Lecture 12

Unconstrained Optimization (contd.)
Constrained Optimization

October 15, 2008

Outline

- Gradient descent algorithm
  - Improvement to result in Lec 11
  - At what rate will it converge?

- Constrained minimization over simple sets

Unconstrained Minimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject} & x \in \Re^n. \end{array}$$

- Assumption 1:
  - The function f is convex and continuously differentiable over $\mathbb{R}^n$
  - The optimal value $f^* = \inf_{x \in \mathbb{R}^n} f(x)$ is finite.

- Gradient descent algorithm

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

Theorem 1: Bounded Gradients

- **Theorem 1** Let Assumption 1 hold, and suppose that the gradients are bounded. Then, the gradient method generates the sequence $\{x_k\}$ such that

$$\liminf_{k\to\infty} f(x_k) \le f^* + \frac{\alpha L^2}{2}$$

- We first proved that for $y \in \mathbb{R}^n$ and for all k,

$$\|x_{k+1} - y\|^2 \le \|x_k - y\|^2 - 2\alpha_k(f(x_k) - f(y)) + \alpha_k^2\|\nabla f(x_k)\|^2.$$

Theorem 2: Lipschitz Gradients

- **Q-approximation Lemma**
  For continuously differentiable function with Lipschitz gradients, we have

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{M}{2} \|y - x\|^2 \qquad \text{for all } x, y \in \mathbb{R}^n,$$

- **Theorem** ~~Let Assumption 1 hold~~, and assume that the gradients of f are Lipschitz. Then, for α with $\alpha < \frac{2}{M}$, we have

$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0.$$

  If in addition, an optimal solution exists [i.e., the $min_x f(x)$ is attained at some x^*], then every accumulation point of the sequence $\{x_k\}$ is optimal.

- *Proof:* Using Q-approximation Lemma with $y = x_{k+1}$ and $x = x_k$, we have

$$\begin{aligned} f(x_{k+1}) &\le f(x_k) - \alpha\|\nabla f(x_k)\|^2 + \frac{\alpha^2 M}{2}\|\nabla f(x_k)\|^2 \\ &= f(x_k) - \frac{\alpha}{2}(2 - \alpha M)\,\|\nabla f(x_k)\|^2 \end{aligned}$$

By summing these relations and using $2 - \alpha M > 0$, we can see that

$$\sum_k \|\nabla f(x_k)\|^2 < \infty,$$

implying $\lim_{k\to\infty} \|f(x_k)\| = 0$.

Suppose a solution exists, and the sequence $\{x_k\}$ has an accumulation point $\bar{x}$. Then, by continuity of the gradient we have $\nabla f(x_k) \to \nabla f(\bar{x})$ along an appropriate subsequence.

Since $\nabla f(x_k) \to 0$, it follows $\nabla f(\bar{x}) = 0$, implying that $\bar{x}$ is a solution.

- Where do we use convexity?
- Correct Theorem 2: Assume that the gradients of f are Lipschitz. Then, for α with $\alpha < \frac{2}{M}$, we have

$$\sum_{k=1}^{\infty} \|\nabla f(x_k)\|^2 < \infty.$$

If in addition, an optimal solution exists [i.e., the $min_x f(x)$ is attained at some x^*], then every accumulation point of the sequence $\{x_k\}$ is optimal.

- $\{x_k\}$ need not converge. Note difference between limit point and accumalation point.
- Example: $f(x) = \frac{1}{1+x^2}$. Satisfies conditions and $x_k \to \infty$.

What does convexity buy?

- Answer: Convergence of the sequence $\{x_k\}$
- Recall/Define X^* as optimal set.
- **Theorem 3:** Let Assumption 1 hold. Further, assume that the gradients of f are Lipschitz and α with $\alpha < \frac{2}{M}$. If X^* is nonempty then

$$\lim_{k\to\infty} x_k = x^*, \quad x^* \in X^*.$$

- Proof (outline): Previous theorem tells us a lot. Every convergent subsequence of $\{x_k\}$ is a point in X^*. Questions
  - Is there a convergent subsequence?
  - Do all the subsequence converge to the same point in X^*?

Conditions of Lemma 1 are satisfied. Fix $y = \tilde{x}$, where $\tilde{x} \in X^*$

$$\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - 2\alpha(f(x_k) - f^*) + \alpha^2 \|\nabla f(x_k)\|^2.$$

Drop the negative term

$$\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 + \alpha^2 \|\nabla f(x_k)\|^2.$$

Note from Theorem 2 that

$$\sum_{l=1}^{\infty} \|\nabla f(x_l)\|^2 < \infty.$$

Therefore adding $\alpha^2 \sum_{l=k+1}^{\infty} \|\nabla f(x_l)\|^2$ to both sides we obtain

$$\|x_{k+1} - \tilde{x}\|^2 + \alpha^2 \sum_{l=k+1}^{\infty} \|\nabla f(x_l)\|^2 \leq \|x_k - \tilde{x}\|^2 + \alpha^2 \sum_{l=k}^{\infty} \|\nabla f(x_l)\|^2.$$

Define

$$u_k = \|x_k - \tilde{x}\|^2 + \alpha^2 \sum_{l=k}^{\infty} \|\nabla f(x_l)\|^2.$$

Thus, equation is

$$u_{k+1} \leq u_k,$$

which implies $\{u_k\}$ converges. We already know $\sum_{l=k}^{\infty} \|\nabla f(x_l)\|^2$ converges as $k \to \infty$. Therefore, $\|x_k - \tilde{x}\|$ converges. Therefore, $\{x_k\}$ is bounded => convergent subsequence exists.

Take a convergent subsequence $\{x_{s_k}\}$ and let limit point be $\bar{x}$. We know

$\bar{x} \in X^*$ from previous Theorem. Thus,

$$\lim_{k\to\infty} \|x_{s_k} - \bar{x}\| = 0.$$

But we just showed that $\lim_{k\to\infty} \|x_k - \bar{x}\|$ exists. Therefore, the limit must be 0 and $\{x_k\}$ converge to $\bar{x}$.

- Quick Summary:
  - Without convexity no convergence of iterates.
  - With convexity convergence of iterates.
- Go back and check $f(x) = \frac{1}{1+x^2}$. It can't be convex.
- General note: What do I get from all this analysis?

Rates of convergence

- Successfully solved the problem. Finish course and go home? Unfortunately, no. Gradient descent has poor rates of convergence. Emperically observed when we simulate.
- See last page for an example.
- Lets consider the class of strongly convex functions. Recall,
  - $\nabla^2 f(x) \succeq mI \qquad$ for all $x \in \mathbb{R}^n$
  - $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}$
  - $\frac{1}{2m} \|\nabla f(x)\|^2 \geq f(x) - f^* \geq \frac{m}{2} \|x - x^*\|^2$

- **Theorem 4:** Let Assumption 1 hold. Further, assume that f is strongly convex, gradients of f are Lipschitz and α with $\alpha < \frac{\min(2,m)}{M}$. Then,

$$\|x_k - x^*\| \leq cq^k, \qquad 0 < q < 1.$$

  Note: Geometric rate of convergence. (Normal people call it exponential;))

- Proof: Strongly convex => Unique minimum. Basic iterate relation with $y = x^*$ gives

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &\le \|x_k - x^*\|^2 - 2\alpha(f(x_k) - f^*) + \alpha\|\nabla f(x_k)\|^2 \\
&\le \|x_k - x^*\|^2 - 2\alpha\frac{m}{2}\|x_k - x^*\|^2 + \alpha\|\nabla f(x_k)\|^2 \\
&= \|x_k - x^*\|^2 - 2\alpha\frac{m}{2}\|x_k - x^*\|^2 + \alpha\|\nabla f(x_k) - \nabla f(x^*)\|^2 \\
&\le \|x_k - x^*\|^2 - \alpha m\|x_k - x^*\|^2 + \alpha^2 M\|x_k - x^*\|^2 \\
&= \left(1 - m\alpha + \alpha^2 M\right)\|x_k - x^*\|^2 \\
&= \left(1 - m\alpha + \alpha^2 M\right)^{k+1}\|x_0 - x^*\|^2
\end{aligned}$$

- Remark: Result holds when $0 < \alpha < \frac{2}{m}$.
- Geometric is not bad. But, how many functions are strongly convex?

Constrained minimization over simple sets

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject} & x \in X. \end{array}$$

- Assumption 2
  - f is continuously differentiable and convex on an open interval containing X
  - X is closed and compact
  - $f^* > -\infty$
- Optimality condition $\nabla f(x^*)^T(x - x^*) \leq 0$
- Simple sets: Easy projection. Ball, hyperplanes, halfspaces
- Assumption 2: The set X is closed and convex.

- Recall projection
  - $\|P_X[x] - x\| \leq \|P_X[x] - y\|$ when $y \in X$
  - $\|P_X[x] - P_X[y]\| \leq \|x - y\|$ when $x, y \in \Re^n$

- Gradient projection algorithm

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

- Can we obtain results similar to Theorems 1-4?

- **Lemma 1 and Theorem 1** Let Assumption 1 and 2 hold, and suppose that the gradients are bounded. Then, for any $y \in X$, and all k

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k(f(x_k) - f(y)) + \alpha_k^2\|\nabla f(x_k)\|^2.$$

and hence for $\alpha_k = \alpha$,

$$\liminf_{k\to\infty} f(x_k) \le f^* + \frac{\alpha L^2}{2}$$

- Proof: By the definition of the method, it follows that for any k,

$$\begin{aligned} \|x_{k+1} - y\|^2 &= \|P_X[x_k - \alpha_k \nabla f(x_k)] - y\|^2 \\ &\le \|x_k - \alpha_k \nabla f(x_k) - y\|^2 \\ &\le \|x_k - y\|^2 - 2\alpha_k \nabla f(x_k)^T (x_k - y) + \alpha_k^2 \|\nabla f(x_k)\|^2. \end{aligned}$$

Rest identical

- **Theorem 2**: Not very interesting. After all, $\nabla f(x^*)$ need not be 0. What we could show is something like $\lim_{k\to\infty} \nabla f(x_k)^T (x - x_k) \le 0$ for all x.

- **Theorem 3:** Let Assumption 2 hold. Further, assume that the gradients of f are Lipschitz and α with $\alpha < \frac{2}{M}$. If X^* is nonempty then

$$\lim_{k\to\infty} x_k = x^*, \quad x^* \in X^*.$$

- **Theorem 4:** Let Assumption 2 hold. Further, assume that f is strongly convex, gradients of f are Lipschitz and α with $\alpha < \frac{\min(2,m)}{M}$. Then,

$$\|x_k - x^*\| \le \overline{c}\overline{q}^k, \qquad 0 < \bar{q} < 1.$$

Proofs

Pack it. I am going home! Prof. Nedić will do it. I think.