Автодифференцирование. Матричное дифференцирование.

Семинар

1 сентября 2025

Определение дифференциала

Функция $f : \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема в точке x, если существует матрица $Df(x) \in \mathbb{R}^{m \times n}$, которая удовлетворяет условию:

$$\lim_{x \in \mathrm{dom} f, z \neq x, z \to x} \frac{\|f(z) - f(x) - Df(x)(z - x)\|_2}{\|z - x\|_2} = 0. \qquad (1)$$

Функцию $Df(x)$ будем называть дифференциалом или Якобианом функции f в точке x.

Градиент функции

Для функции вида $f : \mathbb{R}^n \to \mathbb{R}$ транспонированную матрицу $Df(x)^\top$ будем называть градиентом функции f в точке x:

$$\nabla f(x) = Df(x)^\top$$

Компоненты градиента являются частными производными функции f:

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

, где $i = 1, ..., n$.

Матрица Гессе

Рассмотрим дважды дифференцируемую функцию $f : \mathbb{R}^n \to \mathbb{R}$.
Назовем матрицей Гессе функцию f в точке $x \in \operatorname{int} \operatorname{dom} f$:

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1..n} \tag{2}$$

Приближение функции с помощью дифференциала.

Из определения дифференциала (1) сразу следует одно из применений дифференциала функции, а именно ее аппроксимация.
Аппроксимация первого порядка:

$$f(x) + Df(x)(z - x) \tag{3}$$

Аппроксимация второго порядка:

$$f(x) + Df(x)^\top (z - x) + \frac{1}{2}(z - x)^\top D^2 f(x)(z - x)$$

Пример 1.

Задача (Квадратичная функция.)

Квадратичная функция $f : \mathrm{R}^n \to \mathrm{R}$,

$$f(x) = (1/2)x^T Px + q^T x + r,$$

где $P \in \mathrm{S}^n$ *(симметричная квадратная матрица),* $q \in \mathrm{R}^n$*, и* $r \in \mathrm{R}$.

Решение.

$Df(x) = (1/2)(Px + x^\top P) + q^\top = x^\top P + q^\top$
Мы получили вектор-строку $Df(x)$, а её градиент равен

$$\nabla f(x) = Px + q.$$

$\square$

Пример 2.

Задача (Определитель матрицы)

Пусть $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ *– функция* $f(X) := \mathrm{Det}(X)$. *Рассмотрим произвольную точку* $X \in \mathbb{R}^{n\times n}$ *и произвольное приращение аргумента* $H \in \mathbb{R}^{n\times n}$. *Будем предполагать, что матрица* X *обратима.*

Решение.

Выпишем соответствующее приращение функции:

$$f(X+H)-f(X)=\mathrm{Det}(X+H)-\mathrm{Det}(X)$$
$$=\mathrm{Det}(X(I_n+X^{-1}H))-\mathrm{Det}(X)=\mathrm{Det}(X)\big(\mathrm{Det}(I_n+X^{-1}H)-1\big). \tag{4}$$

Решение (продолжение).

Рассмотрим отдельно $\mathrm{Det}(I_n + X^{-1}H)$:

$$\mathrm{Det}(I_n + X^{-1}H) = \prod_{i=1}^{n}[1 + \lambda_i(X^{-1}H)] =$$

$$1 + \sum_{i=1}^{n} \lambda_i(X^{-1}H) + \left(\sum_{1 \le i < j \le n} \lambda_i(X^{-1}H)\lambda_j(X^{-1}H) + \cdots \right),$$

где многоточие скрывает сумму всевозможных троек $\lambda_i(X^{-1}H)\lambda_j(X^{-1}H)\lambda_k(X^{-1}H)$, всевозможных четверок и т.д.

Решение (продолжение).

Заметим, что выражение, стоящее в скобках, представляет из себя величину $o(\|H\|)$, поэтому

$$\mathrm{Det}(I_n + X^{-1}H) = \prod_{i=1}^{n}[1 + \lambda_i(X^{-1}H)] = 1 + \mathrm{Tr}(X^{-1}H) + o(\|H\|)$$

Подставим полученное выражение обратно в (4) и получим, что для любой обратимой матрицы $X \in \mathbb{R}^{n\times n}$ функция f дифференцируема в точке X с производной

$$df(X)[H] = \mathrm{Det}(X)\mathrm{Tr}(X^{-1}H) = \mathrm{Det}(X)\langle X^{-T}, H\rangle.$$

$\square$

Пример 3.

Задача

$$f(X) = \log\det X,$$

где $\operatorname{dom} f = \mathbf{S}^n_{++}$, $f : \mathbf{S}^n \to \mathbf{R}$

Пример 3.

Решение.

Найдём первую аппроксимацию f в точке $X \in \mathbf{S}^n_{++}$ (3). Пусть $Z \in \mathbf{S}^n_{++}$ близка к X, и пусть $\Delta X = Z - X$. Тогда

$$\begin{aligned}
\log\det Z &= \log\det(X + \Delta X) \\
&= \log\det\left(X^{1/2}(I + X^{-1/2}\Delta X X^{-1/2})X^{1/2}\right) \\
&= \log\det X + \log\det(I + X^{-1/2}\Delta X X^{-1/2}) \\
&= \log\det X + \sum_{i=1}^{n} \log(1 + \lambda_i),
\end{aligned}$$

где λ_i — i-ое собственное значение матрицы $X^{-1/2}\Delta X X^{-1/2}$.

Пример 3.

Решение.

Теперь воспользуемся тем, что ΔX мало, это эквивалентно тому, что собственные числа λ_i так же малы, поэтому $\log(1+\lambda_i) \approx \lambda_i$. Отсюда получаем:

$$\log\det Z \approx \log\det X + \sum_{i=1}^{n} \lambda_i$$

$$= \log\det X + \mathrm{tr}(X^{-1/2}\Delta X X^{-1/2})$$

$$= \log\det X + \mathrm{tr}(X^{-1}\Delta X)$$

$$= \log\det X + \mathrm{tr}\left(X^{-1}(Z-X)\right),$$

Пример 3.

Решение.

Отсюда заключаем, что аппроксимация первого порядка (3) функции f в X имеет вид

$$f(Z) \approx f(X) + \mathrm{tr}(X^{-1}(Z - X))$$

Заметим, что второе слагаемое в правой части равенства выше - это стандартное скалярное умножение матриц (Frobenius inner product) X^{-1} и $(Z - X)$, и отсюда заключаем

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

$\square$

Дифференцирование сложной функции (chain rule).

Пусть $f : \mathbb{R}^n \to \mathbb{R}^m$ дифференцируема в точке $x \in \text{int dom } f$ и функция $g : \mathbb{R}^m \to \mathbb{R}^p$ дифференцируема в точке $f(x) \in \text{int dom } g$. Определим композицию функций $h : \mathbb{R}^m \to \mathbb{R}^p$ как $h(z) = g(f(z))$. Тогда h дифференцируема в точке x с дифференциалом вида:

$$Dh(x) = Dg(f(x))Df(x) \tag{5}$$

В качестве примера предположим, что $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$, и $h(x) = g(f(x))$. Транспонирование $Dh(x) = Dg(f(x))Df(x)$ даёт

$$\nabla h(x) = g'(f(x))\nabla f(x)$$

Пример 4.

Задача

Продифференцируем функцию $f : \mathbb{R}^n \to \mathbb{R}$ *с областью определения* $\operatorname{dom} f = \mathbb{R}^n$, *заданную формулой:*

$$f(x) = \log \sum_{i=1}^{m} \exp(a_i^T x + b_i),$$

, *где* $a_i \in \mathbb{R}^n$ *и* $b_i \in \mathbb{R}$ *для всех* $i = 1, \ldots, m$.

Пример 4.

Решение.

Заметим, что данная функция является композицией аффинной функции $Ax + b$, где $A \in \mathbb{R}^{m \times n}$ со строками $a_1, \ldots, a_m$ и зададим $g : \mathbb{R}^m \to \mathbb{R}$ формулой $g(y) = \log \sum_{i=0}^{m} \exp y_i$ Тогда градиент функции $g(y)$ вычисляется по формуле (5):

$$\nabla g(y) = \frac{1}{\sum_{i=1}^{m} \exp y_i} \begin{bmatrix} \exp y_1 \\ \vdots \\ \exp y_m \end{bmatrix}.$$

Следовательно, используя формулу композиции, получаем градиент функции $f(x)$:

$$\nabla f(x) = \frac{1}{1^T z} A^T z,$$

где $z_i = \exp(a_i^T x + b_i)$ для $i = 1, \ldots, m$. $\square$

Пример 5.

Задача

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

$$f(x) = \log\det(F_0 + x_1 F_1 + \cdots + x_n F_n),$$

где $F_0, \ldots, F_n \in \mathbb{S}^p$, *и*

$$\mathrm{dom}\, f = \{x \in \mathbb{R}^n \mid F_0 + x_1 F_1 + \cdots + x_n F_n > 0\}.$$

Пример 5.

Решение.

Функция f является композицией аффинного отображения из $x \in \mathbb{R}^n$ в $F_0 + x_1 F_1 + \cdots + x_n F_n \in \mathbb{S}^p$ и функции $\log\det X$. Используя правило дифференцирования сложной функции, получаем:

$$\frac{\partial f(x)}{\partial x_i} = \mathrm{tr}(F_i \nabla \log\det(F)) = \mathrm{tr}(F^{-1} F_i),$$

где $F = F_0 + x_1 F_1 + \cdots + x_n F_n$. Таким образом, градиент функции равен:

$$\nabla f(x) = \begin{bmatrix} \mathrm{tr}(F^{-1} F_1) \\ \vdots \\ \mathrm{tr}(F^{-1} F_n) \end{bmatrix}.$$

$\square$

Пример 5.

Решение.

Функция f является композицией аффинного отображения из $x \in \mathbb{R}^n$ в $F_0 + x_1 F_1 + \cdots + x_n F_n \in \mathbb{S}^p$ и функции $\log\det X$. Используя правило дифференцирования сложной функции, получаем:

$$\frac{\partial f(x)}{\partial x_i} = \mathrm{tr}(F_i \nabla \log\det(F)) = \mathrm{tr}(F^{-1} F_i),$$

где $F = F_0 + x_1 F_1 + \cdots + x_n F_n$. Таким образом, градиент функции равен:

$$\nabla f(x) = \begin{bmatrix} \mathrm{tr}(F^{-1} F_1) \\ \vdots \\ \mathrm{tr}(F^{-1} F_n) \end{bmatrix}.$$

$\square$

Ссылки I

Boyd, Stephen P., and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004. (Appendix A.4)