SE/EC/ME 724 Advanced Optimization Theory and Methods

Yannis Paschalidis
yannisp@bu.edu, http://sites.bu.edu/paschalidis/

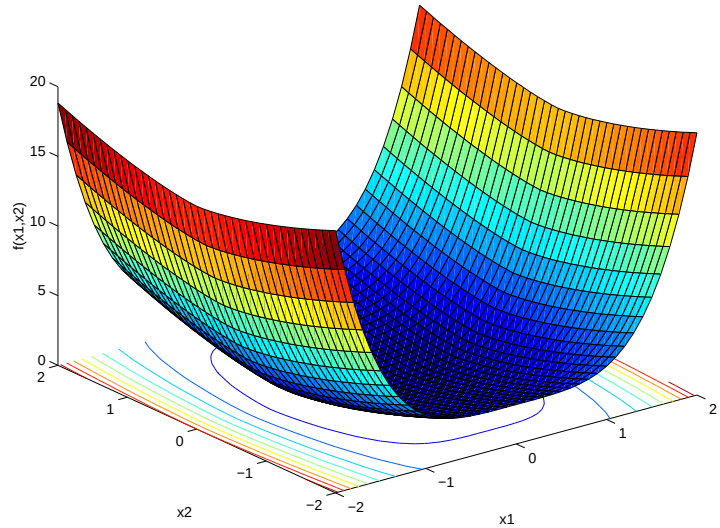
Department of Electrical and Computer Engineering,
Division of Systems Engineering,
Department of Biomedical Engineering
and Center for Information and Systems Engineering,
Boston University

Lecture 6: Outline

1. Rate of convergence for nonquadratic problems, singular problems.
2. Heavy ball and Nesterov's method.
3. Newton's methods.
4. Least squares.
5. Gradient methods with errors.

A singular problem

$$f(x_1, x_2) = |x_1|^4 + |x_2|^{3/2}$$



Heavy Ball method

Using extrapolation:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \nabla f(\mathbf{x}^k) + \beta^k(\mathbf{x}^k - \mathbf{x}^{k-1}).$$

If α^k, β^k are constant, this is the heavy ball method.

Nesterov's method

Extrapolation step:

$$\mathbf{y}^k = \mathbf{x}^k + \beta^k(\mathbf{x}^k - \mathbf{x}^{k-1}).$$

with $\beta^k \to 1$.
Gradient step:

$$\mathbf{x}^{k+1} = \mathbf{y}^k - \alpha \nabla f(\mathbf{y}^k).$$

Newton's method

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$. We seek to solve the system of equations

$$\mathbf{g}(\mathbf{x}) = 0.$$

Newton's method:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (\nabla \mathbf{g}(\mathbf{x}^k)')^{-1}\mathbf{g}(\mathbf{x}^k).$$

Newton's method (cont.)

Theorem

Let $\mathbf{x}^*$ s.t. $\mathbf{g}(\mathbf{x}^*) = 0$ and assume that $\mathbf{g}$ is continuously differentiable in some $\mathscr{S}_\delta = \{\mathbf{x} \mid ||\mathbf{x} - \mathbf{x}^*|| \leq \delta\}$ and $\nabla \mathbf{g}(\mathbf{x}^*)$ invertible.
(a) $\exists \delta > 0$ s.t. if $\mathbf{x}^0 \in \mathscr{S}_\delta$ then $\{\mathbf{x}^k\}$ by Newton's method is in $\mathscr{S}_\delta$ and converges to $\mathbf{x}^*$. Further, $\{||\mathbf{x}^k - \mathbf{x}^*||\}$ converges superlinearly.
(b) If for some $L, M, \delta > 0$ and for all $\mathbf{x}, \mathbf{y} \in \mathscr{S}_\delta$

$$||\nabla \mathbf{g}(\mathbf{x}) - \nabla \mathbf{g}(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||, \quad ||(\nabla \mathbf{g}(\mathbf{x})')^{-1}|| \leq M$$

then if $\mathbf{x}^0 \in \mathscr{S}_\delta$ it follows

$$||\mathbf{x}^{k+1} - \mathbf{x}^*|| \leq \frac{LM}{2}||\mathbf{x}^k - \mathbf{x}^*||^2 \qquad \forall k,$$

so $\{||\mathbf{x}^k - \mathbf{x}^*||\}$ converges superlinearly with order at least 2.

Least squares problems

Let $\mathbf{g} = (\mathbf{g}_1, \ldots, \mathbf{g}_m)$, where $\mathbf{g}_i : \mathbb{R}^n \to \mathbb{R}^{r_i}$

$$\begin{array}{ll} \min & f(\mathbf{x}) = \frac{1}{2}||\mathbf{g}(\mathbf{x})||^2 = \frac{1}{2}\sum_{i=1}^m ||\mathbf{g}_i(\mathbf{x})||^2 \\ \text{s.t.} & \mathbf{x} \in \mathbb{R}^n. \end{array}$$

Note that

$$\nabla f(\mathbf{x}) = \nabla g(\mathbf{x}) g(\mathbf{x}).$$

Gauss-Newton method for least squares:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \left(\nabla g(\mathbf{x}^k) \nabla g(\mathbf{x}^k)'\right)^{-1} \nabla g(\mathbf{x}^k) g(\mathbf{x}^k).$$