Векторное дифференцирование. Семинар 3. 25 сентября 2019 г.

Семинарист: Данилова М.

Векторное дифференцирование

Основные понятия

Пусть $f : \mathrm{D} \to \mathrm{E}$, производная $\frac{\partial f}{\partial x} \in \mathrm{G}$:

	D	E	G	Название
	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	Производная, $f'(x)$
1)	$\mathbb{R}^n$	$\mathbb{R}$	$\mathbb{R}^n$	Градиент, $\frac{\partial f}{\partial x_i}$
2)	$\mathbb{R}^n$	$\mathbb{R}^m$	$\mathbb{R}^{n\times m}$	Матрица Якоби, $\frac{\partial f_i}{\partial x_j}$
3)	$\mathbb{R}^{m\times n}$	$\mathbb{R}$	$\mathbb{R}^{m\times n}$	$\frac{\partial f}{\partial x_{ij}}$

1. (a) $\nabla f(x)$ – градиент

   $\nabla f(x) = \frac{df}{dx} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$ – вектор-столбец $(n \times 1)$

   $\nabla f^\top(x) = \frac{df}{dx^\top} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{pmatrix}$ – вектор-строка $(1 \times n)$

   (b) $\mathrm{H} = [h_{ij}] = \nabla^2 f(x)$ – матрица Гессе, матрица вторых производных

   $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$

   $\nabla^2 f(x) = \frac{d(\nabla f(x))}{dx^\top} = \frac{d(\nabla f^\top(x))}{dx}$

2. Матрица Якоби

   $\frac{df}{dx^\top} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ & \dots & \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \in \mathbb{R}^{m\times n}$

   $\frac{df^\top}{dx} = \left(\frac{df}{dx^\top}\right)^\top \in \mathbb{R}^{n\times m}$

Матричное представление функции:

$$f = x^\top \mathrm{A} x,\ \mathrm{A} = \mathrm{A}^\top$$

Скалярное предсталение функции:

$$f = \sum_{ij} a_{ij} x_i x_j$$

Градиент

$$\nabla f_i = \frac{\partial f(x)}{\partial x_i}$$

Скалярное предсталение градиента:

$$\nabla f_i = 2 \sum_j a_{ij} x_j$$

Матричное представление градиента:

$$\nabla f = 2\mathrm{A}x$$

Формулы дифференцирования сложной функции

1. $f \in \mathbb{R}^1,\ g \in \mathbb{R}^p,\ x \in \mathbb{R}^n$

$$\frac{\partial f(g(x))}{\partial x_j} = \sum_{i=1}^{p} \frac{\partial f}{\partial g_i} \cdot \frac{\partial g_i}{\partial x_j} = \frac{dg^\top}{dx_j} \cdot \frac{df}{dg}$$

$\frac{dg^\top}{dx_j}$ - строка $(1 \times p)$

$\frac{df}{dg}$ - столбец $(p \times 1)$

$\rightarrow$ объединим n чисел $(x \in \mathbb{R}^n)$:

$$\frac{df(g(x))}{dx} = \frac{dg^\top}{dx} \cdot \frac{df}{dg} \quad \left(\frac{df(g(x))}{dx} - (n \times 1),\ \frac{dg^\top}{dx} - (n \times p),\ \frac{df}{dg} - (p \times 1) \right)$$

$$\frac{df(g(x))}{dx^\top} = \frac{df}{dg^\top} \cdot \frac{dg}{dx^\top} \quad \left(\frac{df(g(x))}{dx^\top} - (1 \times n),\ \frac{df}{dg^\top} - (1 \times p),\ \frac{dg}{dx^\top} - (p \times n) \right)$$

2. $f \in \mathbb{R}^m,\ g \in \mathbb{R}^p,\ x \in \mathbb{R}^n$

$\rightarrow$ объединим m столбцов (матрица Якоби)

$$\frac{df^\top(g(x))}{dx} = \frac{dg^\top}{dx} \cdot \frac{df^\top}{dg} \quad \left(\frac{df^\top(g(x))}{dx} - (n \times m),\ \frac{dg^\top}{dx} - (n \times p),\ \frac{df^\top}{dg} - (p \times m) \right)$$

$$\frac{df(g(x))}{dx^\top} = \frac{df}{dg^\top} \cdot \frac{dg}{dx^\top} \quad \left(\frac{df(g(x))}{dx^\top} - (m \times n),\ \frac{df}{dg^\top} - (m \times p),\ \frac{dg}{dx^\top} - (p \times n) \right)$$

Частные случаи

1. $f \in \mathbb{R}^m,\ g \in \mathbb{R}^1,\ x \in \mathbb{R}^n$

$$\frac{df}{dx^\top} = \frac{df}{dg} \cdot \frac{dg}{dx^\top} \quad \left(\frac{df}{dx^\top} - (m \times n),\ \frac{df}{dg} - (m \times 1),\ \frac{dg}{dx^\top} - (1 \times n)\right)$$

2. $f \in \mathbb{R}^1,\ g \in \mathbb{R}^1,\ x \in \mathbb{R}^n$

$$\frac{df}{dx} = \frac{dg}{dx} \cdot \frac{df}{dg} \quad \left(\frac{df}{dx} - (n \times 1),\ \frac{dg}{dx} - (n \times 1),\ \frac{df}{dg} - (1 \times 1)\right)$$

Примеры

1. $u(x),\ v(x) \in \mathbb{R}^m$

$$\frac{d(u^\top v)}{dx} = \frac{du^\top}{dx} v + \frac{dv^\top}{dx} u$$

2. $u(x) \in \mathbb{R}^m,\ v(x) \in \mathbb{R}^l,\ \mathrm{A} \in \mathbb{R}^{m \times l}$

$$\frac{d}{dx}\left(u^\top \mathrm{A} v\right) = \frac{du^\top}{dx}\mathrm{A} v + \frac{dv^\top}{dx}\mathrm{A}^\top u$$

3. $u(x) \in \mathbb{R}^m,\ \mathrm{A} \in \mathbb{R}^{m \times m},\ \mathrm{A} = \mathrm{A}^\top,\ x \in \mathbb{R}^n$

$$\frac{d}{dx}\left(u^\top \mathrm{A} u\right) = \frac{du^\top}{dx}\mathrm{A} u + \frac{du^\top}{dx}\mathrm{A}^\top u = \left(\mathrm{A}^\top = \mathrm{A}\right) = 2\frac{du^\top}{dx}\mathrm{A} u$$

$$\mathrm{A} = \mathrm{E} \ \rightarrow\ \frac{d}{dx}(u^\top u) = 2\frac{du^\top}{dx} u$$

4. $x \in \mathbb{R}^n,\ \mathrm{A} \in \mathbb{R}^{n \times n},\ \mathrm{A}^\top = \mathrm{A},\ c \in \mathbb{R}^n$

$$\frac{dx^\top}{dx} = \frac{dx}{dx^\top} = \mathrm{E}$$

$$\frac{d}{dx}\left(x^\top \mathrm{A}\right) = \mathrm{A}$$

$$\frac{d}{dx^\top}\left(\mathrm{A} x\right) = \mathrm{A}$$

$$\frac{d}{dx}(x^\top c) = c = \frac{d}{dx}(c^\top x)$$

5. $x \in \mathbb{R}^n,\ \mathrm{A} \in \mathbb{R}^{n \times n},\ \mathrm{A} = \mathrm{A}^\top$

$$\frac{d}{dx}\left(x^\top \mathrm{A} x\right) = \left(\mathrm{A}^\top + \mathrm{A}\right) x = 2\mathrm{A} x$$

$$\frac{d}{dx}(x^\top x) = 2x$$

6. $f(x) = x^\top \mathrm{A} x + b^\top x$
$$\nabla f(x) = \left(\mathrm{A} + \mathrm{A}^\top\right) x + b$$
$$\nabla^2 f(x) = \mathrm{A} + \mathrm{A}^\top$$

7. $f(x) = \|x\|_2 = \sqrt{x^\top x}$
   $f(g) \in \mathbb{R}^1,\ g(x) \in \mathbb{R}^1,\ x \in \mathbb{R}^n$

$$\nabla f(g(x)) = \frac{df(g(x))}{dx} = \frac{dg}{dx} \cdot \frac{df}{dg} \quad \left(\frac{df(g(x))}{dx} - (n \times 1),\ \frac{dg}{dx} - (n \times 1),\ \frac{df}{dg} - (1 \times 1)\right)$$
$$\frac{df}{dx} = \frac{d(x^\top x)}{dx} \cdot \frac{1}{2\sqrt{x^\top x}} = \frac{2x}{2\sqrt{x^\top x}} = \frac{x}{\sqrt{x^\top x}} = \frac{x}{\|x\|_2}$$
$$\nabla f = \frac{x}{\|x\|_2} \in \mathbb{R}^n \quad (n \times 1)$$
$$\nabla^2 f = \frac{d(\nabla f(x))}{dx^\top} = \frac{d}{dx^\top}\left(\frac{x}{\|x\|_2}\right) = \frac{1}{\|x\|_2}\mathrm{E} - \frac{xx^\top}{\|x\|_2^3}$$

8. $f(x) = -e^{-x^\top x}$
$$\nabla f = \frac{df}{dx} = -e^{-x^\top x}\frac{d(-x^\top x)}{dx} = -e^{-x^\top x}\left(-\frac{dx^\top}{dx}x - \frac{dx^\top}{dx}x\right) = 2xe^{-x^\top x} \quad (n \times 1)$$
$$\nabla^2 f = \frac{d(\nabla f(x))}{dx^\top} = 2\frac{d}{dx^\top}\left(xe^{-x^\top x}\right) = 2\mathrm{E}e^{-x^\top x} + 2x\frac{de^{-x^\top x}}{dx^\top} = 2\mathrm{E}e^{-x^\top x} - 2x\left(\frac{de^{-x^\top x}}{dx}\right)^\top =$$
$$2\mathrm{E}e^{-x^\top x} - 2x\left(2xe^{-x^\top x}\right)^\top = 2e^{-x^\top x}\left(\mathrm{E} - 2xx^\top\right)$$

9. $f(x) = \frac{1}{2}\|\mathrm{A}x - b\|_2^2$
   $x \in \mathbb{R}^n,\ \mathrm{A} \in \mathbb{R}^{n \times n},\ b \in \mathbb{R}^n$
$$\frac{df}{dx} = \frac{1}{2}\frac{d(\mathrm{A}x - b)^\top}{dx}(\mathrm{A}x - b) + \frac{1}{2}\frac{d(\mathrm{A}x - b)^\top}{dx}(\mathrm{A}x - b) = \mathrm{A}^\top(\mathrm{A}x - b)$$

Замечание 1.
$$\|x\| = \sqrt{x^\top x}$$
$$d\|x\| = \frac{x^\top}{\|x\|}dx = \left(\frac{x}{\|x\|}, dx\right) \text{ – дифференциал}$$
$$\nabla\left(\|x\|\right) = \frac{x}{\|x\|} \text{ – градиент}$$

Скалярное произведение

Определение 1 (Скалярное произведение)**.** Пусть V - вещественное векторное пространство. Функция $\langle\cdot,\cdot\rangle \ : \ \mathrm{V}\times\mathrm{V}\to\mathbb{R}$, ставящая каждой паре векторов $x, y \in \mathrm{V}$ в соответствие вещественное число $\langle x, y\rangle$, называется **скалярным произведением**, если оно удовлетворяет следующим условиям:

1. для любого $x \in \mathrm{V}$ выполнено $\langle x, x\rangle \geq 0$, причем $\langle x, x\rangle = 0$ только при $x = 0$ (положительная определенность);
2. для любых $x, y \in \mathrm{V}$ выполнено $\langle x, y\rangle = \langle y, x\rangle$ (симметричность);
3. для любых $x, y, z \in \mathrm{V}, \ \alpha, \beta \in \mathbb{R}$ выполнено $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ (линейность).

Определение 2 (Стандартное скалярное произведение в $\mathbb{R}^n$)**.** В пространстве $\mathbb{R}^n$ вещественных n-мерных вектор-столбцов стандартное скалярное произведение задается формулой:

$$\langle x, y\rangle = \mathrm{Tr}\left(x^\top y\right) = \sum_{i=1}^n x_i y_i.$$

Стандартное скалярное произведение в $\mathbb{R}^n$ не является единственно возможным.

Определение 3 (Стандартное скалярное произведение в $\mathbb{R}^{m\times n}$)**.** В пространстве $\mathbb{R}^{m\times n}$ матриц вводят **фробениусово** скалярное произведение:

$$\langle \mathrm{A}, \mathrm{B}\rangle = \mathrm{Tr}\left(\mathrm{A}^\top \mathrm{B}\right) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}.$$

Определение 4 (**Стандартное скалярное произведение в** $\mathbb{S}^n$)**.** Так как скалярное произведение можно наследовать на подпространство, получаем **фробениусово** скалярное произведение в $\mathbb{S}^n$:

$$\langle \mathrm{A}, \mathrm{B}\rangle = \mathrm{Tr}\left(\mathrm{A}\mathrm{B}\right).$$

Норма

Определение 5 (Стандартная евклидова норма в $\mathbb{R}^n$)**.**

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{\left(\sum_{i=1}^n x_i^2\right)}$$

Стандартное евклидова норма также называется l_2–норма.

Определение 6 (Стандартная фробениусова норма в $\mathbb{R}^{m\times n}$)**.**

$$\|\mathrm{A}\|_\mathrm{F} = \sqrt{\langle \mathrm{A}, \mathrm{A}\rangle} = \sqrt{\mathrm{Tr}\left(\mathrm{A}^\top \mathrm{A}\right)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$$

Стандартная фробениусова норма также называется нормой Гильберта-Шмидта.

Определение 7 (**Стандартная фробениусова норма в** $\mathbb{S}^n$). Так как норму можно наследовать на подпространство, получаем **фробениусову** норму в $\mathbb{S}^n$:

$$\|\mathrm{A}\|_{\mathrm{F}} = \sqrt{\langle \mathrm{A}, \mathrm{A} \rangle} = \sqrt{\mathrm{Tr}\,(\mathrm{A}^2)}.$$

Замечание 2. Если $\mathrm{A} \in \mathbb{S}^n$, то

$$\|\mathrm{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n} \lambda_i^2(\mathrm{A})},$$

где $\lambda_i(\mathrm{A})$ - собственные значения A.

Определение 8 (Операторная норма в $\mathbb{R}^{m\times n}$). Назовем норму матрицы операторной, если она подчинена некоторой норме векторов:

$$\|\mathrm{A}\| = \sup_{x\in\mathbb{R}^n,\ \|x\|=1} \|\mathrm{A}x\| = \sup_{x\neq 0} \frac{\|\mathrm{A}x\|}{\|x\|}.$$

Также называется подчинённой или индуцированной нормой.

Матричное дифференцирование

Пусть $f(\mathrm{A})$ - **функция матричного аргумента**, где $\mathrm{A} \in \mathbb{R}^{m\times n},\ \mathrm{A} = \|a_{ij}\|,\ f : \mathbb{R}^{m\times n} \to \mathbb{R}^1$

$$\frac{\partial f}{\partial \mathrm{A}} = \begin{pmatrix} \frac{\partial f}{\partial a_{11}} & ... & \frac{\partial f}{\partial a_{1n}} \\ & ... & \\ \frac{\partial f}{\partial a_{m1}} & ... & \frac{\partial f}{\partial a_{mn}} \end{pmatrix}$$

матричный градиент – матрица той же размерности, что и A

Пример 1. $f(\mathrm{A}) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}u_iu_j = u^\top \mathrm{A} u$

$\mathrm{A} \in \mathbb{R}^{n\times n},\ u \in \mathbb{R}^n$

$$\frac{\partial f}{\partial \mathrm{A}} = \frac{\partial u^\top Au}{\partial \mathrm{A}} = \begin{pmatrix} u_1u_1 & ... & u_1u_m \\ & ... & \\ u_mu_1 & ... & u_mu_m \end{pmatrix} = uu^\top \in \mathbb{R}^{m\times m}$$

$$\frac{\partial u^\top \mathrm{A} u}{\partial u_{ij}} = u_iu_j$$

Базовые операции с матрицами

1. $\mathrm{A}(\mathrm{B}+\mathrm{C}) = \mathrm{AB} + \mathrm{AC}$
2. $(\mathrm{A}+\mathrm{B})^\top = \mathrm{A}^\top + \mathrm{B}^\top$
3. $(\mathrm{AB})^\top = \mathrm{B}^\top\mathrm{A}^\top$
4. $(\mathrm{AB})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$
5. $(\mathrm{A}^{-1})^\top = (\mathrm{A}^\top)^{-1}$
6. $(\mathrm{A}+\mathrm{B})^{-1} = \mathrm{A}^{-1} - \mathrm{A}^{-1}\mathrm{B}(\mathrm{B}+\mathrm{B}^{-1}\mathrm{AB})\mathrm{BA}^{-1}$

След и определитель

1. $\det(\mathrm{AB}) = \det\mathrm{A}\ \det\mathrm{B}$
2. $\det(\mathrm{A})^{-1} = \frac{1}{\det A}$
3. $\mathrm{Tr}\mathrm{A} = \sum\limits_j a_{jj} = \sum\limits_j \lambda_j$
4. $\mathrm{Tr}(\mathrm{A}+\mathrm{B}) = \mathrm{Tr}\mathrm{A} + \mathrm{Tr}\mathrm{B}$
5. $\mathrm{Tr}(\mathrm{ABC}) = \mathrm{Tr}(\mathrm{BCA}) = \mathrm{Tr}(\mathrm{CAB})$

Пространство симметричных матриц $\mathbb{S}^n$

Если $\mathrm{X} \in \mathbb{S}^n$ - симметричная матрица, тогда мы работаем в гильбертовом пространстве:

$$\langle \mathrm{A}, \mathrm{B} \rangle = \mathrm{Tr}\,(\mathrm{AB}) \text{ – скалярное произведение}$$

$$\|\mathrm{A}\|_{\mathrm{F}} = \sqrt{\langle \mathrm{A}, \mathrm{A} \rangle} = \sqrt{\sum_{ij} a_{ij}^2} \text{ – фробениусова норма}$$

Правила матричного дифференцирования

$$f(\mathrm{X}) \text{ – функция матричного аргумента}$$

$$d\mathrm{X} \text{ – матрица}$$

$$df(\mathrm{X}) = \langle \nabla f(\mathrm{X}), d\mathrm{X} \rangle \text{ – дифференциал}$$

$$\nabla f(\mathrm{X}) \text{ – градиент (матрица)}$$

Пусть A, B - фиксированные матрицы, α - скаляр, X, Y - матричные функции.

Правила преобразования

1. $d\mathrm{A} = 0$
2. $d\,(\alpha \mathrm{A}) = \alpha\,(d\mathrm{A})$
3. $d\,(\mathrm{AXB}) = \mathrm{A}\,(d\mathrm{X})\,\mathrm{B}$
4. $d\,(\mathrm{X} + \mathrm{Y}) = d\mathrm{X} + d\mathrm{Y}$
5. $d\left(\mathrm{X}^\top\right) = (d\mathrm{X})^\top$
6. $d\,(\mathrm{XY}) = (d\mathrm{X})\,\mathrm{Y} + \mathrm{X}\,(d\mathrm{Y})$
7. $d\,\langle \mathrm{X}, \mathrm{Y} \rangle = \langle d\mathrm{X}, \mathrm{Y} \rangle + \langle \mathrm{X}, d\mathrm{Y} \rangle$

Стандартные производные

1. $d\,\langle \mathrm{A}, \mathrm{X} \rangle = \langle \mathrm{A}, d\mathrm{X} \rangle$
2. $d\,\langle \mathrm{A}x, x \rangle = \left\langle \left(\mathrm{A} + \mathrm{A}^\top\right) x, dx \right\rangle$
3. $d\,(\det(\mathrm{X})) = \det(\mathrm{X})\left\langle \mathrm{X}^{-\top}, d\mathrm{X} \right\rangle$
4. $d(\mathrm{X}^{-1}) = -\mathrm{X}^{-1}\,(d\mathrm{X})\,\mathrm{X}^{-1}$

Пример 2. $f(\mathrm{X}) = \ln \det\mathrm{X},\ \mathrm{X} \in \mathbb{S}^n_{++}$

$$df(\mathrm{X}) = d\left(\ln \det\mathrm{X}\right) = \frac{d(\det\mathrm{X})}{\det\mathrm{X}} = \frac{\det(\mathrm{X})\left\langle \mathrm{X}^{-\top}, d\mathrm{X}\right\rangle}{\det\mathrm{X}} = \left\langle \mathrm{X}^{-1}, d\mathrm{X}\right\rangle$$

$$\nabla f(\mathrm{X}) = \mathrm{X}^{-1}$$

Пример 3. $f(\mathrm{X}) = \|\mathrm{X}\|_{\mathrm{F}},\ \mathrm{X} \in \mathbb{R}^{m\times n}$

$$df(\mathrm{X}) = d\left(\|\mathrm{X}\|_{\mathrm{F}}\right) = d\sqrt{\langle \mathrm{X}, \mathrm{X}\rangle} = \frac{1}{2}\frac{1}{\sqrt{\langle \mathrm{X}, \mathrm{X}\rangle}}d\langle \mathrm{X}, \mathrm{X}\rangle = \frac{1}{2}\frac{1}{\|\mathrm{X}\|_{\mathrm{F}}}2\langle \mathrm{X}, d\mathrm{X}\rangle = \|\mathrm{X}\|_{\mathrm{F}}^{-1}\langle \mathrm{X}, d\mathrm{X}\rangle$$

$$\nabla f(\mathrm{X}) = \|\mathrm{X}\|_{\mathrm{F}}^{-1}\mathrm{X}$$

Формула Тейлора

Пусть $f(\mathrm{X})$ - функция матричного аргумента, а Δ - приращение:

$$f(\mathrm{X} + \Delta) = f(\mathrm{X}) + \langle \nabla f(\mathrm{X}), \Delta\rangle + \mathrm{o}(\Delta)$$

Примеры

1. $f(\mathrm{X}) = \mathrm{Tr}\mathrm{X},\ \mathrm{X} \in \mathbb{S}^n_{++}$

$$f(\mathrm{X} + \Delta) = \mathrm{Tr}(\mathrm{X} + \Delta) = \mathrm{Tr}\mathrm{X} + \mathrm{Tr}\Delta = \mathrm{Tr}\mathrm{X} + \langle \mathrm{E}, \Delta\rangle$$

$$\nabla f(\mathrm{X}) = \mathrm{E}$$

2. $f(\mathrm{X}) = \ln \det\mathrm{X},\ \mathrm{X} \in \mathbb{S}^n_{++}$

$$f(\mathrm{X} + \Delta) = \ln\left(\det\left(\mathrm{X} + \Delta\right)\right) = \ln\left(\det\mathrm{X}\left(\mathrm{E} + \mathrm{X}^{-1}\Delta\right)\right) = \ln\left(\det\mathrm{X}\cdot\det\left(\mathrm{E} + \mathrm{X}^{-1}\Delta\right)\right) =$$

$$\ln\left(\det\mathrm{X}\right) + \ln\left(\det\left(\mathrm{E} + \mathrm{X}^{-1}\Delta\right)\right) = f(\mathrm{X}) + \ln\left(1 + \mathrm{Tr}(\mathrm{X}^{-1}\Delta) + \mathrm{o}(\Delta)\right) =$$

$$f(\mathrm{X}) + \mathrm{Tr}(\mathrm{X}^{-1}\Delta) + \mathrm{o}(\Delta) = f(\mathrm{X}) + \left\langle \mathrm{X}^{-1}, \Delta\right\rangle + \mathrm{o}(\Delta)$$

$$\nabla f(\mathrm{X}) = \mathrm{X}^{-1}$$