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

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

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:

$$\min f(x) + \underline{cp(x)}$$

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 penalty, iff $g(x) \leq 0$]

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

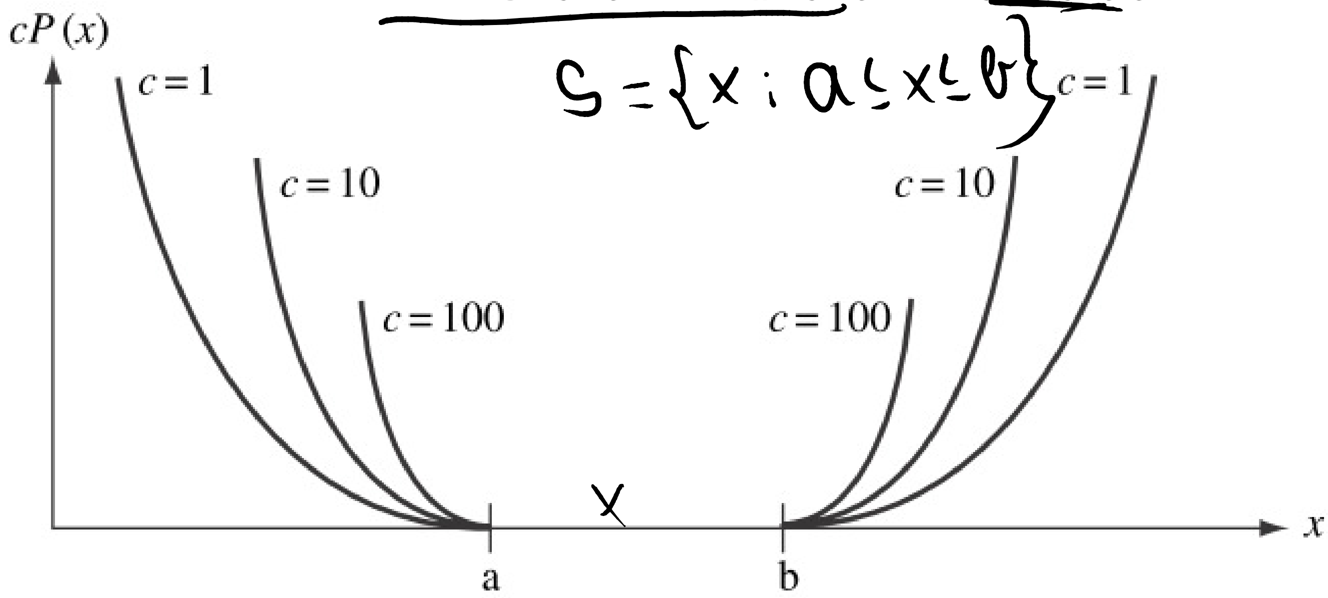
$a \leq x \leq b$

$g_1(x) = x - b$

$g_2(x) = a - x$

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:

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

$$\text{s.t.} \quad g_i(x) \le 0 \qquad i = 1 \ldots m$$

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

Rewrite them as:

$$P(x) = P_0(x) + \sum |h_j(x)|^{2q}$$

$$h_j(x) \le 0$$

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

Penalty Method

Penalty parameter:

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

Penalty program:

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

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

Penalty Lemma:

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

Proof of Penalty Lemma (1)

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

<u>PROOF</u>

$$\begin{aligned} q(c_{k+1}, x_{k+1}) &\doteq f(x_{k+1}) + c_{k+1} P(x_{k+1}) \\ &\geq f(x_{k+1}) + c_k P(x_{k+1}) \\ &\geq f(x_k) + c_k P(x_k) \\ &\doteq q(c_k, x_k) \end{aligned}$$

$\square$

Proof of Penalty Lemma (2)

$$[\; P(x_a) \geq P(x_{a+1})$$

- $f(x_a) + c_a P(x_a) \leq f(x_{a+1}) + c_a P(x_{a+1})$
- $f(x_{a+1}) + c_{a+1} P(x_{a+1}) \leq f(x_a) + c_{a+1} P(x_a)$

$+$

$$c_a P(x_a) + c_{a+1} P(x_{a+1}) \leq c_a P(x_{a+1}) + c_{a+1} P(x_a)$$

$$\underbrace{(c_{a+1} - c_a)}_{>0} P(x_{a+1}) \leq \underbrace{(c_{a+1} - c_a)}_{>0} P(x_a)$$

$$P(x_{a+1}) \leq P(x_a) \quad \text{Q.E.D.}$$

Proof of Penalty Lemma (3)

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

PROOF

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

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

$\square$

Proof of Penalty Lemma (4)

$$x^* = \operatorname*{ARG\,MIN}_{x \in S} f(x) \qquad \text{PROOF:}$$

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

$$f(x^*) = f(x^*) + c_k \underbrace{P(x^*)}_{\substack{0 \\ x^* \in S}} \geq \underbrace{f(x_k) + c_k P(x_k)}_{q(c_k, x_k)}$$

$$\geq f(x_k) \qquad \square$$

Convergence of Penalty Method

Theorem: [Penalty convergence]

- SUPPOSE f, g, P CONTINUOUS
- $x_k = \underset{x \in \mathbb{R}^n}{\text{ARG MIN}} \; f(x) + c_k P(x)$
- $0 < c_1 < c_2 < \ldots < c_k \ldots \to \infty$
- $\bar{x}$ IS AN ARBITRARY LIMIT POINT OF $\{x_k\}_{k=1}^{\infty}$

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

(P) $\quad \text{MIN } f(x)$

$\quad\quad \text{S.T. } g(x) \leq 0$

Proof of Penalty Convergence

$$\boxed{\bar{x} = \lim_{k\in\mathcal{K}} x_k}$$

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

LEMMA 4

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

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

$$\underline{q^* - f(\bar{x})} = \underline{\lim_{k\in\mathcal{K}} \underset{\infty}{\underset{\downarrow}{c_k}}\ \underset{0}{\underset{\downarrow}{P(x_k)}}} \implies P(\bar{x}) = 0, \quad \bar{x} \in S$$

$\square$

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

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

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

$P(x) = g^+(x)^T g^+(x)$ 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]

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

Derivative of the penalty function

Penalty program: $x_k = \text{ARGMIN } f(x) + c_k P(x)$

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

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

$P(x) = \gamma(g^+(x)) \quad \gamma(y) = y^T y$

Assumptions: $f \in C^1$

$g \in C^1$

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

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

Derivative of the penalty function

Difficulties: max is not differentiable

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

$P(x) = \gamma(g^+(x))$

$\gamma(y) \in \mathbb{R}$

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

This is not perfectly correct, because ∇g_i^+ IS NOT CONTINUOUS

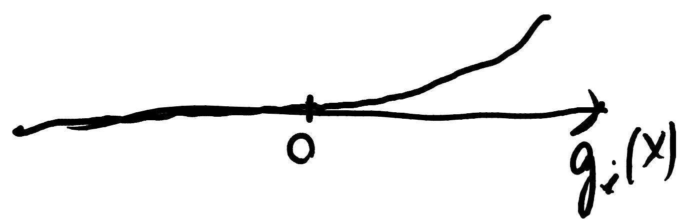
Solution: $\gamma(y)$

$\gamma \quad y_i = 0, \; g_i^+(x) \Rightarrow \frac{\partial \gamma(y)}{\partial y_i} = 0 \Rightarrow P(x)$ IS DIFFERENTIABLE

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

Example:

$$P(x) = \sum_{i=1}^{m} \left[g_i^+(x)\right]^q \qquad q > 1$$



KKT in Penalty methods

Penalty program: $x_k = \underset{x}{\mathrm{ARGMIN}}\ \underline{f(x) + c_k P(x)}$

Penalty function: $P(x) = \gamma\left(g^+(x)\right)$ $\quad g_i^+(x)$

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

1st order condition in local minimum: 0

$$0 = \nabla f(x_k) + \underbrace{\nabla P(x_k)}_{\sum_{i=1}^{m} c_k \frac{\partial \gamma\left(g^+(x_k)\right)}{\partial y_i} \cdot \nabla g_i(x_k)}$$

$$\underbrace{c_k \frac{\partial \gamma\left(g^+(x_k)\right)}{\partial y_i}}_{u_i^k}$$

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

KKT and Penalty method multipliers

Problem (P)

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

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

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

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

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

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