LECTURE 3

TODAY

(1) REVIEW: DERIVATIVE OF A VECTOR-VALUED FUNCTION (Gill & Wright Sec 1.7)

(2) DERIVATION OF OPTIMALITY CONDITIONS FOR UNCONSTRAINED OPTIMIZATION (Gill & Wright Sec 3

DERIVATIVE OF A VECTOR-VALUED FUNCTION

Let $F: \mathbb{R}^3 \to \mathbb{R}^2$ be such that

$$F(x) = \begin{bmatrix} x_1^2 + 2x_2x_3 - x_1x_2 + x_2^2 \\ 3x_1^2 + x_1x_2 + 4x_3^2 \end{bmatrix} \quad \left(\text{where } x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right.$$

Then the Jacobian of F

$$F'(x) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \frac{\partial F_1}{\partial x_3} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \frac{\partial F_2}{\partial x_3} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 - x_2 & 2x_3 - x_1 + 2x_2 & 2x_2 \\ 6x_1 + x_2 & x_1 & 8x_3 \end{bmatrix}$$

(Note:

* $F'(x)$ is a matrix-valued function; $F': \mathbb{R}^3 \to \mathbb{R}^{2\times 3}$
* For each x the Jacobian $F'(x)$ is 2×3

In general let $F: \mathbb{R}^n \to \mathbb{R}^m$ be such that

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix}$$

Jacobian of F (if F is differentiable)

$$F'(x) = \begin{bmatrix} \frac{\partial F_1(x)}{\partial x_1} & \frac{\partial F_1(x)}{\partial x_2} & \cdots & \frac{\partial F_1(x)}{\partial x_n} \\ \frac{\partial F_2(x)}{\partial x_1} & \frac{\partial F_2(x)}{\partial x_2} & \cdots & \frac{\partial F_2(x)}{\partial x_n} \\ \vdots & & & \\ \frac{\partial F_m(x)}{\partial x_1} & \frac{\partial F_m(x)}{\partial x_2} & \cdots & \frac{\partial F_m(x)}{\partial x_n} \end{bmatrix}$$

REMARKS

* $F'(x)$ is $m \times n$.
* $F': \mathbb{R}^n \to \mathbb{R}^{m \times n}$

DEFN (Differentiability)

We say $F: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \mathbb{R}^n$ if there exists an $m \times n$ matrix $F'(x)$ such that

$$\lim_{h \to 0} \frac{1}{\|h\|} \left[(F(x+h) - F(x)) - F'(x)h \right] = 0$$

Scalar-valued functions

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be such that

$$(*) \quad f(x) = x_1^2 + 5x_2x_3 - 2x_2^2 + x_3^2 .$$

Jacobian of f

$$f'(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_3} \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 & 5x_3 - 4x_2 & 5x_2 + 2x_3 \end{bmatrix}$$

Gradient of f (column vector of partial derivatives)

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 5x_3 - 4x_2 \\ 5x_2 + 2x_3 \end{bmatrix} = \left(f'(x)\right)^T$$

REMARK

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then

$$\nabla f(x) = \left(f'(x)\right)^T$$

For example $(*)$ consider the Jacobian of the gradient

$$\left(\nabla f(x)\right)' = \begin{bmatrix} 2x_1 \\ 5x_3 - 4x_2 \\ 5x_2 + 2x_3 \end{bmatrix}' = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -4 & 5 \\ 0 & 5 & 2 \end{bmatrix}$$

Note that

$$\left(\nabla_f(x)\right)' = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3 \partial x_3} \end{bmatrix} = \underbrace{\nabla^2 f(x)}_{\text{Hessian matrix}}$$

REMARK

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then

$$\left(\nabla_f(x)\right)' = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}'$$

$$= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

$$= \nabla^2 f(x) \quad \left(\substack{n \times n \\ \text{Hessian}}\right)$$

EXERCISES

Any quadratic function can be written of the form

$$q(x) = \frac{1}{2} x^T A x + b^T x + c$$

where

* $q : \mathbb{R}^n \to \mathbb{R}$
* $A \in \mathbb{R}^{n \times n}$ * $b \in \mathbb{R}^n$ * $c \in \mathbb{R}$

e.g.

$$f(x) = x_1^2 + 5x_2x_3 - 2x_2^2 + x_3^2 + x_1 + x_2 - 3x_3 + 4$$

$$= \frac{1}{2} x^T \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 5 \\ 0 & 5 & 1 \end{bmatrix}}_{A} x + \underbrace{[1 \;\; 1 \;\; -3]}_{b^T} x +$$

Verify that

(i) $\nabla q(x) = \frac{1}{2}(A + A^T)x + b$ (consequently $q'(x) = \frac{1}{2} x^T(A+A$ $+ b^T$

(ii) $\nabla^2 q(x) = \frac{1}{2}(A + A^T)$

Generalized Chain Rule

Let

$F : \mathbb{R}^n \longrightarrow \mathbb{R}^p$
$G : \mathbb{R}^p \longrightarrow \mathbb{R}^m$ be differentiable.

Consider the composition

$H : \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad H(x) = G(F(x))$

$$\overbrace{\mathbb{R}^n \xrightarrow{F} \mathbb{R}^p \xrightarrow{G} \mathbb{R}^m}^{H}$$

· H is differentiable with the derivative

$$\underbrace{H'(x)}_{m\times n} = \underbrace{G'(F(x))}_{m\times p}\ \underbrace{F'(x)}_{p\times n}$$

EXAMPLE

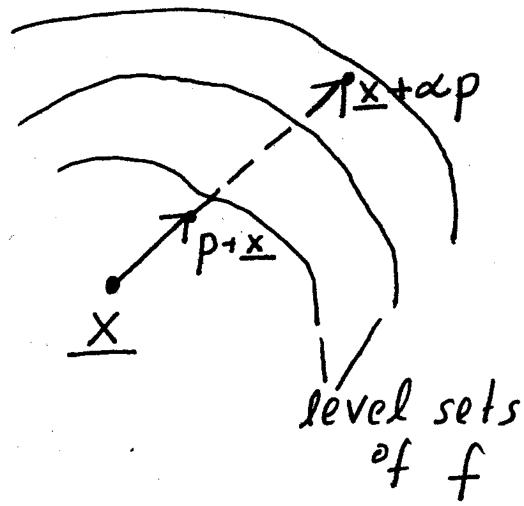
Let $\quad f : \mathbb{R}^n \to \mathbb{R}$

$\qquad \underline{x}, p \in \mathbb{R}^n$

Consider $\phi : \mathbb{R} \to \mathbb{R}$ (line-search function)

$$\phi(\alpha) = f(\underline{x} + \alpha p)$$

In unconstrained optimization



f : function to be minimized

$\underline{x}$: estimate for a local minimizer

p : search direction

α : step-length

e.g. Let

$$f(x) = x_1^2 + x_2^2, \quad \underline{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad p = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Then

$$\phi(\alpha) = f\left(\begin{bmatrix} \alpha \\ 2\alpha \end{bmatrix}\right) = \alpha^2 + (2\alpha)^2$$
$$= 5\alpha^2$$

· Notice that

$$\phi(\alpha) = f(g(\alpha))$$

where

$$f : \mathbb{R}^n \to \mathbb{R}$$

$$g : \mathbb{R} \to \mathbb{R}^n, \quad g(\alpha) = \underbrace{\underset{\sim}{x}}_{\text{fixed}} + \alpha \underbrace{\underset{\sim}{p}}_{\text{fixed}}$$

By the generalized chain rule

$$\phi'(\alpha) = \underbrace{f'(g(\alpha))}_{1\times n} \; \underbrace{g'(\alpha)}_{n\times 1}$$

$$= \nabla f(g(\alpha))^T g'(\alpha)$$

$$= \nabla f(\underline{x} + \alpha p)^T p$$

Note

$$g(\alpha) = \begin{bmatrix} \underline{x}_1 + \alpha p_1 \\ \underline{x}_2 + \alpha p_2 \\ \vdots \\ \underline{x}_n + \alpha p_n \end{bmatrix},$$

$$g'(\alpha) = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = p$$

e.g.

Let

$$f(x) = x_1^2 + x_2^2, \quad \underline{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad p = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Then $\phi'(\alpha) = \dfrac{d f(\underline{x} + \alpha p)}{d\alpha}$

$$= \nabla f(\underline{x} + \alpha p)^T p$$

$$= \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix} \Bigg|_{\substack{x_1 = \alpha \\ x_2 = 2\alpha}} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2\alpha & 4\alpha \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 10\alpha$$

OPTIMALITY CONDITIONS

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth.

OUTLINE FOR FIRST-ORDER NECESSARY CONDITIONS

(1) Suppose $x_* \in \mathbb{R}^n$ is a point such that $\nabla f(x_*) \neq 0$.

(2) The function f decreases at x_* in the direction of $p = -\nabla f(x_*)$

$\left(\text{i.e. } \phi(\alpha) = f(x_* + \alpha p) \text{ for } p = -\nabla f(x_*) \text{ is decreasing at } \alpha = 0 \right)$

(3) x_* cannot be a local minimizer

EXAMPLE

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x) = x_2 e^{x_1}$$

$$\nabla f(x) = \begin{bmatrix} x_2 e^{x_1} \\ e^{x_1} \end{bmatrix} \neq 0$$

f decreases at any x in the direction

$$p = -\nabla f(x) = \begin{bmatrix} -x_2 e^{x_1} \\ -e^{x_1} \end{bmatrix}$$

Since

$$\left.\frac{df(x+\alpha p)}{d\alpha}\right|_{\substack{\alpha=0\\ p=-\nabla f(x)}} = \left.\nabla f(x+\alpha p)^T p\,\right|_{\substack{\alpha=0\\ p=-\nabla f(x)}}$$

$$= \begin{bmatrix} x_2 e^{x_1} & e^{x_1} \end{bmatrix} \begin{bmatrix} -x_2 e^{x_1} \\ -e^{x_1} \end{bmatrix}$$

$$= -\underbrace{x_2^2 e^{2x_1}}_{\geq 0} - \underbrace{e^{2x_1}}_{>0} < 0$$

Consequently x cannot be a local minimizer.

More generally

(1) Suppose $x_* \in \mathbb{R}^n$ is such that

$$\nabla f(x_*) \neq 0$$

(2) Consider the line search function

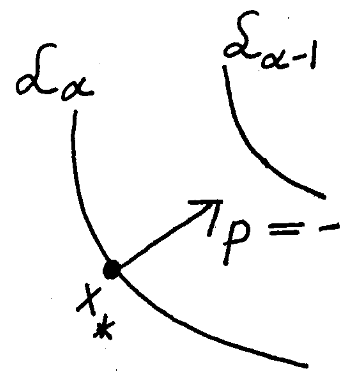
$$\phi(\alpha) = f(x_* + \alpha p)$$

where $p = -\nabla f(x_*)$.

(2b) But then

$$\phi'(\alpha) = \nabla f(x_* + \alpha p)^T p$$

$$\Longrightarrow \phi'(0) = \nabla f(x_*)^T(-\nabla f(x_*)) = -\|\nabla f(x_*)\|_2^2 <$$



$p = -\nabla f(x_*)$ is a direction of decrease

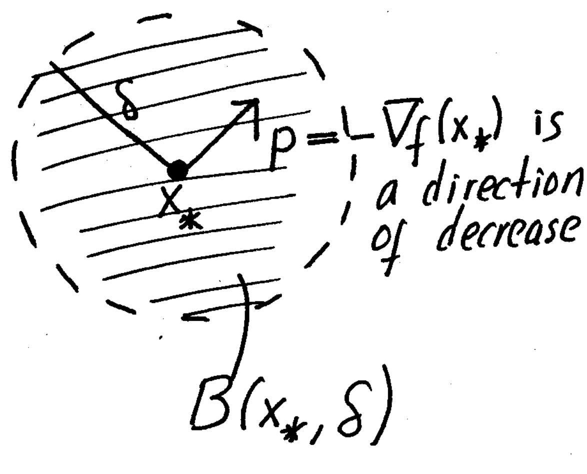
Notice that

$$\phi'(0) < 0 \implies \phi \text{ is decreasing at } \alpha = 0$$

$$\implies \phi(\alpha) < \phi(0) \text{ for all positive } \alpha \text{ sufficiently close to zero}$$

$$\implies f(x_* + \alpha p) < f(x_*) \text{ for all positive } \alpha \text{ sufficiently close to zero.}$$

(3) x_* is not a local minimizer, i.e.



No matter how small δ is chosen, there are points $x_* + \alpha p$ inside $B(x_*, \delta)$ such that

$$f(x_* + \alpha p) < f(x_*)$$

To prove the first-order optimality conditions we need to recall Taylor's thm.

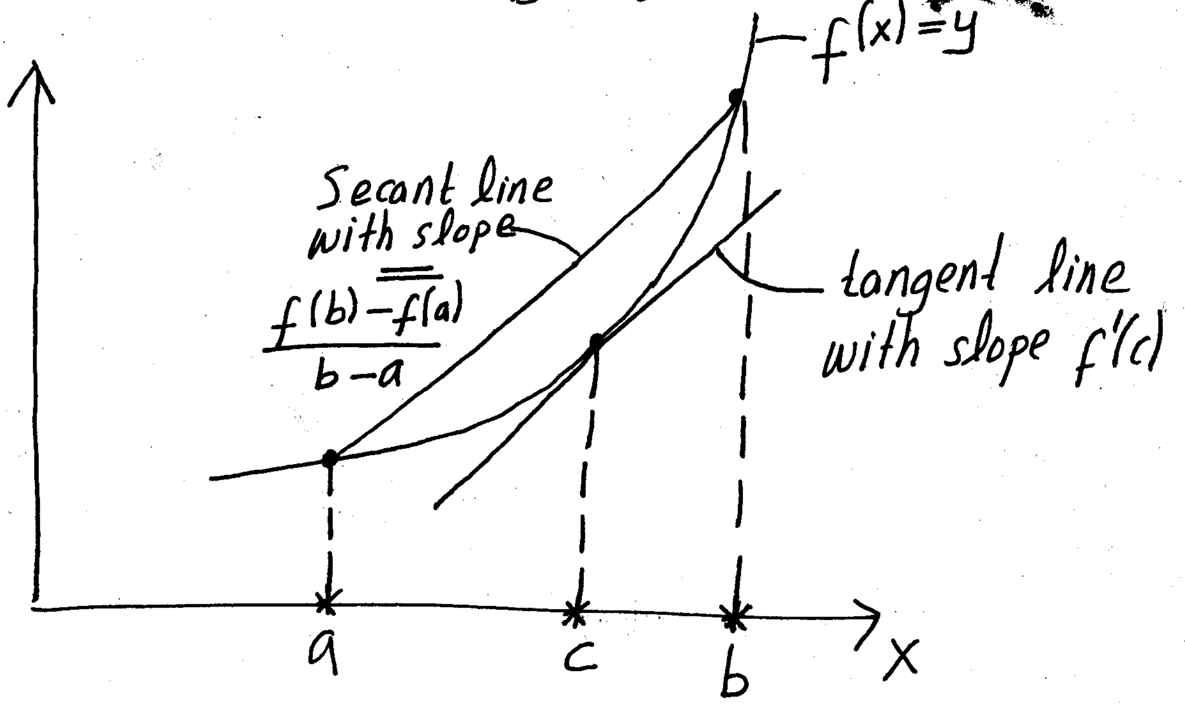
THM (Mean Value)

Let $f:\mathbb{R}\to\mathbb{R}$ be

* continuous on $[a,b]$,
* differentiable on (a,b).

Then there exists a $c\in(a,b)$ such that

$$f'(c) = \frac{f(b)-f(a)}{b-a}$$



EXAMPLE

Consider $f(x) = 4x^2-12x+9$

$$f(1) = f(2) = 1$$

By mean value thm

$$f'(c) = \frac{f(2)-f(1)}{2-1} = 0$$

for some $c\in(1,2)$.

Indeed $f'(x) = 8x-12$; in particular $f'(\frac{3}{2}) = 0$.

THM (Taylor's thm with first-order remainder)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth and $x, p \in \mathbb{R}^n$.
There exists a $t \in (0,1)$ such that

$$f(x+p) = f(x) + \nabla f(x+tp)^T p$$

PROOF
Consider $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(\alpha) = f(x+\alpha p)$.

By the mean value thm there exists a $t \in (0,1)$ such that

$$\frac{\phi(1) - \phi(0)}{1 - 0} = \phi'(t)$$

$$\Longrightarrow f(x+p) - f(x) = \nabla f(x+tp)^T p$$

$$\Longrightarrow f(x+p) = f(x) + \nabla f(x+tp)^T p \qquad \square$$

THM (First Order Necessary Conditions)
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is smooth. Then

$x_* \in \mathbb{R}^n$ is a local minimizer

$$\Longrightarrow$$

$$\nabla f(x_*) = 0$$

PROOF

Let $x_* \in \mathbb{R}^n$ be such that $\nabla f(x_*) \neq 0$.

Notice that for direction $p = -\nabla f(x_*)$ we have

$$\nabla f(x_*)^T p = \nabla f(x_*)^T(-\nabla f(x_*)) = -\|\nabla f(x_*)\|_2^2 < 0$$

Since the derivatives of f are continuous, there exists a positive σ such that

$$(1) \quad \nabla f(x_* + \alpha p)^T p < 0 \quad \text{for all } \alpha \in [0, \sigma)$$

Choose any $\alpha \in [0, \sigma)$. By Taylor's thm

$$f(x_* + \alpha p) = f(x_*) + \nabla f(x_* + tp)^T p$$

$$\xRightarrow[\text{(By inequality (1))}]{} f(x_* + \alpha p) - f(x_*) = \nabla f(x_* + tp)^T p < 0$$

$$\Longrightarrow (2)\ f(x_* + \alpha p) < f(x_*)$$

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

Consider any ball $B(x_*, 2\alpha)$ such that $\alpha < \sigma$. There exists $x_* + \alpha p \in B(x_*, 2\alpha)$ satisfying (2). Consequently x_* is not a local minimizer.

$\square$