



Lecture 8

Math Foundations Team



BITS Pilani

Pilani | Dubai | Goa | Hyderabad



- ▶ Till now we have discussed about Taylor/Maclaurian series, Partial Derivatives and Gradients.
- ▶ Now we are interested in Higher order Derivatives.
- ▶ Multivariate Taylor Series and its uses in the expansion of a function with multivariables.



Consider a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

Notations for Higher-Order Partial Derivatives:

$\frac{\partial^2 f}{\partial x^2}$: Second Partial Derivative of x w.r.t. x

$\frac{\partial^n f}{\partial x^n}$: n^{th} Partial Derivative of x w.r.t. x

$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)$: is the partial derivative obtained by first partial differentiating with respect to x and then with respect to y .

$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)$: is the partial derivative obtained by first partial differentiating by y and then x .

The Hessian is the collection of all second-order partial derivatives. If $f(x, y)$ is a twice (continuously) differentiable function, then $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ i.e., the order of differentiation does not matter, and the corresponding Hessian matrix

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

is symmetric. The Hessian is denoted as $\nabla_{x,y}^2 f(x, y)$



The gradient ∇f of a function f is often used for a locally linear approximation of f around x_0 :

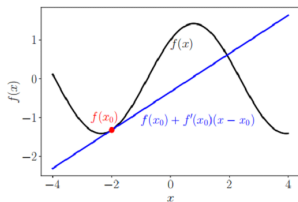
$$f(x) \approx f(x_0) + (\nabla_x f)(x_0)(x - x_0) \quad (1)$$

Here $(\nabla_x f)(x_0)$ is the gradient of f with respect to x , evaluated at x_0 . Figure illustrates the linear approximation of a function f at an input x_0 . The original function is approximated by a straight line.

Linearization and Multivariate Taylor Series...



This approximation is locally accurate, but the farther we move away from x_0 the worse the approximation gets. Equation (1) is a special case of a multivariate Taylor series expansion of f at x_0 , where we consider only the first two terms. We discuss the more general case in the following, which will allow for better approximations.



Linear approximation of a function. The original function f is linearized at $x_0 = -2$ using a first-order Taylor series expansion.



Consider a function $f : \mathbb{R}^D \rightarrow \mathbb{R}$, $x \rightarrow f(x)$,

$$x \in \mathbb{R}^D$$

that is smooth at x_0 . When we define the difference vector $\delta := x - x_0$ the multivariate Taylor series of f at (x_0) is defined as multivariate Taylor series

$$f(x) = \sum_{k=0}^{\infty} \frac{D_x^k f(x_0)}{k!} \delta^k \quad (2)$$

where $D_x^k f(x_0)$ is the k^{th} (total) derivative of f with respect to x , evaluated at x_0 .



The Taylor polynomial of degree n of Taylor polynomial f at x_0 contains the first $n + 1$ components of the series in (2) and is defined as

$$T_n(x) = \sum_{k=0}^n \frac{D_x^k f(x_0)}{k!} \delta^k \quad (3)$$

In (2) and (3), we used the slightly sloppy notation of δ^k , which is not defined for vectors

$$x \in \mathbb{R}^D,$$

$D > 1$, and $k > 1$. Note that both $D_x^k f$ and δ^k are k^{th} order tensors, i.e., k -dimensional arrays.

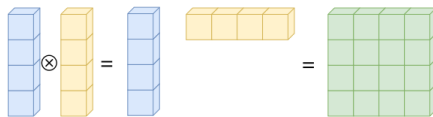


k th-order tensor $\boldsymbol{\delta}^k \in \mathbb{R}^{\overbrace{D \times D \times \dots \times D}^{k \text{ times}}}$ is obtained as a k -fold outer product, denoted by \otimes , of the vector $\boldsymbol{\delta} \in \mathbb{R}^D$. For example,

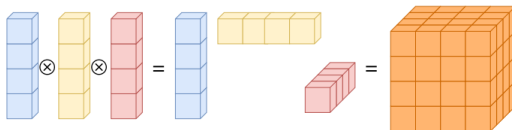
$$\boldsymbol{\delta}^2 := \boldsymbol{\delta} \otimes \boldsymbol{\delta} = \boldsymbol{\delta} \boldsymbol{\delta}^\top, \quad \boldsymbol{\delta}^2[i, j] = \delta[i] \delta[j]$$

$$\boldsymbol{\delta}^3 := \boldsymbol{\delta} \otimes \boldsymbol{\delta} \otimes \boldsymbol{\delta}, \quad \boldsymbol{\delta}^3[i, j, k] = \delta[i] \delta[j] \delta[k].$$

Taylor Polynomial...



(a) Given a vector $\delta \in \mathbb{R}^4$, we obtain the outer product $\delta^2 := \delta \otimes \delta = \delta \delta^T \in \mathbb{R}^{4 \times 4}$ as a matrix.



(b) An outer product $\delta^3 := \delta \otimes \delta \otimes \delta \in \mathbb{R}^{4 \times 4 \times 4}$ results in a third-order tensor ("three-dimensional matrix"), i.e., an array with three indexes.

$$D_x^k f(x_0) \delta^k = \sum_{i_1=1}^D \cdots \sum_{i_k=1}^D D_x^k f(x_0) [i_1, \dots, i_k] \delta[i_1] \cdots \delta[i_k]$$

In general, we obtain the terms in the Taylor series, where $D_x^k f(x_0) \delta^k$ contains k^{th} order polynomials. Now that we defined the Taylor series for vector fields, let us explicitly write down the first terms $D_x^k f(x_0) \delta^k$ of the Taylor series expansion for

$$k = 0, \dots, 3 \text{ and } \delta := x - x_0:$$

$$k = 0 : D_x^0 f(x_0) \delta^0 = f(x_0) \in \mathbb{R}$$

$$k = 1 : D_x^1 f(x_0) \delta^1 = \underbrace{\nabla_x f(x_0)}_{1 \times D} \underbrace{\delta}_{D \times 1} = \sum_{i=1}^D \nabla_x f(x_0)[i] \delta[i] \in \mathbb{R}$$

$$k = 2 : D_x^2 f(x_0) \delta^2 = \text{tr} \left(\underbrace{H(x_0)}_{D \times D} \underbrace{\delta}_{D \times 1} \underbrace{\delta^\top}_{1 \times D} \right) = \delta^\top H(x_0) \delta$$

$$= \sum_{i=1}^D \sum_{j=1}^D H[i, j] \delta[i] \delta[j] \in \mathbb{R}$$

$$k = 3 : D_x^3 f(x_0) \delta^3 = \sum_{i=1}^D \sum_{j=1}^D \sum_{k=1}^D D_x^3 f(x_0)[i, j, k] \delta[i] \delta[j] \delta[k] \in \mathbb{R}$$

Here, $H(x_0)$ is the Hessian of f evaluated at x_0 .

Taylor Series Expansion of a Function with Two Variables



Consider the function $f(x, y) = x^2 + 2xy + y^3$.

We want to compute the Taylor series expansion of f at $(x_0, y_0) = (1, 2)$.

Before we start, let us discuss what to expect: The function in $f(x, y)$ is a polynomial of degree 3. We are looking for a Taylor series expansion, which itself is a linear combination of polynomials. Therefore, we do not expect the Taylor series expansion to contain terms of fourth or higher order to express a third-order polynomial. This means that it should be sufficient to determine the first four terms of $f(x) = \sum_{k=0}^{\infty} \frac{D_x^k f(x_0)}{k!} \delta^k$ for an exact alternative representation of $f(x, y)$. To determine the Taylor series expansion, we start with the constant term and the first-order derivatives, which are given by $f(1, 2) = 13$

Taylor Series Expansion of a Function with Two Variables...



$$\begin{aligned}\frac{\partial f}{\partial x} &= 2x + 2y \implies \frac{\partial f}{\partial x}(1, 2) = 6 \\ \frac{\partial f}{\partial y} &= 2x + 3y^2 \implies \frac{\partial f}{\partial y}(1, 2) = 14.\end{aligned}$$

Therefore, we obtain

$$D_{x,y}^1 f(1, 2) = \nabla_{x,y} f(1, 2) = \left[\frac{\partial f}{\partial x}(1, 2) \quad \frac{\partial f}{\partial y}(1, 2) \right] = [6 \quad 14] \in \mathbb{R}^{1 \times 2}$$

such that

$$\frac{D_{x,y}^1 f(1, 2)}{1!} \delta = [6 \quad 14] \begin{bmatrix} x - 1 \\ y - 2 \end{bmatrix} = 6(x - 1) + 14(y - 2).$$

Therefore, we obtain

$$D_{x,y}^1 f(1, 2) = \nabla_{x,y} f(1, 2) = \left[\frac{\partial f}{\partial x}(1, 2) \quad \frac{\partial f}{\partial y}(1, 2) \right] = [6 \quad 14] \in \mathbb{R}^{1 \times 2}$$

Taylor Series Expansion of a Function with Two Variables...



When we collect the second-order partial derivatives, we obtain the Hessian

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6y \end{bmatrix},$$

such that

$$\mathbf{H}(1, 2) = \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

Therefore, the next term of the Taylor-series expansion is given

$$\begin{aligned} \frac{D_{x,y}^2 f(1, 2)}{2!} \boldsymbol{\delta}^2 &= \frac{1}{2} \boldsymbol{\delta}^\top \mathbf{H}(1, 2) \boldsymbol{\delta} \\ &= \frac{1}{2} [x-1 \quad y-2] \begin{bmatrix} 2 & 2 \\ 2 & 12 \end{bmatrix} \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\ &= (x-1)^2 + 2(x-1)(y-2) + 6(y-2)^2. \end{aligned}$$

Taylor Series Expansion of a Function with Two Variables...



The third-order derivatives are obtained as

$$\begin{aligned} D_{x,y}^3 f &= \begin{bmatrix} \frac{\partial H}{\partial x} & \frac{\partial H}{\partial y} \end{bmatrix} \in \mathbb{R}^{2 \times 2 \times 2}, \\ D_{x,y}^3 f[:, :, 1] &= \frac{\partial H}{\partial x} = \begin{bmatrix} \frac{\partial^3 f}{\partial x^3} & \frac{\partial^3 f}{\partial x^2 \partial y} \\ \frac{\partial^3 f}{\partial x \partial y \partial x} & \frac{\partial^3 f}{\partial x \partial y^2} \end{bmatrix}, \\ D_{x,y}^3 f[:, :, 2] &= \frac{\partial H}{\partial y} = \begin{bmatrix} \frac{\partial^3 f}{\partial y \partial x^2} & \frac{\partial^3 f}{\partial y \partial x \partial y} \\ \frac{\partial^3 f}{\partial y^2 \partial x} & \frac{\partial^3 f}{\partial y^3} \end{bmatrix}. \end{aligned}$$

Taylor Series Expansion of a Function with Two Variables...

innovate

achieve

lead

Since most second-order partial derivatives in the Hessian, are constant, the only nonzero third-order partial derivative is $\frac{\partial^3 f}{\partial y^3} = 6 \implies \frac{\partial^3 f}{\partial y^3}(1, 2) = 6$ Higher-order derivatives and the mixed derivatives of degree 3 (e.g., $\frac{\partial^3 f}{\partial x^2 \partial y}$) vanish, such that

$$D_{x,y}^3 f[:, :, 1] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad D_{x,y}^3 f[:, :, 2] = \begin{bmatrix} 0 & 0 \\ 0 & 6 \end{bmatrix}$$

and

$$\frac{D_{x,y}^3 f(1, 2)}{3!} \delta^3 = (y - 2)^3,$$

Taylor Series Expansion of a Function with Two Variables...



which collects all cubic terms of the Taylor series. Overall, the (exact) Taylor series expansion of f at $(x_0, y_0) = (1, 2)$ is

$$\begin{aligned} f(x) &= f(1, 2) + D_{x,y}^1 f(1, 2) \delta + \frac{D_{x,y}^2 f(1, 2)}{2!} \delta^2 + \frac{D_{x,y}^3 f(1, 2)}{3!} \delta^3 \\ &= f(1, 2) + \frac{\partial f(1, 2)}{\partial x} (x - 1) + \frac{\partial f(1, 2)}{\partial y} (y - 2) \\ &\quad + \frac{1}{2!} \left(\frac{\partial^2 f(1, 2)}{\partial x^2} (x - 1)^2 + \frac{\partial^2 f(1, 2)}{\partial y^2} (y - 2)^2 \right. \\ &\quad \left. + 2 \frac{\partial^2 f(1, 2)}{\partial x \partial y} (x - 1)(y - 2) \right) + \frac{1}{6} \frac{\partial^3 f(1, 2)}{\partial y^3} (y - 2)^3 \\ &= 13 + 6(x - 1) + 14(y - 2) \\ &\quad + (x - 1)^2 + 6(y - 2)^2 + 2(x - 1)(y - 2) + (y - 2)^3. \end{aligned}$$