

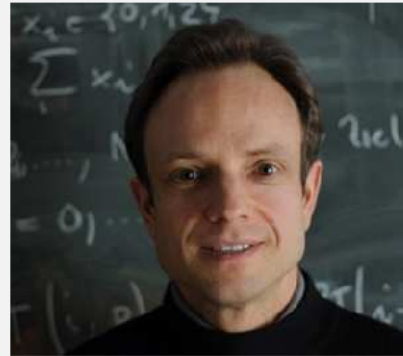
Course Video

§ 8.5.1.15: Multi-dimensional Differentiation

Prof. R. Hiptmair, SAM, ETH Zurich

Date: December 5, 2020

(C) Seminar für Angewandte Mathematik, ETH Zürich



Prerequisites.

- Elementary derivatives
- Multi-dimensional calculus

Dependencies. Eq. (8.5.1.6)

Duration: 35 minutes

VIII Iterative Methods for Non-Linear Systems of Equations

8.5 Newton's Method in $\mathbb{R}^n$

NLSE: Given $C^1 F: \mathcal{D} \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $n \in \mathbb{N}$,
find $\underline{x}^* \in \mathcal{I} : F(\underline{x}^*) = 0$

8.5.1. The Newton Iteration

$$\underline{x}^{(k+1)} := \underline{x}^{(k)} - DF(\underline{x}^{(k)})^{-1} F(\underline{x}^{(k)}), \quad k \in \mathbb{N}_0$$

$$DF(\underline{x}^{(k)}) \in \mathbb{R}^{n,n} \cong \text{Jacobian}$$

$$F = [F_1, \dots, F_n]:$$

$$DF(\mathbf{z}) = \left[\frac{\partial F_i}{\partial x_j}(\mathbf{z}) \right]_{i,j=1}^n = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{z}) & \frac{\partial F_1}{\partial x_2}(\mathbf{z}) & \dots & \dots & \frac{\partial F_1}{\partial x_n}(\mathbf{z}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{z}) & \frac{\partial F_2}{\partial x_2}(\mathbf{z}) & & & \frac{\partial F_2}{\partial x_n}(\mathbf{z}) \\ \vdots & \vdots & & & \vdots \\ \frac{\partial F_n}{\partial x_1}(\mathbf{z}) & \frac{\partial F_n}{\partial x_2}(\mathbf{z}) & \dots & \dots & \frac{\partial F_n}{\partial x_n}(\mathbf{z}) \end{bmatrix}. \quad (8.3.2.8)$$

②

Newton correction $\underline{S} := -DF(x^{(k)})^{-1} F(x^{(k)})$

LSE : $DF(x^{(k)}) \underline{S} = -F(x^{(k)})$

§ 8.5.1.12 (Derivatives)

Derivation $\leftrightarrow$ local linearization

$$F(\underline{x} + \underline{h}) = F(\underline{x}) + \underbrace{DF(\underline{x}) \underline{h}}_{\substack{\uparrow \text{perturbation } \in \mathbb{R}^n \\ \downarrow \\ \text{A linear mapping } \mathbb{R}^n \rightarrow \mathbb{R}^n}} + O(\|\underline{h}\|^2) \quad \text{for } \underline{h} \rightarrow 0$$

Definition 8.5.1.13. Derivative of functions between vector spaces

Let V, W be finite dimensional vector spaces and $F : D \subset V \mapsto W$ a sufficiently smooth mapping. The **derivative** (differential) $DF(\underline{x})$ of F in $\underline{x} \in V$ is the **unique linear mapping** $DF(\underline{x}) : V \mapsto W$ such that there is a $\delta > 0$ and a function $\epsilon : [0, \delta] \rightarrow \mathbb{R}^+$ satisfying $\lim_{\xi \rightarrow 0} \epsilon(\xi) = 0$ such that

$$\|F(\underline{x} + \underline{h}) - F(\underline{x}) - DF(\underline{x})\underline{h}\| = \epsilon(\|\underline{h}\|) \quad \forall \underline{h} \in V, \|\underline{h}\| < \delta. \quad (8.5.1.14)$$

$\hookrightarrow O(\|\underline{h}\|^2)$

$$V = W = \mathbb{R}$$

$$DF(\underline{x}) \leftrightarrow F'(\underline{x}) \in \mathbb{R} \leftrightarrow \text{linear mapping } \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ \xi \mapsto F'(\underline{x}) \xi \end{cases}$$

§ 8.5.1.15 ("High-level" differentiation)

• $F : V \rightarrow W$ linear $\Rightarrow DF = F$!
 ($F(\underline{x}) = A\underline{x}$, $A \in \mathbb{R}^{m,n} \Rightarrow DF(\underline{x})\underline{h} = A\underline{h} = F(\underline{h})$)

• **Chain rule** : $F : V \rightarrow W$, $G : \mathcal{M} \rightarrow V$
 $F(G(\underline{x} + \underline{h})) = F(G(\underline{x}) + DG(\underline{x})\underline{h} + O(\|\underline{h}\|^2))$
 $= F(G(\underline{x})) + \underbrace{DF(G(\underline{x}))DG(\underline{x})\underline{h}}_{= D(F \circ G)(\underline{x})\underline{h}} + O(\|\underline{h}\|^2)$

Chain rule: For $F : V \mapsto W$, $G : U \mapsto V$ sufficiently smooth

$$D(F \circ G)(\underline{x})\underline{h} = DF(G(\underline{x}))(DG(\underline{x})\underline{h}), \quad \underline{h} \in V, \underline{x} \in D. \quad (8.5.1.16)$$

• **Product rule** :

- $F : V \rightarrow W$, $G : V \rightarrow \mathcal{M}$

- $b : W \times \mathcal{M} \rightarrow \mathbb{Z}$ **bilinear**

$$\Rightarrow T : V \rightarrow \mathbb{Z}, \quad T(\underline{x}) := b(F(\underline{x}), G(\underline{x}))$$

③

$$(V = W = U = Z = \mathbb{R} : b(x, y) = x \cdot y)$$

$$T(x) = F(x) \cdot G(x) \Rightarrow T'(x) = F'(x)G(x) + F(x)G'(x)$$

$$\text{As L.M. : } T'(x)h = (F'(x)h) \cdot G(x) + F(x) \cdot (G'(x)h)$$

$$\begin{aligned} T(\mathbf{x} + \mathbf{h}) &= b(F(\mathbf{x} + \mathbf{h}), G(\mathbf{x} + \mathbf{h})) \\ &= b(F(\mathbf{x}) + DF(\mathbf{x})\mathbf{h} + o(\|\mathbf{h}\|), G(\mathbf{x}) + DG(\mathbf{x})\mathbf{h} + o(\|\mathbf{h}\|)) \\ &\stackrel{(*)}{=} b(F(\mathbf{x}), G(\mathbf{x})) + \underbrace{b(DF(\mathbf{x})\mathbf{h}, G(\mathbf{x})) + b(F(\mathbf{x}), DG(\mathbf{x})\mathbf{h})}_{=DT(\mathbf{x})\mathbf{h}} + o(\|\mathbf{h}\|). \end{aligned}$$

$\downarrow$
 $O(\|\mathbf{h}\|^2)$

$\uparrow$
 b bilinear

Product rule: $F : D \subset V \mapsto W, G : D \subset V \mapsto U$ sufficiently smooth, $b : W \times U \mapsto Z$ bilinear, ie., linear in each argument:

$$T(\mathbf{x}) := b(F(\mathbf{x}), G(\mathbf{x})) \Rightarrow \begin{aligned} DT(\mathbf{x})\mathbf{h} &= b(DF(\mathbf{x})\mathbf{h}, G(\mathbf{x})) + b(F(\mathbf{x}), DG(\mathbf{x})\mathbf{h}), \\ \mathbf{h} \in V, \mathbf{x} \in D. \end{aligned} \quad (8.5.1.17)$$

Special notion for $F : \mathbb{R}^n \rightarrow \mathbb{R}$:

Definition 8.5.1.18. Gradient and Hessian

For sufficiently smooth $F : D \subset \mathbb{R}^n \mapsto \mathbb{R}$ the **gradient** $\text{grad } F : D \mapsto \mathbb{R}^n$, and the **Hessian** (matrix) $HF(\mathbf{x}) : D \mapsto \mathbb{R}^{n,n}$ are defined as

$$\text{grad } F(\mathbf{x})^T \mathbf{h} := DF(\mathbf{x})\mathbf{h}, \quad \mathbf{h}_1^T HF(\mathbf{x})\mathbf{h}_2 := D(DF(\mathbf{x})(\mathbf{h}_1))(\mathbf{h}_2), \quad \forall \mathbf{h}, \mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^n.$$

$\uparrow$
 L.M. $\mathbb{R}^n \rightarrow \mathbb{R}$

$$\begin{aligned} &\triangleq D(DF) \triangleq \text{L.M. } \mathbb{R}^n \rightarrow \{\text{L.M. } \mathbb{R}^n \rightarrow \mathbb{R}\} \\ &\triangleq \text{bilinear mapping } \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \end{aligned}$$

Ex 8.5.1.19 (Derivative of a bilinear form)

$$\psi : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \psi(\underline{x}) := \underline{x}^T A \underline{x}, \quad A \in \mathbb{R}^{n,n}$$

Product rule: $F : D \subset V \mapsto W, G : D \subset V \mapsto U$ sufficiently smooth, $b : W \times U \mapsto Z$ bilinear, ie., linear in each argument:

$$T(\mathbf{x}) := b(F(\mathbf{x}), G(\mathbf{x})) \Rightarrow \begin{aligned} DT(\mathbf{x})\mathbf{h} &= b(DF(\mathbf{x})\mathbf{h}, G(\mathbf{x})) + b(F(\mathbf{x}), DG(\mathbf{x})\mathbf{h}), \\ \mathbf{h} \in V, \mathbf{x} \in D. \end{aligned} \quad (8.5.1.17)$$

$$\text{Now: } F, G = \text{Id} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad b(x, y) = x^T A y$$

$$D\psi(\underline{x})\underline{h} = \underline{h}^T A \underline{x} + \underline{x}^T A \underline{h} = (\underline{x}^T A^T + \underline{x}^T A) \underline{h}$$

$$\Rightarrow \text{grad } \psi(\underline{x}) = (A^T + A)\underline{x} \in \mathbb{R}^n$$

"Pedestrian way"

$$\Psi(\mathbf{x}) = \sum_{k=1}^n \sum_{j=1}^n (A)_{k,j} x_k x_j = (A)_{i,i} x_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^n (A)_{i,j} x_i x_j + \sum_{\substack{k=1 \\ k \neq i}}^n (A)_{k,i} x_k x_i + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{j=1 \\ j \neq i}}^n (A)_{k,j} x_k x_j,$$

$$\begin{aligned} \frac{\partial \Psi(\mathbf{x})}{\partial x_i}(\mathbf{x}) &= 2(A)_{i,i} x_i + \sum_{\substack{j=1 \\ j \neq i}}^n (A)_{i,j} x_j + \sum_{\substack{k=1 \\ k \neq i}}^n (A)_{k,i} x_k = \sum_{j=1}^n (A)_{i,j} x_j + \sum_{k=1}^n (A)_{k,i} x_k \\ &= (A\mathbf{x} + A^T \mathbf{x})_i, \quad i = 1, \dots, n. \end{aligned}$$

(4)

Ex 8.5.1.20 (Derivative of Euclidean norm)

$$F: \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}, \quad F(x) = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$

$$F = G \circ H, \quad G(\xi) := \sqrt{\xi}, \quad H(x) = x^T x$$

$$DG(\xi)\eta = \frac{1}{2\sqrt{\xi}}\eta, \quad DH(x)\underline{h} = 2x^T \underline{h}$$

Chain rule: For $F: V \mapsto W$, $G: U \mapsto V$ sufficiently smooth

$$D(F \circ G)(x)\mathbf{h} = DF(G(x))(DG(x)\mathbf{h}), \quad \mathbf{h} \in V, x \in D. \quad (8.5.1.16)$$

$$\text{Here: } F \leftarrow G, \quad G \leftarrow H$$

$$\begin{aligned} \Rightarrow DF(x)\underline{h} &= DG(H(x))DH(x)\underline{h} \\ &= \frac{1}{2\sqrt{x^T x}} \cdot 2x^T \underline{h} \end{aligned}$$

$$\text{grad } F(x) = \frac{x}{\|x\|_2}, \quad x \neq 0$$

§ 8.5.1.22 (Newton iteration via product rule)

We seek solutions of $F(x) = 0$ with $F(x) := b(G(x), H(x))$, where

- ♦ V, W are some vector spaces (finite- or even infinite-dimensional),
- ♦ $G: D \rightarrow V, H: D \rightarrow W, D \subset \mathbb{R}^n$, are continuously differentiable in the sense of Def. 8.5.1.13,
- ♦ $b: V \times W \mapsto \mathbb{R}$ is **bilinear** (linear in each argument).

$$[F: D \subset \mathbb{R}^n \rightarrow \mathbb{R}]$$

$$\text{LSE for Newton correction: } DF(x)\underline{s} = -F(x)$$

Product rule:

$$DF(x)\mathbf{h} = b(DG(x)\mathbf{h}, H(x)) + b(G(x), DH(x)\mathbf{h}), \quad \mathbf{h} \in \mathbb{R}^n. \quad (8.5.1.23)$$

 $\Downarrow$ P.R.

$$\underbrace{b(DG(x^{(k)})\mathbf{s}, H(x^{(k)})) + b(G(x^{(k)}), DH(x^{(k)})\mathbf{s})}_{\cong M\underline{s}, M \in \mathbb{R}^{n,n}} = \underbrace{-b(G(x^{(k)}), H(x^{(k)}))}_{\in \mathbb{R}^n}. \quad (8.5.1.24)$$

