Calculus II

Essential Terminology for Multivariate Vector Functions

Uwe Naumann

Informatik 12:
Software and Tools for Computational Engineering

RWTH Aachen University

Contents

Outline

Objective

- Introduction to essential vector calculus

Learning Outcomes

- You will understand
  - continuity
  - differentiability
  - gradient, Jacobian, Hessian
  - chain rule of differential calculus
  - partial, total, directional derivative
  - directional derivative DAG $\times$ vector product
  - Taylor series.

Outline

Let $\boldsymbol{R}^n$ be the domain of the multivariate scalar function $f : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$. The function f is continuous at a point $\tilde{\mathbf{x}} \in \boldsymbol{R}^n$ if

$$\lim_{\mathbf{x} \rightarrow \tilde{\mathbf{x}}} f(\mathbf{x}) = f(\tilde{\mathbf{x}}) \quad .$$

A multivariate vector function

$$F = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m$$

is continuous if and only if all its component functions f_i, $i = 1, \ldots, m$ are continuous.

Outline

Let $\boldsymbol{R}^n$ be the domain of the multivariate scalar function $f : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$. The function f is differentiable at point $\tilde{\mathbf{x}} \in \boldsymbol{R}^n$ if there is a vector $f' \in \boldsymbol{R}^n$ such that

$$f(\tilde{\mathbf{x}} + \Delta\mathbf{x}) = f(\tilde{\mathbf{x}}) + f' \cdot \Delta\mathbf{x} + r$$

with asymptotically vanishing remainder $r = r(\tilde{\mathbf{x}}, \Delta\mathbf{x}) \in \boldsymbol{R}$, such that

$$\lim_{\Delta\mathbf{x}\rightarrow 0} \frac{r}{\|\Delta\mathbf{x}\|_2} = 0 \;, \quad \text{where } \|\mathbf{v}\|_2 \equiv \sqrt{\mathbf{v}^T \cdot \mathbf{v}} = \sqrt{\sum_{i=0}^{n-1} v_i^2}$$

denotes the Euclidean norm of the vector $\mathbf{v} \in \boldsymbol{R}^n$.

$$f' = f'(\mathbf{x}) = \frac{df}{d\mathbf{x}}(\mathbf{x}) : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^n$$

is called the gradient of f.

Let $f : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ be differentiable at $\tilde{\mathbf{x}} \in \boldsymbol{R}^n$. Then

$$f'(\tilde{\mathbf{x}}) = \begin{pmatrix} \frac{dy}{dx_0} \\ \vdots \\ \frac{dy}{dx_{n-1}} \end{pmatrix}$$

where $y = f(\mathbf{x})$ and

$$\frac{dy}{dx_i} = \frac{dy}{dx_i}(\tilde{\mathbf{x}}) = \lim_{\Delta x \rightarrow \pm 0} \frac{f(\tilde{\mathbf{x}} + \Delta x \cdot \mathbf{e}^i) - f(\tilde{\mathbf{x}})}{\Delta x} < \infty$$

with the i-th Cartesian basis vector in $\boldsymbol{R}^n$ denoted by $\mathbf{e}^i$.

Let

$$y = f(\mathbf{x}) = e^{\sin\left(\|\mathbf{x}\|_2^2\right)} = e^{\sin\left(\mathbf{x}^T \cdot \mathbf{x}\right)} = e^{\sin\left(\sum_{i=0}^{n-1} x_i^2\right)}$$

Differentiation wrt. $\mathbf{x}$ yields the gradient

$$f'(\mathbf{x}) = \left(2 \cdot x_j \cdot \cos\left(\sum_{i=0}^{n-1} x_i^2\right) \cdot e^{\sin\left(\sum_{i=0}^{n-1} x_i^2\right)}\right)_{j=0,\ldots,n-1}$$

Differentiability

$F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m$

Let $\boldsymbol{R}^n$ be the domain of the multivariate vector function $F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m$. The function F is differentiable at point $\tilde{\mathbf{x}} \in \boldsymbol{R}^n$ if there is a matrix $F' \in \boldsymbol{R}^{m \times n}$ such that

$$F(\tilde{\mathbf{x}} + \Delta\mathbf{x}) = F(\tilde{\mathbf{x}}) + F' \cdot \Delta\mathbf{x} + \mathbf{r}$$

with asymptotically vanishing remainder $\mathbf{r} = \mathbf{r}(\tilde{\mathbf{x}}, \Delta\mathbf{x}) \in \boldsymbol{R}^m$, such that

$$\lim_{\Delta\mathbf{x} \rightarrow 0} \frac{\|\mathbf{r}\|_2}{\|\Delta\mathbf{x}\|_2} = 0 \quad .$$

The matrix

$$F' = F'(\mathbf{x}) = \frac{dF}{d\mathbf{x}}(\mathbf{x}) : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^{m \times n}$$

is called the Jacobian of F.

Let $F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^n$ be differentiable at $\tilde{\mathbf{x}} \in \boldsymbol{R}^n$. Then

$$F'(\tilde{\mathbf{x}}) = \begin{pmatrix} \frac{dy_0}{dx_0} & \cdots & \frac{dy_0}{dx_{n-1}} \\ \vdots & & \vdots \\ \frac{dy_{n-1}}{dx_0} & \cdots & \frac{dy_{n-1}}{dx_{n-1}} \end{pmatrix}$$

where $y_j = F_j(\mathbf{x})$ and

$$\frac{dy_j}{dx_i} = \frac{dy_j}{dx_i}(\tilde{\mathbf{x}}) = \lim_{\Delta x \rightarrow \pm 0} \frac{F_j(\tilde{\mathbf{x}} + \Delta x \cdot \mathbf{e}^i) - F_j(\tilde{\mathbf{x}})}{\Delta x} < \infty \; .$$

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable at $\tilde{\mathbf{x}} \in \mathbb{R}^n$. It is twice differentiable at $\tilde{\mathbf{x}}$ if $f' : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is differentiable at $\tilde{\mathbf{x}}$.

The matrix

$$f''(\tilde{\mathbf{x}}) = \begin{pmatrix} \frac{d^2y}{dx_0^2} & \dots & \frac{d^2y}{dx_0\,dx_{n-1}} \\ \vdots & & \vdots \\ \frac{d^2y}{dx_{n-1}\,dx_0} & \dots & \frac{d^2y}{dx_{n-1}^2} \end{pmatrix}$$

is called the Hessian matrix of f at point $\tilde{\mathbf{x}}$.

If f' is continuous [at some point, within some subdomain], then f is called continuously differentiable [at this point, within this subdomain].

If f is twice continuously differentiable at $\tilde{\mathbf{x}}$, then its Hessian is symmetric, i.e, $f''(\tilde{\mathbf{x}}) = f''(\tilde{\mathbf{x}})^T$.

Hessians of multivariate vector functions $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are 3-tensors. So are third derivatives of f, and so forth.

Outline

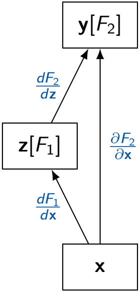
Chain Rule

$F : R^n \rightarrow R^m$

Let $\mathbf{y} = F(\mathbf{x}) : R^n \rightarrow R^m$ be such that

$$\mathbf{y} = F(\mathbf{x}) = F_2(F_1(\mathbf{x}), \mathbf{x}) = F_2(\mathbf{z}, \mathbf{x})$$

with (continuously) differentiable $F_1 : R^n \rightarrow R^p$ and $F_2 : R^p \times R^n \rightarrow R^m$.



Then F is continuously differentiable over R^n and

$$\frac{dF}{d\mathbf{x}}(\tilde{\mathbf{x}}) = \frac{dF_2}{d\mathbf{x}}(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}) = \frac{dF_2}{d\mathbf{z}}(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}) \cdot \frac{dF_1}{d\mathbf{x}}(\tilde{\mathbf{x}}) + \frac{\partial F_2}{\partial \mathbf{x}}(\tilde{\mathbf{z}}, \tilde{\mathbf{x}})$$

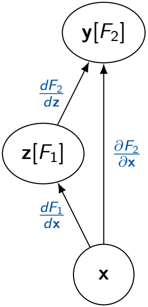
for all $\tilde{\mathbf{x}} \in R^n$ and $\tilde{\mathbf{z}} = F_1(\tilde{\mathbf{x}})$.

Notation: $\frac{\partial F_2}{\partial \mathbf{x}}$ partial derivative; $\frac{dF_2}{d\mathbf{x}}$ total derivative

A composite function $\mathbf{y} = F(\mathbf{x})$ such as

$$\mathbf{z} = F_1(\mathbf{x})$$
$$\mathbf{y} = F_2(\mathbf{z}, \mathbf{x})$$

induces a directed acyclic graph (DAG) $G = (V, E)$ with vertices in V representing variables (e.g, $\mathbf{x}$, $\mathbf{z}$ and $\mathbf{y}$) and with local (partial) derivatives associated with the edges in E.



$$F'(\mathbf{x}) \equiv \frac{d\mathbf{y}}{d\mathbf{x}} = \sum_{\text{path}\in\text{DAG}} \prod_{(i,j)\in\text{path}} \frac{\partial \mathbf{v}_j}{\partial \mathbf{v}_i} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} + \frac{d\mathbf{y}}{d\mathbf{z}} \cdot \frac{d\mathbf{z}}{d\mathbf{x}}$$

The directional derivative (Jacobian × vector product)

$$\mathbf{y}^{(1)} = \frac{dF}{d\mathbf{x}}(\tilde{\mathbf{x}}) \cdot \mathbf{x}^{(1)}$$

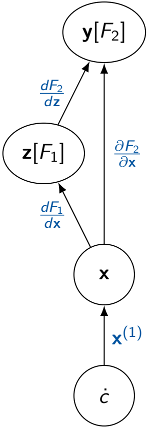
of $\mathbf{y} = F(\mathbf{x})$ at $\tilde{\mathbf{x}}$ can be represented as the derivative of $\mathbf{y} = \mathbf{y}(\mathbf{x}(\dot{c}))$ with respect to (wrt.) an auxiliary variable $\dot{c} \in \boldsymbol{R}$ at $\tilde{\mathbf{x}}$ such that

$$\frac{d\mathbf{x}}{d\dot{c}} = \mathbf{x}^{(1)} .$$

The chain rule yields

$$\mathbf{y}^{(1)} \equiv \frac{dF}{d\dot{c}} = \frac{dF}{d\mathbf{x}}(\tilde{\mathbf{x}}) \cdot \frac{d\mathbf{x}}{d\dot{c}} = \frac{dF}{d\mathbf{x}}(\tilde{\mathbf{x}}) \cdot \mathbf{x}^{(1)} .$$

Directional derivatives are marked with the superscript $*^{(1)}$.



Outline

In scientific computing the multivariate vector functions

$$F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m : \mathbf{y} = F(\mathbf{x})$$

of interest are implemented as differentiable computer programs.

Such programs decompose into sequences of $q = p + m$ differentiable elemental functions φ_j evaluated as a single assignment code[1]

$$v_j = \varphi_j(v_k)_{k \prec j} \quad \text{for } j = n, \ldots, n + q - 1$$

and where $v_i = x_i$ for $i = 0, \ldots, n-1$, $y_k = v_{n+p+k}$ for $k = 0, \ldots, m-1$ and $k \prec j$ if v_k is an argument of φ_j.

A DAG $G = (V, E)$ is induced. Partial derivatives of the elemental functions wrt. their arguments are associated with the corresponding edges.

[1] Variables are written once.

Software and Tools for Computational Engineering — RWTH AACHEN UNIVERSITY

$$y = f(\mathbf{x}) = e^{\sin\left(\|\mathbf{x}\|_2^2\right)} = e^{\sin\left(\mathbf{x}^T \cdot \mathbf{x}\right)} = e^{\sin\left(\sum_{i=0}^{n-1} x_i^2\right)}, \quad n = 2$$

$$v_0 = x_0$$

$$v_1 = x_1$$

$$v_2 = v_0^2; \qquad \frac{dv_2}{dv_0} = 2 \cdot v_0$$

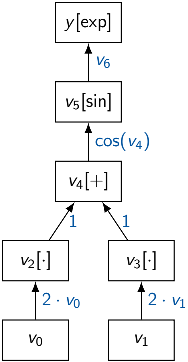
$$v_3 = v_1^2; \qquad \frac{dv_3}{dv_1} = 2 \cdot v_1$$

$$v_4 = v_2 + v_3; \qquad \frac{dv_4}{dv_2} = \frac{dv_4}{dv_3} = 1$$

$$v_5 = \sin(v_4); \qquad \frac{dv_5}{dv_4} = \cos(v_4)$$

$$v_6 = e^{v_5}; \qquad \frac{dv_6}{dv_5} = v_6$$

$$y = v_6$$



Outline

The DAG $G = G(\tilde{\mathbf{x}})$ of F induces a linear mapping (generalized Jacobian × Vector Product)

$$G : R^n \rightarrow R^m : \quad \mathbf{y}^{(1)} = G \cdot \mathbf{x}^{(1)}$$

defined by the chain rule applied to $F(\mathbf{x}(\dot{c}))$ at $\mathbf{x} = \tilde{\mathbf{x}}$ and for

$$\frac{d\mathbf{x}}{d\dot{c}} \equiv \mathbf{x}^{(1)} \in R^n \ .$$

This DAG × vector product is evaluated as

$$v_i^{(1)} = \sum_{j \prec i} \frac{d\varphi_i(v_k)_{k \prec i}}{dv_j} \cdot v_j^{(1)} \quad \text{for } i = n, \ldots, n+q-1$$

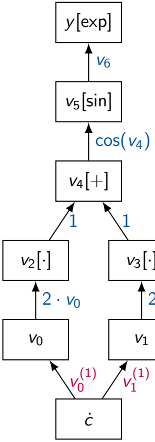
and where $v_i^{(1)} = x_i^{(1)}$ for $i = 0, \ldots, n-1$ and $y_k^{(1)} = v_{p+k}^{(1)}$ for $k = 0, \ldots, m-1$.

DAG × Vector Product

Example

$$y = f(\mathbf{x}) = e^{\sin\left(\|\mathbf{x}\|_2^2\right)} = e^{\sin\left(\mathbf{x}^T \cdot \mathbf{x}\right)} = e^{\sin\left(\sum_{i=0}^{n-1} x_i^2\right)}, \quad n = 2$$

$$\begin{aligned}
v_0 &= x_0 & v_0^{(1)} &= x_0^{(1)} \\
v_1 &= x_1 & v_1^{(1)} &= x_1^{(1)} \\
v_2 &= v_0^2 & v_2^{(1)} &= 2 \cdot v_0 \cdot v_0^{(1)} \\
v_3 &= v_1^2 & v_3^{(1)} &= 2 \cdot v_1 \cdot v_1^{(1)} \\
v_4 &= v_2 + v_3 & v_4^{(1)} &= v_2^{(1)} + v_3^{(1)} \\
v_5 &= \sin(v_4) & v_5^{(1)} &= \cos(v_4) \cdot v_4^{(1)} \\
v_6 &= e^{v_5} & v_6^{(1)} &= v_6 \cdot v_5^{(1)} \\
y &= v_6 & y^{(1)} &= v_6^{(1)}
\end{aligned}$$



Outline

- $f : \boldsymbol{R}^n \to \boldsymbol{R}$

$$f(\mathbf{x} + \Delta\mathbf{x}) = f(\mathbf{x}) + f'(\mathbf{x})^T \cdot \Delta\mathbf{x} + \frac{1}{2} \cdot \Delta\mathbf{x}^T \cdot f''(\mathbf{x}) \cdot \Delta\mathbf{x} + O(\|\Delta\mathbf{x}\|_2^3)$$

- $F : \boldsymbol{R}^n \to \boldsymbol{R}^m$

$$F(\mathbf{x} + \Delta\mathbf{x}) = F(\mathbf{x}) + F'(\mathbf{x}) \cdot \Delta\mathbf{x} + O(\|\Delta\mathbf{x}\|_2^2)$$

Higher-order terms are omitted to avoid tensor notation.

A function $F : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^m$ is linear if

- $F(\mathbf{a}+\mathbf{b}) = F(\mathbf{a}) + F(\mathbf{b})$
- $F(\alpha \cdot \mathbf{a}) = \alpha \cdot F(\mathbf{a})$

for all $\mathbf{a}, \mathbf{b} \in \boldsymbol{R}^n$ and $\alpha \in \boldsymbol{R}$.

Example: $F(\mathbf{x}) = M \cdot \mathbf{x}$ with $M \in \boldsymbol{R}^{m \times n}$ is linear.

$$F(\mathbf{a}+\mathbf{b}) = M \cdot (\mathbf{a}+\mathbf{b}) = M \cdot \mathbf{a} + M \cdot \mathbf{b} = F(\mathbf{a}) + F(\mathbf{b})$$
$$F(\alpha \cdot \mathbf{a}) = M \cdot \alpha \cdot \mathbf{a} = \alpha \cdot M \cdot \mathbf{a} = \alpha \cdot F(\mathbf{a})$$

Functions $F(\mathbf{x}) = M \cdot \mathbf{x} + \mathbf{v}$ with $\mathbf{v} \in \boldsymbol{R}^m$ are called affine.

Affine functions define linear systems $m = n$ as well as linear least-squares problems $m \neq n$.

Convexity and Concavity

$f : \mathbf{R}^n \rightarrow \mathbf{R}$

A function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is convex if and only if its Hessian f'' is positive semi-definite for all $\mathbf{x} \in \mathbf{R}^n$, i.e, $\forall 0 \neq \mathbf{v} \in \mathbf{R}^n$

$$\mathbf{v}^T \cdot f''(\mathbf{x}) \cdot \mathbf{v} \geq 0 \; .$$

One can show that f is strictly convex over $\mathbf{R}^n$ if f'' is positive definite for all $\mathbf{x} \in \mathbf{R}^n$, i.e,

$$\mathbf{v}^T \cdot f''(\mathbf{x}) \cdot \mathbf{v} > 0 \; .$$

The other direction does not hold in general.

Similarly, concavity is defined in terms of negative (semi-)definiteness of the Hessian.

The concepts can be generalized for multivariate vector functions $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$.

Outline

Summary

- continuity
- differentiability
- gradient, Jacobian, Hessian
- chain rule of differential calculus
- partial, total, directional derivative
- directional derivative as DAG $\times$ vector product
- Taylor series

Next Steps

- practice DAG $\times$ vector product
- Continue the course to find out more ...