Reading for This Lecture

• Primary Reading
  – Chapter 6, Sections 2-3
Saddle Point Optimality
Lagrangian Saddle Points

• Recall the Lagrangian function

\[ \Phi(x, \mu, v) = f(x) + \mu^T g(x) + v^T h(x) \]

• A point \((x^*, \mu^*, v^*)\) with \(x^* \in X, \mu^* \geq 0\) is a \textit{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 \((x^*, \mu^*, v^*)\) with \(x^* \in X, \mu^* \geq 0\) is a saddle point for \(\Phi(x, \mu, v)\) if and only if
  - \(\Phi(x^*, \mu^*, v^*) = \min\{\Phi(x, \mu^*, v^*) : x \in X\}\)
  - \(g(x^*) \leq 0, h(x^*) = 0\), and
  - \(\mu^*^T g(x^*) = 0\).

- Furthermore, \((x^*, \mu^*, v^*)\) is a saddle point if and only if \(x^*\) and \((\mu^*, v^*)\) are the optimal solutions to \(P\) and \(D\) with no duality gap, i.e., \(f(x^*) = \Theta(\mu^*, v^*)\).
Corollary 1. Suppose

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

If $x^*$ is an optimal solution to the primal problem $P$, then there exists a vector $(\mu^*, v^*)$ with $\mu^* \geq 0$ such that $(x^*, \mu^*, v^*)$ is a saddle point.
Corollary 2. Suppose

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

Then $(x^*, \mu^*, v^*)$ is a Lagrangian saddle point.

Conversely, if $(x^*, \mu^*, v^*)$ is a Lagrangian saddle point with $x^* \in \text{int}(X)$, then $x^*$ is feasible for $P$ and $(x^*, \mu^*, v^*)$ satisfies the KKT conditions.
Saddle Points and the Perturbation Function

• Recall the perturbation function

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

• 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 \( \nu \) at the point \((0, \nu(0))\).
Properties of the Dual Function
**Properties of the Dual Function**

**Theorem 1.** If $X$ is a nonempty compact set in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$ 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 $\Theta$. 
Differentiability of $\Theta$

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 $\Theta$ is differentiable at $w^*$ with $\nabla \Theta(w^*) = \beta(x^*)$. 
Subgradients of $\Theta$

**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 $\Theta$ at $w^*$.

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

- A vector $d$ is called an *ascent direction* of $\Theta$ 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 $\Theta$ at $w$ if
  \[ \Theta'(w; d^*) = \max\{\Theta'(w; d) : \|d\| \leq 1\} \]


**Direction of Steepest Ascent for \( \Theta \)**

**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 \( \Theta \) at \( w \) is given by

\[
d^* = \begin{cases} 
0 & \text{if } \xi^* = 0 \\
\xi^*/\|\xi^*\| & \text{if } \xi^* \neq 0
\end{cases}
\]

where \( \xi^* \) is the subgradient of \( \Theta \) at \( w \) with the smallest norm.