Методы оптимизации. Семинар 2. Матрично-векторное дифференцирование.

Корнилов Никита Максимович

МФТИ ФИВТ

11 сентября 2025г

Дифференцируемость по Фреше

Работаем в множествах U, V линейных нормированных полных (банаховых) пространств.

Definition

Функция $f : U \to V$ дифференцируема по Фреше во внутренней точке $x \in \mathrm{int}U$, если существует линейный оператор $df(x) : U \to V$, т.ч.

$$f(x+h) = f(x) + df[h] + o(\|h\|), \|h\| \to 0.$$

$df(x)$ называется производной f в точке x.

Если точка x не является внутренней, то понятие дифференцируемости не определено.

Definition

Приращение дифференцируемой функции f в точке x с приращением h называется дифференциалом $df(x)[h] \in V$. Часто направление h обозначают как dx, а дифференциал как $df(x)[dx]$.

Производная по направлению

В одномерном случае $f : \mathbb{R} \to \mathbb{R}$, показателем скорости изменения f в точке x вдоль числовой прямой является производная:

$$f'(x) := \lim_{t \to 0} \frac{f(x+t) - f(x)}{t}.$$

В многомерном случае $f : U \to V$, направлений изменения не два, а бесконечно много. Производные по направлению отвечают за изменения функции вдоль одного направления $h \in U$:

Definition

Производной по направлению $h \in U$ функции $f : U \to V$ во внутреней точке $x \in \mathrm{int}\, U$ называется

$$\frac{\partial f}{\partial h}(x) := \lim_{t \to +0} \frac{f(x+th) - f(x)}{t}. \tag{1}$$

$= \partial f(x)[h] + o(\|h\|)$

Если для любого $h \in U$ определена производная по направлению $\frac{\partial f}{\partial h}(x)$, то функция f дифференцируема по Гато в точке x.

Связь определений

Lemma

Если функция f дифференцируема по Фреше в x, то производная по направлению $\frac{\partial f}{\partial h}(x)$ существует, линейна по h и равна дифференциалу $df(x)[h]$.

В матанализе показывается, что из дифференцируемости по Фреше следует существование производных по всем направлением. Однако обратное неверно. Достаточным условием будет непрерывность всех частных производных.

Градиент по вектору

- В случае $f : \mathbb{R}^n \to \mathbb{R}$ дифференциал $df(x)[dx] \in \mathbb{R}$ можно представить в виде

$$df(x)[dx] = \langle \nabla f(x), dx \rangle, \quad \text{где вектор } \nabla f(x) \in \mathbb{R}^n \text{ зависит от } x.$$

Вектор $\nabla f(x)$ называется **градиентом** функции. Взяв $h = e_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0) \in \mathbb{R}^n$, получим формулу градиента в стандартном базисе

$$\langle \nabla f(x), e_i \rangle = \sum_{j=1}^{n} \nabla f(x)_j \, e_{ij}$$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^\top \in \mathbb{R}^n, \tag{2}$$

где $\frac{\partial f}{\partial x_i}(x) := \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t}$ — частные производные по i-ой координате.

Градиент по матрице

U V

- В случае $f : \mathbb{R}^{n\times m} \to \mathbb{R}$ дифференциал $df(X)[dX] \in \mathbb{R}$ можно представить в виде

$$df(X)[dX] = \langle \nabla f(X), dX \rangle,$$

где матрица $\nabla f(X) \in \mathbb{R}^{n\times m}$ зависит от X.
Матрица $\nabla f(X)$ также называется **градиентом** функции.
Аналогично взяв $h = e_{ij}$, получим формулу градиента в стандартном базисе

$$\nabla f(X) = \left(\frac{\partial f}{\partial x_{ij}}(X)\right)_{i,j} \in \mathbb{R}^{n\times m}. \tag{3}$$

$$e_{ij} = \begin{array}{cc} & \begin{array}{cc} & j \end{array} \\ i & \begin{pmatrix} 0 & 0 \\ & 1 \\ 0 & 0 \end{pmatrix} \end{array}$$

Матрица Якоби

- В случае $f : \overset{U}{\underbrace{\mathbb{R}^n}} \to \overset{V}{\underbrace{\mathbb{R}^m}}$ дифференциал $df(x)[dx] \in \mathbb{R}^m$ можно представить в виде

$$df(x)[dx] = \underbrace{J_f(x)dx}, \quad \text{где матрица } J_f(x) \in \underbrace{\mathbb{R}^{m\times n}} \text{ зависит от } x.$$

Матрица $J_f(x)$ называется матрицей Якоби.
Аналогично взяв $h = e_i$, получим формулу матрицы Якоби в стандартном базисе

$j=1$

$i=1$

$$\boxed{J_f(x) \equiv \frac{\partial f}{\partial x} := \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{ij} \in \mathbb{R}^{m\times n}.} \qquad (4)$$

Заметим, что $\nabla f(x) = J_x^\top$. $\to \mathbb{R}^{1\times n}$

$\mathbb{R}^n$

$J_f = \left(\frac{\partial f}{\partial x_1}, \ldots, \right)$

Таблица канонических видов

$\langle x, Ax \rangle,\quad r \in S^d_{++} \Rightarrow \langle \partial f(x), Ax \rangle$

Вход \ Выход	Скаляр	Вектор
Скаляр	$df(x) = f'(x)dx$ $f'(x)$ скаляр, dx скаляр.	-
Вектор	$df(x) = \langle \nabla f(x), \underbrace{dx} \rangle$ $f(x)$ вектор, dx вектор	$df(x) = J_x \underbrace{dx}$ J_x матрица, dx вектор
Матрица	$df(X) = \langle \nabla f(X), \underbrace{dX} \rangle$ $\nabla f(X)$ мат, dX мат	-

Стоит отметить, что данная таблица верна и для произвольных скалярных произведений, а не только для стандартного.

Подходы к вычислению производных

$$f(x) = f(x_1, \ldots, x_n)$$

1. **Прямой подход**
   Идея: выразить функцию $f(x)$ через скалярную зависимость от каждой координаты x_i и напрямую искать частную производную $\frac{\partial f(x)}{\partial x_i}$.
2. **Дифференциальный подход**
   Идея: Используя правила вычисления дифференциалов, получить канонический вид из Таблицы (8) и выделить градиенты функций, гессиан или матрицу Якоби.

Правила преобразования
$d(\alpha X) = \alpha dX$ $d(AXB) = AdXB$ $d(X+Y) = dX + dY$ $d(X^T) = (dX)^T$
$d(XY) = (dX)Y + X(dY)$ $d\langle X, Y\rangle = \langle dX, Y\rangle + \langle X, dY\rangle$
$d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
$d(g(f(x))) = dg(f)[df(x)]$ $J_{g(f)} = J_g J_f \iff \frac{\partial g}{\partial x} = \frac{\partial g}{\partial f}\frac{\partial f}{\partial x}$
$df(x,y) = J_x dx + J_y dy$

$\alpha \in \mathbb{R}$

$A, B = const$

$\to XY$

$\langle X, Y\rangle$

$\phi: U \to \mathbb{R}$

- скалярная

ϕ, $+$, T

$\frac{X}{\varphi} =$ покоорд

Стоит отметить, что данная таблица верна и для произвольных скалярных произведений и гельдоровых ЛНП.

Таблица стандартных производных
$dA = 0$ $\quad d(ax^2) = 2ax\,dx$ $d\langle A, X\rangle = \langle A, dX\rangle$ $d\langle Ax, x\rangle = \langle (A + A^\top)x, dx\rangle$ $\quad A \in \mathbb{R}^{n\times n}$ $d\,\mathrm{Tr}(X) = \mathrm{Tr}(dX)$
$d(\det(X)) = \det(X)\,\mathrm{Tr}(X^{-1}dX)$
$d(X^{-1}) = -X^{-1}(dX)X^{-1}$

Стоит отметить, что данная таблица верна и для произвольных скалярных произведений и гельдоровых ЛНП.

Hint. Для запоминания формулы $d(X^{-1})$

$$
\begin{aligned}
&I = XX^{-1},\\
&dI = 0 = d(XX^{-1}) = (dX)X^{-1} + Xd(X^{-1}),\\
&d(X^{-1}) = -X^{-1}(dX)X^{-1}.
\end{aligned}
$$

Однако это не является доказательством **существования** дифференциала.

Квадратичная функция

Example

Найдите первый дифференциал и градиент функции $\nabla f(x)$ для

$$f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c,$$

где $A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n, c \in \mathbb{R}$.

$$f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c$$

$$df(x) = d\left(\frac{1}{2}\langle Ax, x\rangle\right) + d\langle b, x\rangle + dc$$

$$d\langle b, x\rangle = \underbrace{\langle db, x\rangle}_{=0} + \langle b, dx\rangle, \qquad dc = 0$$

$$d\left(\frac{1}{2}\langle Ax, x\rangle\right) = \frac{d\langle Ax, x\rangle}{2} = \frac{\langle (A+A^T)x, dx\rangle}{2}$$

$$d\langle Ax, x\rangle = g(x+dx) - g(x) =$$

$$\langle A(x+dx), x+dx\rangle - \langle Ax, x\rangle =$$

$$\langle Ax, x\rangle + \langle A\,dx, x\rangle + \langle Ax, dx\rangle + \langle A\,dx, dx\rangle$$

$$- \langle Ax, x\rangle =$$

$$\underbrace{\langle A\,dx, x\rangle}_{\langle dx, A^T x\rangle} + \langle Ax, dx\rangle + \langle A\,dx, dx\rangle$$

$$= \langle (A^T + A)x, dx\rangle + \langle A\,dx, dx\rangle$$

$$\langle A\,dx, dx\rangle = o(\|dx\|)$$

$$|\langle A\,dx, dx\rangle| \le \|A\,dx\|\,\|dx\| \le \|A\|\,\|dx\|^2$$

$$\|dx\| \to 0$$

$$g(x+dx) = f(x) + \langle (A+A^T)x, dx\rangle + o(\|dx\|)$$

$$d(\underbrace{A}_{}\underbrace{x}_{\mathbb{R}^{n\times 1}}) = A\,dx$$

$$df(x) = \left\langle \frac{(A+A^T)x}{2}, dx \right\rangle + \langle b, dx \rangle$$

$$= \left\langle \underbrace{\frac{(A+A^T)}{2}x + b}, dx \right\rangle = \langle \underline{\nabla f(x)}, dx \rangle$$

$$\nabla f(x) = \frac{(A+A^T)}{2}x + b$$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_i} \right)_{i=1}$$

$$f(\ldots, x_i \ldots) = \frac{\langle Ax, x \rangle}{2} + \langle b, x \rangle + c$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{n} \frac{a_{kj} x_k x_j}{2} + \sum_{k=1}^{n} b_k x_k + c$$

$$\frac{\partial}{\partial x_i}$$

Вторая производная

Пусть $f : U \to V$ дифференцируема в каждой точке $x \in U$. Рассмотрим дифференциал функции $df(x)[h_1]$ при фиксированном приращении $h_1 \in U$ как функцию от x:

$$g(x) = df(x)[h_1].$$

Definition (Вторая производная)

Если в некоторой точке $x \in U$ функция $g : U \to V$ дифференцируема, то второй дифференциал $d^2 f(x)[h_1, h_2] : U \times U \to V$ имеет вид

$$d^2 f(x)[h_1, h_2] := d(df[h_1])(\cancel{x})[h_2]. \tag{5}$$

Можно показать, что $d^2 f(x)[h_1, h_2]$ билинейная функция по h_1, h_2. По аналогии определяется третий дифференциал $d^3 f(x)[h_1, h_2, h_3]$, четвёртый и так далее.

Гессиан

В случае $f : \mathbb{R}^n \to \mathbb{R}$ второй дифференциал, как и любую билинейную функцию, можно представить с помощью матрицы

$$d^2 f(x)[dx_1, dx_2] = \langle \nabla^2 f(x) dx_1, dx_2 \rangle.$$

Матрица $\nabla^2 f(x)$ называется **гессианом** функции. В стандартном базисе гессиан имеет вид

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{ij} \in \mathbb{R}^{n \times n}.$$

Напомним, что для дважды непрерывно дифференцируемой функции гессиан - симметричная матрица. В общем случае, удобно считать гессиан как

$$\nabla^2 f(x) = (J_{\nabla f})^\top.$$

$d(\nabla f): R^n \to R^n$

$J_{\nabla f}$

Квадратичная функция

$$J\nabla f = \left(\frac{\partial \nabla f_i}{\partial x_j}\right) = \left(\frac{\partial^2 f}{\partial x_j \, \partial x_i}\right)$$

Дважды непр. диффер $\Rightarrow$ $(J\nabla f) = \nabla^2 f(x)$

Example

Найдите второй дифференциал и гессиан функции $\nabla^2 f(x)$ для

$$f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c,$$

где $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

$$d f(x)[dx_1] = \left\langle \frac{(A+A^T)}{2}x + b,\ dx_1 \right\rangle$$

$$d\big(df(x)[dx_1]\big) = d\left\langle \frac{(A+A^T)}{2}x + b,\ dx_1 \right\rangle$$

$$= \left\langle d\left(\frac{(A+A^T)}{2}x + b\right),\ dx_1 \right\rangle$$

$$= \left\langle \frac{(A+A^T)}{2}dx_2 + 0,\ dx_1 \right\rangle$$

$$= \left\langle \frac{(A+A^T)}{2} dx_2,\ dx_1 \right\rangle$$

$$= \left\langle \nabla^2 f(x)\, dx_1,\ dx_2 \right\rangle$$

$$= \left\langle dx_2,\ \left(\frac{(A+A^T)}{2}\right)^T dx_1 \right\rangle$$

$$\left(\frac{A+A^T}{2}\right)^T = \frac{A+A^T}{2}$$

$$\nabla f(x) = \frac{(A+A^T)}{2}x + b$$

$$d\nabla f(x) = \frac{(A+A^T)}{2}dx = J_{\nabla f}\, dx$$

$$\nabla^2 f(x) = (J_{\nabla f})^T$$

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \ln\langle Ax, x\rangle$$

где $x \in \mathbb{R}^n, A \in \mathbb{S}^n_{++}$.

$$f(x) = \ln\langle Ax, x\rangle \qquad A \in S^n_{++}$$

$$d f(x) = d(\ln\langle Ax, x\rangle) = \quad y = \langle Ax, x\rangle$$

$$= d(\ln y) = \frac{1}{y}\, dy = \frac{d\langle Ax, x\rangle}{\langle Ax, x\rangle}$$

$$= \frac{\langle (\underbrace{A + A^T}_{=2A})x, dx\rangle}{\langle Ax, x\rangle} = \left\langle \frac{2Ax}{\langle Ax, x\rangle}, dx\right\rangle$$

$$= \langle \nabla f(x), dx\rangle \Rightarrow \nabla f(x) = \frac{2Ax}{\langle Ax, x\rangle}$$

$$d\nabla f(x) = d\left(\frac{2Ax}{\langle Ax, x\rangle}\right) = (d(2Ax)\langle Ax, x\rangle$$

$$- 2Ax\, d\langle Ax, x\rangle) / \langle Ax, x\rangle^2$$

$$= \frac{2A\,dx\,\langle Ax, x\rangle - 2Ax \cdot \langle 2Ax, dx\rangle}{\langle Ax, x\rangle^2} - J_{\nabla f}\, dx$$

$$4Ax\langle Ax, dx\rangle = 4Ax(Ax)^T dx =$$

$$4Axx^T A^T dx$$

$$= \underbrace{\left(\frac{2A}{\langle Ax, x\rangle} - \frac{4\,A x x^T A}{\langle Ax, x\rangle^2} \right)}_{J_{\nabla f} \,=\, \nabla f^2(x)} dx$$

$$\nabla f^2(x) = J_{\nabla f}^T =$$

$$(A x x^T A)^T = A x x^T A$$

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \frac{1}{3}\|x\|_2^3, \quad x \in \mathbb{R}^n \setminus \{0\}.$$

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \ln(1 + \exp(\langle a, x \rangle)),$$

где $a \in \mathbb{R}^n$.

Softmax

Example

Найдите матрицу Якоби функции $s(x) = \mathrm{softmax}(x)$

$$\mathrm{softmax}(x) := \left(\frac{\exp(x_1)}{\sum_{i=1}^{n} \exp(x_i)}, \dots, \frac{\exp(x_n)}{\sum_{i=1}^{n} \exp(x_i)} \right)^\top.$$

Фробениусова норма

Example

Найти градиент $\nabla f(X)$ и дифференциал функции $f(X)$

$$f(X) = \|AX - B\|_F, \quad X \in \mathbb{R}^{k \times n},$$

где $A \in \mathbb{R}^{m \times k}$, $B \in \mathbb{R}^{m \times n}$.

$$f(X) = \|AX-B\|_F =$$

$$= \sqrt{\langle AX-B, AX-B\rangle}$$

$$d\sqrt{\langle AX-B, AX-B\rangle} = d\sqrt{z} =$$

$$z = \langle AX-B, AX-B\rangle = \|AX-B\|_F^2$$

$$\frac{1}{2\sqrt{z}} dz = \frac{d\langle AX-B, AX-B\rangle}{2\|AX-B\|_F} =$$

$$Y = AX-B = \frac{d\langle Y, Y\rangle}{2\|AX-B\|_F} =$$

$$\langle Y, dY\rangle + \langle dY, Y\rangle = 2\langle Y\, dY\rangle =$$

$$= \frac{2\langle Y, dY\rangle}{2\|AX-B\|_F}, \quad \frac{\langle AX-B, d(AX-B)\rangle}{\|AX-B\|_F}$$

$$= \left\langle \frac{AX-B}{\|AX-B\|_F}, A\,dX \right\rangle = \langle \nabla f(X), dX\rangle$$

$$=$$

$$\langle A, B \rangle = Tr(A^T B)$$

$$\Rightarrow \frac{Tr\left(\overbrace{(AX-B)^T \cdot A} \cdot dX \right)}{\|AX-B\|_F} =$$

$$= \frac{\left\langle ((AX-B)^T A)^T, dX \right\rangle}{\|AX-B\|_F} =$$

$$= \left\langle \frac{{}_n \overset{m}{A}{}^{T} \left({}_m \overset{n}{A} \overset{n}{X} - B\right)}{\|AX-B\|_F}, dX \right\rangle$$

$$\nabla f(X) \in R^{n \times n}$$

Практика

Example

Найдите первый дифференциал и градиент $\nabla f(X)$ функции $f(X)$

$$f(X) = \det(AX^{-1}B),$$

где A, X, B – такие матрицы с нужными размерностями, что $AX^{-1}B$ обратима.

$$d\left(\det(AX^{-1}B)\right) =$$

$$\det(AX^{-1}B)\,\mathrm{Tr}\left((AX^{-1}B)^{-1} d(AX^{-1}B)\right)$$

$$= \underbrace{(AX^{-1}B)^{-1}}\ \cancel{=B^{-1}XA^{-1}}$$

$$d(AX^{-1}B) = A\,dX^{-1}B =$$

$$-AX^{-1}dX\,X^{-1}B$$

$$= \det(AX^{-1}B)\,\cdot$$

$$\cdot\,\mathrm{Tr}\left(-(AX^{-1}B)^{-1}AX^{-1}\underbrace{dX}X^{-1}B\right)$$

$$= \langle \mathrm{ot}(X), dX\rangle = \underbrace{\mathrm{Tr}(\mathrm{ot}^{T}dX)}$$

$Tr(A_1 \dots A_n) = Tr(A_n A_1 \dots)$

A_i — произвольные

$Tr((AX^{-1}B)^{-1} AX^{-1} dX X^{-1} B)$

$\cdot \det(AX^{-1}B) =$

$= Tr(X^{-1}B(AX^{-1}B)^{-1} AX^{-1} dX)$

$\nabla f(X) = \det(AX^{-1}B)$

$(X^{-1})^T A^T ((AX^{-1}B)^{-1})^T B^T (X^{-1})^T =$

$\underbrace{(X^{-1})^T = (X^T)^{-1} = X^{-T}}$

$= X^{-T} A^T (AX^{-1}B)^{-T} B^T X^{-T}$

Логарифм определителя

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(X)$ функции $f(X)$
$$f(X) = \ln(\det(X))$$
заданной на множестве $X \in \mathbb{S}^n_{++}$ в пространстве $\mathbb{S}^n$.

Практика

Example

Найти градиент $\nabla f(X)$ и дифференциал функции $f(X)$

$$f(X) = \mathrm{Tr}(AX^\top X).$$