LECTURE 20

LECTURE OUTLINE

• Approximation methods
• Cutting plane methods
• Proximal minimization algorithm
• Proximal cutting plane algorithm
• Bundle methods

All figures are courtesy of Athena Scientific, and are used with permission.
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).
Consider minimization of a convex function $f : \mathbb{R}^n \to \mathbb{R}$, 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

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

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

• 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 $x_k \to x$ then $F_k(x_k) \to f(x)$, [otherwise there would exist a hyperplane strictly separating $\text{epi}(f)$ and $(x, \lim_{k \to \infty} F_k(x_k))]$. This implies that $f(x) \leq \lim_{k \to \infty} F_k(x) \leq f(x)$ for all $x$. **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.

\[\text{diagram}\]

• **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
- **Variant III:** Proximal methods - to be discussed next.
PROXIMAL/BUNDLE METHODS

- Aim to reduce the instability problem at the expense of solving a more difficult subproblem.
- A general form:

\[ x_{k+1} \in \arg \min_{x \in X} \{ F_k(x) + p_k(x) \} \]

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

\[ p_k(x) = \frac{1}{2c_k} \| x - y_k \|^2 \]

where \( c_k \) is a positive scalar parameter.
- We refer to \( p_k(x) \) as the *proximal term*, and to its center \( y_k \) as the *proximal center*. 

![Graph illustrating proximal/bundle methods](image)
PROXIMAL MINIMIZATION ALGORITHM

• Starting point for analysis: A general algorithm for convex function minimization

\[ x_{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2c_k} \| x - x_k \|^2 \right\} \]

- \( f : \mathbb{R}^n \mapsto (-\infty, \infty] \) is closed proper convex
- \( c_k \) is a positive scalar parameter
- \( x_0 \) is arbitrary starting point

• Convergence mechanism:

\[ \gamma_k = f(x_{k+1}) + \frac{1}{2c_k} \| x_{k+1} - x_k \|^2 < f(x_k). \]

Cost improves by at least \( \frac{1}{2c_k} \| x_{k+1} - x_k \|^2 \), and this is sufficient to guarantee convergence.
RATE OF CONVERGENCE I

• Role of penalty parameter $c_k$:

• Role of growth properties of $f$ near optimal solution set:
RATE OF CONVERGENCE II

- Assume that for some scalars $\beta > 0$, $\delta > 0$, and $\alpha \geq 1$,

$$f^* + \beta (d(x))^{\alpha} \leq f(x), \quad \forall \ x \in \mathbb{R}^n \text{ with } d(x) \leq \delta$$

where

$$d(x) = \min_{x^* \in X^*} \|x - x^*\|$$

i.e., growth of order $\alpha$ from optimal solution set $X^*$.

- If $\alpha = 2$ and $\lim_{k \to \infty} c_k = \bar{c}$, then

$$\lim_{k \to \infty} \sup \frac{d(x_{k+1})}{d(x_k)} \leq 1$$

linear convergence.

- If $1 < \alpha < 2$, then

$$\lim_{k \to \infty} \sup \left(\frac{d(x_{k+1})}{d(x_k)}\right)^{1/(\alpha - 1)} < \infty$$

superlinear convergence.
FINITE CONVERGENCE

• Assume growth order $\alpha = 1$:

$$f^* + \beta d(x) \leq f(x), \quad \forall \ x \in \mathbb{R}^n,$$

e.g., $f$ is polyhedral.

• Method converges finitely (in a single step for $c_0$ sufficiently large).
PROXIMAL CUTTING PLANE METHODS

• Same as proximal minimization algorithm, but $f$ is replaced by a cutting plane approximation $F_k$:

$$x_{k+1} \in \arg \min_{x \in X} \left\{ F_k(x) + \frac{1}{2c_k} \| x - x_k \|^2 \right\}$$

where

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

• Drawbacks:

  (a) **Hard stability tradeoff:** For large enough $c_k$ and polyhedral $X$, $x_{k+1}$ is the exact minimum of $F_k$ over $X$ in a single minimization, so it is identical to the ordinary cutting plane method. For small $c_k$ convergence is slow.

  (b) **The number of subgradients used in $F_k$ may become very large:** the quadratic program may become very time-consuming.

• These drawbacks motivate algorithmic variants, called *bundle methods*. 
**BUNDLE METHODS**

- Allow a proximal center $y_k \neq x_k$:

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

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

  $$p_k(x) = \frac{1}{2c_k} \|x - y_k\|^2$$

- **Null/Serious test** for changing $y_k$: For some fixed $\beta \in (0, 1)$

  $$y_{k+1} = \begin{cases} 
  x_{k+1} & \text{if } f(y_k) - f(x_{k+1}) \geq \beta \delta_k, \\
  y_k & \text{if } f(y_k) - f(x_{k+1}) < \beta \delta_k,
  \end{cases}$$

  $$\delta_k = f(y_k) - (F_k(x_{k+1}) + p_k(x_{k+1})) > 0$$
6.253 Convex Analysis and Optimization
Spring 2010

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