TMA4180 Optimization: Nonsmooth Penalty Functions

Elisabeth Köbis

NTNU, TMA4180, Optimering, Spring 2021

March 12th, 2021

Nonsmooth Exact Penalty Functions

Some penalty functions are exact, which means that, for certain choices of their penalty parameters, a single minimization with respect to x can yield the exact solution of the nonlinear programming problem. This property is desirable because it makes the performance of penalty methods less dependent on the strategy for updating the penalty parameter. The quadratic penalty function is not exact because its minimizer is generally not the same as the solution of the nonlinear program for any positive value of μ. Now, we discuss nonsmooth exact penalty functions, which have proved to be useful in a number of practical contexts.

ℓ_1 Penalty Function

$$\begin{aligned} &\boxed{\min_{x\in\mathbb{R}^n} f(x)} \\ \text{subject to } &c_i(x) = 0, \quad i \in \mathcal{E}, \\ &c_i(x) \geq 0, \quad i \in \mathcal{I}. \end{aligned} \tag{P}$$

Let $\mu > 0$ be a *penalty parameter*. We define

$$\min_{x\in\mathbb{R}^n} \; \Phi_1(x;\mu) := f(x) + \mu \sum_{i\in\mathcal{E}} |c_i(x)| + \mu \sum_{i\in\mathcal{I}} [c_i(x)]^-, \tag{$\mathsf{P}_{\mathsf{penalty}}$}$$

where again $[y]^- := \max(-y, 0)$. Note that $\Phi_1(x;\mu)$ is not differentiable at some x, because of the presence of the absolute value and $[\cdot]^-$ functions.

ℓ_1 Penalty Function

Theorem

Suppose that x^* is a strict local minimizer of the nonlinear programming problem (P) at which the first-order necessary conditions of Theorem 12.1 in N&W are satisfied with Lagrange multipliers λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$. Then x^* is a local minimizer of ($\text{P}_{\text{penalty}}$) for all $\mu > \mu^*$, where

$$\mu^* = ||\lambda^*||_\infty = \max_{i \in \mathcal{E} \cup \mathcal{I}} |\lambda_i^*|.$$

If, in addition, the second-order sufficient conditions of Theorem 12.6 in N&W hold and $\mu > \mu^*$, then x^* is a strict local minimizer of ($\text{P}_{\text{penalty}}$).

For the proof, see Theorem 4.4 in S. P. Han, O. L. Mangasarian: Exact penalty functions in nonlinear programming, Mathematical Programming, 17 (1979), pp. 251–269.

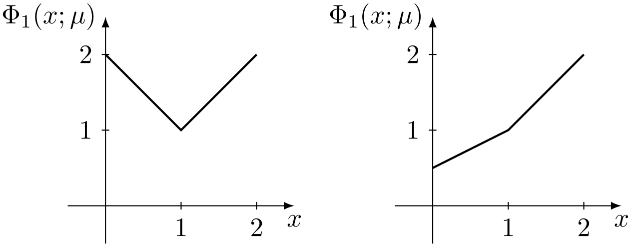
ℓ_1 Penalty Function

Example

$$\min_{x\in\mathbb{R}} x \quad \text{subject to } x-1\geq 0$$

with the minimizer $x^*=1$. We have

$$\Phi_1(x;\mu) = f(x)+\mu[c_1(x)]^- = \begin{cases} (1-\mu)x+\mu & (x\leq 1) \\ x & (x>1) \end{cases}$$



Left: $\mu>1$. Right: $\mu<1$.

ℓ_1 Penalty Function

Recall the directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ in the direction $p \in \mathbb{R}^n$:

$$D(f(x); p) := \lim_{\epsilon \to 0} \frac{f(x + \epsilon p) - f(x)}{\epsilon}.$$

ℓ_1 Penalty Function

Definition: Stationary Points

A point $\hat{x} \in \mathbb{R}^n$ is a *stationary point for the penalty function* $\Phi_1(x;\mu)$ if

$$\forall p \in \mathbb{R}^n : \quad D(\Phi_1(\hat{x};\mu);p) \geq 0.$$

Let the *measure of infeasibility* be defined as

$$h(x) := \sum_{i\in\mathcal{E}} |c_i(x)| + \sum_{i\in\mathcal{I}} [c_i(x)]^-.$$

Then we call $\hat{x} \in \mathbb{R}^n$ is a *stationary point of the measure of infeasibility* if

$$\forall p \in \mathbb{R}^n : \quad D(h(\hat{x});p) \geq 0.$$

If a point is infeasible for (P) but stationary with respect to the infeasibility measure h, we say that it is an *infeasible stationary point*.

ℓ_1 Penalty Function

Example

$$\min_{x \in \mathbb{R}} x \quad \text{subject to } x - 1 \geq 0$$

with the minimizer $x^* = 1$. We have

$$\Phi_1(x;\mu) = f(x) + \mu[c_1(x)]^- = \begin{cases} (1-\mu)x + \mu & (x \leq 1) \\ x & (x > 1). \end{cases}$$

We have

$$D(\Phi_1(x;\mu);p) = \begin{cases} p & (p \geq 0) \\ (1-\mu)p & (p < 0). \end{cases}$$

It follows that when $\mu > 1$, we have $D(\Phi_1(x;\mu);p) \geq 0$ for all $p \in \mathbb{R}$.

ℓ_1 Penalty Function

Theorem

Suppose that $\hat{x}$ is a stationary point of the penalty function $\Phi_1(x;\mu)$ for all μ greater than a certain threshold $\hat{\mu} > 0$. Then, if $\hat{x}$ is feasible for the nonlinear program (P), it satisfies the KKT conditions. Is $\hat{x}$ is not feasible for (P), it is an infeasible stationary point.

Proof: Suppose that $\hat{x}$ is feasible. Then

$$D(\Phi_1(\hat{x},\mu);p) \underset{\substack{\uparrow \\ \text{exercise}}}{=} \nabla f(\hat{x})^T p + \mu \sum_{i\in\mathcal{E}} |\nabla c_i(\hat{x})^T p| + \mu \sum_{i \in \mathcal{I} \cap \mathcal{A}(\hat{x})} [\nabla c_i(\hat{x})^T p]^- .$$

Consider any direction p in the linearized feasible direction set $\mathcal{F}(\hat{x})$. (Recall: $\mathcal{F}(\hat{x}) := \{ d \mid d^T \nabla c_i(\hat{x}) = 0 \ \forall i \in \mathcal{E},\ d^T \nabla c_i(\hat{x}) \geq 0 \ \forall i \in \mathcal{A}(\hat{x}) \cap \mathcal{I} \}$)

ℓ_1 Penalty Function

we have:

$$|\nabla c_i(\hat{x})^T p| + \sum_{i \in \mathcal{I} \cap \mathcal{A}(\hat{x})} [\nabla c_i(\hat{x})^T p]^- = 0.$$

Because $\hat{x}$ is a stationary point of $\phi_1(x;\mu)$, we have:

$$0 \leq D(\phi_1(\hat{x};\mu);p) = \nabla f(\hat{x})^T p \qquad \forall p \in \mathcal{F}(\hat{x}).$$

Applying Farkas Lemma yields

$$\nabla f(\hat{x})^T p = \sum_{i \in \mathcal{A}(\hat{x})} \hat{\lambda}_i \nabla c_i(\hat{x})$$

for some $\hat{\lambda}_i$ with $\hat{\lambda}_i \geq 0 \quad \forall i \in \mathcal{I} \cap \mathcal{A}(\hat{x})$

$\Rightarrow$ KKT conditions hold

- second part of the proof: exercise. ∎

ℓ_1 Penalty Function

ℓ_1 Penalty Function

ℓ_1 Penalty Function

Algorithm: Classical ℓ_1 Penalty Method

Input: $\mu_0 > 0$, tolerance $\tau > 0$, starting point x_0^s
for $k = 0, 1, 2, \ldots$
Find an approximate minimizer x_k of $\Phi_1(\cdot; \mu_k)$, starting at x_k^s **if** $h(x_k) \le \tau$
stop with approximate solution x_k;
end if
Choose new penalty parameter $\mu_{k+1} > \mu_k$;
Choose new starting point x_{k+1}^s;
end for

A General Class of Nonsmooth Penalty Methods

Exact nonsmooth penalty functions can be defined in terms of norms other than the ℓ_1 norm. We can write

$$\Phi(x;\mu) := f(x) + \mu||c_{\mathcal{E}}(x)|| + \mu||[c_{\mathcal{I}}(x)]^-||,$$

where $||\cdot||$ is any vector norm, and all the equality and inequality constraints have been grouped in the vector functions $c_{\mathcal{E}}$ and $c_{\mathcal{I}}$, respectively.

Most common norms:

- ℓ_1;
- ℓ_∞;
- ℓ_2 (not squared).

A General Class of Nonsmooth Penalty Methods

We show now that penalty functions of the type considered so far *must* be nonsmooth to be exact: Consider the case of a single equality constraint $c_1(x) = 0$. Let the penalty function be given by

$$\boxed{\Phi(x;\mu) = f(x) + \mu h(c_1(x)),}$$

where $h : \mathbb{R} \to \mathbb{R}$ is a function satisying the properties

1. $h(y) \geq 0$ for all $y \in \mathbb{R}$ and
2. $h(0) = 0$.

Suppose that h is continuously differentiable. Since h has a minimizer at zero, we have $\nabla h(0) = 0$. If x^* is a local minimizer of (P), we have from x^*'s feasibility that $c_1(x^*) = 0$, thus, $\nabla h(c_1(x^*)) = 0$. If x^* is a local minimizer of ($\mathsf{P}_{\mathsf{penalty}}$), we therefore have

$$0 = \nabla\Phi(x^*;\mu) = \nabla f(x^*) + \mu\nabla c_1(x^*)\overbrace{\nabla h(c_1(x^*))}^{=0} = \nabla f(x^*).$$

However, $\nabla f(x^*) = 0$ does not generally hold for a *constrained* problem. Thus, $\Phi(\cdot,\mu)$ cannot be smooth.