SE524/EC524 Optimization Theory and Methods

Yannis Paschalidis
yannisp@bu.edu, http://ionia.bu.edu/

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

Lecture 15: Outline

1. Introduction to Nonlinear Programming (NLP).
2. Some NLP formulations.
3. Unconstrained optimization.
4. Gradient methods.
5. Stepsize selection.

Some background material for NLP

- Norms $||\cdot||$ on $\mathbb{R}^n$.
- Euclidean norm: $||\mathbf{x}|| = \sqrt{\mathbf{x}'\mathbf{x}}$.
- Open Ball around $\mathbf{a}$ with radius r: $\{\mathbf{y} \mid ||\mathbf{y}-\mathbf{a}|| < r\}$.
- $\mathscr{A} \subset \mathbb{R}^n$ is compact **iff** closed and bounded (Heine-Borel).
- Consider function $f : \mathscr{A} \to \mathbb{R}^n$:
  - **continuous** at $\mathbf{x} \in \mathscr{A}$ if $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{y}) = f(\mathbf{x})$.
  - **right-continuous** if $\lim_{\mathbf{y}\downarrow\mathbf{x}} f(\mathbf{y}) = f(\mathbf{x})$.
  - **left-continuous** if $\lim_{\mathbf{y}\uparrow\mathbf{x}} f(\mathbf{y}) = f(\mathbf{x})$.
  - **lower-semicontinuous** if $f(\mathbf{x}) \leq \liminf_{k\to\infty} f(\mathbf{x}_k)$ for every sequence $\mathbf{x}_k \to \mathbf{x}$.
  - **upper-semicontinuous** if $f(\mathbf{x}) \geq \limsup_{k\to\infty} f(\mathbf{x}_k)$ for every sequence $\mathbf{x}_k \to \mathbf{x}$.
  - **coercive** if $\lim_{k\to\infty} f(\mathbf{x}_k) = \infty$ for every sequence satisfying $||\mathbf{x}_k|| \to \infty$.

Some background material for NLP (cont.)

Theorem

(**Weierstrass**) Let $\mathscr{A} \subset \mathbb{R}^n$ where $\mathscr{A}$ is closed and non-empty. Let $f : \mathscr{A} \to \mathbb{R}^n$ be lower-semicontinuous for all $\mathbf{x} \in \mathscr{A}$.

(a) If $\mathscr{A}$ is compact then $\exists \mathbf{x} \in \mathscr{A}$ s.t. $f(\mathbf{x}) = \inf_{\mathbf{z}\in\mathscr{A}} f(\mathbf{z})$.

(b) If f is coercive, then $\exists \mathbf{x} \in \mathscr{A}$ s.t. $f(\mathbf{x}) = \inf_{\mathbf{z}\in\mathscr{A}} f(\mathbf{z})$.

- Gradient:
  - $f : \mathbb{R}^n \to \mathbb{R} \Rightarrow \nabla f(\mathbf{x}) = (\frac{\vartheta f(\mathbf{x})}{\vartheta x_1}, \ldots, \frac{\vartheta f(\mathbf{x})}{\vartheta x_n})$.
  - $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m \Rightarrow \nabla f(\mathbf{x}) = [\nabla f_1(\mathbf{x}) \cdots \nabla f_m(\mathbf{x})]$.
- Hessian:
  - $f : \mathbb{R}^n \to \mathbb{R} \Rightarrow \nabla^2 f(\mathbf{x}) = \nabla(\nabla f(\mathbf{x})) = \left(\frac{\vartheta^2 f(\mathbf{x})}{\vartheta x_i \vartheta x_j}\right)$.
- Taylor expansion:

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + \mathbf{y}'\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{y}'\nabla^2 f(\mathbf{x})\mathbf{y} + o(||\mathbf{y}||^2).$$

Formulation and definitions

- Unconstrained optimization problem:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \mathbb{R}^n \end{array}$$

- $\mathbf{x}^*$ is a **local minimum** if $\exists \epsilon > 0$ s.t. $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $\forall \mathbf{x}$ with $||\mathbf{x} - \mathbf{x}^*|| < \epsilon$.
- $\mathbf{x}^*$ is a **global minimum** if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $\forall \mathbf{x} \in \mathbb{R}^n$.

Necessary Conditions

Proposition

Let $\mathbf{x}^*$ be an unconstrained local min and $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable in an open set $\mathscr{S}$ containing $\mathbf{x}^*$. Then

$$\nabla f(\mathbf{x}^*) = \mathbf{0}. \qquad (\textbf{1st order})$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable within $\mathscr{S}$ then

$$\nabla^2 f(\mathbf{x}^*) \succeq 0. \qquad (\textbf{2nd order})$$

Convexity

Proposition

Let $f : \mathscr{C} \to \mathbb{R}$ be convex over the convex set $\mathscr{C} \subset \mathbb{R}^n$.

- (a) A local min of f over $\mathscr{C}$ is also a global min over $\mathscr{C}$. If f is strictly convex, $\exists$ at most one global min.
- (b) If f is convex and $\mathscr{C}$ is open, the condition

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

is **necessary and sufficient** for $\mathbf{x}^* \in \mathscr{C}$ to be a global min of f over $\mathscr{C}$.

Sufficient Conditions

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open set $\mathscr{S} \subset \mathbb{R}^n$. Let also $\mathbf{x}^* \in \mathscr{S}$ s.t.

$$\nabla f(\mathbf{x}^*) = \mathbf{0},$$

$$\nabla^2 f(\mathbf{x}^*) \succ 0.$$

Then $\mathbf{x}^*$ is a **strict** unconstrained local min of f, that is, $\exists \gamma, \epsilon > 0$ s.t.

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \frac{\gamma}{2}||\mathbf{x} - \mathbf{x}^*||^2, \qquad \forall \mathbf{x} \text{ with } ||\mathbf{x} - \mathbf{x}^*|| < \epsilon.$$

Gradient Methods

- Generic gradient method:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$$

  such that if $\nabla f(\mathbf{x}^k) \neq \mathbf{0}$ then $\mathbf{d}^k$ is chosen so that $\nabla f(\mathbf{x}^k)'\mathbf{d}^k < 0$ (**descent direction**).

- An interesting class of gradient methods:

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

  - **Steepest descent**: $\mathbf{D}^k = \mathbf{I}$.
  - **Newton's method**: $\mathbf{D}^k = (\nabla^2 f(\mathbf{x}^k))^{-1}$.
  - **Diagonally scaled steepest descent**: $\mathbf{D}^k = \text{diag}\left(\left(\frac{\vartheta^2 f(\mathbf{x}^k)}{(\vartheta x_1)^2}\right)^{-1}, \ldots, \left(\frac{\vartheta^2 f(\mathbf{x}^k)}{(\vartheta x_n)^2}\right)^{-1}\right)$.
  - **Modified Newton's method**: $\mathbf{D}^k = (\nabla^2 f(\mathbf{x}^0))^{-1}$.

Least squares problems

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

Note that

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

Gauss-Newton method for least squares:

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

Stesize Selection

- **Minimization rule**:

$$f(\mathbf{x}^k + \alpha^k \mathbf{d}^k) = \min_{\alpha \geq 0} f(\mathbf{x}^k + \alpha \mathbf{d}^k).$$

- **Limited minimization rule**:

$$f(\mathbf{x}^k + \alpha^k \mathbf{d}^k) = \min_{\alpha \in [0,s]} f(\mathbf{x}^k + \alpha \mathbf{d}^k).$$

- **Constant stepsize**: $\alpha^k = s$.
- **Diminishing stepsize**:

$$\alpha^k \to 0 \qquad \text{with} \qquad \sum_{k=0}^{\infty} \alpha^k = \infty.$$