = \infty$

$$x_\infty = \min_{x \in S} \left(-\sum_{i=1}^{n} \log x_i \right)$$

- If c is such that the problem has a unique optimal solution x^*, the central path ends at x^* (since every limit point of a sequence generated by a barrier method is a global minimum of the original constrained problem).
- If c is such that the linear optimization problem has multiple optimal solutions x^*, it can be shown that the central path ends at one of the optimal solutions

Following approximately the central path

Implementation of the logarithmic barrier method

- The most strightforward way to implement the logarithmic barrier method is to use some iterative algorithm to minimize the function $F(x, \epsilon_k)$ for some decreasing sequence $\epsilon_k \to 0$.
- This is equivalent to finding a sequence $\{x(\epsilon^k)\}$ of points on the central path.
- This approach is inefficient because it requires an infinite number of iterations to compute each point $x(\epsilon^k)$.
- A far more efficient approach , in which each minimization is done approximately through a few iterations, is obtained using the **constrained version of Newton's method**:
  - For a fixed ϵ and a given $x \in S$, this method replaces x by

    $$\tilde{x} = x + \alpha(\overline{x} - x)$$

    where α is a stepsize selected by some rule and $\overline{x}$ is the pure Newton iterate defined as the optimal solution of the quadratic problem in the point $z \in \mathbb{R}^n$

    $$\begin{array}{ll} \text{minimize} & \nabla F(x,\epsilon)^T(z-x) + \frac{1}{2}(z-x)^T \nabla^2 F(x,\epsilon)(z-x) \\ \text{subject to} & Ax = b. \end{array}$$

Following approximately the central path

- The value of $\overline{x}$ is

$$\overline{x} = x - (1/\epsilon)X^2(c - \epsilon x^{-1} - A^T\lambda),$$

where X denotes the diagonal matrix with the coordinates x_i along the diagonal, x^{-1} denotes the vector with coordinates $1/x_i$

$$X = \begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & x_n \end{pmatrix}, \quad \begin{pmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{pmatrix},$$

and

$$\lambda = (AX^2A^T)^{-1}AX^2(c - \epsilon x^{-1}).$$

- The above formula for $\overline{x}$ can also be written as

$$\overline{x} = x - X\, q(x, \epsilon) = x - X\left((1/\epsilon)X^2(c - \epsilon x^{-1} - A^T\lambda)\right).$$

- Since $q(x, \epsilon) = X^{-1}(\overline{x} - x)$, we can consider $\|q(x, \epsilon)\|$ as a measure of proximity of the current point x to the point $x(\epsilon)$ on the central path. In particular, it can be seen that $q(x, \epsilon) = 0$ if and only if $x(\epsilon) = x$.

Following approximately the central path

- The key point is that for the convergence of the logarithmic barrier method it is sufficient to stop the minimization of $F(x, \epsilon^k)$ and decrease ϵ^k to ϵ^{k+1} once the current iterate satisfies

$$\|q(x^k, \epsilon^k)\| < 1.$$

This is: if a sequence of pairs $x^k, \epsilon^k)$ satisfies

$$\|q(x^k, \epsilon^k)\| < 1, \quad 0 < \epsilon^{k+1} < \epsilon^k, \quad k = 0, 1, \ldots \quad \epsilon^k \to 0,$$

then every limit point of $\{x^k\}$ is an optimal solution of the linear programming problem.