Lecture 20

Methods for Dual Problems

November 11, 2008
Outline

• Interpretation of Polyak’s stepsize

• Convergence and Convergence Rate

• Subgradient Methods for Dual Problems
**Interpretation of Polyak’s stepsize**

A vector \( s \) is a subgradient of a convex function \( f : \mathbb{R}^n \mapsto \mathbb{R} \) at \( \hat{x} \in \text{dom } f \) when **subgradient inequality holds**

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

- We have interpreted the subgradient inequality in terms of a **hyperplane** in \( \mathbb{R}^{n+1} \) supporting the **epigraph** \( \text{epi } f \) at \((\hat{x}, f(\hat{x}))\)
Polyak stepsize can be interpreted by looking at the projection of the epigraph and the hyperplane on the set
\[ \{ (x, w) \mid x \in \mathbb{R}^n, w = f^* \} \]

The projection of the hyperplane is given by
\[ \tilde{H} = \{ (z, w) \in \mathbb{R}^{n+1} \mid f(\tilde{x}) + s^T (z - \tilde{x}) = f^*, w = f^* \} \]
• By looking only at $x$-variables (since $w = f^*$), at $\hat{x} = x_k$, the resulting hyperplane in the reduced space becomes

$$H = \{ z \in \mathbb{R}^n \mid f(x_k) + s_k^T(z - x_k) = f^* \}$$

• With Polyak’s stepsize, the iterate $x_{k+1}$ is the projection of $x_k$ on $H$

• To see that note that the projection of $x_k$ on $H$ can be determined by looking at the intersection of the ray $\{ z \mid z = x_k + ts_k, \ t \geq 0 \}$ with $H$, which gives

$$f(x_k) + t^*\|s_k\|^2 = f^*.$$ 

Solving for $t^*$ yields $t^* = \frac{f^* - f(x_k)}{\|s_k\|^2}$.

• The next iterate is

$$x_{k+1} = x_k + t^*s_k = x_k - \frac{f(x_k) - f^*}{\|s_k\|^2}s_k$$

The Polyak stepsize $\alpha_k = \frac{f(x_k) - f^*}{\|s_k\|^2}$ is equal to the distance from $x_k$ to the hyperplane $H$, i.e., $\alpha_k = |t^*|$.
Convergence Rate

• The convergence rate of the subgradient method with Polyak’s stepsize is linear (at best)
  
  • For a function $f$ with **sharp minima**, i.e., such that for some $\eta > 0$
    
    $$f(x) - f^* \geq \eta \text{dist}(x, X^*)$$
    
    for all $x$

  • **The rate is linear** (HW8)
    
    $$\|x_k - \tilde{x}^*\| \leq c_k \|x_0 - \tilde{x}^*\|$$
    
    for all $k \geq 0$

    where $\tilde{x}^* \in X^*$ is the limit point of $\{x_k\}$ and

    $$c = \sqrt{1 - \frac{\eta^2}{L^2}}$$

    and $L$ is an upper bound on the subgradient norms $\|s_k\|$.

• The rate is important for general understanding of the method

• It is rare that we can take advantage of this result in practice
Comments

- Subgradient methods considered so far:
  - Use any subgradient that is available at a given iterate
  - Simple for implementation
  - Convergence rate is at best linear
  - Useful in large-scale and decentralized computations
  - **Main criticism:** There are no general stopping rules
    - A specific criteria have been designed within particular applications

- Alternative methods exist: **Bundle methods**
  - Use a carefully selected subgradient at a given iterate
  - More sophisticated for implementation
  - There is a general stopping criteria