Advanced Mathematical Programming
IE417

Lecture 10

Dr. Ted Ralphs

Reading for This Lecture

- Primary Reading
  - Chapter 6, Sections 2-3

Saddle Point Optimality

Lagrangian Saddle Points

- Recall the Lagrangian function

$$\Phi(x, \mu, v) = f(x) + \mu^T \mathbf{g}(x) + v^T \mathbf{h}(x)$$

- A point (x^*, μ^*, v^*) with $x^* \in X, \mu^* \geq 0$ is a *saddle point* for $\Phi(x, \mu, v)$ if

$$\Phi(x^*, \mu, v) \leq \Phi(x^*, \mu^*, v^*) \leq \Phi(x, \mu^*, v^*), \forall x \in X, (\mu, v), \mu \geq 0.$$

Saddle Point Optimality

- A point (x^*, μ^*, v^*) with $x^* \in X, \mu^* \geq 0$ is a saddle point for $\Phi(x, \mu, v)$ if and only if
  - $\Phi(x^*, \mu^*, v^*) = \min\{\Phi(x, \mu^*, v^*) : x \in X\}$
  - $\mathbf{g}(x^*) \leq 0, \mathbf{h}(x^*) = 0$, and
  - ${\mu^*}^T \mathbf{g}(x^*) = 0$.
- Furthermore, (x^*, μ^*, v^*) is a saddle point if and only if x^* and (μ^*, v^*) are the optimal solutions to P and D with no duality gap, i.e., $f(x^*) = \Theta(\mu^*, v^*)$.

Saddle Points and Convexity

Corollary 1. *Suppose*

1. X, f, *and* $\mathbf{g}$ *are convex, and* $\mathbf{h}$ *is affine,*
2. $\boldsymbol{0} \in int\ \mathbf{h}(X)$, *and*
3. $\exists x' \in X$ *such that* $\mathbf{g}(x') < \boldsymbol{0}$ *and* $\mathbf{h}(x') = 0$.

If x^* *is an optimal solution to the primal problem* P, *then there exists a vector* (μ^*, v^*) *with* $\mu^* \geq 0$ *such that* (x^*, μ^*, v^*) *is a saddle point.*

Saddle Points and KKT

Corollary 2. *Suppose*

1. x^* *is a KKT point with multipliers* (μ^*, v^*),
2. f, g_i *for* $i \in I$ *are convex at* x^**, and*
3. h_i *is affine if* $v_i^* \neq 0$.

Then (x^*, μ^*, v^*) *is a Lagrangian saddle point.*

Conversely, if (x^*, μ^*, v^*) *is a Lagrangian saddle point with* $x^* \in int(X)$*, then* x^* *is feasible for* P *and* (x^*, μ^*, v^*) *satisfies the KKT conditions.*

Saddle Points and the Perturbation Function

- Recall the perturbation function

$$\nu(y, z) = \min\{f(x) : \mathbf{g}(x) \leq y, \mathbf{h}(x) = z, x \in X\}$$

- The following are equivalent:
  - the absence of a duality gap,
  - the existence of a saddle point solution, and
  - The existence of a supporting hyperplane for the epigraph of v at the point $(\mathbf{0}, \nu(\mathbf{0}))$.

Properties of the Dual Function

Properties of the Dual Function

Theorem 1. *If X is a nonempty compact set in $\mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\beta : \mathbb{R}^n \rightarrow \mathbb{R}^{m+l}$ are continuous, then*

$$\Theta(w) = inf\{f(x) + w^T\beta(x) : x \in X\}$$

is concave.

- This means we should be able to maximize Θ.

Differentiability of Θ

Consider the following set of optimal solutions

$$X(w) = \{y : y \text{ minimizes } \Theta(w)\}$$

Theorem 2. *Suppose X is a nonempty compact set in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$ are continuous. Let $w^* \in \mathbb{R}^{m+l}$ be given such that $X(w^*) = \{x^*\}$. Then Θ is differentiable at w^* with $\bigtriangledown\Theta(w^*) = \beta(x^*)$.*

Subgradients of Θ

Theorem 3. *Suppose X is a nonempty compact set in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$ are continuous. If $x^* \in X(w^*)$, then $\beta(x^*)$ is a subgradient of Θ at w^*.*

Theorem 4. *Under the same conditions as above, ξ is a subgradient of Θ at w^* if and only if ξ belongs to the convex hull if $\{\beta(y) : y \in X(w^*)\}$.*

Ascent Directions for Θ

- A vector d is called an *ascent direction* of Θ at w if there exists $\delta > 0$ such that

$$\Theta(w + \lambda d) > \Theta(w), \forall \lambda \in (0, \delta)$$

- A vector d^* is called a *steepest ascent direction* of Θ at w if

$$\Theta^{'}(w; d^*) = \max\{\Theta^{'}(w; d) : \|d\| \leq 1\}$$

Direction of Steepest Ascent for Θ

Theorem 5. *Suppose* X *is a nonempty compact set in* $\mathbb{R}^n$ *and* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$ *are continuous. The direction of steepest ascent* d^* *of* Θ *at* w *is given by*

$$d^* = \begin{cases} 0 & \textit{if} \quad \xi^* = 0 \\ \xi^*/\|\xi^*\| & \textit{if} \quad \xi^* \neq 0 \end{cases}$$

where is the subgradient of Θ *at* w *with the smallest norm.*