LECTURE 21

LECTURE OUTLINE

• We enter a series of lectures on advanced topics
  – Gradient projection
  – Variants of gradient projection
  – Variants of proximal and combinations
  – Incremental subgradient and proximal methods
  – Coordinate descent methods
  – Interior point methods, etc

• Today’s lecture on gradient projection

• Application to differentiable problems

• Iteration complexity issues

**************

• Reference: The on-line chapter of the textbook
GRADIENT PROJECTION METHOD

- Let $f$ be continuously differentiable, and $X$ be closed convex.
- **Gradient projection method:**

  $$x_{k+1} = P_X \left( x_k - \alpha_k \nabla f(x_k) \right)$$

- A specialization of subgradient method, but **cost function descent comes into play**

- $x_{k+1} - x_k$ is a feasible descent direction (by the projection theorem)

- $f(x_{k+1}) < f(x_k)$ if $\alpha_k$: sufficiently small (unless $x_k$ is optimal)

- $\alpha_k$ may be constant or chosen by cost descent-based stepsize rules
CONNECTION TO THE PROXIMAL ALGORITHM

- Linear approximation of $f$ based on $\nabla f(x)$:

$$\ell(y; x) = f(x) + \nabla f(x)'(y - x), \quad \forall \ x, y \in \mathbb{R}^n$$

- For all $x \in X$ and $\alpha > 0$, we have

$$\frac{1}{2\alpha} \| y - (x - \alpha \nabla f(x)) \|^2 = \ell(y; x) + \frac{1}{2\alpha} \| y - x \|^2 + \text{constant}$$

so

$$P_X (x - \alpha \nabla f(x)) \in \arg \min_{y \in X} \left\{ \ell(y; x) + \frac{1}{2\alpha} \| y - x \|^2 \right\}$$

- Three-term inequality holds: For all $y \in \mathbb{R}^n$,

$$\| x_{k+1} - y \|^2 \leq \| x_k - y \|^2 - 2\alpha_k \left( \ell(x_{k+1}; x_k) - \ell(y; x_k) \right) - \| x_k - x_{k+1} \|^2$$
CONSTANT STEPSIZE - DESCENT LEMMA

- Consider constant $\alpha_k$: $x_{k+1} = P_X \left( x_k - \alpha \nabla f(x_k) \right)$
- We need the gradient Lipschitz assumption

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|, \quad \forall x, y \in X$$

- Descent Lemma: For all $x, y \in X$,

$$f(y) \leq \ell(y; x) + \frac{L}{2} \| y - x \|^2$$

- Proof idea: The Lipschitz constant $L$ serves as an upper bound to the “curvature” of $f$ along directions, so $\frac{L}{2} \| y - x \|^2$ is an upper bound to $f(y) - \ell(y; x)$. 
CONSTANT STEPSIZE - CONVERGENCE RESULT

- Assume the gradient Lipschitz condition, and \( \alpha \in (0, 2/L) \) (no convexity of \( f \)). Then \( f(x_k) \downarrow f^* \) and every limit point of \( \{x_k\} \) is optimal.

**Proof:** From the projection theorem, we have

\[
(x_k - \alpha \nabla f(x_k) - x_{k+1})' (x - x_{k+1}) \leq 0, \quad \forall \ x \in X,
\]

so by setting \( x = x_k \),

\[
\nabla f(x_k)' (x_{k+1} - x_k) \leq -\frac{1}{\alpha} \left\| x_{k+1} - x_k \right\|^2
\]

- Using this relation and the descent lemma,

\[
f(x_{k+1}) \leq \ell(x_{k+1}; x_k) + \frac{L}{2} \left\| x_{k+1} - x_k \right\|^2
\]

\[
= f(x_k) + \nabla f(x_k)' (x_{k+1} - x_k) + \frac{L}{2} \left\| x_{k+1} - x_k \right\|^2
\]

\[
\leq f(x_k) - \left( \frac{1}{\alpha} - \frac{L}{2} \right) \left\| x_{k+1} - x_k \right\|^2
\]

so \( \alpha \in (0, 2/L) \) reduces the cost function value.

- If \( \alpha \in (0, 2/L) \) and \( \overline{x} \) is the limit of a subsequence \( \{x_k\}_k \), then \( f(x_k) \downarrow f(\overline{x}) \), so \( \|x_{k+1} - x_k\| \to 0 \). This implies \( P_X (\overline{x} - \alpha \nabla f(\overline{x})) = \overline{x} \). **Q.E.D.**
STEPSIZE RULES

- **Eventually constant stepsize.** Deals with the case of an unknown Lipschitz constant $L$. Start with some $\alpha > 0$, and keep using $\alpha$ as long as

$$f(x_{k+1}) \leq \ell(x_{k+1}; x_k) + \frac{1}{2\alpha} \|x_{k+1} - x_k\|^2$$

is satisfied (this guarantees cost descent). When this condition is violated at some iteration, we reduce $\alpha$ by a certain factor, and repeat. (Satisfied once $\alpha \leq 1/L$, by the descent lemma.)

- **A diminishing stepsize $\alpha_k$, satisfying**

$$\lim_{k \to \infty} \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty.$$ 

Does not require Lipschitz condition or differentiability of $f$, only convexity of $f$.

- **Stepsize reduction and line search rules - Armijo rules.** These rules are based on cost function descent, and ensure that through some form of line search, we find $\alpha_k$ such that $f(x_{k+1}) < f(x_k)$, unless $x_k$ is optimal. Do not require Lipschitz condition, only differentiability of $f$. 

ARMIJO STEPSIZE RULES

- Search along the projection arc: $\alpha_k = \beta^{m_k} s$, where $s > 0$ and $\beta \in (0, 1)$ are fixed scalars, and $m_k$ is the first integer $m$ such that

$$f(x_k) - f(x_k(\beta^m s)) \geq \sigma \nabla f(x_k)'(x_k - x_k(\beta^m s)),$$

with $\sigma \in (0, 1)$ being some small constant, and

$$x_k(\alpha) = P_X(x_k - \alpha \nabla f(x_k))$$

- Similar rule searches along the feasible direction
CONVERGENCE RATE - $\alpha_K \equiv 1/L$

- Assume $f$: convex, the Lipschitz condition, $X^* \neq \emptyset$, and the eventually constant stepsize rule. Denote $d(x_k) = \min_{x^* \in X^*} \|x_k - x^*\|$. Then

$$\lim_{k \to \infty} d(x_k) = 0, \quad f(x_k) - f^* \leq \frac{Ld(x_0)^2}{2k}$$

**Proof:** Let $x^* \in X^*$ be such that $\|x_0 - x^*\| = d(x_0)$. Using the descent lemma and the three-term inequality,

$$f(x_{k+1}) \leq \ell(x_{k+1}; x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2$$

$$\leq \ell(x^*; x_k) + \frac{L}{2} \|x^* - x_k\|^2 - \frac{L}{2} \|x^* - x_{k+1}\|^2$$

$$\leq f(x^*) + \frac{L}{2} \|x^* - x_k\|^2 - \frac{L}{2} \|x^* - x_{k+1}\|^2$$

Let $e_k = f(x_k) - f(x^*)$ and note that $e_k \downarrow$. Then

$$\frac{L}{2} \|x^* - x_{k+1}\|^2 \leq \frac{L}{2} \|x^* - x_k\|^2 - e_{k+1}$$

Use this relation with $k = k - 1, k - 2, \ldots$, and add

$$0 \leq \frac{L}{2} \|x^* - x_{k+1}\|^2 \leq \frac{L}{2} d(x_0)^2 - (k + 1)e_{k+1}$$
GENERALIZATION - EVENTUALLY CONST. $\alpha_K$

- Assume $f$: convex, the Lipschitz condition, $X^* \neq \emptyset$, and any stepsize rule such that
  $$\alpha_k \downarrow \alpha,$$
  for some $\alpha > 0$, and for all $k$,
  $$f(x_{k+1}) \leq \ell(x_{k+1}; x_k) + \frac{1}{2\alpha_k} \|x_{k+1} - x_k\|^2.$$

Denote $d(x_k) = \min_{x^* \in X^*} \|x_k - x^*\|$. Then
  $$\lim_{k \to \infty} d(x_k) = 0, \quad f(x_k) - f^* \leq \left(\frac{d(x_0)^2}{2\alpha k}\right).$$

**Proof:** Show that
  $$f(x_{k+1}) \leq f(x_k) - \frac{1}{2\alpha_k} \|x_{k+1} - x_k\|^2,$$
and generalize the preceding proof. **Q.E.D.**

- Applies to eventually constant stepsize rule.

- **Error complexity $O(1/k)$**, $(k$ iterations produce $O(1/k)$ cost error), i.e., $\min_{\ell \leq k} f(x_\ell) \leq f^* + \frac{\text{const}}{k}$

- **Iteration complexity $O(1/\epsilon)$**, $(O(1/\epsilon)$ iterations produce $\epsilon$ cost error), i.e., $\min_{k \leq \text{const}/\epsilon} f(x_k) \leq f^* + \epsilon$
SHARPNESS OF COMPLEXITY ESTIMATE

- Unconstrained minimization of

\[ f(x) = \begin{cases} \frac{c}{2}|x|^2 & \text{if } |x| \leq \epsilon, \\ c\epsilon|x| - \frac{c\epsilon^2}{2} & \text{if } |x| > \epsilon \end{cases} \]

- With stepsize \( \alpha = 1/L = 1/c \) and any \( x_k > \epsilon \),

\[ x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k) = x_k - \frac{1}{c}\epsilon x_k = x_k - \epsilon \]

- The number of iterations to get within an \( \epsilon \)-neighborhood of \( x^* = 0 \) is \(|x_0|/\epsilon\).

- The number of iterations to get to within \( \epsilon \) of \( f^* = 0 \) is proportional to \( 1/\epsilon \) for large \( x_0 \).