MATH 4301/558
FALL 2015

LECTURE 2

LINE SEARCH METHODS

$$x^{(k+1)} = x^{(k)} + \underbrace{\alpha_k}_{\substack{\text{step}\\ \text{length}}} \; \underbrace{p_k}_{\substack{\text{search}\\ \text{direction}}}$$

Search Direction

(Pure) Newton direction : $\quad p_k = -\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$

BFGS direction : $\quad p_k = -B_k^{-1} \nabla f(x^{(k)})$

It is essential that p_k is a descent direction, that is

$$\phi'(0) = \nabla f(x^{(k)})^T p_k < 0$$

where

$$\phi : \mathbb{R} \to \mathbb{R}, \quad \phi(\alpha) := f(x^{(k)} + \alpha p_k).$$

Newton direction is descent if $\nabla^2 f(x^{(k)})$ is PD, i.e.,

$$\nabla^2 f(x^{(k)}) \text{ is PD} \iff \left[\nabla^2 f(x^{(k)})\right]^{-1} \text{ is PD}$$

$$\Longrightarrow \nabla f(x^{(k)})^T p_k = - \nabla f(x^{(k)})^T \left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)}) < 0$$

Similarly BFGS direction is descent if B_k is PD.

Wolfe Conditions (Selection of Step Length)

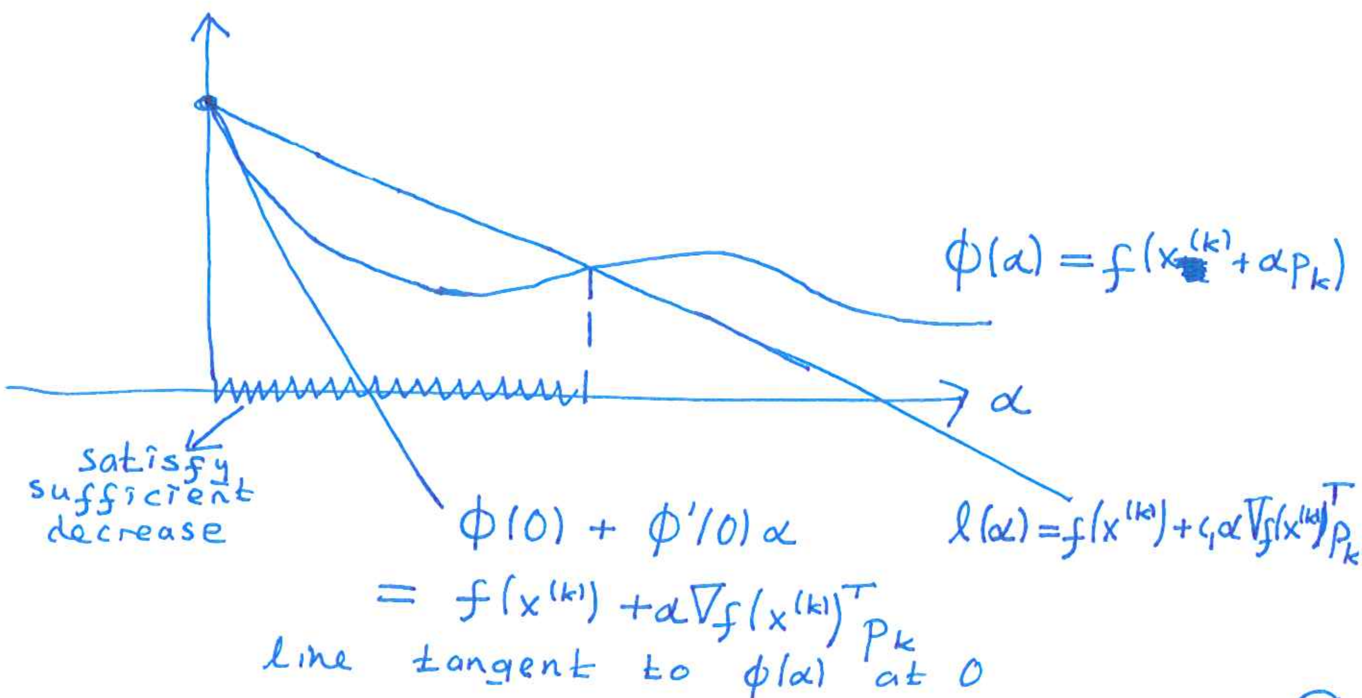
Choose α_k so that

sufficient decrease $\leftarrow$ $f(x^{(k)} + \alpha_k p_k) \leq f(x^{(k)}) + c_1 \alpha_k \nabla f(x^{(k)})^T p_k$

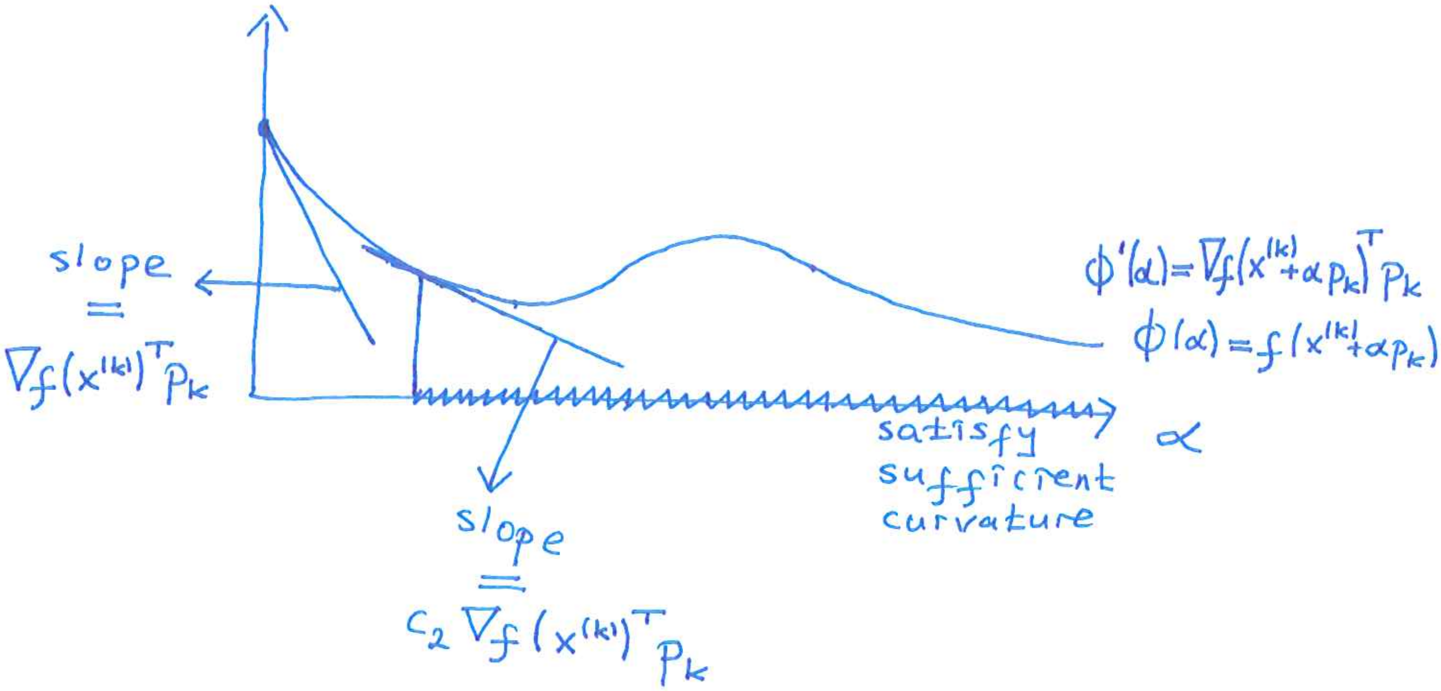
sufficient curvature $\leftarrow$ $\left[\nabla f(x^{(k)} + \alpha_k p_k)\right]^T p_k \geq c_2 \nabla f(x^{(k)})^T p_k$

c_1, c_2 are given parameters, $0 < c_1 < c_2 < 1$

Sufficient decrease



Sufficient curvature



THM

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is bounded below, twice continuously differentiable, and p_k is a descent direction. There exist step-lengths that satisfy Wolfe conditions.

PROOF

For all $\alpha > 0$ small enough

$$f(x^{(k)} + \alpha p_k) = f(x^{(k)}) + \alpha \nabla f(x^{(k)})^T p_k + O(\alpha^2)$$

$$\leq f(x^{(k)}) + \alpha c_1 \nabla f(x^{(k)})^T p_k := \ell(\alpha)$$

Furthermore, $\phi(\alpha)$ is bounded below, whereas $\ell(\alpha) \to -\infty$ as $\alpha \to \infty$ meaning

$$\phi(\alpha) \geq \ell(\alpha)$$

for α large enough. By continuity there exists $\alpha > 0$ such that $\phi(\alpha) = \ell(\alpha)$. Let $\bar{\alpha}$ be the smallest such α.

Sufficient decrease holds on $[0, \underline{\alpha}]$.
By the mean value thm there exists
$\underline{\underline{\alpha}} \in (0, \underline{\alpha})$ such that

$$c_1 \nabla f(x^{(k)})^T p_k = \frac{\phi(\underline{\alpha}) - \phi(0)}{\underline{\alpha}} = \phi'(\underline{\underline{\alpha}})$$
$$= \nabla f(x^{(k)} + \underline{\underline{\alpha}}\, p_k)^T p_k$$

$$\underset{\substack{\text{Recall} \\ c_1 < c_2 \\ \nabla f(x^{(k)})^T p_k}}{\Longrightarrow} \quad \nabla f(x^{(k)} + \underline{\underline{\alpha}}\, p_k)^T p_k > c_2 \nabla f(x^{(k)})^T p_k .$$

By continuity of $\nabla f(x)$ there exists an interval $I \subseteq [0, \underline{\alpha}]$ containing $\underline{\underline{\alpha}}$ such that

$$c_2 \nabla f(x^{(k)})^T p_k < \nabla f(x^{(k)} + \alpha p_k)^T p_k \quad \forall \alpha \in I.$$

$\square$

Strong Wolfe Conditions

$$f(x^{(k)} + \alpha_k p_k) \leq f(x^{(k)}) + c_1 \alpha_k \nabla f(x^{(k)})^T p_k$$
$$\left| \left[\nabla f(x^{(k)} + \alpha_k p_k) \right]^T p_k \right| \leq c_2 \left| \nabla f(x^{(k)})^T p_k \right|$$

Global Convergence

THM (Zoutendijk)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is bounded below and has Lipschitz continuous gradients, that is there exists $L > 0$ such that

$$\| \nabla f(x) - \nabla f(y) \|_2 \leq L \, \| x - y \|_2 \, . \ \forall x, y .$$

A line search method with descent directions and step-lengths satisfying Wolfe conditions would generate $\{ x^{(k)} \}$ such that

$$\lim_{k \to \infty} \cos^2 \theta_k \, \| \nabla f(x^{(k)}) \|_2^2 = 0$$

where

$$\theta_k := \arccos \left(\frac{\nabla f(x^{(k)})^T p_k}{\| \nabla f(x^{(k)}) \|_2 \, \| p_k \|_2} \right) .$$

PROOF

Sufficient curvature condition and Lipschitz continuity of $\nabla f(x)$ would imply

$$\left[\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) \right]^T p_k \geq (c_2 - 1) \, \nabla f(x^{(k)})^T p_k ,$$

$$\left[\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) \right]^T p_k \leq L \, \| p_k \|_2^2 \, \alpha_k$$

$$\Longrightarrow \quad \alpha_k \geq \frac{(c_2 - 1) \, \nabla f(x^{(k)})^T p_k}{L \, \| p_k \|_2^2} .$$

Exploiting the sufficient decrease condition

$$
\begin{aligned}
f(x^{(k+1)}) &\leq f(x^{(k)}) + c_1 \alpha_k \nabla f(x^{(k)})^T p_k \\
&\leq f(x^{(k)}) - \frac{c_1 (1-c_2) \, |\nabla f(x^{(k)})^T p_k|^2}{L \, \|p_k\|_2^2} \\
&= f(x^{(k)}) - \frac{c_1(1-c_2)}{L} \cos^2 \theta_k \, \|\nabla f(x^{(k)})\|_2^2
\end{aligned}
$$

$$
\Longrightarrow \sum_{k=0}^{K} \cos^2 \theta_k \, \|\nabla f(x^{(k)})\|_2^2 \leq \frac{L}{c_1(1-c_2)} \left(f(x^{(0)}) - f(x^{(K)}) \right)
$$

Since f is bounded below, the sum on the left is bounded meaning

$$
\lim_{k \to \infty} \cos^2 \theta_k \, \|\nabla f(x^{(k)})\|_2^2 = 0.
$$

$\square$

Remark

According to Zoutendijk's thm, if there exists $\delta > 0$ such that

$$
|\cos \theta_k| \geq \delta \quad \text{for all } k,
$$

then

$$
\lim_{k \to \infty} \nabla f(x^{(k)}) = 0.
$$

Newton's method

suppose

* $\nabla^2 f(x^{(k)})$ is PD for $k = 0, 1, 2, \dots$
  (this also ensures p_k is a descent direction)

* there exists M such that
$$\| \nabla^2 f(x^{(k)}) \|_2 \; \| [\nabla^2 f(x^{(k)})]^{-1} \|_2 \leq M$$
  for $k = 0, 1, 2, \dots$, equivalently
$$\frac{\lambda_{max}(\nabla^2 f(x^{(k)}))}{\lambda_{min}(\nabla^2 f(x^{(k)}))} \leq M$$
  for $k = 0, 1, 2, \dots$

It follows that

$$|\cos \theta_k| = \frac{|\nabla f(x^{(k)})^T p_k|}{\|\nabla f(x^{(k)})\|_2 \; \|p_k\|_2}$$

$$= \frac{+|\nabla f(x^{(k)})^T [\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})|}{\|\nabla f(x^{(k)})\|_2 \; \| [\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)}) \|_2}$$

$$\geq \frac{\lambda_{min}([\nabla^2 f(x^{(k)})]^{-1}) \; \|\nabla f(x^{(k)})\|_2^2}{\lambda_{max}([\nabla^2 f(x^{(k)})]^{-1}) \; \|\nabla f(x^{(k)})\|_2^2}$$

$$= \frac{\lambda_{min}(\nabla^2 f(x^{(k)}))}{\lambda_{max}(\nabla^2 f(x^{(k)}))} \geq \frac{1}{M}.$$

Consequently,

$$\lim_{k \to \infty} \nabla f(x^{(k)}) = 0.$$

Similar holds for BFGS assuming B_k are PD and $\|B_k\|_2 \, \|B_k^{-1}\|_2$ remain bounded.