

V. Linear and Non-linear Least Squares

Goals: - solve overdetermined linear and nonlinear systems of equations

- linear: Normal equations and orthogonal decomposition method

- nonlinear: Newton and Gauss-Newton methods

- only the numerical aspects (statistics later...)

Why: - curve fitting

In practice, one often has to determine some parameters of a given function (from natural laws or model assumptions) through a series of measurements.

Usually, the number of measurements m is much larger than the number of parameters n , that is $m \gg n$.

Ex.: (1) Model ① $g(t) = a_1 \cdot t + a_2$ (linear)

Model ② $g(t) = a_2 \cdot e^{a_1 t}$ (nonlinear)

no slides

One distinguishes:

$$\text{(linear)} \quad A \vec{x} = \vec{b}, \quad A \in \mathbb{R}^{m \times n}, \quad \vec{b} \in \mathbb{R}^m$$

$$\text{(nonlinear)} \quad \vec{f}(\vec{x}) = \vec{b}, \quad \vec{f}: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ vector of unknown parameters}$$

However, these equations do in general not admit a solution in the common sense. Instead one looks for so-called least-squares solutions

$$\text{(linear)} \quad \min_{\vec{x} \in D} \| A\vec{x} - \vec{b} \|_2 = \| \vec{v} \|_2$$

$$\text{(nonlinear)} \quad \min_{\vec{x} \in D} \| \vec{f}(\vec{x}) - \vec{b} \|_2 = \| \vec{v} \|_2$$

where $D \dots$ admissible parameter domain

$$\| \cdot \|_2 \dots \text{Euclidean norm} \quad \| \vec{x} \|_2 = \sqrt{x_1^2 + \dots + x_n^2}$$

$$\| \vec{v} \|_2 \dots \text{residual} \quad = \sqrt{\vec{x}^T \vec{x}} \quad \text{transpose}$$

So one minimizes the deviation

no slides

V.1 Linear Least-Squares

Let's focus on the linear case:

$$\min_{\vec{x}} \phi(\vec{x})$$

where $\phi(\vec{x}) = \frac{1}{2} \|A\vec{x} - \vec{b}\|_2^2$

and $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, $\vec{x} \in \mathbb{R}^n$.

Moreover, we assume that A has full column rank n , i.e. the columns of A are linearly independent.

Have as usual: $m \dots$ number of measurements
 $n \dots$ number of parameters \downarrow

Rem.: The assumption on the rank implies that the parameters are not ambiguous

V.1.1 Normal equations

Let's rewrite $\phi(\vec{x})$ as:

$$\begin{aligned} \phi(\vec{x}) &= \frac{1}{2} \|A\vec{x} - \vec{b}\|_2^2 \\ &= \frac{1}{2} (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) = \frac{1}{2} (\vec{x}^T A^T - \vec{b}^T) (A\vec{x} - \vec{b}) \\ &= \frac{1}{2} \left(\vec{x}^T A^T A \vec{x} - \underbrace{\vec{x}^T A^T \vec{b}^T}_{\substack{= (\vec{b}^T A \vec{x})^T = \vec{x}^T A^T \vec{b} \\ \text{scalar!}}} - \vec{b}^T A \vec{x} + \vec{b}^T \vec{b} \right) \\ &= \frac{1}{2} \left(\vec{x}^T A^T A \vec{x} - 2 \vec{x}^T A^T \vec{b} + \vec{b}^T \vec{b} \right) \end{aligned}$$

$$\left[\sim \frac{1}{2} (x a a x - 2 x a b + b b) = \frac{1}{2} (a^2 x^2 - 2 a b x + b^2) \right] \quad \&$$

At an extremum:

$$\vec{\nabla} \phi(\vec{x}) = A^T A \vec{x} - A^T \vec{b} \stackrel{!}{=} 0$$

These are the so-called normal equations and their solution is the sought least-squares solution

Rem.: (i) The solution to the normal equations is indeed a minimum of $\phi(\vec{x})$ and it is unique:

① $A^T A$ is a positive definite matrix

$$\|A\vec{x}\|_2^2 = \underbrace{\vec{x}^T A^T A \vec{x}}_{\text{Def. of positive definite}} > 0 \text{ for all } x \in \mathbb{R}^n, \vec{x} \neq 0$$

$$\textcircled{2} \vec{x}^T A^T A \vec{x} = 0 \Leftrightarrow \vec{x} = 0$$

" \Leftarrow ": clear

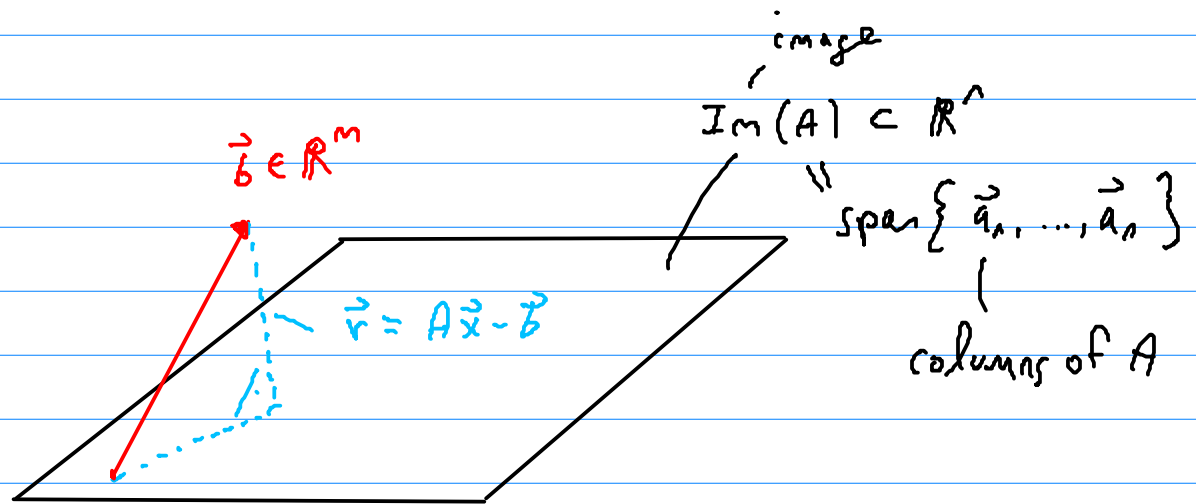
$$\text{"} \Rightarrow \text{"}: \underbrace{\vec{x}^T}_{(A\vec{x})^T} \underbrace{A^T A}_{=0} \vec{x} = 0$$

$= 0 \Leftrightarrow \vec{x} = 0$ because $\text{rank}(A) = n$!
 $(A\vec{x})^T \nearrow \checkmark$

This is just a generalization of the well known fact $\frac{1}{2}(a a x^2 - 2 a b x + b^2) \dots a \cdot a > 0$ for a minimum!

$$(ii) A^T A \vec{x} = A^T \vec{b}: \underbrace{\begin{pmatrix} A^T \\ n \times m \end{pmatrix}}_{\begin{pmatrix} A^T A \\ n \times n \end{pmatrix}} \underbrace{\begin{pmatrix} A \\ m \times n \end{pmatrix}}_{\begin{pmatrix} \vec{x} \\ n \times 1 \end{pmatrix}} = \underbrace{\begin{pmatrix} A^T \\ n \times m \end{pmatrix}}_{\begin{pmatrix} A^T \vec{b} \\ n \times 1 \end{pmatrix}} \underbrace{\begin{pmatrix} \vec{b} \\ m \times 1 \end{pmatrix}}_{\begin{pmatrix} \vec{x} \\ n \times 1 \end{pmatrix}} = \begin{pmatrix} A^T \vec{b} \\ n \times 1 \end{pmatrix}$$

(iii) Geometric interpretation



minimize $\|A\vec{x} - \vec{b}\|_2 \leadsto A\vec{x} - \vec{b}$ is normal
to $Im(A)$
(Hence normal equations)

Ex.: (2) Linear model no slides

However, it turns out that solving the normal equations can be problematic, i.e. ill-conditioned

(a small change in \vec{b} can lead to a large change in \vec{x})

Therefore the (following) orthogonal decomposition method is preferred

V.1.2 The orthogonal decomposition method

Def.: an orthogonal matrix Q is a real square matrix whose columns are orthonormal:

$$Q^T Q = Q Q^T = I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

Rem.: - $Q^{-1} = Q^T$

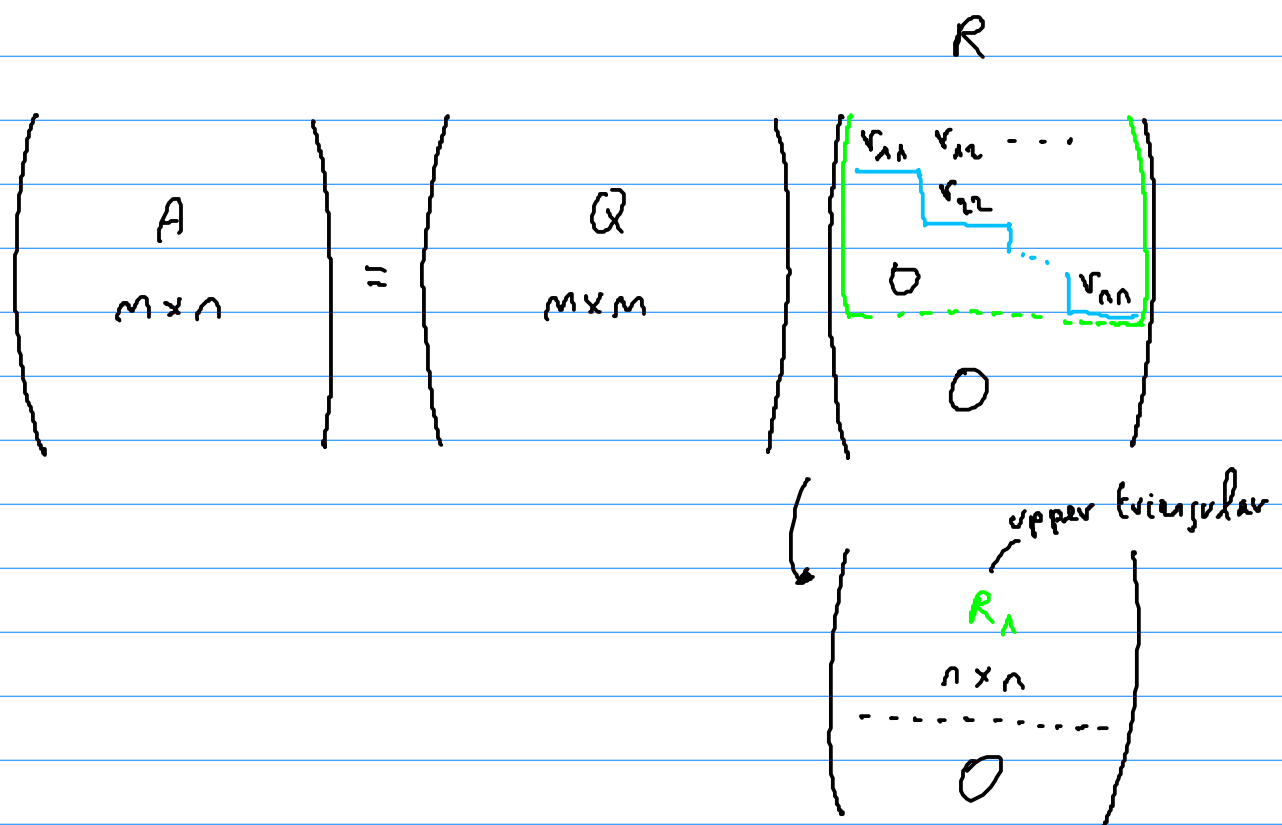
An important property of orthogonal matrices is that they don't change the Euclidean norm (i.e. length):

$$\|Q\vec{x}\|_2^2 = \vec{x}^T \underbrace{Q^T Q}_{I} \vec{x} = \vec{x}^T \vec{x} = \|\vec{x}\|_2^2$$

fact: every matrix $A \in \mathbb{R}^{m \times n}$ with full column rank n (linearly independent columns) has a so-called QR-decomposition

$$A = QR$$

where $Q \in \mathbb{R}^{m \times m}$ is an orthogonal matrix and $R \in \mathbb{R}^{m \times n}$ an upper triangular matrix with $r_{ii} \neq 0$ ($i = 1, \dots, n$)



MATLAB: $[Q, R] = \text{qr}(A)$

With this: Let $A = QR$

$$\|A\vec{x} - \vec{b}\|_2^2 = \|\vec{v}\|_2^2$$

(Q ortho!) $\|Q^T(A\vec{x} - \vec{b})\|_2^2 = \|Q^T\vec{v}\|_2^2 = \|\vec{v}\|_2^2$

$$\|Q^T(QR\vec{x} - \vec{b})\|_2^2 = \|\vec{v}\|_2^2$$

$$\|R\vec{x} - Q^T\vec{b}\|_2^2 = \|\vec{v}\|_2^2$$

$$\left\| \begin{pmatrix} R_1 \\ \dots \\ 0 \end{pmatrix} \vec{x} - \begin{pmatrix} \vec{c} \\ \dots \\ \vec{d} \end{pmatrix} \right\|_2^2 = \|\vec{v}\|_2^2$$

split

$$\|R_1\vec{x} - \vec{c}\|_2^2 + \|\vec{d}\|_2^2 = \|\vec{v}\|_2^2$$

n equations in n unknowns $\leadsto \vec{x} = R_1^{-1} \vec{c}$
invertible because $r_{ii} \neq 0$ ($i=1, \dots, n$)

In summary: ① $A = QR$, $Q^T \vec{b} = \begin{pmatrix} \vec{c} \\ \vec{d} \end{pmatrix}$

② $\vec{x} = R^{-1} \vec{c}$

③ $\|\vec{x}\|_2 = \|\vec{c}\|_2$

Ex.: (3) Same as ex. (2) but with QR

no slides

(Of course, we get the same result!)

(4) Same as ex. (2) & (3) but with the MATLAB

all-in-one solution: $\vec{x} = A \backslash \vec{b}$

↖ backslash operator

V.2 Non-linear Least-Squares

Let's now consider the case of overdetermined nonlinear systems of equations

$$\begin{aligned}
 f_1(x_1, \dots, x_n) &= b_1 \\
 f_2(x_1, \dots, x_n) &= b_2 \\
 &\vdots \\
 f_m(x_1, \dots, x_n) &= b_m
 \end{aligned}
 \quad \text{or short} \quad
 \begin{aligned}
 & \vec{f}: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m \\
 & \vec{f}(\vec{x}) = \vec{b} \quad \vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m \\
 & \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n
 \end{aligned}$$

m (nonlinear) equations in n unknowns.

Usually $m \gg n$ (many more measurements than parameters!)

Least-squares solution:

$$\min_{\vec{x} \in D} \phi(\vec{x})$$

where $\phi(\vec{x}) = \frac{1}{2} \|\vec{f}(\vec{x}) - \vec{b}\|_2^2$

let's rewrite $\phi(\vec{x})$:

$$\begin{aligned} \phi(\vec{x}) &= \frac{1}{2} \|\vec{f}(\vec{x}) - \vec{b}\|_2^2 \\ &= \frac{1}{2} \sum_{i=1}^m (f_i(\vec{x}) - b_i)^2 \end{aligned}$$

(not sufficient... see below)

A necessary condition for a minimum is that the gradient of ϕ vanishes:

$$\vec{\nabla} \phi = \begin{pmatrix} \frac{\partial \phi}{\partial x_1} \\ \vdots \\ \frac{\partial \phi}{\partial x_n} \end{pmatrix} \stackrel{!}{=} 0$$

$$\begin{aligned} \text{Then } \frac{\partial \phi}{\partial x_j} &= \frac{\partial}{\partial x_j} \left(\frac{1}{2} \sum_{i=1}^m (f_i(x_1, \dots, x_j, \dots, x_n) - b_i)^2 \right) \\ &= \sum_{i=1}^m (f_i(\vec{x}) - b_i) \cdot \frac{\partial f_i}{\partial x_j} \stackrel{!}{=} 0, \quad j=1, \dots, n \end{aligned}$$

These are n (nonlinear) equations for the n unknown parameters

The vanishing gradient is only a necessary condition for a minimum (it could also be a maximum or a saddle point)

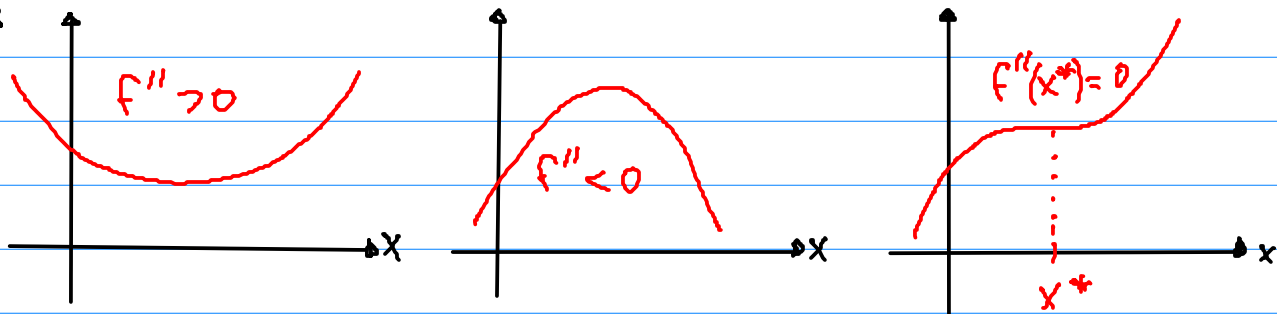
We need one further condition:

$$H = \underset{\text{Hessian matrix}}{\text{Hess } \phi} = \left(\frac{\partial^2 \phi}{\partial x_i \partial x_j} \right)_{i,j} \text{ positiv definit}$$

i.e. $\vec{x}^T H \vec{x} > 0$ for all $\vec{x} \in \mathbb{R}^n, \vec{x} \neq 0$.

In one variable, this reduces to the familiar

cases:



V.2.1 Newton method

Idea: apply Newton's method directly to

$$\vec{\nabla} \phi \stackrel{!}{=} 0$$

Linearize $\vec{\nabla} \phi$ around some given $\vec{x}^{(k)}$:

$$\vec{\nabla} \phi(\vec{x}^{(k)}) \approx \vec{\nabla} \phi(\vec{x}^{(k)}) + (\text{Hess } \phi)(\vec{x}^{(k)}) (\vec{x} - \vec{x}^{(k)})$$

Define $\vec{x}^{(k+1)}$ as

$$\vec{\nabla} \phi(\vec{x}^{(k)}) + (\text{Hess } \phi)(\vec{x}^{(k)}) \underbrace{(\vec{x}^{(k+1)} - \vec{x}^{(k)})}_{\vec{s} \text{ (call it } \vec{s} \text{!)}} = 0$$

Then

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \vec{s}, \quad k = 0, 1, \dots, \text{ convergence}$$

$$\text{solve } (\text{Hess } \phi)(\vec{x}^{(k)}) \vec{s} = -\vec{\nabla} \phi(\vec{x}^{(k)})$$

Rem.: (i) if close enough to a minimum, it

converges quadratically $p=2$ { existence & uniqueness

(ii) needs the Hessian matrix

discussion in
chapter II!

Rem.: (i) if it converges, then it converges with order $p = 1 \dots 2$

(ii) it fails if $A^{(k)}$ is very badly conditioned

\rightsquigarrow Levenberg-Marquardt methods ...

Ex.: (5) Nonlinear least-squares \rightsquigarrow Slides