Advanced Mathematical Programming IE417

Lecture 19

Dr. Ted Ralphs

Reading for this lecture

- Sections 9.1-9.2

Constrained Optimization

- In Chapter 9, we look at methods based on applying unconstrained methods to constrained problems.
- Idea: Penalize violations of the constraints in the objective function.
- Consider

$$\min f(x)$$
$$s.t. \; h(x) = 0$$

- Try $\min\{f(x) + [h(x)]^2\}$. Will this work?

Penalty Functions

- A *suitable penalty function* α is

$$\alpha(x) = \Sigma\phi(g_i(x)) + \Sigma\psi(h_i(x))$$

  where

  - $\phi(y) = 0 \; if \; y \leq 0$
  - $\phi(y) > 0 \; if \; y > 0$
  - $\psi(y) = 0 \; if \; y = 0$
  - $\psi(y) > 0 \; if \; y \neq 0$

Performance of Penalty Methods

- Suppose we simply solve $\min\{f(x) + \mu\alpha(x) : x \in X\}$ for some suitable penalty function and some $\mu > 0$.

- Consider solving the following problem:

$$\max \Theta(\mu)$$
$$s.t. \quad \mu \geq 0$$

  where $\Theta(\mu) = inf\{f(x) + \mu\alpha(x) : x \in X\}$.

- What will the solution be?

Main result on Penalty Methods

- As long as the following assumptions are satisfied:
  - The problem is feasible,
  - All the functions involved are continuous,
  - $\{x_\mu\}$ is contained in a compact set,

  then

$$\inf\{f(x) : x \in X, g(x) \leq 0, h(x) = 0\} = \sup\{\Theta(\mu) : \mu \geq 0\} \\ = \lim\{\Theta(\mu)\}$$

- Also, if x_μ is the min point for a particular penalty μ, then every limit point of $\{x_\mu\}$ is optimal, and
  - $f(x_\mu)$ is a nondecreasing function of μ,
  - $\Theta(\mu)$ is a nondecreasing function of μ,
  - $\alpha(x_\mu)$ is a nonincreasing function of μ,
  - $\mu\alpha(x_\mu) \to 0$ as $\mu \to \infty$.
- Note that this implies that if $\alpha(x_\mu) = 0$ for some μ, then x_μ is optimal.

Penalty Functions and KKT

- Under certain conditions, we can use the solution of the penalty problem to derive Lagrange multipliers.
- Suppose $\{x_\mu\}$ converges to a *regular solution* x^*. For each x_μ, we have

$$\nabla f(x_\mu) + \Sigma \mu \phi'(g_i(x_\mu)) \nabla g_i(x_\mu) + \Sigma \mu \psi'(h_i(x_\mu)) \nabla h_i(x_\mu) = 0$$

- Furthermore, $g_i(x_\mu) < 0 \; \forall \; i \notin I$ if μ is sufficiently large.
- If we define $(u_\mu)_i \equiv \mu \phi'(g_i(x_\mu))$ and $(v_\mu)_i \equiv \mu \psi'(h_i(x_\mu))$, these values converge to the optimal multipliers.

Implementing Penalty Methods

- Initialization: Choose a termination scalar $\varepsilon > 0$, an initial point x_1, an initial penalty parameter $\mu_1 > 0$, and a scalar $\beta > 0$. Set $k = 1$.
- Loop
  - Minimize $f(x) + \mu_k \alpha(x)$ subject to $x \in X$ to obtain x_{k+1}.
  - If $\mu_k \alpha(x_{k+1}) < \varepsilon$, then STOP. Otherwise, let $\mu_{k+1} = \beta \mu_k$, replace k by $k + 1$ and iterate.
- Notice the relationship to solving the Lagrangian dual.

Computational Issues

- We can get arbitrarily close to an optimal solution with a large enough μ.
- However, large multipliers can cause serious numerical difficulties.
- Intuitively, this is because feasiblity tends to dominate optimality.
- More formally, large values of μ result in a severely ill-conditioned Hessian.
- Quasi-Newton methods can help overcome this problem.