# SE/EC 724 Advanced Optimization Theory and Methods

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

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

## Lecture 5: Outline

1. Convergence.
2. Rate of convergence.



## Gradient related condition

Consider the gradient method

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \qquad \text{where } \nabla f(\mathbf{x}^k)'\mathbf{d}^k < 0.$$

**Definition**

We say that $\mathbf{d}^k$ is **gradient related** to $\{\mathbf{x}^k\}$ if for every subsequence $\{\mathbf{x}^k\}_{k\in\mathscr{K}}$ converging to a non-stationary point, the corresponding subsequence $\{\mathbf{d}^k\}_{k\in\mathscr{K}}$ is bounded and satisfies

$$\lim_{k\to\infty, k\in\mathscr{K}} \sup \nabla f(\mathbf{x}^k)'\mathbf{d}^k < 0.$$

**Typical convergence result**: Every limit point of $\{\mathbf{x}^k\}$ is a stationary point.



## Convergence with a constant stepsize

**Theorem**

Let $\mathbf{x}^k$ be generated by the gradient method $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha\mathbf{d}^k$, where for all $k$ and some $c_1, c_2 > 0$, $\mathbf{d}^k$ satisfies

$$c_1||\nabla f(\mathbf{x}^k)||^2 \leq -\nabla f(\mathbf{x}^k)'\mathbf{d}^k, \quad ||\mathbf{d}^k|| \leq c_2||\nabla f(\mathbf{x}^k)||.$$

Furthermore, for some $L > 0$

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||, \qquad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Assume $0 < \alpha < \frac{2c_1}{Lc_2^2}$. Then either $f(\mathbf{x}^k) \to -\infty$ or else $f(\mathbf{x}^k)$ converges to a finite value and $\lim_{k\to\infty} \nabla f(\mathbf{x}^k) = 0$. Furthermore, every limit point of $\{\mathbf{x}^k\}$ is a stationary point of $f(\cdot)$.

## Convergence with a diminishing stepsize

**Theorem**

Let $\mathbf{x}^k$ be generated by the gradient method $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$, where for all $k$ and some $c_1, c_2 > 0$, $\mathbf{d}^k$ satisfies

$$c_1 ||\nabla f(\mathbf{x}^k)||^2 \leq -\nabla f(\mathbf{x}^k)' \mathbf{d}^k, \quad ||\mathbf{d}^k|| \leq c_2 ||\nabla f(\mathbf{x}^k)||.$$

Furthermore, for some $L > 0$

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L ||\mathbf{x} - \mathbf{y}||, \qquad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Assume $\alpha^k \to 0$ and $\sum_k \alpha^k = \infty$. Then either $f(\mathbf{x}^k) \to -\infty$ or else $f(\mathbf{x}^k)$ converges to a finite value and $\lim_{k\to\infty} \nabla f(\mathbf{x}^k) = 0$. Furthermore, every limit point of $\{\mathbf{x}^k\}$ is a stationary point of $f(\cdot)$.



## Rate of convergence

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x}$ where $\mathbf{Q} \succ 0$ and consider the steepest-descent method

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

Then, for all $k$

$$\frac{||\mathbf{x}^{k+1}||}{||\mathbf{x}^k||} \leq \frac{M - m}{M + m},$$

where $M$ and $m$ are the largest and smallest eigenvalues of $\mathbf{Q}$, respectively. Furthermore, if $\alpha^k$ is selected according to the minimization rule , for all $k$

$$f(\mathbf{x}^{k+1}) \leq \left(\frac{M - m}{M + m}\right)^2 f(\mathbf{x}^k).$$

The quantity $M/m$ is known as the **condition number**.