# LA2 5

KYB

Thrn, it's a Fact

*mathrnfact@gmail.com*

October 13, 2020

# Overview

# 7.3 Optimization and the Hessian matrix

## Optimization of quadratic functions

A quadratic function $q : \mathbb{R}^n \to \mathbb{R}$ is a multi-variable polynomial of the form

$$\begin{aligned} q(x) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j + \sum_{i=1}^{n} b_i x_i + c \\ &= \frac{1}{2} x \cdot Ax + b \cdot x + c. \end{aligned}$$

## Step0, Ex7.3.1

Suppose $A \in \mathbb{R}^{n\times n}$ and define

$$A_{\text{sym}} = \frac{1}{2}(A + A^T).$$

Then $A_{\text{sym}}$ is symmetric and $x \cdot Ax = x \cdot A_{\text{sym}}x$.

## Remark

By this results, we may assume $A$ is symmetric.

## Step1

Suppose $A$ is diagonal and $b = 0$.

$$q(x) = \frac{1}{2}x \cdot Ax + c = c + \frac{1}{2}\sum_{i=1}^{n} A_{ii}x_i^2.$$

- If $A_{ii} > 0$ for all $i$, $q$ has the unique global minimizer $x^* = 0$.
- If $A_{ii} < 0$ for some $i$, $q(\alpha e_i) = c + \frac{1}{2}A_{ii}\alpha^2 \to -\infty$ as $\alpha \to \infty$.
- If $A_{ii} \geq 0$ and at least one $A_{ii} = 0$, $q$ has global minimizer but it is not unique.

$\Rightarrow$ If $A$ is diagonal and positive semidefinite, $q$ has minimizer.

## Step2

Suppose $A$ is diagonalizable, say $A = UDU^T$, where $U$ is orthogonal and $D$ is diagonal. Write $y = U^T x$, and then

$$x \cdot Ax = x \cdot UDU^T x = (U^T x) \cdot D(U^T x) = y \cdot Dy.$$

By the result of Setp1, if $A$ is positive semidefinite, $c + \frac{1}{2} y \cdot Dy$ has minimizer $y^*$. Since $U$ is orthogonal, the corresponding $x \to y$ is bijective. Thus $x^* = Uy^*$ is a minimizer of $q$.

## Step3

Suppose $A$ is invertible and diagonalizable, and $q(x) = \frac{1}{2}x \cdot Ax + b \cdot x + c$. Define $x^* = -A^{-1}b$, or $b = -Ax^*$.

$$\begin{aligned}
q(x) &= \frac{1}{2}x \cdot Ax + b \cdot x + c \\
&= \frac{1}{2}x \cdot Ax - (Ax^*) \cdot x + c \\
&= \frac{1}{2}(x - x^*) \cdot A(x - x^*) + c - \frac{1}{2}x^* \cdot Ax^* \\
&= \frac{1}{2}(x - x^*) \cdot A(x - x^*) + \tilde{c}, \tilde{c} = c - \frac{1}{2}x^* \cdot Ax^*.
\end{aligned}$$

If we write $y = U^T(x - x^*)$, $q(x) = \frac{1}{2}y \cdot Dy + \tilde{c}$.

## Step4, Ex7.3.2

Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive semidefinite, but not positive definite. Define $q(x) = \frac{1}{2}x \cdot Ax + b \cdot x + c$. Prove that if $b \in \operatorname{col}(A)$ and $x^*$ is a solution to $Ax = -b$, then every vector in $x^* + \mathcal{N}(A)$ is a global minimizer of $q$.

## Proof

It suffices to show that for every $n \in \mathcal{N}(A)$, $q(x^*) = q(x^* + n)$.

$$\begin{aligned}
q(x^*+n) &= \frac{1}{2}(x^*+n)\cdot A(x^*+n) + b\cdot(x^*+n) + c \\
&= \frac{1}{2}(x^*+n)\cdot Ax^* - (Ax^*)\cdot(x^*+n) + c \\
&= \frac{1}{2}A(x^*+n)\cdot x^* - x^*\cdot A(x^*+n) + c \\
&= \frac{1}{2}x^*\cdot Ax^* - Ax^*\cdot x^* + c = q(x^*)
\end{aligned}$$

Since $x^*$ is a local minimizer of $q$ and $q(x^*+n) = q(x^*)$, every vector in $x^* + \mathcal{N}(A)$ is a global minimizer of $q$.

## Step5, Ex7.3.3

Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive semidefinite, but not positive definite. Define $q(x) = \frac{1}{2}x \cdot Ax + b \cdot x + c$. Suppose $b \notin \mathrm{col}(A)$. Find $x^*, n \in \mathbb{R}^n$ such that

$$q(x^* + \alpha n) \to -\infty \text{ as } \alpha \to \infty.$$

## Proof

Let $b^* = \mathrm{proj}_{\mathrm{col}(A)}\, b$ and $n = b^* - b$. Then $n \cdot Ax = 0$ for all $x \in \mathbb{R}^n$.

$$\begin{aligned} q(x) &= \frac{1}{2}x \cdot Ax + b \cdot x + c \\ &= \frac{1}{2}x \cdot Ax + (b^* - n) \cdot x + c \\ &= \frac{1}{2}x \cdot Ax + b^* \cdot x + c - n \cdot x \end{aligned}$$

(continued)

## Proof

Let $x^*$ be a minimizer of $\frac{1}{2}x \cdot Ax + b^* \cdot x + c$. Since $A$ is symmetric, $n \cdot Ax = An \cdot x = 0$ implies $An = 0$. By Ex7.3.2, $x^* + \alpha n$ is a minimizer of $\frac{1}{2}x \cdot Ax + b^* \cdot x + c$.

$$\begin{aligned} q(x^* + \alpha n) &= c - \frac{1}{2}x^* \cdot Ax^* - n \cdot (x^* + \alpha n) \\ &= c - \frac{1}{2}x^* \cdot Ax^* - n \cdot x^* - \alpha n \cdot n \end{aligned}$$

## Summary

WLOG, assume $A$ is symmetric. Let $q(x) = \frac{1}{2}x \cdot Ax + b \cdot x + c$.

- If $A$ is not positive semidefinite, $q(x)$ has no minimizer.
- If $A$ is positive definite, then $A$ is invertible. And $q(x)$ has a minimizer.
- Suppose $A$ is positive semidefinite but not positive definite.
  - If $b \in \mathrm{col}(A)$, $q(x)$ has a minimizer.
  - If $b \notin \mathrm{col}(A)$, $q(x)$ has no minimizer.

## Rudin - Principles of Mathematical Analysis

### Differentiation

For $f : \mathbb{R} \to \mathbb{R}$, we say $f$ is differentiable at $x = a$ if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists.
Let $A = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$. Then

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - Ah|}{h} = 0.$$

## Partial Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ be function. The partial derivative of $f$ at $a \in \mathbb{R}^n$ with respect to the $i$-th component is

$$\lim_{h \to 0} \frac{f(\cdots, a_i + h, \cdots) - f(\cdots, a_i, \cdots)}{h}.$$

If the limit exsits, write

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(\cdots, a_i + h, \cdots) - f(\cdots, a_i, \cdots)}{h}.$$

Using vector notation,

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}.$$

## Differentiation

Let $E \subset \mathbb{R}^n$ be open (or, $E$ is $\epsilon$-ball for some $\epsilon > 0$). Suppose $\boldsymbol{f} : E \to \mathbb{R}^m$ is a function. Choose any norm $\|\cdot\|$ on $\mathbb{R}^n$ and $\mathbb{R}^m$. (In chapter 10, we will show that a choice of norm does not matter.) We say $\boldsymbol{f}$ is differentiable at $\boldsymbol{x} = \boldsymbol{a}$ if there is $A \in \mathbb{R}^{m \times n}$ such that

$$\lim_{\|\boldsymbol{h}\| \to 0} \frac{\|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{a}) - A\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0.$$

And denote $D\boldsymbol{f}(\boldsymbol{a}) = A$.

## Remark

Let $\boldsymbol{f} = (f_1, \cdots, f_m)$ where $f_i : E \to \mathbb{R}$. If we take $\boldsymbol{h} = \epsilon \boldsymbol{e}_i$, we get

$$A_i = \begin{bmatrix} \frac{\partial}{\partial x_i} f_1(\boldsymbol{a}) \\ \vdots \\ \frac{\partial}{\partial x_i} f_m(\boldsymbol{a}) \end{bmatrix}.$$

So

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}.$$

However, the existence of $\frac{\partial f_j}{\partial x_i}(\boldsymbol{a})$ for all $i, j$ does not guarantee the existence $D\boldsymbol{f}(\boldsymbol{a})$.

$D\boldsymbol{f}(\boldsymbol{a})$ is called the Jacobian matrix.

## Notation

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then the gradient of $f$ is

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}.$$

So if $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$,

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} Df_1(\boldsymbol{x}) \\ \vdots \\ Df_m(\boldsymbol{x}) \end{bmatrix} = \left[\nabla f_1(\boldsymbol{x}) \mid \cdots \mid \nabla f_m(\boldsymbol{x})\right]^T = (\nabla \boldsymbol{f}(\boldsymbol{x}))^T.$$

If $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$, $Df(p) \in \mathbb{R}^{n \times 1}$. In this case, write

$$D\boldsymbol{f} = \boldsymbol{f}' = \left(\frac{df_1}{dt}, \cdots, \frac{df_n}{dt}\right) = \begin{bmatrix} \frac{df_1}{dt} \\ \vdots \\ \frac{df_n}{dt} \end{bmatrix}$$

## The Hessian matrix

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at least second times. Note that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. Then the second derivative of $f$ is

$$
\begin{aligned}
D(\nabla f) &= D\left[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_1}\right]^T \\
&= \begin{bmatrix}
\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_1}\right) \\
\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_2}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_n}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_n}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_n}\right)
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}
\end{bmatrix}
\end{aligned}
$$

## The Hessian matrix

Then the Hessian matrix is

$$
\begin{aligned}
\boldsymbol{H}(f) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\boldsymbol{x}) \end{bmatrix} \\
&= \nabla^2 f(\boldsymbol{x}) = (D\nabla f(\boldsymbol{x}))^T.
\end{aligned}
$$

$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(\boldsymbol{x})\right)\right)$

If $f \in C^2(\mathbb{R}^n)$, $\boldsymbol{H}(f)$ is symmetric because

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{x}).$$

## Chain rule

Let $f:\mathbb{R}^n \to \mathbb{R}^m$ and $g:\mathbb{R}^m \to \mathbb{R}^p$ be differentiable at $p \in \mathbb{R}^n$ and $f(p) = q \in \mathbb{R}^m$, respectively. Then $g \circ f$ is also differentiable at $p$ and

$$D(g \circ f)(p) = (Dg(q))\,(Df(p))\,.$$

Note that $Df(p) \in \mathbb{R}^{m \times n}$ and $Dg(q) \in \mathbb{R}^{p \times m}$. So $D(g \circ f)(p) \in \mathbb{R}^{m \times p}$.

## Vector calculus

Let $f, g : \mathbb{R} \to \mathbb{R}^n$. Then $f \cdot g : \mathbb{R} \to \mathbb{R}$ by $f(t) \cdot g(t)$. By the Leibniz rule, $\frac{d}{dt}(f_i g_i) = f_i' g_i + f_i g_i'$. Then

$$\frac{d}{dt}(f \cdot g) = \frac{d}{dt}\left(\sum f_i g_i\right) = \sum (f_i' g_i + f_i g_i') = f' \cdot g + f \cdot g'.$$

## Mean Value Thoerem

Suppose $f : [a, b] \to \mathbb{R}$ is a function such that

- $f$ is continuous on $[a, b]$,
- $f'$ exists for all $t \in (a, b)$.

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

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

## Intermediate Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be differentiable. Suppose $f'(a) < \lambda < f'(b)$. Then there is $c \in (a, b)$ such that

$$f'(c) = \lambda$$

## The Taylor's theorem for one variable

Suppose $f : [a, b] \to \mathbb{R}$ is a function such that

- $f^{(n-1)}$ is continuous on $[a, b]$,
- $f^n$ exists for all $t \in (a, b)$.

Let $x, \alpha \in [a, b]$. Define

$$P(t) = \sum_{k=0}^{n-1} \frac{f^k(\alpha)}{k!}(t-\alpha)^k$$

Then there is $\beta$ between $x$ and $\alpha$ such that

$$f(x) = P(x) + \frac{f^n(x)}{k!}(\beta-\alpha)^n.$$

## The Taylor's theorem for one variable

The last term $\frac{f^n(\beta)}{n!}(\beta-\alpha)^n$ is error term. If there is $M>0$ such that $|f^n(x)|<M$ for all $x\in(a,b)$,

$$|f(x)-P(x)|\le\left|\frac{f^n(\beta)}{n!}(\beta-\alpha)^n\right|<\frac{M}{n!}|\beta-\alpha|^n\le\frac{M}{n!}|x-\alpha|^n$$

If $x\to\alpha$, $\frac{M}{n!}|x-\alpha|^n\to 0$. This means for given $\epsilon>0$, we can find $r>0$ such that

$$|f(x)-P(x)|<\epsilon \text{ for all } |x-\alpha|<r.$$

If $n=3$ and $\alpha=0$,

$$\begin{aligned} f(x)&=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\text{error}\\ &\approx f(0)+f'(0)x+\frac{f''(0)}{2!}x^2. \end{aligned}$$

## Big O notation

Let $f, g$ be real valued functions. We write

$$f(x) = O(g(x)) \text{ near } x = a$$

if there are $\delta, M > 0$ such that

$$|f(x)| \leq Mg(x) \text{ for } 0 < |x - a| < \delta.$$

## Example

Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_k x^k$ where $a_n, a_k \neq 0$ and $n > k$. Then $p(x) = O(x^k)$ near $x = 0$.
Similarly, $p(x) = O(x^n)$ as $x \to \infty$.

## The Taylor's theorem for multi-variable

For $x^*, p \in \mathbb{R}^n$, define $\phi(\alpha) = f(x^* + \alpha p)$. Then $\phi : I \to \mathbb{R}$ is a function on some open interval $I$ in $\mathbb{R}$ containing $0$. Since $x^* + \alpha p = (x_1^* + \alpha p_1, \cdots, x_n^* + \alpha p_n)$ is smooth, $\phi$ is also smooth. Now we have the quadratic approximation $\phi$,

$$\phi(\alpha) = \phi(0) + \phi'(0)\alpha + \frac{1}{2!}\phi''(0)\alpha^2 + O(\alpha^3).$$

If we write $g(\alpha) = x^* + \alpha p$, $g_i(\alpha) = x_i^* + \alpha p_i$. So $\phi = f \circ g$. Moreover, $\frac{dg_i}{d\alpha} = p_i$ implies $g'(\alpha) = p$.

## The Taylor's theorem

Let $\beta = g(\alpha)$. By the chain rule,

$$\phi'(\alpha) = Df(\beta)Dg(\alpha) = (\nabla f(\beta))^T p = \nabla f(x^* + \alpha p) \cdot p$$

$$\begin{aligned}
\phi''(\alpha) &= \frac{d}{d\alpha}(\nabla f(g(\alpha)) \cdot p) \\
&= (D(\nabla f(g(\alpha)))) \cdot p + \nabla f(g(\alpha)) \cdot p' \\
D(\nabla f(g(\alpha))) &= (D\nabla f)\,(x^* + \alpha p)(\frac{d}{d\alpha}g)(\alpha) = (\nabla^2 f(x^* + \alpha p))^T p \\
\phi''(\alpha) &= p \cdot \nabla^2 f(x^* + \alpha p)p.
\end{aligned}$$

So

$$\phi(0) = f(x^*), \phi'(0) = \nabla f(x^*), \phi''(0) = p \cdot \nabla^2 f(x^*)p.$$

## First- and second-order optimality conditions

Finally, take $\alpha = 1$ and $p = x - x^*$. By the Taylor's theorem,

$$f(x) \approx f(x^*) + \nabla f(x^*) \cdot (x - x^*) + \frac{1}{2}(x - x^*) \cdot \nabla^2 f(x^*)(x - x^*), x \text{ near } x^*.$$

Then our optimality condition says

- $\nabla f(x^*) = 0$.
- $\nabla^2 f(x^*)$ is positive semidefinite.

## Ex7.3.5

$q(x) = (1/2)x \cdot (Ax) + b \cdot x$, where

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, b = \begin{bmatrix} -5 \\ -4 \end{bmatrix}.$$

Find all global minimizers of $q$, if it has any, or explain why none exist.

## Proof

$3$ and $-1$ are eigenvalues of $A$ with coresponding eigenvectors $(1,1)$ and $(1,-1)$, respectively. Since $A$ is indefinite, $q$ has no global minimizer or maximizer.

## Ex7.3.8

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x) = 100x_2^2 - 200x_1^2x_2 + 100x_1^4 + x_1^2 - 2x_1 + 1.$$

Show that $x^* = (1,1)$ is a stationary point of $f$. If possible, determine whether it is a local minimizer, a local maximizer, or a saddle point.

## Proof

$$\nabla f(x) = \begin{bmatrix} -400x_1x_2 + 400x_1^3 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{bmatrix}$$

Since $\nabla f(1,1) = 0$, $x^*$ is a stationary point of $f$.
(continued)

## Proof

$$\nabla^2 f(x) = \begin{bmatrix} -400x_2 + 1200x_1^2 + 2x_1 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

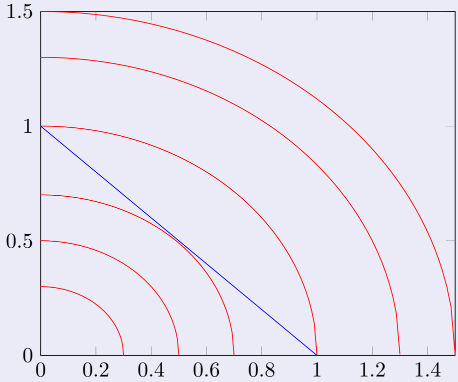
$$\nabla^2 f(1,1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} = A$$

Consider $p_A(r) = r^2 - 1002r + 400$. Then $r = 501 \pm \sqrt{501^2 - 4 \cdot 400}$. Since $0 \leq 501^2 - 4 \cdot 400 < 501^2$, $r = 501 \pm \sqrt{501^2 - 4 \cdot 400} > 0$. Hecne $A$ has two positive eigenvalues, and so $A$ is positive definite.

# 7.4 Lagrange multipliers

## The equality constrained nonlinear program

Let $f(x, y) = \sqrt{x^2 + y^2}$. We want to find a minimizer where $g(x, y) = x + y - 1 = 0$.



## The equality constrained nonlinear program

$$\begin{aligned} &\min \ f(x) \\ &s.t. \ g(x) = 0 \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$.

## Lagrange multiplier

If $x^*$ is a local minimizer of $f(x)$ subject to the constraint $g(x) = 0$, then $\nabla f(x^*) \in \mathrm{col}(\nabla g(x^*))$, that is, there is $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) = \nabla g(x^*)\lambda^*$$

Michael T. Heath - Scientific Computing 6.2.3

## Constrained Optimality Conditions

Define $\mathcal{L} : \mathbb{R}^{n+m} \to \mathbb{R}$ by

$$\mathcal{L}(x, \lambda) = f(x) - \lambda \cdot g(x) = f(x) - \sum_{k=1}^{m} \lambda_k g_k(x).$$

$\mathcal{L}$ is called the Lagrangian function and $\lambda$ is called the Lagrangian multiplier. Then

$$\nabla \mathcal{L}(x, \lambda) = \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda) \\ \nabla_\lambda \mathcal{L}(x, \lambda) \end{bmatrix} = \begin{bmatrix} \nabla f(x) - \nabla g(x)\lambda \\ g(x) \end{bmatrix}$$

## Constrained Optimality Conditions

The first optimality condition says

$$\nabla \mathcal{L}(x, \lambda) = 0,$$

or

$$\begin{cases} \nabla f(x) & = \nabla g(x)\lambda \\ g(x) & = 0 \end{cases}$$

So if $(x^*, \lambda^*)$ satisfies the above condition, it is a candidate of minimizer or maximizer.

## Ex7.4.1

Explain why, if $x^*$ is a local minimizer of

$$\begin{aligned} &\min\ f(x) \\ &s.t.\ g(x) = 0, \end{aligned}$$

and a regular point of the constraint $g(x) = 0$, then the Lagrange multiplier $\lambda^*$ is unique. ($g(x^*) = 0$ and $\nabla g(x^*)$ has full rank.)

## Proof

Let $A = \nabla g(x^*)$ and $b = \nabla f(x^*)$. Since $x^*$ is a minimizer of the system, $\nabla f(x^*) \in \text{col}(A)$. So $Ax = b$ has a solution. Since $A$ has full rank, this solution is unique.

## 7.4.2

$$f(x) = x_3, g(x) = \begin{bmatrix} x_1 + x_2 + x_3 - 12 \\ x_1^2 + x_2^2 - x_3 \end{bmatrix}.$$

## Proof

$$\nabla f(x) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \nabla g(x) = \begin{bmatrix} 1 & 2x_1 \\ 1 & 2x_2 \\ 1 & -1 \end{bmatrix}.$$

So we get

$$0 = \lambda_1 + 2\lambda_2 x_1, 0 = \lambda_1 + 2\lambda_2 x_2, 1 = \lambda_1 - \lambda_2$$
$$x_1 + x_2 + x_3 - 12 = 0, x_1^2 + x_2^2 - x_3 = 0$$

## Proof

$\lambda_1 = 1 + \lambda_2$, $\lambda_2(x_2 - x_1) = 0$.
1) If $\lambda_2 = 0$, then $\lambda_1 = 1$ and $0 = \lambda_1 + 2\lambda_2 x_1 = \lambda_1$. (condtradiction).
2) So $x_1 = x_2$. Then $x_1 = 2$ or $x_1 = -3$ and $x_3 = 8$ or $x_3 = 18$, respectively.

$$(x^*, \lambda^*) = (2, 2, 8, 4/5, -1/5), (-3, -3, 18, 6/5, 1/5).$$

Consider $g_1(x) + g_2(x) = 0 = x_1^2 + x_1 + x_2^2 + x_2 - 12$, or

$$(x_1 + 1/2)^2 + (x_2 + 1/2)^2 = 25/2.$$

This implies $\{x \mid g(x) = 0\}$ is closed and bounded. So it is compact. Now $f$ is a continuous function on compact set. Hence it must have a minimum and maximum.

## Remark

In fact, we can caluate directly. Write $x_1 = -\frac{1}{2} + \frac{5}{\sqrt{2}} \cos t$, $x_2 = -\frac{1}{2} + \frac{5}{\sqrt{2}} \sin t$. So $x_3 = 12 - x_1 - x_2 = 13 + \frac{5}{\sqrt{2}}(\cos t + \sin t)$. Since $\frac{1}{\sqrt{2}} \cos t + \frac{1}{\sqrt{2}} \sin t = \cos(t - \frac{\pi}{4})$,

$$x_3 = 13 + 5\cos(t - \tfrac{\pi}{4}).$$

Hence $x_3 = 8$ is the minimum and $x_3 = 18$ is the maximum.

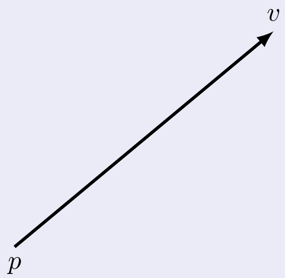
# Tangent Space

Chapter 3 in John M. Lee - Introduction to Smooth Manifolds(2012).

## Tangent vectors

A tangent vector $v_p$ to $\mathbb{R}^n$ at $p$ consits of two parts,

- vector part $v \in \mathbb{R}^n$,
- application point $p \in \mathbb{R}^n$.



## Tangent space

For each $p \in \mathbb{R}^n$, the tangent space $\mathbb{R}^n$ at $p$ is the set of all tangent vector that have $p$ as an application point, denoted by $T_p\mathbb{R}^n$. If we define operations by

$$v_p + w_p = (v + w)_p, \alpha v_p = (\alpha v_p),$$

$T_p\mathbb{R}^n$ is a vector space. If $p \neq q$, $v_p \neq w_q$ and there is no addition $v_p + v_q$. Define $T\mathbb{R}^n = \bigcup_{p\in\mathbb{R}^n} T_p\mathbb{R}^n$. Then $T\mathbb{R}^n$ is called the tangent space of $\mathbb{R}$. Note that $T\mathbb{R}^n$ is not a vector space. In fact, every element of $T\mathbb{R}^n$ is of the form $(p, v) \in \mathbb{R}^n \times \mathbb{R}^n$, simply $v_p$.

## Corresponding Derivation

For each $v_p \in T_p\mathbb{R}^n$, there is a corresponding derivative such that

$$D(v_p)[f] := \frac{d}{dt} f(p+tv)\Big|_{t=0}.$$

$v_p$ satisfies

- $D(v_p)[af+bg] = aD(v_p)[f] + bD(v_p)[g]$ where $a, b \in \mathbb{R}$ and $f, g \in C^\infty(\mathbb{R}^n)$. (linearlity)
- $D(v_p)[fg] = D(v_p)[f]g(p) + f(p)D(v_p)[g]$. (Leibniz rule).

## Derivation

Let $p \in \mathbb{R}^n$. $v : C^\infty(\mathbb{R}^n) \to \mathbb{R}$ is called a derivation if $v$ satisfies

- $v[af+bg] = av[f] + bv[g]$ where $a, b \in \mathbb{R}$ and $f, g \in C^\infty(\mathbb{R}^n)$. (linearlity)
- $v[fg] = v[f]g(p) + f(p)v[g]$. (Leibniz rule).

## Set of derivation

Let $\mathcal{D}_p$ be the set of all derivation at $p$. Then we can define operations by

$$(v+w)[f] = v[f] + w[f], (\alpha v)[f] = \alpha v[f].$$

Thus $\mathcal{D}_p$ is a vector space. Moreover, we have a linear map $D : T_p\mathbb{R}^n \to \mathcal{D}$.

## Properties of derivations

Suppose $p \in \mathbb{R}^n$ and $v \in \mathcal{D}_p$ and $f, g \in C^\infty(\mathbb{R}^n)$.

- If $f$ is constant, $v[f] = 0$.
- If $f(p) = g(p) = 0$, then $v[fg] = 0$.

## Isomorphism of tangent space and derivations

$D : T_p\mathbb{R}^n \to \mathcal{D}$ is isomorphism.

## Proof

Take $f(x) = x_i$. Then

$$D(v_p)[f] = \frac{d}{dt}(p_i + tv_i)\Big|_{t=0} = v_i = 0.$$

So $v = 0$.
Let $v_p \in T\mathbb{R}^n$. Then by the chain rule,

$$D(v_p)[f] = \frac{d}{dt}f(p + tv)\Big|_{t=0} = \sum_i \frac{\partial f}{\partial x_i}(p)v_i.$$

## Proof

To prove surjectivity, for $v \in \mathcal{D}$, define $w = \sum v[x_i]e_i$. We will show that $D(w_p) = v$. For $f \in C^\infty(\mathbb{R}^n)$, apply Taylor's theorem,

$$f(x) = f(p) + \sum_i \frac{\partial f}{\partial x_i}(p)(x_i - p_i)$$
$$+ \sum_{i,j}(x_i - p_i)(x_j - p_j)\int_0^1 (1-t)\frac{\partial^2}{\partial x_i \partial x_j}(p + t(x-p))dt.$$

Then the last sum vanishes at $p$.

$$\begin{aligned} v[f] &= v[f(p)] + \sum_i v[\frac{\partial f}{\partial x_i}(p)(x_i - p_i)] \\ &= \sum_i \frac{\partial f}{\partial x_i}(p)(v[x_i] - v[p_i]) \\ &= \sum_i \frac{\partial f}{\partial x_i}(p)w_i = D(w_p)[f]. \end{aligned}$$

## Notation

Now we can identify $v_p \in T\mathbb{R}^n$ with $D(v_p)$, i.e.

$$v_p[f] := D(v_p)[f].$$

Let $e_i \in \mathbb{R}^n$. Then

$$(e_i)_p[f] = \left(\frac{\partial}{\partial x_i} f\right)(p).$$

So we can identify $(e_i)_p$ with $\left.\frac{\partial}{\partial x_i}\right|_p$. Then for all $v_p \in T\mathbb{R}^n$,

$$v_p = \sum v_i \left.\frac{\partial}{\partial x_i}\right|_p.$$

Hence $\{\left.\frac{\partial}{\partial x_i}\right|_p\}_{i=1}^n$ is a basis for $T_p\mathbb{R}^n$.

## Fact

We can define a directional derivative ($D(v_p)$) by another way. Let $\gamma : (-1,1) \to \mathbb{R}^n$ be a smooth curve such that $\gamma(0) = p$ and $\gamma'(0) = v$. Define

$$v_p[f] := \frac{d}{dt}(f \circ \gamma)(0).$$

In fact, $v_p[f]$ is invariant under a choice of curve. So if we take $\gamma(t) = p + tv$, $v_p[f]$ is the same value as in the first definition.

## Vector field

A function $X : \mathbb{R}^n \to \mathbb{R}^n$ is called a vector field. For each $p \in \mathbb{R}^n$, we can identify $X(p) \in T_p\mathbb{R}^n$. Then we can define a derivative with respect to $X$ by

$$X[f](p) := X(p)[f].$$

If $f : \mathbb{R}^n \to \mathbb{R}$, $fX$ is also vector field. Then the following properties hold:

- $X[af + bg] = aX[f] + bX[g]$.
- $X[fg] = X[f]g + fX[g]$.
- $(fV + gW)[h] = fV[h] + gW[h]$.

## Vector field

Consider $\frac{\partial}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\frac{\partial}{\partial x_i}(p) = \left.\frac{\partial}{\partial x_i}\right|_p .$$

This is a vector field. $\{\frac{\partial}{\partial x_i}\}_{i=1}^n$ is called a frame field. Then every vector field $X$ can be written as

$$X = \sum X_i \frac{\partial}{\partial x_i}$$

where $X_i : \mathbb{R}^n \to \mathbb{R}$ a function such that $X(p) = (X_1(p), \cdots, X_n(p))$.

## Example

In $\mathbb{R}^2$, the polar coordinates induce tangent vectors $\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \theta}$ where $x = r\cos\theta$ and $y = r\sin\theta$. Then

$$\begin{aligned}\partial_r &= \cos\theta\partial_x + \sin\partial_y \\ \partial_\theta &= -r\sin\theta\partial_x + r\cos\theta\partial_y\end{aligned}$$

or

$$\begin{aligned}\partial_x &= \cos\theta\partial_r - \frac{1}{r}\sin\theta\partial_\theta \\ \partial_y &= \sin\theta\partial_r + \frac{1}{r}\cos\theta\partial_\theta\end{aligned}$$

# The End