6.252 NONLINEAR PROGRAMMING

LECTURE 16: PENALTY METHODS

LECTURE OUTLINE

- Quadratic Penalty Methods
- Introduction to Multiplier Methods

****

- Consider the equality constrained problem

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X, \qquad h(x) = 0,$$

where $f : \Re^n \to \Re$ and $h : \Re^n \to \Re^m$ are continuous, and X is closed.

- The quadratic penalty method:

$$x^k = \arg \min_{x \in X} L_{c^k}(x, \lambda^k) \equiv f(x) + \lambda^{k\prime} h(x) + \frac{c^k}{2} \|h(x)\|^2$$

where the $\{\lambda^k\}$ is a bounded sequence and $\{c^k\}$ satisfies $0 < c^k < c^{k+1}$ for all k and $c^k \to \infty$.

TWO CONVERGENCE MECHANISMS

- Taking λ^k close to a Lagrange multiplier vector
  - Assume $X = \Re^n$ and (x^*, λ^*) is a local min-Lagrange multiplier pair satisfying the 2nd order sufficiency conditions
  - For c suff. large, x^* is a strict local min of $L_c(\cdot, \lambda^*)$
- Taking c^k very large
  - For large c and any λ

$$L_c(\cdot, \lambda) \approx \begin{cases} f(x) & \text{if } x \in X \text{ and } h(x) = 0 \\ \infty & \text{otherwise} \end{cases}$$

- Example:

$$\text{minimize } f(x) = \tfrac{1}{2}(x_1^2 + x_2^2)$$

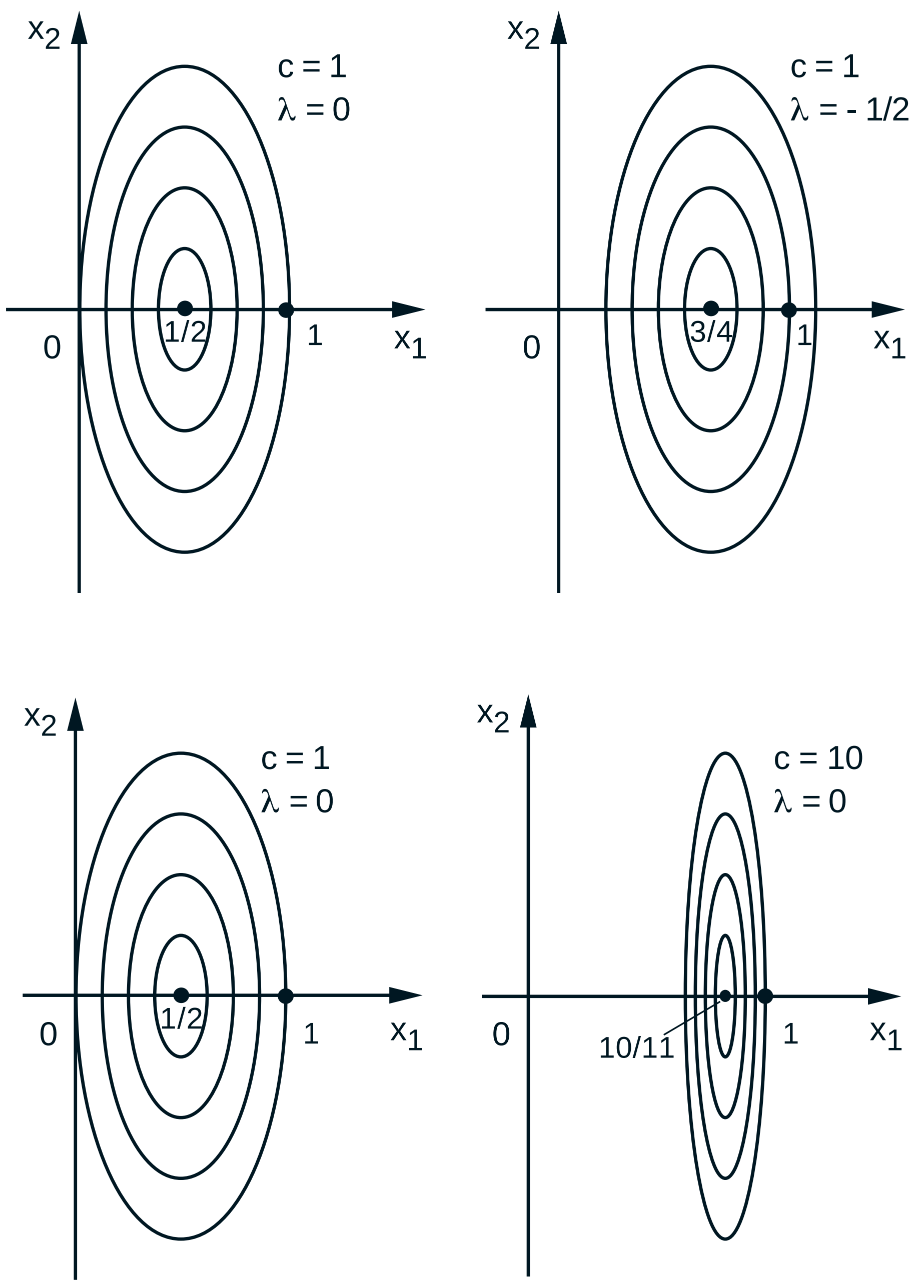
$$\text{subject to } x_1 = 1$$

$$L_c(x, \lambda) = \tfrac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + \frac{c}{2}(x_1 - 1)^2$$

$$x_1(\lambda, c) = \frac{c - \lambda}{c + 1}, \qquad x_2(\lambda, c) = 0$$

EXAMPLE CONTINUED

$$\min_{x_1=1} x_1^2 + x_2^2, \qquad x^* = 1, \quad \lambda^* = -1$$



GLOBAL CONVERGENCE

- Every limit point of $\{x^k\}$ is a global min.

Proof: The optimal value of the problem is $f^* = \inf_{h(x)=0,\, x\in X} L_{c^k}(x,\lambda^k)$. We have

$$L_{c^k}(x^k,\lambda^k) \le L_{c^k}(x,\lambda^k), \qquad \forall\, x\in X$$

so taking the inf of the RHS over $x\in X$, $h(x)=0$

$$L_{c^k}(x^k,\lambda^k) = f(x^k) + \lambda^{k\prime} h(x^k) + \frac{c^k}{2}\|h(x^k)\|^2 \le f^*.$$

Let $(\bar{x},\bar{\lambda})$ be a limit point of $\{x^k,\lambda^k\}$. Without loss of generality, assume that $\{x^k,\lambda^k\}\to(\bar{x},\bar{\lambda})$. Taking the limsup above

$$f(\bar{x}) + \bar{\lambda}' h(\bar{x}) + \limsup_{k\to\infty} \frac{c^k}{2}\|h(x^k)\|^2 \le f^*. \qquad (*)$$

Since $\|h(x^k)\|^2 \ge 0$ and $c^k\to\infty$, it follows that $h(x^k)\to 0$ and $h(\bar{x})=0$. Hence, $\bar{x}$ is feasible, and since from Eq. (*) we have $f(\bar{x})\le f^*$, $\bar{x}$ is optimal. **Q.E.D.**

LAGRANGE MULTIPLIER ESTIMATES

- Assume that $X = \Re^n$, and f and h are cont. differentiable. Let $\{\lambda^k\}$ be bounded, and $c^k \to \infty$. Assume x^k satisfies $\nabla_x L_{c^k}(x^k, \lambda^k) = 0$ for all k, and that $x^k \to x^*$, where x^* is such that $\nabla h(x^*)$ has rank m. Then $h(x^*) = 0$ and $\tilde{\lambda}^k \to \lambda^*$, where

$$\tilde{\lambda}^k = \lambda^k + c^k h(x^k), \qquad \nabla_x L(x^*, \lambda^*) = 0.$$

Proof: We have

$$\begin{aligned} 0 = \nabla_x L_{c^k}(x^k, \lambda^k) &= \nabla f(x^k) + \nabla h(x^k)\big(\lambda^k + c^k h(x^k)\big) \\ &= \nabla f(x^k) + \nabla h(x^k)\tilde{\lambda}^k. \end{aligned}$$

Multiply with

$$\big(\nabla h(x^k)'\nabla h(x^k)\big)^{-1}\nabla h(x^k)'$$

and take lim to obtain $\tilde{\lambda}^k \to \lambda^*$ with

$$\lambda^* = -\big(\nabla h(x^*)'\nabla h(x^*)\big)^{-1}\nabla h(x^*)'\nabla f(x^*).$$

We also have $\nabla_x L(x^*, \lambda^*) = 0$ and $h(x^*) = 0$ (since $\tilde{\lambda}^k$ converges).

PRACTICAL BEHAVIOR

- Three possibilities:
  - The method breaks down because an x^k with $\nabla_x L_{c^k}(x^k, \lambda^k) \approx 0$ cannot be found.
  - A sequence $\{x^k\}$ with $\nabla_x L_{c^k}(x^k, \lambda^k) \approx 0$ is obtained, but it either has no limit points, or for each of its limit points x^* the matrix $\nabla h(x^*)$ has rank $< m$.
  - A sequence $\{x^k\}$ with with $\nabla_x L_{c^k}(x^k, \lambda^k) \approx 0$ is found and it has a limit point x^* such that $\nabla h(x^*)$ has rank m. Then, x^* together with λ^* [the corresp. limit point of $\left\{\lambda^k + c^k h(x^k)\right\}$] satisfies the first-order necessary conditions.

- Ill-conditioning: The condition number of the Hessian $\nabla^2_{xx} L_{c^k}(x^k, \lambda^k)$ tends to increase with c^k.

- To overcome ill-conditioning:
  - Use Newton-like method (and double precision).
  - Use good starting points.
  - Increase c^k at a moderate rate (if c^k is increased at a fast rate, $\{x^k\}$ converges faster, but the likelihood of ill-conditioning is greater).

INEQUALITY CONSTRAINTS

- Convert them to equality constraints by using squared slack variables that are eliminated later.
- Convert inequality constraint $g_j(x) \le 0$ to equality constraint $g_j(x) + z_j^2 = 0$.
- The penalty method solves problems of the form

$$\min_{x,z} \bar{L}_c(x, z, \lambda, \mu) = f(x)$$

$$+ \sum_{j=1}^{r} \left\{ \mu_j \left(g_j(x) + z_j^2 \right) + \frac{c}{2} |g_j(x) + z_j^2|^2 \right\},$$

for various values of μ and c.

- First minimize $\bar{L}_c(x, z, \lambda, \mu)$ with respect to z,

$$L_c(x, \lambda, \mu) = \min_z \bar{L}_c(x, z, \lambda, \mu) = f(x)$$

$$+ \sum_{j=1}^{r} \min_{z_j} \left\{ \mu_j \left(g_j(x) + z_j^2 \right) + \frac{c}{2} |g_j(x) + z_j^2|^2 \right\}$$

and then minimize $L_c(x, \lambda, \mu)$ with respect to x.

MULTIPLIER METHODS

- Recall that if (x^*, λ^*) is a local min-Lagrange multiplier pair satisfying the 2nd order sufficiency conditions, then for c suff. large, x^* is a strict local min of $L_c(\cdot, \lambda^*)$.
- This suggests that for $\lambda^k \approx \lambda^*$, $x^k \approx x^*$.
- Hence it is a good idea to use $\lambda^k \approx \lambda^*$, such as

$$\lambda^{k+1} = \tilde{\lambda}^k = \lambda^k + c^k h(x^k)$$

This is the (1st order) method of multipliers.

- Key advantages to be shown:
  - Less ill-conditioning: It is not necessary that $c^k \to \infty$ (only that c^k exceeds some threshold).
  - Faster convergence when λ^k is updated than when λ^k is kept constant (whether $c^k \to \infty$ or not).