Convex Optimization
CMU-10725

Penalty Methods

Barnabás Póczos & Ryan Tibshirani

MACHINE LEARNING DEPARTMENT
Carnegie Mellon.
School of Computer Science

Outline

- Penalty functions

Books to Read

David G. Luenberger, Yinyu Ye: Linear and Nonlinear Programming

Boyd and Vandenberghe: Convex Optimization

Penalty Methods

Penalty Methods

$$\min_{x \in \mathcal{S}} f(x), \qquad\qquad (P)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, and $\mathcal{S}$ is a constraint set in $\mathbb{R}^n$

Penalty program: replace (P) with the unconstrained problem:

$$\min f(x) + cp(x)$$

where $c > 0$

Penalty $p : \mathbb{R}^n \to \mathbb{R}$:

(i) p is continuous,

(ii) $p(x) \geq 0$ for all $x \in \mathbb{R}^n$,

(iii) $p(x) = 0$ if and only if $x \in \mathcal{S}$

Penalty term: high cost for violation of the constraints

Inequality Constraints

$$\min_{x \in \mathcal{S}} f(x) \qquad \mathcal{S} = \{x : g_i(x) \leq 0, i = 1, 2, \ldots, p\}$$

A useful penalty function in this case is:

$$P(x) = \frac{1}{2} \sum_{i=1}^{p} (\max[0, g_i(x)])^2$$

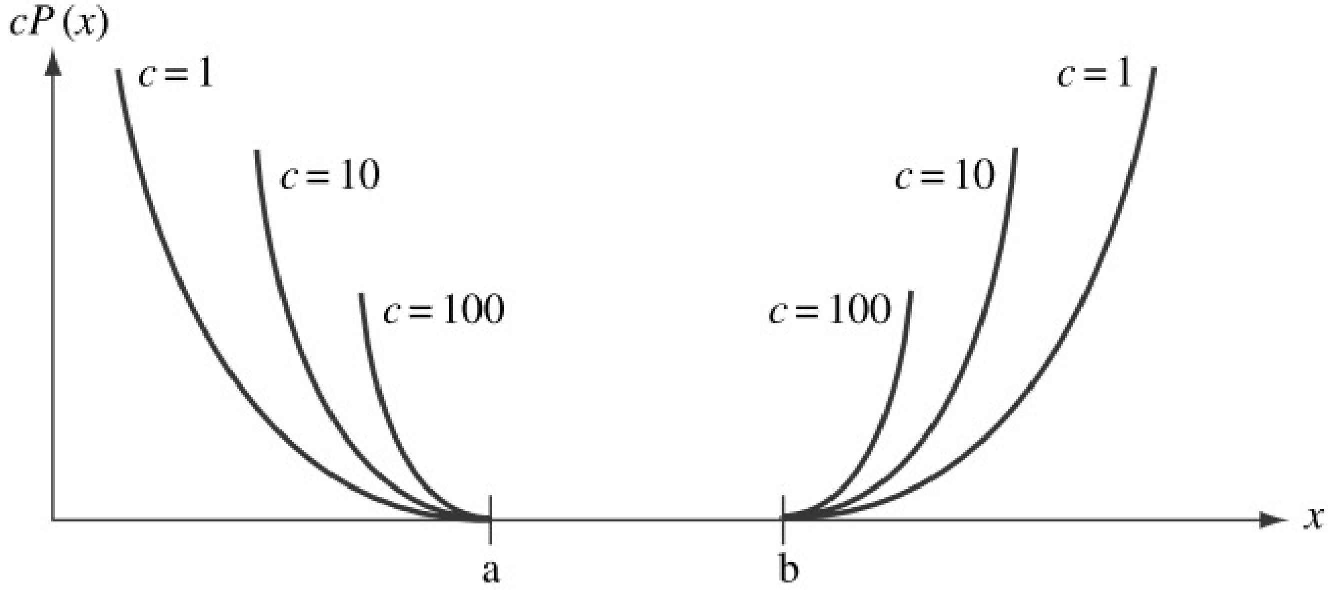
[No penaltiy, iff $g(x) \leq 0$]

Example: $g_1(x) = x - b$, $g_2(x) = x - a$

$$a \leq x, \; x \leq b$$

Penalty Methods

Penalty program: $\min f(x) + cp(x) \quad (P(c))$



For large c the minimum point of problem $(P(c))$ is in a region where penalty p is small.

We will prove: as $c \to \infty$ the solution point of the penalty problem will converge to a solution of the constrained problem (P).

Inequality and Equality Constraints

Inequality and Equality constraints:

$$\min_{x} f(x)$$

SUBJECT TO $\quad g_i(x) \le 0 \quad i = 1 \ldots m$

$$h_j(x) = 0 \quad j = 1 \ldots k$$

Rewrite them as:

$$h_j(x) \le 0$$

$$-h_j(x) \le 0$$

Penalty Method

Penalty parameter:

LET $0 < c_1 < c_2 < \ldots < c_k < c_{k+1} < \ldots \to \infty$

Penalty program:

$$q(c, x) \doteq f(x) + c\,P(x)$$

$$x_k = \underset{x \in \mathbb{R}^n}{\text{ARG MIN}}\; q(c_k, x) = \underset{x \in \mathbb{R}^n}{\text{ARG MIN}}\; f(x) + c_k P(x)$$

Penalty Lemma:

1. $q(c_k, x_k) \le q(c_{k+1}, x_{k+1})$
2. $P(x_k) \ge P(x_{k+1})$
3. $f(x_k) \le f(x_{k+1})$
4. $f(x^*) \ge q(c_k, x_k) \ge f(x_k)$

Proof of Penalty Lemma (1)

(1) PROVE $q(c_k, x_k) \le q(c_{k+1}, x_{k+1})$

PROOF:

$$q(c_{k+1}, x_{k+1}) \doteq f(x_{k+1}) + c_{k+1} P(x_{k+1}) \geq f(x_{k+1}) + c_k P(x_{k+1})$$

$c_{k+1} \geq c_k$

$$\geq f(x_k) + c_k P(x_k) \doteq q(c_k, x_k)$$

x_k IS OPTIMAL WITH c_k

$\square$

Proof of Penalty Lemma (2)

② PROVE $P(x_k) \geq P(x_{k+1})$

PROOF:

$$f(x_k) + c_k P(x_k) \leq f(x_{k+1}) + c_k P(x_{k+1}) \qquad (*_1)$$

x_k IS OPTIMAL WITH c_k

$$f(x_{k+1}) + c_{k+1} P(x_{k+1}) \leq f(x_k) + c_{k+1} P(x_k) \qquad (*_2)$$

x_{k+1} IS OPTIMAL WITH c_{k+1}

$(*_1) + (*_2) \Rightarrow$

$$c_k P(x_k) + c_{k+1} P(x_{k+1}) \leq c_k P(x_{k+1}) + c_{k+1} P(x_k)$$

$$\Rightarrow \underbrace{(c_{k+1} - c_k)}_{>0} P(x_{k+1}) \leq \underbrace{(c_{k+1} - c_k)}_{>0} P(x_k)$$

$$\Rightarrow P(x_{k+1}) \leq P(x_k) \quad \square$$

Proof of Penalty Lemma (3)

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

PROOF :

$$f(x_{k+1}) + c_k P(x_{k+1}) \geq f(x_k) + c_k P(x_k)$$

$\uparrow$ x_k IS OPTIMAL WITH c_k

$$\geq f(x_k) + c_k P(x_{k+1})$$

$\uparrow$ $P(x_k) \geq P(x_{k+1})$

$$\Rightarrow f(x_{k+1}) \geq f(x_k) \qquad \square$$

Proof of Penalty Lemma (4)

LET $x^* = \underset{x \in S}{\text{ARGMIN}}\ f(x)$

PROVE THAT

ALREADY PROVED IN LEMMA (1)

$$f(x^*) \geq q(c_{k+1}, x_{k+1}) \geq q(c_k, x_k) \geq f(x_k) \quad \forall k$$

WHERE $q(c, x) = f(x) + c\,P(x)$

$$x_{k+1} = \underset{x}{\text{ARGMIN}}\ f(x) + c_k P(x)$$

PROOF

$$f(x^*) = f(x^*) + c_k \underbrace{P(x^*)}_{\overset{\shortparallel}{0}} \geq \overbrace{f(x_k) + c_k \underbrace{P(x_k)}_{\geq 0}}^{q(c_k, x_k)} \geq f(x_k)$$

SINCE $x^* \in S$

x_k IS OPTIMAL WITH c_k

$\square$

Convergence of Penalty Method

Theorem: [Penalty convergence]

- SUPPOSE f, g, p ARE CONTINOUS FUNCTIONS
- $x_k = \arg\min_{x \in \mathbb{R}^n} f(x) + c_k\, p(x)$ ← PENALTY FUNCTION
- $0 < c_1 < c_2 < \dots < c_k < c_{k+1} < \dots \to \infty$
- $\bar{x}$ IS AN ARBITRARY LIMIT POINT OF $\{x_k\}_{k=1}^{\infty}$

$\Rightarrow$ $\bar{x}$ SOLVES (P)

$$(P)\ \begin{cases} \min\limits_{x} & f(x) \\ \text{s.t.} & g(x) \leq 0 \end{cases}$$

Proof of Penalty Convergence

LIMIT POINT: $\lim_{k \in \mathcal{K}} x_k = \bar{x}$ $\quad [x_k \to \bar{x}]$

SINCE f IS CONTINUOUS $\lim_{k \in \mathcal{K}} f(x_k) = f(\bar{x})$

$q^* \doteq \lim_{k \in \mathcal{K}} \underbrace{q(c_k, x_k)}_{f(x_k) + c_k P(x_k)} \underset{\text{LEMMA (4)}}{\le} f(x^*) \doteq f^*$ $\quad x^*$: SOLUTION OF (P)

$\Rightarrow q^* = \underbrace{\lim_{k \in \mathcal{K}} f(x_k)}_{f(\bar{x})} + \lim_{k \in \mathcal{K}} c_k P(x_k) \le f(x^*) \Rightarrow f(\bar{x}) \le f(x^*)$

$\Rightarrow \underbrace{q^* - f(\bar{x})}_{<\infty} = \lim_{k \in \mathcal{K}} \underset{\infty}{\underset{\downarrow}{c_k}} \; \underset{0}{\underset{\vee|}{P(x_k)}} \Rightarrow \lim_{k \in \mathcal{K}} P(x_k) = 0$

$\xrightarrow{P \text{ IS CONTINUOUS}} \; \Rightarrow P(\lim_{k \in \mathcal{K}} x_k) = P(\bar{x}) = 0$

$\Rightarrow \bar{x} \in S$ [FEASIBLE] $\quad \left[\begin{array}{l} f(\bar{x}) \le f(x^*) \\ x \in S \end{array} \right. \Rightarrow f(\bar{x}) = f(x^*)$ $\quad \square$

Penalty functions

Often used penalty functions

Polynomial penalty: $P(x) = \sum_{i=1}^{m} [\mathrm{MAX}\{0, g_i(x)\}]^q \quad q \geq 1$

Linear penalty: $q = 1$: $P(x) = \sum_{i=1}^{m} \mathrm{MAX}\{0, g_i(x)\}$

Quadratic penalty: $q = 2$: $P(x) = \sum_{i=1}^{m} [\mathrm{MAX}\{0, g_i(x)\}]^2$

$g_i^+(x) = \mathrm{MAX}\{0, g_i(x)\}$

$g^+(x) = [g_1^+(x), \ldots, g_m^+(x)]^T$

$P(x) = g^+(x)^T g^+(x)$ OR $P(x) = g^+(x)^T \Gamma g^+(x)$

WHERE $\Gamma > 0$

Inequality and Equality constraints

Problem (P)

$$P: \min_{x} f(x)$$

$$\text{S.T.} \quad g(x) \leq 0$$
$$h(x) = 0$$
$$x \in \mathbb{R}^n$$

Definition [Penalty function]

$$P(x) = 0 \quad \text{IF } g(x) \leq 0 \text{ AND } h(x) = 0$$
$$P(x) > 0 \quad \text{IF } g(x) > 0 \text{ OR } h(x) \neq 0$$

Example [Penalty function]

$$P(x) = \sum_{i=1}^{m} \left[\text{MAX}\{0, g_i(x)\}\right]^q + \sum_{i=1}^{k} |h_i(x)|^q \quad q \geq 1$$

Derivative of the penalty function

Penalty program: $x_k = \underset{x \in \mathbb{R}^n}{\mathrm{ARGMIN}} \; f(x) + c_k P(x)$

Penalty function: $g_i^+(x) = \max\{0, g_i(x)\}$

$g^+(x) = [g_1^+(x), \dots g_m^+(x)]^T$

$P(x) = \gamma(g^+(x))$ WHERE $\gamma : \mathbb{R}^m \to \mathbb{R}_{\oplus}$

FOR EXAMPLE $\gamma(y) = y^T y$

Assumptions: $f \in C^1$ OBJECTIVE

$g \in C^1$ CONSTRAINTS

$P(\cdot) = \gamma(g^+(\cdot)) \in C^1$ PENALTY FUNCTION

Derivatives: $\dfrac{\partial P(x)}{\partial x} = ?$ g^+ IS NOT DIFFERENTIABLE!

Derivative of the penalty function

Difficulties: max is not differentiable

??!

$$\frac{\partial P(x)}{\partial x} = \sum_{i=1}^{m} \frac{\partial \gamma(g^+(x))}{\partial g_i} \frac{\partial g_i^+(x)}{\partial x}$$

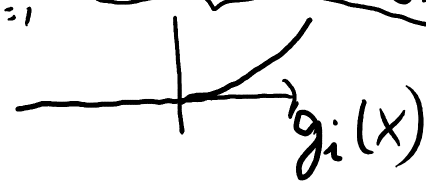
$$\nabla g_i^+(x) = \frac{\partial g_i^+(x)}{\partial x} = \frac{\partial \max[0, g_i(x)]}{\partial x} = \begin{cases} \frac{\partial g_i(x)}{\partial x} & \text{IF } g_i(x) \geq 0 \\ 0 & \text{IF } g_i(x) < 0 \end{cases}$$

This is not perfectly correct, because ∇g_i^+ IS NOT CONTINUOUS WHERE $g_i(x) = 0$!

Solution: IF $\underset{\overset{\|}{g_i^+(x)}}{g_i} = 0 \Rightarrow \frac{\partial \gamma(0)}{\partial g_i} = 0$ THEN $P(x)$ IS STILL DIFFERENTIABLE

Example: $P(x) = \sum_{i=1}^{m} [g_i^+(x)]^q \quad q > 1$

$$\frac{\partial P(x)}{\partial x} = \sum_{i=1}^{m} \underbrace{q[g_i^+(x)]^{q-1} \frac{\partial g_i^+(x)}{\partial x}}$$



KKT in Penalty methods

Penalty program: $x_k = \underset{x \in \mathbb{R}^n}{\text{ARGMIN}} \; f(x) + c_k P(x)$

Penalty function: $P(x) = \gamma(g^+(x))$

Derivatives: $\nabla P(x) = \sum_{i=1}^{m} \frac{\partial \gamma(g^+(x))}{\partial y_i} \nabla g_i^+(x) = \sum_{i=1}^{m} \frac{\partial \gamma(g^+(x))}{\partial y_i} \nabla g_i(x)$

NO NEED FOR $\nabla g_i^+(x)$

1st order condition in local minimum:

$$0 = \nabla f(x_k) + c_k \nabla P(x_k)$$

$$= \nabla f(x_k) + c_k \sum_{i=1}^{m} \underbrace{\frac{\partial \gamma(g^+(x_k))}{\partial y_i}}_{u_i^k / c_k} \nabla g_i(x_k)$$

$$= \nabla f(x_k) + u_i^k \nabla g_i(x_k) = 0$$

BEHAVES LIKE LAGRANGE MULTIPLIERS IN KKT!

KKT and Penalty method multipliers

Problem (P)

$$P: \min_{x} f(x) \quad \text{s.t.} \quad g(x) \leq 0$$

Penalty program:

$$x_k = \underset{x \in \mathbb{R}^n}{\mathrm{ARGMIN}}\ f(x) + c_k P(x)$$

$$u_i^k = c_k \frac{\partial \gamma(g^+(x_k))}{\partial y_i}$$

$$\nabla f(x_k) + (u^k)^T \nabla g(x_k) = 0$$

KKT multipliers: u^* IN A LOCAL OPT SOLUTION x^* OF (P)

Theorem: Under some mild conditions IF $x_k \to x^* \Rightarrow u^k \to u^*$