

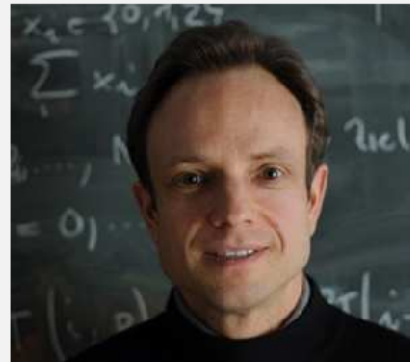
## Course Video

Section 8.7.1: Non-linear Least Squares:  
(Damped) Newton Method

Prof. R. Hiptmair, SAM, ETH Zurich

Date: December 6, 2020

(C) Seminar für Angewandte Mathematik, ETH Zürich



## Prerequisites.

- Multi-dimensional differentiation

Dependencies. Section 8.5.1

Duration: 19 minutes

# VIII Iterative Methods for Non-Linear Systems of Equations

## 8.7 Non-Linear Least Squares

## Non-linear least squares problem

Given:  $F : D \subset \mathbb{R}^n \mapsto \mathbb{R}^m$ ,  $m, n \in \mathbb{N}$ ,  $m > n$ .Find:  $\mathbf{x}^* \in D$ :  $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in D} \Phi(\mathbf{x})$ ,  $\Phi(\mathbf{x}) := \frac{1}{2} \|F(\mathbf{x})\|_2^2$ . (8.7.0.7)

Special case: Linear least-squares problem

$$F(\mathbf{x}) = A\mathbf{x} - \underline{b} \quad \Rightarrow \quad \Phi(\mathbf{x}) = \|A\mathbf{x} - \underline{b}\|_2^2$$

$$A \in \mathbb{R}^{m,n}, \quad \underline{b} \in \mathbb{R}^m$$

## 8.7.1 (Damped) Newton Method

Assume :  $F \in C^2$

$$[\phi: \mathbb{R}^n \rightarrow \mathbb{R}]$$

$$\Phi(x^*) = \min \Rightarrow \underbrace{\text{grad } \Phi(x) = 0}_{n \times n \text{ NLSSE}}, \text{ where } \text{grad } \Phi(x) := \left[ \frac{\partial \Phi}{\partial x_1}(x), \dots, \frac{\partial \Phi}{\partial x_n}(x) \right]^T \in \mathbb{R}^n.$$

necessary condition

$n \times n$  NLSSE

↓

Newton iteration for  $G(x) = 0$  :  $x^{(k+1)} = x^{(k)} - DG(x^{(k)})^{-1} G(x^{(k)})$

$$x^{(k+1)} = x^{(k)} - H\Phi(x^{(k)})^{-1} \text{grad } \Phi(x^{(k)}), \quad (8.7.1.2)$$

where  $H\Phi(x) \in \mathbb{R}^{n,n}$  is the **Hessian matrix** of  $\Phi$

$$H\Phi(x) = \left[ \frac{\partial^2 \Phi}{\partial x_i \partial x_j} \right]_{i,j=1}^n \in \mathbb{R}^{n,n}.$$

•  $\text{grad } \phi$  :  $\phi(x) = \frac{1}{2} F(x)^T \cdot F(x)$

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

$$T(x) := b(H(x), G(x)) \Rightarrow DT(x)h = b(DH(x)h, G(x)) + b(H(x), DG(x)h), \quad (8.5.1.17) \\ h \in V, x \in D.$$

with  $H = G := F$ ,  $b(x, y) = \frac{1}{2} x^T y$

$$D\phi(x)h = \frac{1}{2} (DF(x)h)^T F(x) + \frac{1}{2} F(x)^T (DF(x)h) \\ \uparrow \text{Jacobian}$$

$$= F(x)^T DF(x)h$$

### Definition 8.5.1.18. Gradient and Hessian

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

$$\text{grad } F(x)^T h := DF(x)h, \quad h_1^T HF(x)h_2 := D(DF(x)(h_1))(h_2), \quad \forall h, h_1, h_2 \in \mathbb{R}^n.$$

$$\Rightarrow \text{grad } \phi(x) = DF(x)^T F(x)$$

•  $H\phi$  :

$$\iff D\{x \in D \subset \mathbb{R}^n \longrightarrow F(x) DF(x)h_1 \in \mathbb{R}\}$$

Again: **product rule**

$$D\{x \mapsto F(x)^T DF(x)h_1\}(x)h_2 \\ = (DF(x)h_2)^T (DF(x)h_1) + F(x)^T \cdot D(DF(x)h_1)h_2$$

③  $\rightarrow \in \mathbb{R}^{n,n}$  Jacobian

$$H\Phi(\mathbf{x}) := D(\text{grad } \Phi)(\mathbf{x}) = DF(\mathbf{x})^T DF(\mathbf{x}) + \sum_{j=1}^m F_j(\mathbf{x}) D^2 F_j(\mathbf{x}),$$

 $\Leftrightarrow$ 

$$(H\Phi(\mathbf{x}))_{i,k} = \sum_{j=1}^n \left\{ \frac{\partial^2 F_j}{\partial x_i \partial x_k}(\mathbf{x}) F_j(\mathbf{x}) + \frac{\partial F_j}{\partial x_k}(\mathbf{x}) \frac{\partial F_j}{\partial x_i}(\mathbf{x}) \right\}.$$

▷ LSE for Newton connection  $\underline{s} \in \mathbb{R}^n$

$$\underbrace{\left( DF(\mathbf{x}^{(k)})^T DF(\mathbf{x}^{(k)}) + \sum_{j=1}^m F_j(\mathbf{x}^{(k)}) D^2 F_j(\mathbf{x}^{(k)}) \right)}_{=H\Phi(\mathbf{x}^{(k)})} \mathbf{s} = - \underbrace{DF(\mathbf{x}^{(k)})^T F(\mathbf{x}^{(k)})}_{=\text{grad } \Phi(\mathbf{x}^{(k)})}.$$



Newton method  $F(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$   
 $A \in \mathbb{R}^{m,n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  ?

→ Normal equation method!

(8.7.1.5)

Rem 8.7.1.7. (Newton method and minimization of quadratic functional)

Task :  $\phi(\mathbf{x}) \rightarrow \min$ ,  $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$

Idea: Taylor approximation at  $\underline{x}^{(k)}$

$$\Phi(\mathbf{x}) \approx Q(\mathbf{s}) := \Phi(\mathbf{x}^{(k)}) + \text{grad } \Phi(\mathbf{x}^{(k)})^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T H\Phi(\mathbf{x}^{(k)}) \mathbf{s}, \quad (8.7.1.8)$$

$$(8.7.1.6) \quad \text{grad } Q(\mathbf{s}) = 0 \Leftrightarrow H\Phi(\mathbf{x}^{(k)}) \mathbf{s} + \text{grad } \Phi(\mathbf{x}^{(k)}) = 0 \Leftrightarrow (8.7.1.4).$$

▷ Review questions 8.7.0.10  
 Review questions 8.7.1.9

4

5



7