Nonlinear Optimization Lecture 15

Garrick Aden-Buie

Tuesday, March 8, 2016

Review HW 4

Problem 1

$$
\begin{aligned}
\min \quad & x_1^2 + x_2^2 \\
\text{s.t} \quad & -x_1 - x_2 + 4 \leq 0 && \leftarrow g_1 \\
& -x_1 \leq 0 && \leftarrow g_2 \\
& -x_2 \leq 0 && \leftarrow g_3
\end{aligned}
$$

Write the Lagrangian dual function where $X = \{x\colon x_1, x_2 \geq 0\}$

$$
\begin{aligned}
&\theta(u) = \inf_{x \in X} \{x_1^2 + x_2^2 + u(-x_1 - x_2 = 4)\} \\
&\frac{\partial}{\partial x_1} = 2x_1 - u = 0 \\
&\frac{\partial}{\partial x_2} = 2x_2 - u = 0
\end{aligned}
$$

$x_1 = x_2 = \frac{u}{2}$ valid only if $u \geq 0$. In the first case where $u \geq 0$, then $x_1 = x_2 = \frac{u}{2}$ and $\theta(u) = -\frac{u^2}{2} + 4u$. In the second case when $u < 0$, we know that $x_1 = x_2 = 0$, so $\theta(u) = 4u$. In summary,

$$
\theta(u) = \begin{cases} -\frac{u^2}{2} + 4u & u \geq 0 \\ 4u & u < 0 \end{cases}
$$

Both functions are differentiable everywhere, but the point of contention is $u = 0$. But $\theta'(u) = 4$ for both, so it is differentiable everywhere.

There is no duality gap because $\theta(u^*) = f(\bar{x})$.

Problem 2

$\min x$ s.t. $g(x) \leq 0$.

$$\text{(a)} \quad g(x) = \begin{cases} -\frac{1}{x} & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$$

$$\text{(b)} \quad g(x) = \begin{cases} -\frac{1}{x} & \text{for } x \neq 0 \\ -1 & \text{for } x = 0 \end{cases}$$

Weird function, but the goal is to find the subgradients of θ at $u = 0$.

$$\theta(u) = \min_{x \geq 0}\{x + ug(x)\}$$
$$= \min\left\{\min_{x>0}\left\{x + u(-\frac{1}{x})\right\}, 0\right\}$$
$$\text{If } u \geq 0 \to \theta(u) \to -\infty$$
$$\text{If } u < 0 \to \theta(u) = 0$$
$$\text{If } u = 0 \to \theta(u) = 0$$
$$\theta(u) = \begin{cases} -\infty & u > 0 \\ 0 & u \leq 0 \end{cases}$$

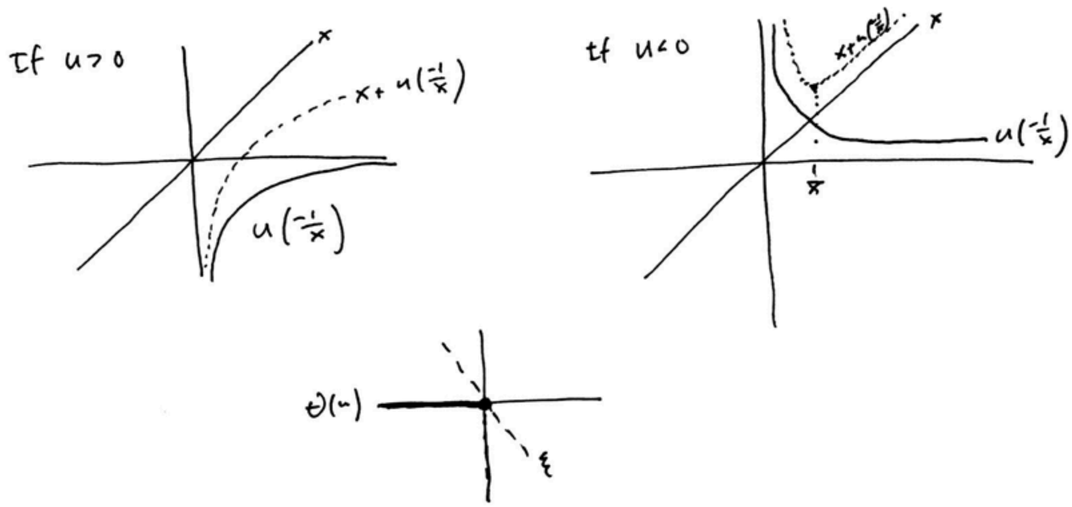

Figure 1: (a) when $u > 0$ and $u < 0$

At the point $u = 0$, any line with negative slope is a subgradient of $\theta(u)$.

Line Search with Derivative

Bisection Method

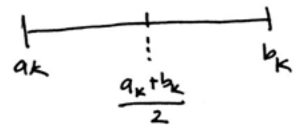

Figure 2: Bisection method example

f is assumed to be pseudoconvex.

Three cases

(1) $f'(\frac{a_k+b_k}{2}) = 0 \Rightarrow \frac{a_k+b_k}{2} = x^*$ optimal.

(2) $f'(\frac{a_k+b_k}{2}) > 0$. Then for all $x > \frac{a_k+b_k}{2}$, $f'(\frac{a_k+b_k}{2})(x - \frac{a_k+b_k}{2}) > 0$ implies, by pseudoconvexity, that $f(x) \geq f(\frac{a_k+b_k}{2})$.

Then let

$$\begin{aligned} a_{k+1} &= a_k \\ b_{k+1} &= \frac{a_k + b_k}{2} \end{aligned}$$

(3) $f'(\frac{a_k+b_k}{2}) > 0 \Rightarrow a_{k+1} = \frac{a_k+b_k}{2}, b_{k+1} = b_k$.

Numerical Differentiation, Finite Difference

Note that $f(x)$ can be approximated by

$$\begin{aligned} f' &\approx \frac{f(x+\Delta) - f(x)}{\Delta} && \text{forward} \\ &\approx \frac{f(x) - f(x-\Delta)}{\Delta} && \text{backward} \\ &\approx \frac{f(x+\Delta) - f(x-\Delta)}{2\Delta} && \text{central} \end{aligned}$$

Newton's Method (Successive Quadratic Approximation)

$\min f(x)$, use the quadratic approximation at x_k.

$$f(x) \approx q_k(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2$$

From this approximation, we minimize $q_k(x)$ and let the minimum be x_{k+1}.

$$q_k'(x) = f'(x_k) + f''(x_k)(x - x_k) = 0 \Rightarrow x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Note that f'' cannot be $= 0$. Stop once $x_{k+1} = x_k + \epsilon$. Note also that this technique is equivalent to finding a root of $f'(x) = 0$.

Multi-Dimensional Search

There are two variations: without derivative and with derivative.

Cyclic Search

Deal with each dimension one-by-one. Consider you have $f(x_1, x_2)$ and you are at (x_1^k, x_2^k), then $\min_{x_1} f(x_1, x_2^k)$. Then switch, and search over x_2 setting x_1^{k+1} fixed as the solution from the previous step.

This method tends to be slow, but generally works as a fall-back.

Steepest Descent

At x^k, we want direction $d^k = -\nabla f(x^k)$

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k)$$

λ_k: step size, $\lambda_k > 0$.

$$\min_{\lambda \geq 0} f\left(x^k - \lambda \nabla f(x^k)\right)$$

From here we can use the line search algorithm to solve the above equation, because x^k and $\nabla f(x^k)$ are fixed, only λ is variable. So use line search to solve the above problem, at every step to find a new λ.

This algorithm is also quite slow; it may be fast at the beginning but then it's slow in the final steps.

Some other methods for unconstrained problems

- BFGS
  - A class of methods similar to steepest descent with additional terms added
- Conjugate Gradient (CG)

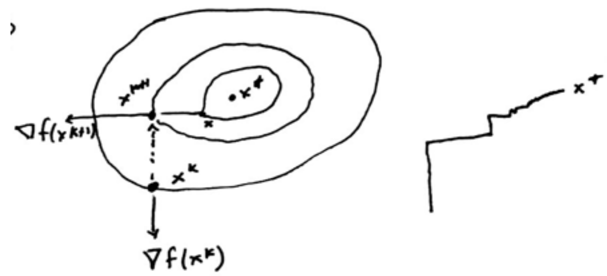

Figure 3: Illustration of Steepest Descent algorithm

Constrained Optimization

Penalty Function Method

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & x \in X \end{aligned}$$

Define a penalty function $p(x)$ such that

(1) $p(x)$ continuous $\forall x$
(2) $p(x) \geq 0 \ \forall x$
(3) $p(x) = 0 \Leftrightarrow x \in X$

Examples

$$\begin{aligned} g_i(x) \leq 0 &\Rightarrow p(x) = [\max\{0, g_i(x)\}]^P \\ h_i(x) = 0 &\Rightarrow p(x) = |h_i(x)|^p \end{aligned}$$

where p is a positive number, usually $p = 2$.

$p = 1$ $\quad p(x) = \max\{0, g_i(x)\}$ non differentiable
$p = 2$ $\quad p(x)$ is differentiable

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_j(x) = 0 \quad j = 1, \ldots, l \end{aligned} \quad \Rightarrow \quad \begin{aligned} \min \quad & f(x) + \mu p(x) \\ \text{where} \quad & p(x) = \sum_{i=1}^{m} [\max\{0, g_i(x)\}]^p + \sum_{j=1}^{l} |h_j(x)|^p \end{aligned}$$

Where μ small infeasible and μ large feasible.

SUMT: Successive Unconstrained Minimization Technique

- Begin with small $\mu_k > 0$
- Solve min $f(x) + \mu_k p(x)$ with initial solution being x_k.
- Obtain x_{k+1}
- Increase μ_k by some factor and set $k = k + 1$
- Repeat.

Example

$$\begin{aligned} \min \quad & x^2 \\ \text{s.t} \quad & x \geq 1 \end{aligned}$$

$$\begin{aligned} \min \quad & x^2 + \mu \left[\max(0, 1-x)\right]^2 \\ & = \begin{cases} x^2 & x \geq 1 \\ x^2 + \mu(1-x)^2 & x < 1 \end{cases} \end{aligned}$$

Let $r(x) = x^2 + \mu(1-x)^2$, then $r'(x) = 2x - 2\mu(1-x) = 0 \Rightarrow x_\mu = \frac{\mu}{1-\mu} < 1$. And $x_\mu \to 1$ as $\mu \to \infty$.

Note that in this method, you never have a feasible solution. The way this method works, you start outside the feasible set and work your way to a feasible solution.