LECTURE 9

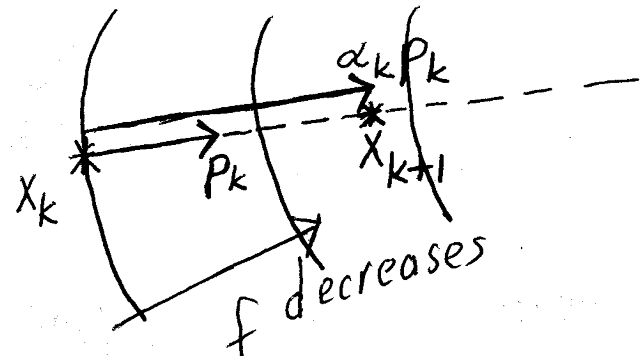
LINE-SEARCH ALGORITHMS (Gill & Wright 3.4-5)

Generates a sequence $\{x_k\}$ such that

$$x_{k+1} = x_k + \alpha_k p_k$$

(NEXT LECT GW 3.4) $p_k \in \mathbb{R}^n$: descent search direction

(TODAY GW 3.5) $\alpha_k > 0$: step-length

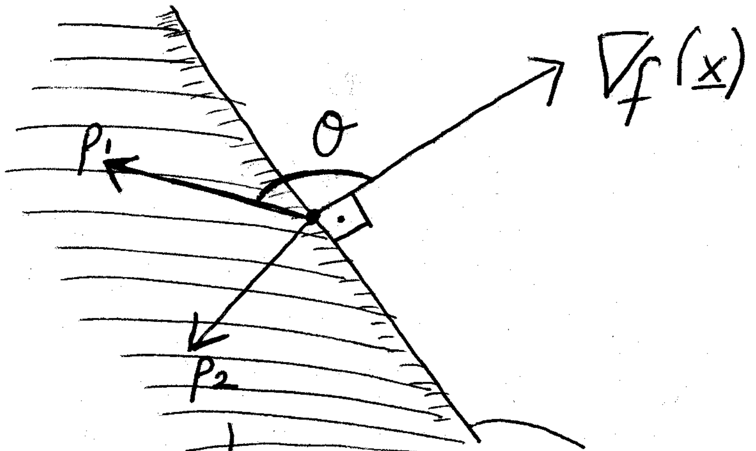


DEFN (Descent direction)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and $\underline{x}, p \in \mathbb{R}^n$. The vector p is called a ~~direction~~ descent direction ~~of decrease~~ for f at $\underline{x}$ if

$$\nabla f(\underline{x})^T p < 0.$$

$$\nabla f(\underline{x})^T p = \|\nabla f(\underline{x})\| \, \|p\| \cos\theta$$

θ: angle between $\nabla f(\underline{x})$ and p.



$\{p \in \mathbb{R}^n : \nabla f(\underline{x})^T p = 0\}$

HYPERPLANE OF ORTHOGONAL VECTORS

HALF-SPACE OF DESCENT DIRECTIONS

$\{p \in \mathbb{R}^n : \nabla f(\underline{x})^T p < 0\}$

e.g. $p_1, p_2 \in \mathbb{R}^n$ are descent directions.

REMARK

p is a descent direction at $\underline{x}$

$$\Longleftrightarrow$$

$$\nabla f(\underline{x})^T p < 0$$

$$\Longleftrightarrow$$

$$\theta > \pi/2$$

EXAMPLE

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

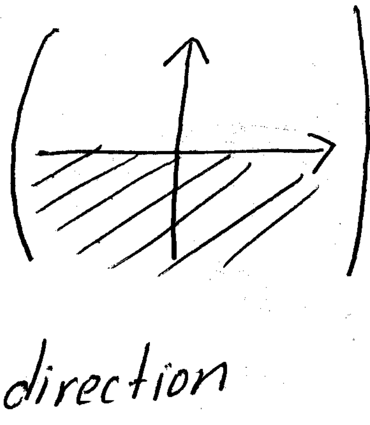
$$\nabla_f(x) = \begin{bmatrix} x_2 e^{x_1} \\ e^{x_1} \end{bmatrix}, \quad \nabla_f(\underline{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Halfspace of descent directions at $\underline{x}$

$$\left\{ \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2 : \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} < 0 \right\}$$

$$=$$

$$\left\{ \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2 : b < 0 \right\}$$



Suppose $p \in \mathbb{R}^n$ is a descent direction at $\underline{x}$. By Taylor's thm

$$f(\underline{x} + \alpha p) = f(\underline{x}) + \nabla_f(\underline{x})^T (\alpha p) + \frac{1}{2} (\alpha p)^T \nabla_f^2 (\underline{x} + tp)(\alpha p$$

$$\Longrightarrow$$

$$f(\underline{x} + \alpha p) - f(\underline{x}) = \alpha \Big(\underbrace{\nabla_f(\underline{x})^T p}_{<0} + \underbrace{\frac{1}{2} \alpha p^T \nabla_f^2 (\underline{x} + t}_{\text{as small as you like as } \alpha \to}$$

Consequently for all sufficiently small $\alpha > 0$

$$f(\underline{x} + \alpha p) - f(\underline{x}) < 0.$$

DEFN (Direction of decrease)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\underline{x}, p \in \mathbb{R}^n$. The vector $p \in \mathbb{R}^n$ is called a direction of decrease at $\underline{x}$ for f, if there exists a $\sigma > 0$ such that

$$f(\underline{x} + \alpha p) < f(\underline{x}) \text{ for all } \alpha \in (0, \sigma$$

REMARK

$p \in \mathbb{R}^n$ is a descent direction at $\underline{x}$

$$\Longrightarrow$$

$p \in \mathbb{R}^n$ is a direction of decrease at $\underline{x}$

ALGORITHM (Generic Line Search)

Given $x_0 \in \mathbb{R}^n$, $k = 0$

While $\|\nabla f(x_k)\| > \epsilon$

1. Choose a descent direction p_k at x_k
2. Choose a step-length α_k such that $f(x_k + \alpha_k p_k)$ is sufficiently smaller than $f(x_k)$

$x_{k+1} = x_k + \alpha_k p_k$, $k = k+1$

end

SELECTION OF STEP-LENGTH

What do we mean by sufficient decrease?

Not sufficient to choose α_k such that

$$f(x_k + \alpha_k p_k) < f(x_k)$$

This does not guarantee convergence to a local minimizer as

$$\{f(x_k)\} \text{ is decreasing } \not\Longrightarrow \nabla f(x_k) \to 0 \text{ as } k \to \infty$$

EXAMPLE

$$f(x) = x^2$$

① Choose p_k as the pure Newton step for root finding

$$p_k = -\frac{f(x_k)}{f'(x_k)} = \frac{-x_k^2}{2x_k} = -\frac{x_k}{2}$$

* p_k is a descent direction i.e.

$$\nabla f(x_k)^T p_k = (2x_k)\left(\frac{-x_k}{2}\right) = -x_k^2 < 0$$

(unless $x_k = 0$)

* consequently p_k is a direction of decrease

② Choose $\alpha_k = 2^{-k}$

$$
\begin{aligned}
x_{k+1} &= x_k + \alpha_k p_k \\
&= x_k - \alpha_k \frac{x_k}{2} \\
&= x_k \left(1 - 2^{-k-1}\right)
\end{aligned}
$$

Note $f(x_{k+1}) < f(x_k)$ for all k

Suppose $x_0 = 2$. Then

$$x_{k+1} = \prod_{j=1}^{k+1} \left(1 - 2^{-j}\right) x_0 .$$

Indeed

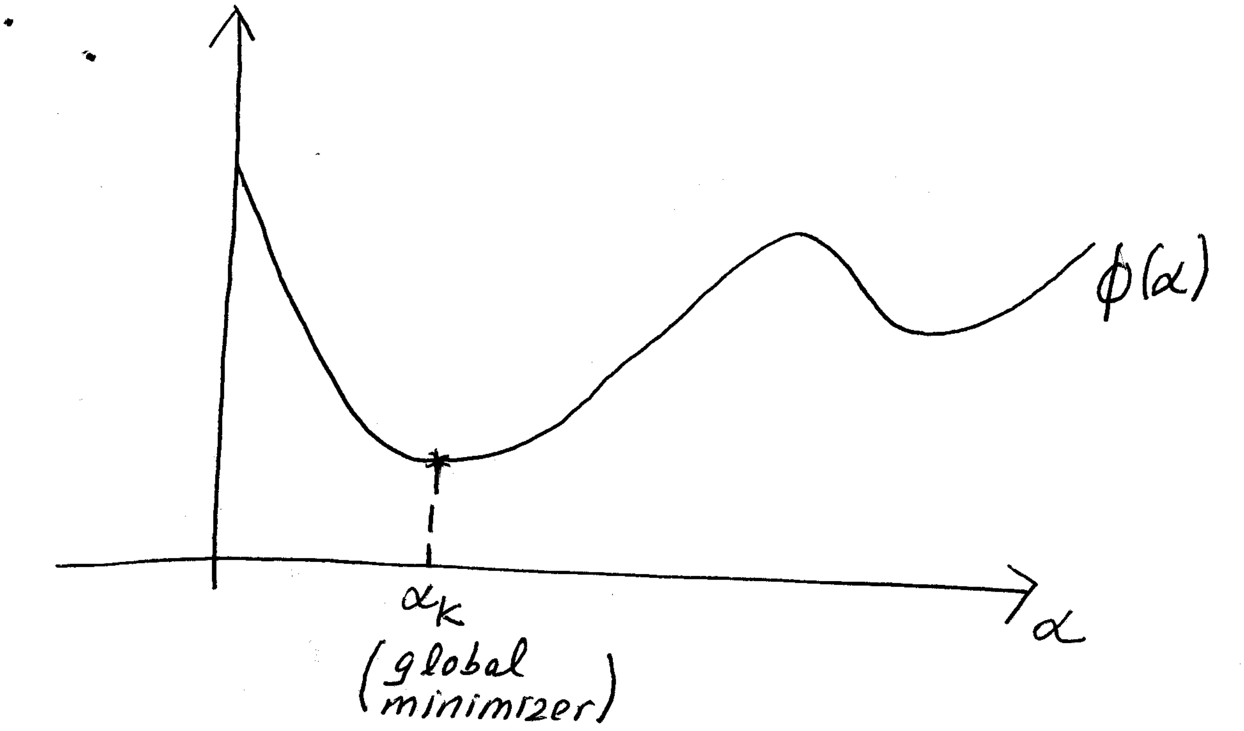
$$\lim_{k \to \infty} x_{k+1} \approx 0.5776 > 0$$

Even though $\{f(x_k)\}$ is a decreasing sequence, $\{x_k\}$ does not converge to the unique local minimizer $x_* = 0$.

EXACT LINE SEARCH

Choose α_k as the global mimizer of

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



In general calculation of a global minimizer of $\phi(\alpha)$ (numerically) is too expensive.

For very special cases there may be an analytic formula.

<u>EXAMPLE</u> (Quadratic Polynomials)

$$q : \mathbb{R}^n \rightarrow \mathbb{R}, \quad q(x) = \frac{1}{2} x^T A x + b^T x + c$$

where $A \succ 0$, and symmetric.

Line search function

$$\phi(\alpha) = \frac{1}{2}(x_k + \alpha p_k)^T A (x_k + \alpha p_k) + b^T (x_k + \alpha p_k) + c$$

Its derivative

$$\phi'(\alpha) = \nabla f(x_k + \alpha p_k)^T p_k = (A(x_k + \alpha p_k) + b)^T p_k$$

- If α_k is a minimizer, then

$$0 = \phi'(\alpha_k) = (A\cancel{x}_k + b)^T p_k + \alpha_k p_k^T A p_k$$

$$\Longrightarrow \boxed{\alpha_k = -\frac{(Ax_k + b)^T p_k}{p_k^T A p_k} = -\frac{\nabla f(x_k)^T p_k}{p_k^T A p_k}}$$