Penalty and barrier methods

Recherche opérationnelle
Dimitri Watel - ENSIIE

2024

The penalty and barrier methods are two approaches that allow transforming a constraint optimization problem into a series of unconstrained optimization problems, so that solving these problems leads to finding an optimal solution to the original problem. It will be assumed that we have a black box capable of efficiently solving an unconstrained optimization problem.

We want to solve the following problem (O).

$$\begin{array}{llll} (O): \min & f(x) & x \in \mathbb{R}^n \\ s.c. & g_i(x) \leq 0 & \forall\, i \in [\![1; m]\!] \\ & h_j(x) = 0 & \forall\, j \in [\![1; p]\!] \end{array}$$

We write S the set of feasible solutions. Another way to define the problem is the following:

$$(O): \min_{x \in S} f(x)$$

In the continuation of this course, we will assume that f, g_i, and h_j are $\mathcal{C}^1$. We also assume that these functions are defined on all of $\mathbb{R}^n$. We can deduce that S is closed. This property is used for one of the methods described in this chapter. These constraints can be relaxed as long as it does not jeopardize the use of the black box to optimize the unconstrained versions of our problems.

1 Principle and definition of the methods

The principle of penalty and barrier methods is to integrate the constraints into the objective function so that not satisfying these constraints results in a high objective function value. The two methods have a different way of achieving this. The penalty method penalizes values x outside of S. The barrier method prevents x from leaving S by penalizing it as it approaches the boundary.

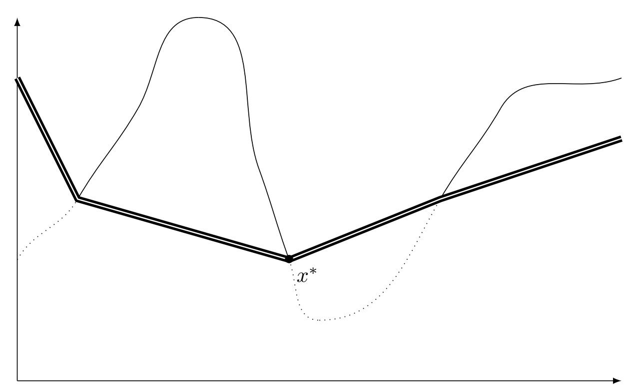
1.1 Penalty Method

For this method, we assume that we have a positive continuous function $P : \mathbb{R}^n \to \mathbb{R}$ such that $P(x) = 0$ if and only if $x \in S$. We then solve the following problem P_μ:

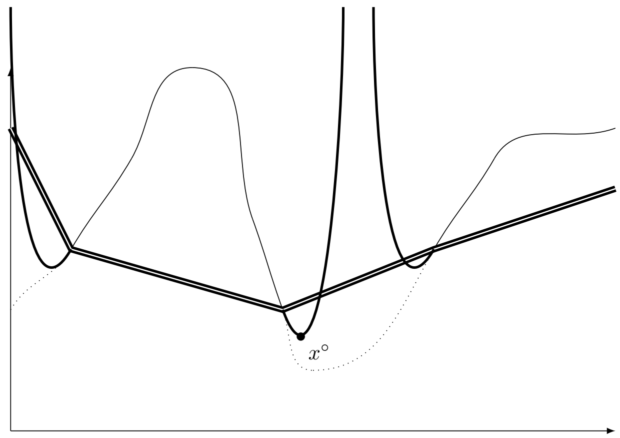
$$(P_\mu): \min_{x \in \mathbb{R}^n} f(x) + \mu \cdot P(x)$$

As we can see, we are now minimizing a function without constraints on $\mathbb{R}^n$. Assuming that μ is large enough, then the function $\mu P(x)$ will heavily penalize any value of x that is not in S. If x is in S, since $P(x) = 0$, the function f is not affected, and the optimal solution of (P_μ) will naturally tend to be an optimal solution of (O) as we increase μ. The complicated question is what value to assign to μ so that the penalty is sufficient?

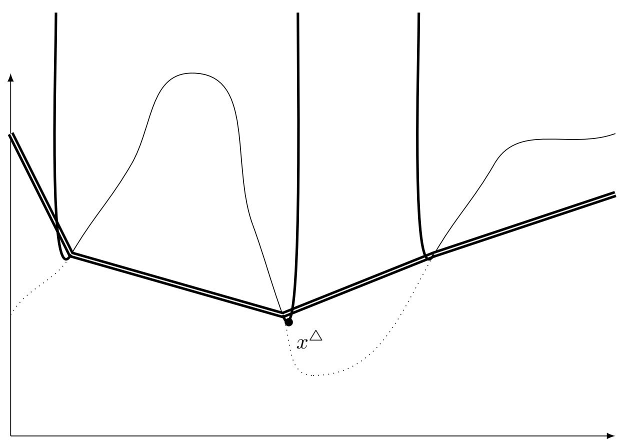
Let's take the following example where the constraint is to be above the function with the double line. The optimal solution x^* of (O) is indicated in the drawing.



A penalty is applied. We obtain a new function that can be optimized without constraints. We find the optimal solution x° below. We notice that this solution is not in S. This is normal; in the unconstrained problem, we can return a solution outside of S. The goal of the penalty is to be sufficiently high to avoid it.



If we increase the penalty, we obtain a new optimal solution $x^{\triangle}$, closer to S. The more we increase the penalty, the closer we will get to S until we converge to x^*.



A classic penalty function is the Courant-Beltrami penalty, described below. If the functions g_i and h_j are $\mathcal{C}^1$, then P is also $\mathcal{C}^1$.

$$P(x) = \sum_{i=1}^{m} \max(0, g_i(x))^2 + \sum_{j=1}^{p} h_j(x)^2$$

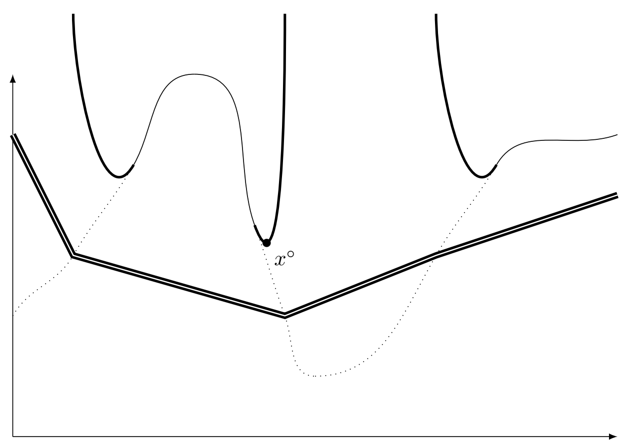
1.2 Barrier Method

We denote $\mathring{S}$ as the interior of S. We assume $\mathring{S}$ is non-empty and that there exists an optimal solution x^* such that every neighborhood of x^* intersects $\mathring{S}$. We assume we have a function $B : \mathring{S} \to \mathbb{R}$ that is continuous and positive, and such that $B(x) \to +\infty$ as x approaches the boundary of S. We then solve the following problem B_μ:

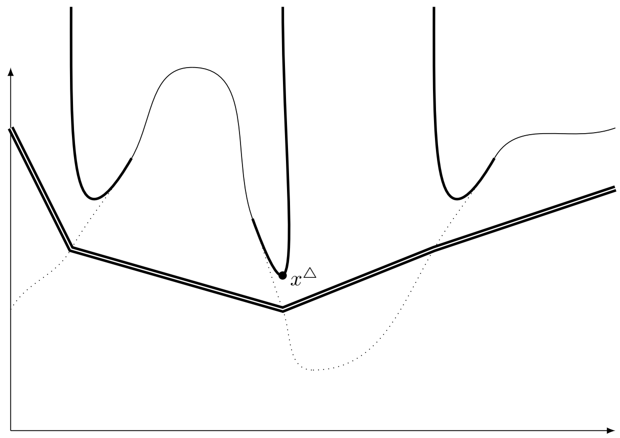
$$(B_\mu) : \min_{x \in \mathring{S}} f(x) + \mu \cdot P(x)$$

We are now minimizing a function without constraints on $\mathring{S}$. Since $B(x)$ tends to infinity as x approaches the boundary of S, an optimal solution of B_μ cannot be on the edge of S. Thus, if the optimal solution x^* of (O) is central in S, B_μ makes it easy to find. Assuming that μ is small enough, the function $\mu B(x)$ will allow us to find x^* even if this solution is close to the boundary. If x^* is right on the boundary itself, then the optimal solution of (B_μ) will naturally tend towards x^* as we decrease μ. The complex question is what value should be assigned to μ so that the barrier is not too penalizing?

Let's take the same example as for the penalty method, and apply a barrier. We obtain the new function to optimize. The optimal solution x° is marked in the drawing. As explained earlier, since x^* is on the boundary, the barrier prevents us from reaching it.



If we decrease μ, then the barrier collapses and the optimal solution approaches x^*.



Here are two classic barrier functions, namely the logarithmic barrier and the inverse barrier.

If $-1 < g_i(x) < 0$ iff $x \in \mathring{S}$, $B(x) = \sum_{i=1}^{m} -\log(-g_i(x))$

If $g_i(x) < 0$ iff $x \in \mathring{S}$, $B(x) = \sum_{i=1}^{m} \frac{-1}{g_i(x)}$

In general, the penalty method is easier to implement. This is because optimization occurs over the entire $\mathbb{R}^n$, whereas it occurs over $\mathring{S}$ in the case of the barrier method. Furthermore, classical barriers have constraints that make it not always possible to apply them. However, the barrier method has an advantage in that for each value of μ, optimizing B_μ produces a feasible solution to the problem (O). As can be seen in the previous example, this is not the case with the penalty method, which can produce infeasible solutions.

2 Convergence properties of methods

The convergence properties are similar in both methods. We will show, in both cases, that these methods, if they converge, always converge to an optimal solution of (O). We will then provide a case of a function f where we can guarantee this convergence.

2.1 Penalty Method

Theorem 2.1. *Let P be a penalty function. We assume that f is continuous and that (O) has an optimal solution x^*. We also assume that, for any $\mu > 0$, (P_μ) has an optimal solution x_μ. Let $(\mu_p)_{p\in\mathbb{N}}$ be a sequence such that $\lim_{p\to+\infty} \mu_p = +\infty$. If $(x_{\mu_p})_{p\in\mathbb{N}}$ converges, then it converges to an optimal solution of (O).*

Proof. We set $q(x, \mu) = f(x) + \mu P(x)$. Let $x_\infty = \lim_{p\to+\infty} x_{\mu_p}$. Since the sequence $(x_{\mu_p})_{p\in\mathbb{N}}$ converges to x_∞ and since f is continuous on $\mathbb{R}^n$, then

$$\lim_{p\to+\infty} f(x_{\mu_p}) = f(x_\infty) \tag{1}$$

We will show that $f(x_\infty) \leq f(x^*)$ and that x_∞ is feasible.

As $x_\mu = \arg\min_{x\in\mathbb{R}^n} q(x, \mu)$

$$q(x_{\mu_p}, \mu_p) \leq q(x^*, \mu_p)$$

However $x^* \in S$, donc $P(x^*) = 0$

$$\begin{aligned} q(x_{\mu_p}, \mu_p) &\leq f(x^*) \\ f(x_{\mu_p}) + \mu_p \cdot P(x_{\mu_p}) &\leq f(x^*) \qquad (2) \\ P(x_{\mu_p}) &\leq \frac{f(x^*) - f(x_{\mu_p})}{\mu_p} \end{aligned}$$

However $\lim_{p\to+\infty} \mu_p = +\infty$ et By (1)

$$\lim_{p\to+\infty} \frac{f(x^*) - f(x_{\mu_p})}{\mu_p} = 0$$

Since the function P is positive, the sequence $P(x_{\mu_p})_{p\in\mathbb{N}}$ is bounded between two sequences converging to 0. Therefore, this sequence also converges to 0. Since P is continuous on $\mathbb{R}^n$, then

$$\lim_{p\to+\infty} P(x_{\mu_p}) = P(x_\infty) = 0$$

So $x_\infty \in S$. Since $P(x_{\mu_p})$ and μ_p are positive, and by (2)

$$f(x_{\mu_p}) \leq f(x^*) \tag{3}$$

So, since f is continuous

$$f(x_\infty) \leq f(x^*) \tag{4}$$

$$\tag{5}$$

Then x_∞ is optimal solution of (O).

□

We provide below a case where convergence is guaranteed. We recall that a coercive function is a function such that $\lim_{\|x\|\to+\infty} f(x) = +\infty$.

Theorem 2.2. *If f is coercive and if (O) admits an optimal solution x^*, then, for all μ, (P_μ) admits an optimal solution x_μ and there exists a subsequence of $(x_p)_{p\in\mathbb{N}}$ that converges.*

Proof. We set $q(x, \mu) = f(x) + \mu P(x)$. If f is coercive, since $q(x, \mu) \geq f(x)$, then q is also a coercive function with respect to x, for any fixed μ. Moreover, every continuous coercive function has a global minimum, so (P_μ) has an optimal solution x_μ.

Furthermore, due to coercivity, the norm of x_μ is bounded. Applying the Bolzano-Weierstrass theorem, we obtain the desired result. □

2.2 Barrier Method

Theorem 2.3. *Let B be a barrier function. We assume that f is continuous and that (O) has an optimal solution x^*. We also assume that for every $\mu > 0$, (B_μ) has an optimal solution x_μ. Let $(\mu_p)_{p\in\mathbb{N}}$ be a strictly decreasing sequence converging to 0. If $(x_{\mu_p})_{p\in\mathbb{N}}$ converges, then it converges to an optimal solution of (O).*

Proof. We define $q(x,\mu) = f(x) + \mu B(x)$. Let $x_0 = \lim_{p\to+\infty} x_{\mu_p}$. Since the sequence $(x_{\mu_p})_{p\in\mathbb{N}}$ converges to x_0 and since f is continuous on $\mathbb{R}^n$, then

$$\lim_{p\to+\infty} f(x_{\mu_p}) = f(x_0) \tag{6}$$

Given that the barrier method optimizes on $\mathring{S}$, we have $x_{\mu_p} \in \mathring{S}$. Recall that S is closed and thus $x_0 \in S$. It remains to demonstrate that $f(x_0) = f(x^*)$.

First, let's show that there is a limit to $(q(x_{\mu_p},\mu_p))_{p\in\mathbb{N}}$ as p tends to infinity. To do this, we will show that this sequence is bounded and decreasing.

$$\begin{aligned} q(x_{\mu_p},\mu_p) &= f(x_{\mu_p}) + \mu_p B(x_{\mu_p}) \\ &\geq f(x^*) + \mu_p B(x_{\mu_p}) \geq f(x^*) \end{aligned} \tag{7}$$

It is indeed bounded, let's show that it is decreasing.

$$q(x_{\mu_p},\mu_p) = f(x_{\mu_p}) + \mu_p B(x_{\mu_p})$$

Since $(\mu_p)_{p\in\mathbb{N}}$ is decreasing.

$$\geq f(x_{\mu_p}) + \mu_{p+1} B(x_{\mu_p}) = q(x_{\mu_p},\mu_{p+1})$$

As $x_{\mu_{p+1}} = \arg\min_{x\in\mathring{S}} q(x,\mu_{p+1})$

$$\geq q(x_{\mu_{p+1}},\mu_{p+1}) \tag{8}$$

By (7) and (8), there exists a limit Q to the sequence $(q(x_{\mu_p},\mu_p))_{p\in\mathbb{N}}$. Now let $x \in \mathring{S}$.

$$q(x_{\mu_p},\mu_p) \leq f(x) + \mu_p B(x)$$

Since $(\mu_p)_{p\in\mathbb{N}}$ converges to 0 as p tends to infinity.

$$Q \leq f(x) \tag{9}$$

Since f is continuous on $\mathbb{R}^n$, for every $\varepsilon > 0$, there exists a neighborhood V of x^* such that, for all $x \in V$, $f(x) \leq f(x^*) + \varepsilon$. At the beginning of the presentation of the method, in section 1.2, we assumed that every neighborhood of x^* intersects $\mathring{S}$. Therefore, there exists $x \in \mathring{S}$ such that $f(x) \leq f(x^*) + \varepsilon$. From (7) and (9), we deduce that, for all $\varepsilon > 0$,

$$\begin{aligned} f(x^*) &\leq Q \leq f(x^*) + \varepsilon \\ f(x^*) &= Q \end{aligned}$$

To conclude

$$\mu_p B(x_{\mu_p}) = q(x_{\mu_p},\mu_p) - f(x_{\mu_p})$$

By 6

$$\lim_{p\to+\infty} \mu_p B(x_{\mu_p}) = Q - f(x_0) = f(x^*) - f(x_0)$$

However $\mu_p B(x_{\mu_p}) \geq 0$ for all $p \in \mathbb{N}$, so

$$f(x^*) - f(x_0) \geq 0$$

Then x_0 is an optimal solution of (O).

□

Remark 1. The continuity of B is not used in this proof. But it is easier to use unconstrained optimization algorithms whith this property.

Theorem 2.4. *If f is coercive then $\exists$ a subsequence of $(x_{1/p})_{p\in\mathbb{N}}$ which is convergent.*

Proof. The proof is the same than the one for Theorem 2. □