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.

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

$df(x)$ 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 $x \in \operatorname{int} U$ называется

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

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

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

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

- 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

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

Матрица Якоби

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

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

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

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

  Заметим, что $\nabla f(x) = J_x^\top$.

Таблица канонических видов

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

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

Подходы к вычислению производных

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.

Дифференциальное исчисление: правила

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

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

Дифференциальное исчисление: табличные производные

Таблица стандартных производных
$dA = 0$ $d\langle A, X\rangle = \langle A, dX\rangle$ $d\langle Ax, x\rangle = \langle (A + A^\top)x, dx\rangle$ $d\,\mathrm{Tr}(X) = \mathrm{Tr}(dX)$
$d(\det(X)) = \det(X)\,\mathrm{Tr}(X^{-1}dX)$
$d(X^{-1}) = -X^{-1}(dX)X^{-1}$

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

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

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

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

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

Квадратичная функция

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

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

Евклидова норма

Example

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

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

Логистическая регрессия

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 $a \in \mathbb{R}^n$.

Example

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

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

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

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.

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 $f(X)$

$$f(X) = \mathrm{Tr}(AX^\top BX^{-1}), \quad A, B, X \in \mathbb{R}^{n\times n}.$$

Дифференцирование скаляра

Example

Выразите первую $\phi'(\alpha)$ и вторую производную $\phi''(\alpha)$ функции $\phi : \mathbb{R} \to \mathbb{R}$:

$$\phi(\alpha) = f(x + \alpha p), \quad x, p \in \mathbb{R}^n,$$

через градиент ∇f и гессиан $\nabla^2 f$ функции $f : \mathbb{R}^n \to \mathbb{R}$.

Поэлементные функции

Example

Найдите градмент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции $f : \mathbb{R}^n \to \mathbb{R}$:

$$f = h(g(x)),$$

где $g : \mathbb{R}^n \to \mathbb{R}^n$ действует поэлементно

$$g(x) = \sin(x),$$

а функция $h : \mathbb{R}^n \to \mathbb{R}$ суммирует вектор

$$h(u) = \sum_{i=1}^{n} u_i.$$