Nonlinear Optimization Lecture 13

Garrick Aden-Buie

Thursday, February 25, 2016

Theorem

$$
\begin{aligned}
\min \quad & f(x) \\
& g(x) \leq 0 \\
\text{s.t} \quad & h(x) = 0 \\
& x \in X
\end{aligned}
\qquad\qquad
\begin{aligned}
w &= \begin{bmatrix} u \\ v \end{bmatrix} \\
\beta(x) &= \begin{bmatrix} g(x) \\ h(x) \end{bmatrix} \\
\theta(w) &= \inf_{x \in X} \left\{ f(x) + w^T \beta(x) \right\}
\end{aligned}
$$

Def. *Lagrangian Dual Function* is the $\theta(w) = \inf_{x \in X} \left\{ f(x) + w^T \beta(x) \right\}$.

Rewriting the theorem from last lecture:

- **If** X is a non-empty compact set.
- $X(w) \left\{ \bar{x} \colon f(\bar{x}) + w^T \beta(\bar{x}) = \inf\{ f(x) + w^T \beta(x) \} \right\}$
- Suppose $X(\bar{w})$ is the singleton $\{\bar{x}\}$
- **Then** $\theta(w)$ is differentiable at $\bar{w}$ and $\nabla \theta(\bar{w}) = \beta(\bar{x})$.

Theorem

- X: non-empty, compact
- f, β are continuous
- $X(\bar{w})$ is *not empty* for any $\bar{w}$
- If $\bar{x} \in X(\bar{w})$, then $\beta(\bar{w})$ is a *subgradient* of θ at $\bar{w}$.

Proof. $\theta(w)$ is a concave function $\Rightarrow \exists$ a subgradient for all w.

$$
\begin{aligned}
\theta(w) &= \inf_{x \in X} \{ f(x) + w^T \beta(x) \} \\
&\geq f(\bar{x}) + w^T \beta(x) \\
&= f(\bar{x}) + (w - \bar{w})^T \beta(\bar{x}) + \bar{w}^T \beta(\bar{x}) \\
&= [f(\bar{x}) + \bar{w}^T \beta(\bar{x})] + (w - \bar{w})^T \beta(\bar{x}) \\
&= \theta(\bar{w}) + \beta(\bar{x})^T (w - \bar{w}) \\
\Rightarrow & \beta(\bar{x}) \text{ is a subgradient at } \bar{w}
\end{aligned}
$$

Example

$$
\begin{aligned}
\min \quad & -x_1 - x_2 \\
\text{s.t} \quad & x_1 + 2x_2 - 3 \leq 0 \\
& x_1, x_2 \in \{0, 1, 2, 3\}
\end{aligned}
$$

$$
\begin{aligned}
\theta(u) &= \inf_{x \in X} \{-x_1 - x_2 + u(x_1 + 2x_2 - 3)\} \\
&= \inf_{x \in X} \{(u-1)x_1 + (2u-1)x_2 - 3u\} \\
&= \begin{cases} -6 + 6u & \text{if } u \leq \frac{1}{2} \\ -3 & \text{if } \frac{1}{2} \leq u \leq 1 \\ -3u & \text{if} u \geq 1 \end{cases}
\end{aligned}
$$

Let $\bar{u} = \frac{1}{2}$. Then $X(\bar{u}) = \arg\min_{x \in X}\{f(x) + \bar{u}g(x)\}$.

$$
\begin{aligned}
\min \quad & -x_1 - x_2 + \frac{1}{2}(x_1 + x_2 - 3) = -\frac{1}{2}x_1 - \frac{3}{2} \\
\text{s.t} \quad & x_1, x_2 \in \{0, 1, 2, 3\}
\end{aligned}
$$

Then $X(\frac{1}{2}) = \{(3,0), (3,1), (3,2), (3,3)\}$.

The subgradients of $\theta(u)$ at $u = \frac{1}{2}$. From the theorem, $\beta(\bar{u})\ \forall \bar{x} \in X(\bar{w})$...

$$
\begin{aligned}
g(3,0) &= 3 - 3 &&= 0 \\
g(3,1) &= 3 + 2 - 3 &&= 2 \\
g(3,2) &= 3 + 4 - 3 &&= 4 \\
g(3,3) &= 3 + 6 - 3 &&= 6
\end{aligned}
$$

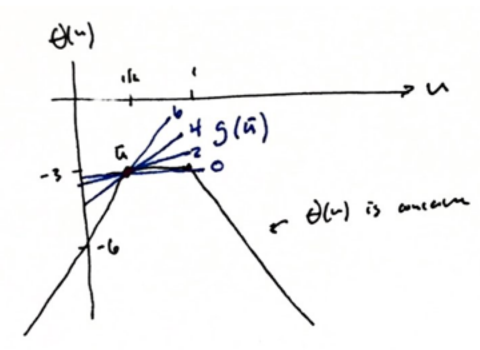

Figure 1:

Note that there are infinite subgradients at $\bar{u}$, for any line with slope between 0 and 6. The theorem states that *some* of the subgradients are given in the form above, but not all.

Theorem

- X: non-empty compact
- f, β: continuous
- ξ is a subgradient of θ at $\bar{w}$ if and only if $\xi \in$ convex hull of $\{\beta(\bar{x})\colon \bar{x} \in X(\bar{w})\}$.

Line Search without Derivative

$\min f(x)$, $f\colon \mathbb{R} \to \mathbb{R}$. Let f be strictly quasiconvex (monotonically decreasing, and then monotonically increasing).

Strictly quasiconvex function: $f(\lambda\bar{x} + (1-\lambda)\hat{x}) < \max\{f(\bar{x}), f(\hat{x})\}$, $\forall\lambda \in (0,1)$, $f(\bar{x}) \neq f(\hat{x})$.

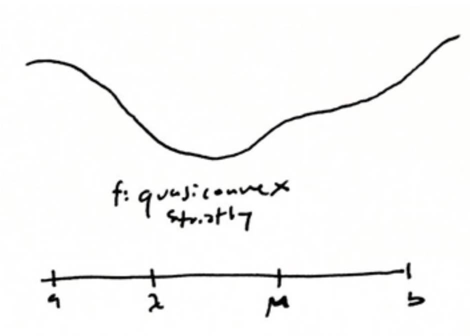

Figure 2: Quasiconvex function illustration and first line search algorithm layout.

Theorem

(1) If $f(\lambda) \leq f(\mu)$, then $f(x) \geq f(\lambda)\ \forall x \in (\mu, b]$.

(2) If $f(\lambda) \geq f(\mu)$, then $f(x) \geq f(\mu)\ \forall x \in [a, \lambda)$.

Point (1) states that if $f(b)$ is highest and $f(\mu)$ and $f(\lambda)$ are lower (in that order), then the inflection point is definitely not between μ and b. Point (2) says the same thing but on the side of $f(a)$, discarding the points between a and λ.

Proof. Suppose not: assume $\exists\bar{x} \in (\mu, b]$ such that $f(\bar{x}) < f(\lambda)$. Then $f(\lambda) < f(\mu) < f(\bar{x}) \leq f(b)$. Consider the definition of strong quasiconvex functions:

$$\begin{aligned} f(\mu) &< \max\{f(\lambda), f(\bar{x})\} \\ &= f(\lambda) \end{aligned}$$

But this is a contradiction.

Dichotomous Search

Intuition: We would like to maximize the search area that is being abandoned in each step. In the above line search if λ, μ, a, b are all highly separated, then each iteration discards a small portion of the search space.

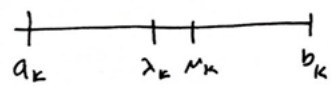

Figure 3:

Let $\lambda_k = \frac{a_k+b_k}{2} - \epsilon$ and $\mu_k = \frac{a_k+b_k}{2} + \epsilon$

Step 0 Choose an interval $[a_1, b_1]$ that contains an optimal solution. Choose $\epsilon > 0, \delta > 0$. Set $k = 1$.

Step 1 Compute λ_k, μ_k.

Step 2 If $f(\lambda_k) < f(\mu_k)$ let

$$a_{k+1} = a_k$$
$$b_{k+1} = \mu_k$$

Otherwise, let

$$a_{k+1} = \lambda_k$$
$$b_{k+1} = b_k$$

Step 3 If $b_{k+1} - a_{k+1} < \delta$, then stop:

$$x^* \approx \frac{a_{k+1} + b_{k+1}}{2}$$

Otherwise, set $k = k + 1$ and go to **Step 1**.

Note that we ned $\epsilon < \frac{\delta}{2}$ for this to work. But that this algorithm requires a significant number of function evalutions, which will add computation time. This leads us to the next algorithm.

Golden Section Search

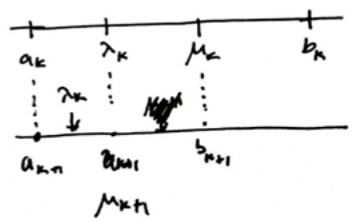

Figure 4:

In the previous algorithm, new function evaluations are needed for every point evaluated at every iteration. In the golden section search, we want to re-use previous function evaluations, and only evaluate one new point at each search.

Position so that $\mu_{k+1} = \lambda_k$ or $\lambda_{k+1} = \mu_k$.

$$\begin{aligned}
\lambda_k &= \alpha a_k + (1-\alpha) b_k \\
\mu_k &= (1-\alpha) a_k + \alpha b_k
\end{aligned}$$

Find α so that $\mu_{k+1} = \lambda_k$

$$\begin{aligned}
\mu_{k+1} &= (1-\alpha) a_{k+1} + \alpha b_{k+1} \\
&= (1-\alpha) a_k + \alpha \mu_k \\
&= (1-\alpha) a_k + \alpha((1-\alpha) a_k + \alpha b_k) \\
&= (1-\alpha^2) a_k + \alpha^2 b_k \\
\lambda_k &= \alpha a_k + (1-\alpha) b_k
\end{aligned}$$

So we want $1 - \alpha^2 = \alpha \Rightarrow \alpha = \frac{-1+\sqrt{5}}{2} \approx 0.618$

If we do n function evaluations the length of the interval is reduced by $(0.618)^{n-1}$. Dichotomous search is only $\approx (0.5 - \epsilon)^{\frac{n}{2}}$.