LECTURE 16

LECTURE OUTLINE

• Approximate subgradient methods
• Approximation methods
• Cutting plane methods
APPROXIMATE SUBGRADIENT METHODS

• Consider minimization of

\[ f(x) = \sup_{z \in Z} \phi(x, z) \]

where \( Z \subset \mathbb{R}^m \) and \( \phi(\cdot, z) \) is convex for all \( z \in Z \) (dual minimization is a special case).

• To compute subgradients of \( f \) at \( x \in \text{dom}(f) \), we find \( z_x \in Z \) attaining the supremum above. Then

\[ g_x \in \partial \phi(x, z_x) \quad \Rightarrow \quad g_x \in \partial f(x) \]

• Potential difficulty: For subgradient method, we need to solve exactly the above maximization over \( z \in Z \).

• We consider methods that use “approximate” subgradients that can be computed more easily.
\(\epsilon\)-SUBDIFFERENTIAL

- For a proper convex \(f : \mathbb{R}^n \mapsto (-\infty, \infty]\) and \(\epsilon > 0\), we say that a vector \(g\) is an \(\epsilon\)-subgradient of \(f\) at a point \(x \in \text{dom}(f)\) if

\[
f(z) \geq f(x) + (z - x)'g - \epsilon, \quad \forall \ z \in \mathbb{R}^n
\]

- The \(\epsilon\)-subdifferential \(\partial_\epsilon f(x)\) is the set of all \(\epsilon\)-subgradients of \(f\) at \(x\). By convention, \(\partial_\epsilon f(x) = \emptyset\) for \(x \notin \text{dom}(f)\).

- We have \(\cap_{\epsilon \downarrow 0} \partial_\epsilon f(x) = \partial f(x)\) and

\[
\partial_\epsilon_1 f(x) \subset \partial_\epsilon_2 f(x) \quad \text{if} \ 0 < \epsilon_1 < \epsilon_2
\]
CALCULATION OF AN $\epsilon$-SUBGRADIENT

- Consider minimization of

$$ f(x) = \sup_{z \in Z} \phi(x, z), \quad (1) $$

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $Z$ is a subset of $\mathbb{R}^m$, and $\phi : \mathbb{R}^n \times \mathbb{R}^m \mapsto (-\infty, \infty]$ is a function such that $\phi(\cdot, z)$ is convex and closed for each $z \in Z$.

- How to calculate $\epsilon$-subgradient at $x \in \text{dom}(f)$?

- Let $z_x \in Z$ attain the supremum within $\epsilon \geq 0$ in Eq. (1), and let $g_x$ be some subgradient of the convex function $\phi(\cdot, z_x)$.

- For all $y \in \mathbb{R}^n$, using the subgradient inequality,

$$ f(y) = \sup_{z \in Z} \phi(y, z) \geq \phi(y, z_x) \geq \phi(x, z_x) + g_x'(y - x) \geq f(x) - \epsilon + g_x'(y - x) $$

i.e., $g_x$ is an $\epsilon$-subgradient of $f$ at $x$, so

$$ \phi(x, z_x) \geq \sup_{z \in Z} \phi(x, z) - \epsilon \text{ and } g_x \in \partial \phi(x, z_x) $$

$$ \implies g_x \in \partial_\epsilon f(x) $$
ε-SUBGRADIENT METHOD

- Uses an ε-subgradient in place of a subgradient.
- **Problem:** Minimize convex $f : \mathbb{R}^n \mapsto \mathbb{R}$ over a closed convex set $X$.
- **Method:**

  $$x_{k+1} = P_X(x_k - \alpha_k g_k)$$

where $g_k$ is an $\epsilon_k$-subgradient of $f$ at $x_k$, $\alpha_k$ is a positive stepsize, and $P_X(\cdot)$ denotes projection on $X$.

- Can be viewed as subgradient method with “errors”.


CONVERGENCE ANALYSIS

- **Basic inequality:** If \( \{x_k\} \) is the \( \epsilon \)-subgradient method sequence, for all \( y \in X \) and \( k \geq 0 \)

\[
\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y) - \epsilon_k) + \alpha_k^2 \|g_k\|^2
\]

- Replicate the entire convergence analysis for subgradient methods, but carry along the \( \epsilon_k \) terms.

- **Example:** Constant \( \alpha_k \equiv \alpha \), constant \( \epsilon_k \equiv \epsilon \). Assume \( \|g_k\| \leq c \) for all \( k \). For any optimal \( x^* \),

\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\alpha (f(x_k) - f^* - \epsilon) + \alpha^2 c^2,
\]

so the distance to \( x^* \) decreases if

\[
0 < \alpha < \frac{2(f(x_k) - f^* - \epsilon)}{c^2}
\]

or equivalently, if \( x_k \) is outside the level set

\[
\left\{ x \mid f(x) \leq f^* + \epsilon + \frac{\alpha c^2}{2} \right\}
\]

- **Example:** If \( \alpha_k \to 0, \sum k \alpha_k \to \infty \), and \( \epsilon_k \to \epsilon \), we get convergence to the \( \epsilon \)-optimal set.
INCREMENTAL SUBGRADIENT METHODS

- Consider minimization of sum
  \[ f(x) = \sum_{i=1}^{m} f_i(x) \]

- Often arises in duality contexts with \( m: \) very large (e.g., separable problems).

- Incremental method moves \( x \) along a subgradient \( g_i \) of a component function \( f_i \) NOT the (expensive) subgradient of \( f \), which is \( \sum_i g_i \).

- View an iteration as a cycle of \( m \) subiterations, one for each component \( f_i \).

- Let \( x_k \) be obtained after \( k \) cycles. To obtain \( x_{k+1} \), do one more cycle: Start with \( \psi_0 = x_k \), and set \( x_{k+1} = \psi_m \), after the \( m \) steps
  \[ \psi_i = P_X(\psi_{i-1} - \alpha_k g_i), \quad i = 1, \ldots, m \]

  with \( g_i \) being a subgradient of \( f_i \) at \( \psi_{i-1} \).

- Motivation is faster convergence. A cycle can make much more progress than a subgradient iteration with essentially the same computation.
CONNECTION WITH $\epsilon$-SUBGRADIENTS

- **Neighborhood property:** If $x$ and $\bar{x}$ are “near” each other, then subgradients at $\bar{x}$ can be viewed as $\epsilon$-subgradients at $x$, with $\epsilon$ “small.”

- If $g \in \partial f(\bar{x})$, we have for all $z \in \mathbb{R}^n$,

\[
    f(z) \geq f(\bar{x}) + g'(z - \bar{x}) \\
    \geq f(x) + g'(z - x) + f(\bar{x}) - f(x) + g'(x - \bar{x}) \\
    \geq f(x) + g'(z - x) - \epsilon,
\]

where $\epsilon = |f(\bar{x}) - f(x)| + \|g\| \cdot \|\bar{x} - x\|$. Thus, $g \in \partial_\epsilon f(x)$, with $\epsilon$: small when $\bar{x}$ is near $x$.

- The incremental subgradient iter. is an $\epsilon$-subgradient iter. with $\epsilon = \epsilon_1 + \cdots + \epsilon_m$, where $\epsilon_i$ is the “error” in $i$th step in the cycle ($\epsilon_i$: Proportional to $\alpha_k$).

- Use

\[
    \partial_{\epsilon_1} f_1(x) + \cdots + \partial_{\epsilon_m} f_m(x) \subset \partial_\epsilon f(x),
\]

where $\epsilon = \epsilon_1 + \cdots + \epsilon_m$, to approximate the $\epsilon$-subdifferential of the sum $f = \sum_{i=1}^{m} f_i$.

- Convergence to optimal if $\alpha_k \to 0$, $\sum_k \alpha_k \to \infty$. 

 APPROXIMATION APPROACHES

• Approximation methods replace the original problem with an approximate problem.
• The approximation may be iteratively refined, for convergence to an exact optimum.
• A partial list of methods:
  – Cutting plane/outer approximation.
  – Simplicial decomposition/inner approximation.
  – Proximal methods (including Augmented Lagrangian methods for constrained minimization).
  – Interior point methods.
• A partial list of combination of methods:
  – Combined inner-outer approximation.
  – Bundle methods (proximal-cutting plane).
  – Combined proximal-subgradient (incremental option).
SUBGRADIENTS-OUTER APPROXIMATION

- Consider minimization of a convex function \( f : \mathbb{R}^n \mapsto \mathbb{R} \), over a closed convex set \( X \).
- We assume that at each \( x \in X \), a subgradient \( g \) of \( f \) can be computed.
- We have

\[
f(z) \geq f(x) + g'(z - x), \quad \forall z \in \mathbb{R}^n,
\]

so each subgradient defines a plane (a linear function) that approximates \( f \) from below.
- The idea of the outer approximation/cutting plane approach is to build an ever more accurate approximation of \( f \) using such planes.
CUTTING PLANE METHOD

• Start with any $x_0 \in X$. For $k \geq 0$, set

$$x_{k+1} \in \arg \min_{x \in X} F_k(x),$$

where

$$F_k(x) = \max \{ f(x_0) + (x-x_0)'g_0, \ldots, f(x_k) + (x-x_k)'g_k \}$$

and $g_i$ is a subgradient of $f$ at $x_i$.

![Graph showing cutting plane method](image)

• Note that $F_k(x) \leq f(x)$ for all $x$, and that $F_k(x_{k+1})$ increases monotonically with $k$. These imply that all limit points of $x_k$ are optimal.

**Proof:** If $x_k \to \bar{x}$ then $F_k(x_k) \to f(\bar{x})$, [otherwise there would exist a hyperplane strictly separating $\text{epi}(f)$ and $(\bar{x}, \lim_{k \to \infty} F_k(x_k))]$. This implies that $f(\bar{x}) \leq \lim_{k \to \infty} F_k(x) \leq f(x)$ for all $x$.  \textbf{Q.E.D.}
CONVERGENCE AND TERMINATION

• We have for all $k$

$$F_k(x_{k+1}) \leq f^* \leq \min_{i \leq k} f(x_i)$$

• Termination when $\min_{i \leq k} f(x_i) - F_k(x_{k+1})$ comes to within some small tolerance.

• For $f$ polyhedral, we have finite termination with an exactly optimal solution.

• Instability problem: The method can make large moves that deteriorate the value of $f$.

• Starting from the exact minimum it typically moves away from that minimum.
VARIANTS

- **Variant I**: Simultaneously with $f$, construct polyhedral approximations to $X$.
- **Variant II**: Central cutting plane methods
6.253 Convex Analysis and Optimization
Spring 2012

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.