Lecture 18

Subgradients

November 3, 2008
Outline

• Existence of Subgradients

• Subdifferential Properties

• Optimality Conditions
**Convex-Constrained Non-differentiable Minimization**

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in C
\end{align*}
\]

- **Characteristics:**
  - The function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex and possibly non-differentiable
  - The set \( C \subseteq \mathbb{R}^n \) is nonempty closed and convex
  - The optimal value \( f^* \) is finite

- Our focus here is *non-differentiability*
Definition of Subgradient and Subdifferential

Def. A vector $s \in \mathbb{R}^n$ is a **subgradient of $f$ at $\hat{x} \in \text{dom } f$** when

$$f(x) \geq f(\hat{x}) + s^T(x - \hat{x}) \quad \text{for all } x \in \text{dom } f$$

Def. A **subdifferential of $f$ at $\hat{x} \in \text{dom } f$** is the set of all subgradients $s$ of $f$ at $\hat{x} \in \text{dom } f$

- The **subdifferential of $f$ at $\hat{x}$** is denoted by $\partial f(\hat{x})$

- When $f$ is differentiable at $\hat{x}$, we have $\partial f(\hat{x}) = \{\nabla f(\hat{x})\}$ (the subdifferential is a singleton)

- Examples

\[
 f(x) = |x|, \quad \partial f(0) = \begin{cases} 
 \text{sign}(x) & \text{for } x \neq 0 \\
 [-1, 1] & \text{for } x = 0 
\end{cases}
\]

\[
 f(x) = \begin{cases} 
 x^2 + 2|x| - 3 & \text{for } |x| > 1 \\
 0 & \text{for } |x| \leq 1 
\end{cases}
\]
Subgradients and Epigraph

• Let $s$ be a subgradient of $f$ at $\hat{x}$:
  \[ f(x) \geq f(\hat{x}) + s^T(x - \hat{x}) \quad \text{for all } x \in \text{dom } f \]

• The subgradient inequality is equivalent to
  \[ -s^T\hat{x} + f(\hat{x}) \leq -s^Tx + f(x) \quad \text{for all } x \in \text{dom } f \]

• Let $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. Then
  \[ \text{epi } f = \{(x, w) \mid f(x) \leq w, x \in \mathbb{R}^n \} \]
  Thus, $-s^T\hat{x} + f(\hat{x}) \leq -s^Tx + w$ for all $(x, w) \in \text{epi } f$, equivalent to
  \[
  \begin{bmatrix}
  -s \\
  1
  \end{bmatrix}^T
  \begin{bmatrix}
  \hat{x} \\
  f(\hat{x})
  \end{bmatrix}
  \leq
  \begin{bmatrix}
  -s \\
  1
  \end{bmatrix}^T
  \begin{bmatrix}
  x \\
  w
  \end{bmatrix}
  \quad \text{for all } (x, w) \in \text{epi } f
  \]

Therefore, the hyperplane
\[ H = \{(x, \gamma) \in \mathbb{R}^{n+1} \mid (-s, 1)^T(x, \gamma) = (-s, 1)^T(\hat{x}, f(\hat{x})) \} \]
supports \text{epi } f at the vector $(\hat{x}, f(\hat{x}))$
Subdifferential Set Properties

**Theorem 1** A subdifferential set $\partial f(\hat{x})$ is convex and closed

*Proof* H7.

**Theorem 2** *(Existence)* Let $f$ be convex with a nonempty $\text{dom } f$. Then:

(a) For $x \in \text{relint}(\text{dom } f)$, we have $\partial f(x) \neq \emptyset$.

(b) $\partial f(x) \neq \emptyset$ is nonempty and bounded if and only if $x \in \text{int}(\text{dom } f)$.

**Implications**

- The subdifferential $\partial f(\hat{x})$ is nonempty compact convex set for every $\hat{x}$ in the interior of $\text{dom } f$.

- When $\text{dom } f = \mathbb{R}^n$, $\partial f(x)$ is nonempty compact convex set for all $x$
Partial Proof: If $\hat{x} \in \text{int}(\text{dom } f)$, then $\partial f(\hat{x})$ is nonempty and bounded.

- $\partial f(\hat{x})$ Nonempty.
  Let $\hat{x}$ be in the interior of $\text{dom } f$. The vector $(\hat{x}, f(\hat{x}))$ does not belong to the interior of $\text{epi } f$. The epigraph $\text{epi } f$ is convex and by the Supporting Hyperplane Theorem, there is a vector $(d, \beta) \in \mathbb{R}^{n+1}$, $(d, \beta) \neq 0$ such that

  $$d^T \hat{x} + \beta f(\hat{x}) \leq d^T x + \beta w \quad \text{for all } (x, w) \in \text{epi } f$$

  We have $\text{epi } f = \{(x, w) \mid f(x) \leq w, x \in \text{dom } f\}$. Hence,

  $$d^T \hat{x} + \beta f(\hat{x}) \leq d^T x + \beta w \quad \text{for all } x \in \text{dom } f, f(x) \leq w$$

  We must have $\beta \geq 0$. We cannot have $\beta = 0$ (it would imply $d = 0$). Dividing by $\beta$, we see that $-d/\beta$ is a subgradient of $f$ at $\hat{x}$.
• $\partial f(\hat{x})$ Bounded.

By the subgradient inequality, we have

$$f(x) \geq f(\hat{x}) + s^T(x - \hat{x}) \quad \text{for all } x \in \text{dom } f$$

Suppose that the subdifferential $\partial f(\hat{x})$ is unbounded. Let $s_k$ be a sequence of subgradients in $\partial f(\hat{x})$ with $\|s_k\| \to \infty$.

Since $\hat{x}$ lies in the interior of domain, there exists a $\delta > 0$ such that $\hat{x} + \delta y \in \text{dom } f$ for any $y \in \mathbb{R}^n$. Letting $x = \hat{x} + \delta \frac{s_k}{\|s_k\|}$ for any $k$, we have

$$f \left( \hat{x} + \delta \frac{s_k}{\|s_k\|} \right) \geq f(\hat{x}) + \delta \|s_k\| \quad \text{for all } k$$

As $k \to \infty$, we have $f \left( \hat{x} + \delta \frac{s_k}{\|s_k\|} \right) - f(\hat{x}) \to \infty$.

However, this relation contradicts the continuity of $f$ at $\hat{x}$. [Recall, a convex function is continuous over the interior of its domain.]

**Example** Consider $f(x) = -\sqrt{x}$ with $\text{dom } f = \{ x \mid x \geq 0 \}$. We have $\partial f(0) = \emptyset$. Note that 0 is not in the interior of the domain of $f$
Boundedness of the Subdifferential Sets

**Theorem 2** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex and let $X$ be a bounded set. Then, the set

$$\bigcup_{x \in X} \partial f(x)$$

is bounded.

**Proof**
Assume that there is an unbounded sequence of subgradients $s_k$, i.e.,

$$\lim_{k \to \infty} \|s_k\| = \infty,$$

where $s_k \in \partial f(x_k)$ for some $x_k \in X$. The sequence $\{x_k\}$ is bounded, so it has a convergent subsequence, say $\{x_k\}_K$ converging to some $x \in \mathbb{R}^n$. Consider $d_k = \frac{s_k}{\|s_k\|}$ for $k \in K$. This is a bounded sequence, and it has a
convergent subsequence, say \( \{d_k\}_{K'} \) with \( K' \subseteq K \). Let \( d \) be the limit of \( \{d_k\}_{K'} \).

Since \( s_k \in \partial f(x_k) \), we have for each \( k \),

\[
    f(x_k + d_k) \geq f(x_k) + s_k^T d_k = f(x_k) + \|s_k\|.
\]

By letting \( k \to \infty \) with \( k \in K' \), we see that

\[
    \limsup_{k \to \infty} [f(x_k + d_k) - f(x_k)] \geq \limsup_{k \to \infty} \|s_k\|.
\]

By continuity of \( f \),

\[
    \limsup_{k \to \infty} [f(x_k + d_k) - f(x_k)] = f(x + d) - f(x),
\]

hence finite, implying that \( \|s_k\| \) is bounded. This is a contradiction (\( \{s_k\} \) was assumed to be unbounded).
Continuity of the Subdifferential

**Theorem 3** Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let $\{x_k\}$ converge to some $x \in \mathbb{R}^n$. Let $s_k \in \partial f(x_k)$ for all $k$. Then, the sequence $\{s_k\}$ is bounded and every of its limit points is a subgradient of $f$ at $x$.

*Proof* H7.
Subdifferential and Directional Derivatives

Definition The directional derivative $f'(x; d)$ of $f$ at $x$ along direction $d$ is the following limiting value

$$f'(x; d) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$ 

- When $f$ is convex, the ratio $\frac{f(x + \alpha d) - f(x)}{\alpha}$ is nondecreasing function of $\alpha > 0$, and as $\alpha$ decreases to zero, the ratio converges to some value or decreases to $-\infty$. (HW)

Theorem 4 Let $x \in \text{int}(\text{dom } f)$. Then, the directional derivative $f'(x; d)$ is finite for all $d \in \mathbb{R}^n$. In particular, we have

$$f'(x; d) = \max_{s \in \partial f(x)} s^T d.$$
Proof
When \( x \in \int(\text{dom } f) \), the subdifferential \( \partial f(x) \) is nonempty and compact. Using the subgradient defining relation, we can see that \( f'(x; d) \geq s^T d \) for all \( s \in \partial f(x) \). Therefore,

\[
f'(x; d) \geq \max_{s \in \partial f(x)} s^T d.
\]

To show that actually equality holds, we rely on Separating Hyperplane Theorem. Define

\[
C_1 = \{(z, w) \mid z \in \text{dom } f, f(z) < w\},
\]

\[
C_2 = \{(y, v) \mid y = x + \alpha d, v = f(x) + \alpha f'(x; d), \alpha \geq 0\}.
\]

These sets are nonempty, convex, and disjoint (HW). By the Separating
Hyperplane Theorem, there exists a nonzero vector \((a, \beta) \in \mathbb{R}^{n+1}\) such that

\[
a^T(x + \alpha d) + \beta(f(x) + \alpha f'(x; d)) \leq a^T z + \beta w, \tag{1}
\]

for all \(\alpha \geq 0\), \(z \in \text{dom } f\), and \(f(z) < w\). We must have \(\beta \geq 0\) - why? We cannot have \(\beta = 0\) - why?

Thus, \(\beta > 0\) and we can divide by \(\beta\) the relation in (1), and obtain with \(\tilde{a} = a/\beta\),

\[
\tilde{a}^T(x + \alpha d) + f(x) + \alpha f'(x; d) \leq \tilde{a}^T z + w, \tag{2}
\]

for all \(\alpha \geq 0\), \(z \in \text{dom } f\), and \(f(z) < w\). Choosing \(\alpha = 0\) and letting \(w \downarrow f(z)\), we see

\[
\tilde{a}^T x + f(x) \leq \tilde{a}^T z + f(z),
\]

implying that \(f(x) - \tilde{a}^T(z - x) \leq f(z)\) for all \(z \in \text{dom } f\). Therefore \(-\tilde{a} \in \partial f(x)\).
Letting $z = x$, $w \downarrow f(z)$ and $\alpha = 1$ in (2), we obtain

$$\tilde{a}^T(x + d) + f(x) + f'(x; d) \leq \tilde{a}^T x + f(x),$$

implying $f'(x; d) \leq -\tilde{a}^T d$. In view of $f'(x; d) \geq \max_{s \in \partial f(x)} s^T d$, it follows that

$$f'(x; d) = \max_{s \in \partial f(x)} s^T d,$$

where the maximum is attained at the “constructed” subgradient $-\tilde{a}$.
Optimality Conditions: Unconstrained Case

Unconstrained optimization

\[ \text{minimize } f(x) \]

**Assumption**

- The function \( f \) is convex (non-differentiable) and proper
  \([f \text{ proper means } f(x) > -\infty \text{ for all } x \text{ and } \text{dom } f \neq \emptyset]\)

**Theorem** Under this assumption, a vector \( x^* \) minimizes \( f \) over \( \mathbb{R}^n \) if and only if

\[ 0 \in \partial f(x^*) \]

- The result is a generalization of \( \nabla f(x^*) = 0 \)

- **Proof** \( x^* \) is optimal if and only if \( f(x) \geq f(x^*) \) for all \( x \), or equivalently
  \[ f(x) \geq f(x^*) + 0^T(x - x^*) \quad \text{for all } x \in \mathbb{R}^n \]

  Thus, \( x^* \) is optimal if and only if \( 0 \in \partial f(x^*) \)
Examples

• The function $f(x) = |x|

\[ \partial f(0) = \begin{cases} \text{sign}(x) & \text{for } x \neq 0 \\ [-1, 1] & \text{for } x = 0 \end{cases} \]

The minimum is at $x^* = 0$, and evidently $0 \in \partial f(0)$

• The function $f(x) = \|x\|

\[ \partial f(x) = \begin{cases} \frac{x}{\|x\|} & \text{for } x \neq 0 \\ \{s \mid \|s\| \leq 1\} & \text{for } x = 0 \end{cases} \]

Again, the minimum is at $x^* = 0$ and $0 \in \partial f(0)$
The function \( f(x) = \max\{x^2 + 2x - 3, x^2 - 2x - 3, 4\} \)

\[
f(x) = \begin{cases} 
  x^2 - 2x - 3 & \text{for } x < -1 \\
  4 & \text{for } x \in [-1, 1] \\
  x^2 + 2x - 3 & \text{for } x > 1 
\end{cases}
\]

\[
\partial f(x) = \begin{cases} 
  2x - 2 & \text{for } x > 1 \\
  [-4, 0] & \text{for } x = -1 \\
  0 & \text{for } x \in (-1, 1) \\
  [0, 4] & \text{for } x = 1 \\
  2x + 2 & \text{for } x > 1 
\end{cases}
\]

The optimal set is \( X^* = [-1, 1] \)

For every \( x^* \in X^* \), we have \( 0 \in \partial f(x^*) \)
Optimality Conditions: Constrained Case

Constrained optimization

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in C
\end{align*}
\]

Assumption

- The function \( f \) is convex (non-differentiable) and proper

- The set \( C \) is nonempty closed and convex

Theorem

Under this assumption, a vector \( x^* \in C \) minimizes \( f \) over the set \( C \) if and only if there exists a subgradient \( d \in \partial f(x^*) \) such that

\[
d^T(x - x^*) \geq 0 \quad \text{for all } x \in C
\]

- The result is a generalization of \( \nabla f(x^*)^T(x - x^*) \geq 0 \) for \( x \in C \)