Multivariate Taylor Series

Jack David Carson - February 2, 2025

The gradient of a function can be used as a locally linear approximation of function f around $\boldsymbol{x}_0$ such as

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_0) + \boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)$$

However, this only works near the center at $\boldsymbol{x}$. We can generalize this with the **Multivariate Taylor Series**

Multivariate Taylor Series DEFINITION 0.0.1

Consider a function

$$\begin{aligned} f : \mathbb{R}^D &\to \mathbb{R} \\ \boldsymbol{x} &\mapsto f(\boldsymbol{x}) \end{aligned}$$

that is smooth at $\boldsymbol{x}_0$. If we define the difference vector $\boldsymbol{\delta} \coloneqq \boldsymbol{x} - \boldsymbol{x}_0$ then the taylor series of f at $\boldsymbol{x}_0$ is defined as

$$f(\boldsymbol{x}) = \sum_{k=0}^{\infty} \frac{D_{\boldsymbol{x}}^k f(\boldsymbol{x}_0)}{k!} \boldsymbol{\delta}^k$$

where $D_{\boldsymbol{x}}^k f(\boldsymbol{x}_0)$ is the k-th total derivative of f with respect to $\boldsymbol{x}$ at $\boldsymbol{x}_0$

Finally $\boldsymbol{\delta}^k$ is actually specially defined as the k-th order tensor $\boldsymbol{\delta}^k \in \mathbb{R}^{D \times D \times \cdots \times D}$ k times. It is obtained as the k-fold outer product by the $\otimes$ operator.

This last point is not as confusing as it sounds. It is mostly a formalization of an idea intuitive in programming. E.g.

$$\boldsymbol{\delta}^2 = \boldsymbol{\delta} \otimes \boldsymbol{\delta} = \boldsymbol{\delta}\boldsymbol{\delta}^{\mathsf{T}} \text{ (Outer Product)}$$

where $\boldsymbol{\delta}^2[i, j] = \delta[i]\delta[j]$ as we expect in programming. So,

$$\begin{aligned} k = 0 \quad &: \quad D_{\boldsymbol{x}}^0 f(\boldsymbol{x}_0)\boldsymbol{\delta}^0 = f(\boldsymbol{x}_0) \in R \\ k = 1 \quad &: \quad D_{\boldsymbol{x}}^0 f(\boldsymbol{x_0}\boldsymbol{\delta}^1) = \underbrace{\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}_0)}_{1 \times D} \underbrace{\boldsymbol{\delta}}_{D \times 1} = \sum_{i=1}^{D} \boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x_0})[i]\delta[i] \in \mathbb{R} \end{aligned}$$