6.252 NONLINEAR PROGRAMMING

LECTURE 17: AUGMENTED LAGRANGIAN METHODS

LECTURE OUTLINE

- Multiplier Methods

****

- Consider the equality constrained problem

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

where $f : \Re^n \to \Re$ and $h : \Re^n \to \Re^m$ are continuously differentiable.

- The (1st order) multiplier method finds

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

and updates λ^k using

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

CONVEX EXAMPLE

- Problem: $\min_{x_1=1}(1/2)(x_1^2 + x_2^2)$ with optimal solution $x^* = (1, 0)$ and Lagr. multiplier $\lambda^* = -1$.
- We have

$$x^k = \arg\min_{x \in \Re^n} L_{c^k}(x, \lambda^k) = \left(\frac{c^k - \lambda^k}{c^k + 1}, 0 \right)$$

$$\lambda^{k+1} = \lambda^k + c^k \left(\frac{c^k - \lambda^k}{c^k + 1} - 1 \right)$$

$$\lambda^{k+1} - \lambda^* = \frac{\lambda^k - \lambda^*}{c^k + 1}$$

- We see that:
  - $\lambda^k \to \lambda^* = -1$ and $x^k \to x^* = (1, 0)$ for every nondecreasing sequence $\{c^k\}$. It is NOT necessary to increase c^k to ∞.
  - The convergence rate becomes faster as c^k becomes larger; in fact $\left\{|\lambda^k - \lambda^*|\right\}$ converges superlinearly if $c^k \to \infty$.

NONCONVEX EXAMPLE

- Problem: $\min_{x_1=1}(1/2)(-x_1^2+x_2^2)$ with optimal solution $x^* = (1,0)$ and Lagr. multiplier $\lambda^* = 1$.
- We have

$$x^k = \arg\min_{x\in\Re^n} L_{c^k}(x,\lambda^k) = \left(\frac{c^k-\lambda^k}{c^k-1}, 0\right)$$

provided $c^k > 1$ (otherwise the min does not exist)

$$\lambda^{k+1} = \lambda^k + c^k\left(\frac{c^k-\lambda^k}{c^k-1} - 1\right)$$

$$\lambda^{k+1} - \lambda^* = -\frac{\lambda^k-\lambda^*}{c^k-1}$$

- We see that:
  - No need to increase c^k to ∞ for convergence; doing so results in faster convergence rate.
  - To obtain convergence, c^k must eventually exceed the threshold 2.

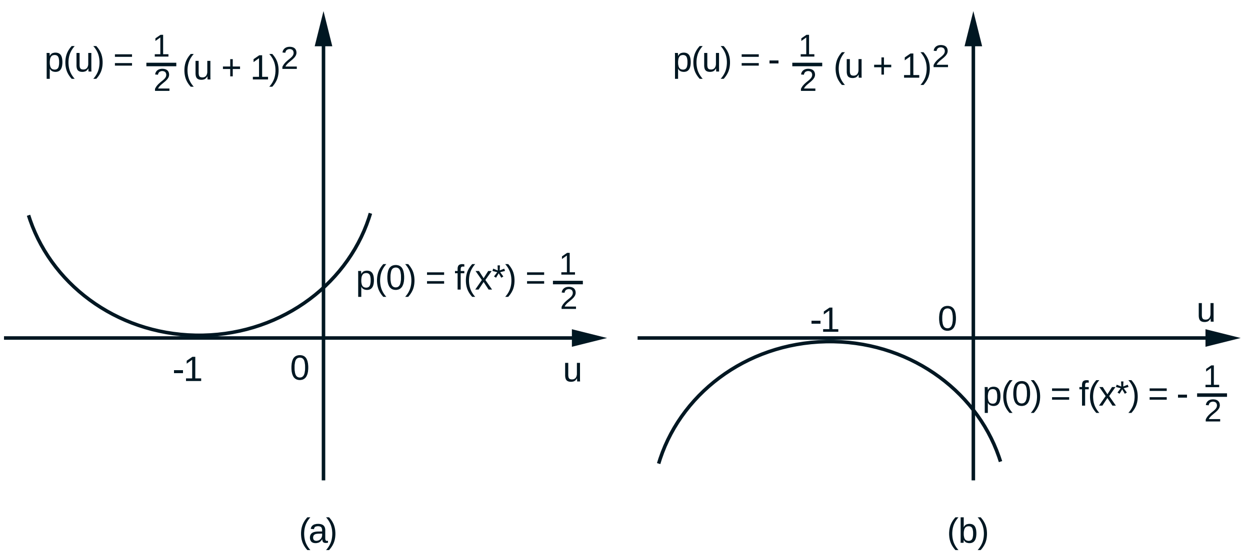
THE PRIMAL FUNCTIONAL

- Let (x^*, λ^*) be a regular local min-Lagr. pair satisfying the 2nd order suff. conditions are satisfied.
- The primal functional

$$p(u) = \min_{h(x)=u} f(x),$$

defined for u in an open sphere centered at $u = 0$, and we have

$$p(0) = f(x^*), \qquad \nabla p(0) = -\lambda^*,$$



$$p(u) = \min_{x_1 - 1 = u} \tfrac{1}{2}(x_1^2 + x_2^2), \qquad p(u) = \min_{x_1 - 1 = u} \tfrac{1}{2}(-x_1^2 + x_2^2)$$

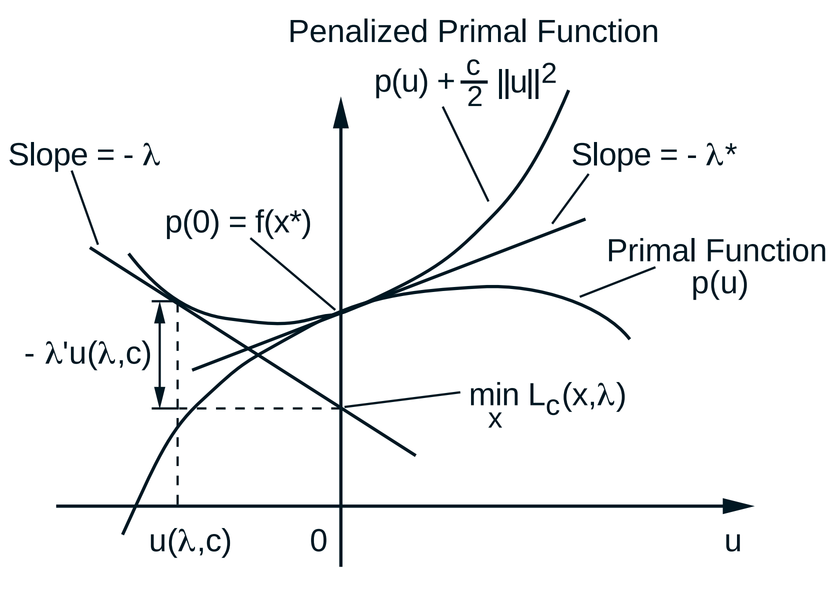
AUGM. LAGRANGIAN MINIMIZATION

- Break down the minimization of $L_c(\cdot, \lambda)$:

$$\min_x L_c(x, \lambda) = \min_u \min_{h(x)=u} \left\{ f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2 \right\}$$
$$= \min_u \left\{ p(u) + \lambda' u + \frac{c}{2} \|u\|^2 \right\},$$

where the minimization above is understood to be local in a neighborhood of $u = 0$.

- Interpretation of this minimization:

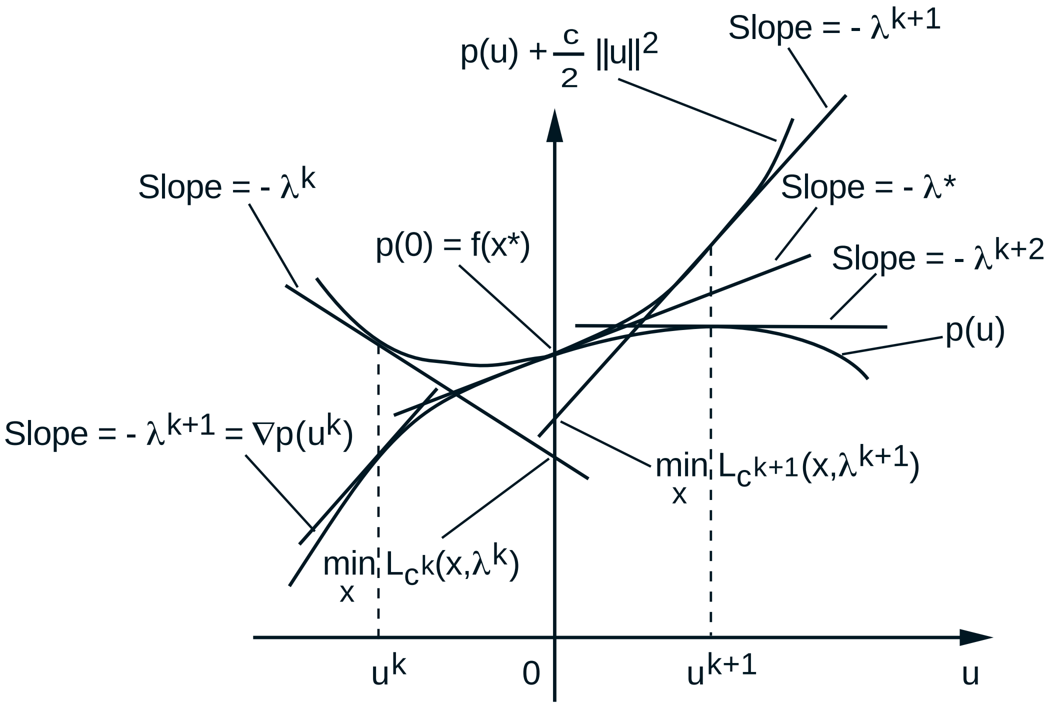


- If c is suf. large, $p(u) + \lambda' u + \frac{c}{2} \|u\|^2$ is convex in a neighborhood of 0. Also, for $\lambda \approx \lambda^*$ and large c, the value $\min_x L_c(x, \lambda) \approx p(0) = f(x^*)$.

INTERPRETATION OF THE METHOD

- Geometric interpretation of the iteration

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



- If λ^k is sufficiently close to λ^* and/or c^k is suf. large, λ^{k+1} will be closer to λ^* than λ^k.

- c^k need not be increased to ∞ in order to obtain convergence; it is sufficient that c^k eventually exceeds some threshold level.

- If $p(u)$ is linear, convergence to λ^* will be achieved in one iteration.

COMPUTATIONAL ASPECTS

- Key issue is how to select $\{c^k\}$.
  - c^k should eventually become larger than the "threshold" of the given problem.
  - c^0 should not be so large as to cause ill-conditioning at the 1st minimization.
  - c^k should not be increased so fast that too much ill-conditioning is forced upon the unconstrained minimization too early.
  - c^k should not be increased so slowly that the multiplier iteration has poor convergence rate.

- A good practical scheme is to choose a moderate value c^0, and use $c^{k+1} = \beta c^k$, where β is a scalar with $\beta > 1$ (typically $\beta \in [5, 10]$ if a Newton-like method is used).

- In practice the minimization of $L_{c^k}(x, \lambda^k)$ is typically inexact (usually exact asymptotically). In some variants of the method, only one Newton step per minimization is used (with safeguards).

DUALITY FRAMEWORK

- Consider the problem

$$\begin{aligned}&\text{minimize} \quad f(x) + \frac{c}{2}\|h(x)\|^2 \\ &\text{subject to} \quad \|x - x^*\| < \epsilon, \quad h(x) = 0,\end{aligned}$$

where ϵ is small enough for a local analysis to hold based on the implicit function theorem, and c is large enough for the minimum to exist.

- Consider the dual function and its gradient

$$q_c(\lambda) = \min_{\|x-x^*\|<\epsilon} L_c(x, \lambda) = L_c\big(x(\lambda, c), \lambda\big)$$

$$\begin{aligned}\nabla q_c(\lambda) &= \nabla_\lambda x(\lambda, c) \nabla_x L_c\big(x(\lambda, c), \lambda\big) + h\big(x(\lambda, c)\big) \\ &= h\big(x(\lambda, c)\big).\end{aligned}$$

We have $\nabla q_c(\lambda^*) = h(x^*) = 0$ and $\nabla^2 q_c(\lambda^*) > 0$.

- The multiplier method is a steepest ascent iteration for maximizing q_{c^k}

$$\lambda^{k+1} = \lambda^k + c^k \nabla q_{c^k}(\lambda^k),$$