# LECTURE 10

## INEXACT LINE SEARCH

To define sufficient decrease condition we again use the linear approximation

$$\ell(\alpha) = \phi(0) + \alpha \phi'(0)$$
$$= f(x_k) + \alpha \nabla f(x_k)^T p_k$$

where

$$\phi(\alpha) = f(x_k + \alpha p_k).$$

**THM:** Suppose $\phi'(0) \neq 0$. Then

$$\lim_{\alpha \to 0} \frac{\phi(0) - \phi(\alpha)}{\ell(0) - \ell(\alpha)} = 1$$

**PROOF**

By Taylor's thm

$$\phi(\alpha) = \phi(0) + \alpha \phi'(0) + \frac{\alpha^2}{2} \phi''(t)$$

where $t \in (0, \alpha)$.

~~But then~~

$$\cancel{\lim \;\; -\alpha\phi'(0) - \alpha^2/2}$$

But then

$$\lim_{\alpha \to 0} \frac{\phi(0) - \phi(\alpha)}{l(0) - l(\alpha)}$$

$$=$$

$$\lim_{\alpha \to 0} \frac{\cancel{-\alpha}\, \phi'(0) - \overset{-\alpha/2}{(\cancel{\alpha^2/2})}\, \phi''(t)}{\cancel{-\alpha}\, \phi'(0)} = 1$$

$\square$

Therefore for any $\mu \in (0,1)$

$$\frac{\phi(0) - \phi(\alpha)}{l(0) - l(\alpha)} \geq \mu$$

for all $\alpha > 0$ sufficiently close to $0$.

Equivalently for all $\alpha > 0$ close to $0$

$$\phi(0) - \phi(\alpha) \geq \mu\,(l(0) - l(\alpha))$$

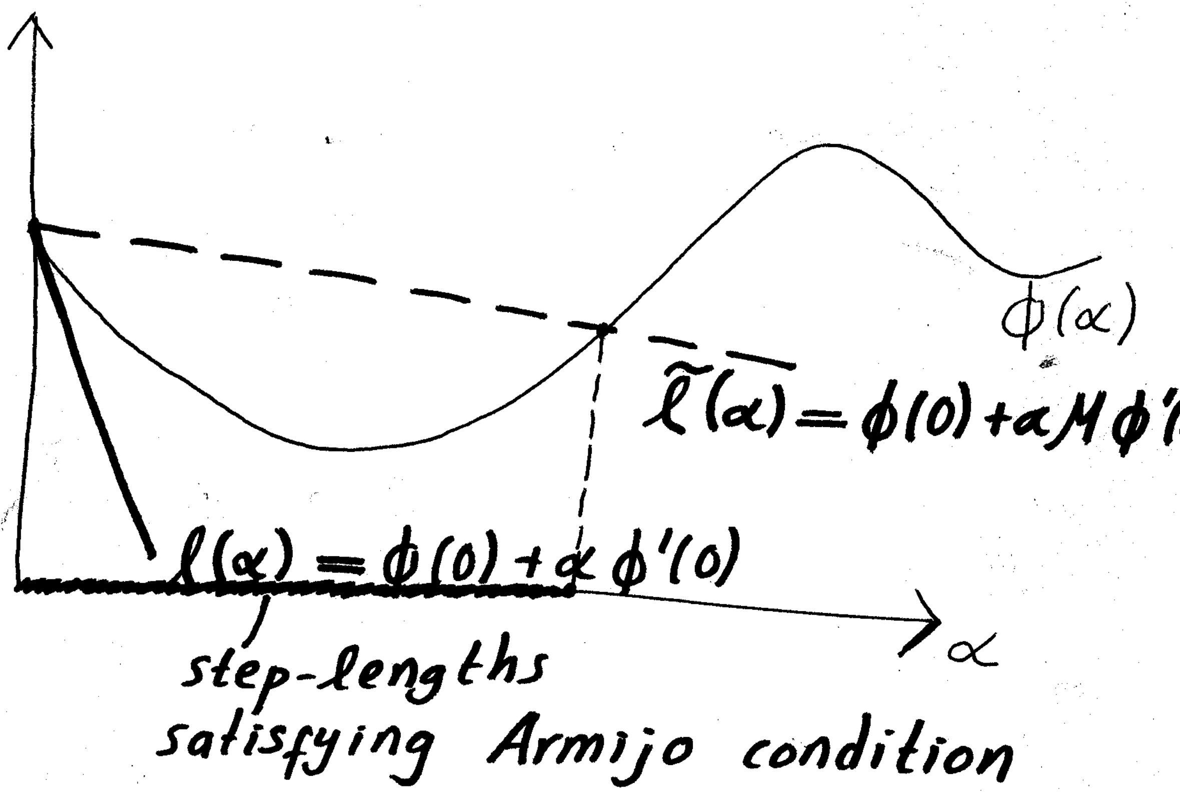
$$\Longrightarrow \phi(0) - \phi(\alpha) \geq \mu(-\alpha\,\phi'(0))$$

$$\Longrightarrow \underbrace{\phi(0) + \mu\alpha\,\phi'(0)}_{\tilde{l}(\alpha)} \geq \phi(\alpha)$$

$$\Longrightarrow f(x_k) + \mu\alpha\, \nabla f(x_k)^T p_k \geq f(x_k + \alpha p_k)$$

## ARMIJO SUFFICIENT DECREASE CONDITION

$$f(x_k) + \alpha \mu \nabla f(x_k)^T p_k \geq f(x_k + \alpha p_k)$$




In Goldstein-Armijo backtracking line search we start with $\alpha = 1$ and halve $\alpha$ until Armijo condition is satisfied.

<u>ALGORITHM</u> (Goldstein-Armijo backtracking line search)

Given $x_k, p_k \in \mathbb{R}^n$, $\alpha_k = 1$

While $f(x_k + \alpha_k p_k) > f(x_k) + \alpha_k \mu \nabla f(x_k)^T p_k$

$\qquad \alpha_k = \alpha_k / 2$

end

return $\alpha_k$

# GENERATION OF A DESCENT DIRECTION

## Basic idea

* Represent $f(x)$ by a quadratic approximation $Q(x)$ about $x_k$.

* Find the local minimizer $x_*$ of $Q(x)$.

* Define $p_k := x_* - x_k$

## Quadratic Approximation

By Taylor's thm

$$f(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k + t(x - x_k))(x - x_k)$$

for some $t \in (0,1)$.

Approximate $f$ by

$$Q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T B_k (x - x_k)$$

where $B_k$ is symmetric.

Assume $B_k \succ 0$. Local minimizer $x_*$ of $Q(x)$ satisfies

$$0 = \nabla Q(x_*) = \nabla f(x_k) + B_k(x_* - x_k)$$

$$\Longrightarrow$$

$$B_k (x_* - x_k) = -\nabla f(x_k)$$

Search direction $p_k$ satisfies

$$\boxed{B_k p_k = -\nabla f(x_k)}$$

equivalently

$$\boxed{p_k = -B_k^{-1} \nabla f(x_k)}$$

## STEEPEST DESCENT

Choose $B_k = I$, i.e.

$$Q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T (x - x_k$$

Then the steepest descent direction

$$p_k^{SD} = -\nabla f(x_k)$$

## NEWTON'S METHOD

Choose $B_k = \nabla^2 f(x_k)$, i.e.

$$Q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

Then the Newton direction satisfies

$$\nabla^2 f(x_k)\, p_k^N = -\nabla f(x_k)$$

equivalently

$$p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

## EXAMPLE

$$f(x) = x_2 e^{x_1} \quad \text{at} \quad x_k = (0, 0)$$

Gradient

$$\nabla f(x) = \begin{bmatrix} x_2 e^{x_1} \\ e^{x_1} \end{bmatrix}, \quad \nabla f(x_k) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Hessian

$$\nabla^2 f(x) = \begin{bmatrix} x_2 e^{x_1} & e^{x_1} \\ e^{x_1} & 0 \end{bmatrix}, \quad \nabla^2 f(x_k) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Steepest descent direction

$$p_k^{SD} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Newton direction satisfies

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} p_k^N = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\Longrightarrow$$

$$p_k^N = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Steepest descent direction is a descent direction

$$\nabla f(x_k)^T p_k^{SD} = -1 < 0$$

Newton direction is not a descent direction

$$\nabla f(x_k)^T p_k^N = 0$$

## THM

Suppose $B_k \succ 0$. Then the direction $p_k$ such that $\nabla Q(p_k) = 0$ is a descent direction.

## PROOF

Notice that

$$B_k \succ 0 \iff B_k^{-1} \succ 0 \quad \left(\text{see remarks below}\right)$$

Consequently

$$\nabla f(x_k)^T p_k = -\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) < 0$$

$\square$

## REMARK

* $Ax = \lambda x \iff A^{-1} x = (1/\lambda) x$
* $\lambda$ is an eigenvalue of $A$ $\iff$ $(1/\lambda)$ is an eigenvalue of $A^{-1}$

(i.e. $AA^{-1} = I$)

e.g.

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

eigvals $\lambda_1 = 3$, $\lambda_2 = 1$ eigvals $\lambda_1 = 1/3$, $\lambda_2 = 1$

* Consequently

$$A \succ 0 \iff A^{-1} \succ 0$$

Both $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $A^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$

are positive definite.