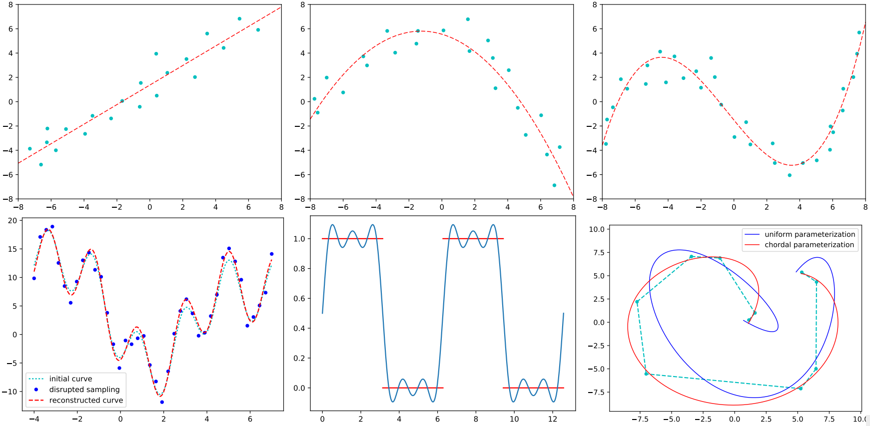
Chapter 6

Least squares approximation

- *Method of least squares* → standard approach to approximate solution of over-determined systems (systems of equations in which there are more equations than unknowns).
- "*Least squares*" → means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation.
- A major application consists in data fitting.



What is a good approximation ?

As an example, assume we want to approximate the function $f(x) = x^2$ over the interval $[0, 1]$ by a simple function, e.g. by a polynomial of degree one : $p(x) = ax + b$.
Of course we need a tool to measure this approximation, i.e. the distance (the error) between f and p on $[0, 1]$. Consider the following three cases.

- *Continuous least squares* — The minimization of

$$\int_0^1 \Big(f(x) - p(x)\Big)^2 \mathrm{d}x \;=\; \int_0^1 \Big(x^2 - (ax+b)\Big)^2 \mathrm{d}x \qquad \text{leads to} \quad p(x) = x - \frac{1}{6}$$

- *Discrete least squares* — Considering the 3 points $x_0 = 0, x_1 = 1/2, x_2 = 1$, the minimization of

$$\sum_{i=0}^{2} \Big(f(x_i) - p(x_i)\Big)^2 \;=\; \sum_{i=0}^{2} \Big(x_i^2 - (ax_i+b)\Big)^2 \qquad \text{leads to} \quad p(x) = x - \frac{1}{12}$$

- *Absolute values* — Considering the 3 points $x_0 = 0, x_1 = 1/2, x_2 = 1$, the minimization of

$$\sum_{i=0}^{2} \Big|f(x_i) - p(x_i)\Big| \;=\; \sum_{i=0}^{2} \Big|x_i^2 - (ax_i+b)\Big| \qquad \text{leads to} \quad p(x) = x$$

Consequently, one can see that the choice of the tool (the norm) for measuring the error (the approximation level) is therefore essential.

Analysis & implementation

Then, after choosing an approximation criterion, we must consider the following questions.

- Is there a solution ? i.e., does such a polynomial p exist ?
- Uniqueness ?
- How can we characterize this solution p ?
- How can we calculate this solution p ?

Problem 1 : Linear system.

Consider the linear system

$$\begin{pmatrix} 1 & 1 \\ 4 & 1 \\ 7 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix}$$

that can be written in the matrix form $Ax = b$.
This linear system does not admit an exact solution.

We propose to minimize the quantity $||A\,x - b||^2$ which can be written as the quadratic form

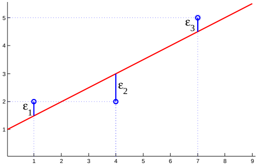
$$x^T(A^TA)x - 2\,(A^Tb)^T\,x + b^Tb.$$

Problem 2 : Projection in an Euclidean space.
Consider the vector $v = (2,2,5) \in \mathbb{R}^3$ and the 2D linear space Π of $\mathbb{R}^3$ spanned by the two vectors $u_1 = (1,4,7)$ and $u_2 = (1,1,1)$.
We are looking for a vector $\hat{v} = \alpha\, u_1 + \beta\, u_2 \in \Pi$ which minimizes the distance between v and Π.

The solution consists in the orthogonal projection of v in the plane Π characterized by $v - \hat{v} \perp u_1$ and $v - \hat{v} \perp u_2$, i.e.,
$\langle v - \hat{v}, u_1 \rangle = 0$ and
$\langle v - \hat{v}, u_2 \rangle = 0$.

Problem 3 : Curve fitting.

Given the three points $(1,2)$, $(4,2)$, $(7,5)$, we are looking for a straight line $Y = \alpha\, X + \beta$ which passes (as close as possible) through these 3 points.

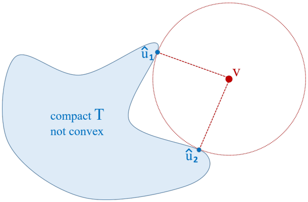


For this purpose, we minimize the sum of the square of the errors ϵ_i, i.e., the quantity $\sum \epsilon_i^2$.

In a normed vector space

Let $(V, ||.||)$ a normed vector space and $T \subset V$ an arbitrary subset. Given an element $v \in V$ we look for an element $u \in T$ which is as close as possible of v. Precisely, $\hat{u} \in T$ is called a *best approximation* of v in T if

$$||v - \hat{u}|| = \inf_{u \in T} ||v - u||$$



Proposition 6.1

Let $T \subset V$ be a compact subset, then for every $v \in V$ there exists a best approximation $\hat{u} \in T$ of v.

Proposition 6.2 (Uniqueness)

Let $T \subset V$ be a compact and strictly convex subset of a normed vector space V. Then for every $v \in V$, there exists a unique best approximation $\hat{u} \in T$ of v.

Proposition 6.3

Let U be a finite dimensional vector subspace of a normed vector space V. Then for every $v \in V$, there exists at least one best approximation $\hat{u} \in U$ of v.

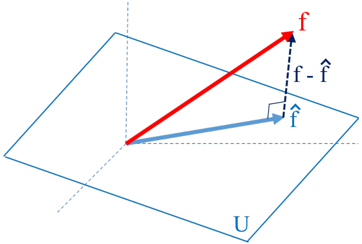
In a pre-Hilbert space

Let V be a vector space equipped with the inner product $f, g \in V \mapsto \langle f, g\rangle$ and let $||f|| = \langle f, f\rangle^{1/2}$ be the induced norm. In addition, let U be a finite dimensional vector subspace of this pre-Hilbert space.
For any given element $f \in V$, there exists a unique best approximation $\hat{f} \in U$ of f.
We now consider a useful characterization of this best approximation.

Proposition 6.4 (Characterization)

$\hat{f} \in U$ is the best approximation of $f \in V$ if and only if

$$\langle f - \hat{f}, g\rangle = 0 \quad \textit{for all } g \in U.$$



Polynomial of best uniform approximation

Proposition 6.5

Consider a function $f \in C[a,b]$. Then for each integer $n \in \mathbb{N}$, there exists a unique polynomial q_n of degree less than or equal to n such that

$$||f - q_n|| = \min_{p \in \mathbb{R}_n[x]} ||f - p||.$$

This polynomial q_n is called the polynomial of best uniform approximation of f of order n.

Proposition 6.6 (Weierstrass)

The space of polynomials $\mathbb{R}[x]$ is dense in the space $C[a,b]$ endowed with the uniform norm. As a result, for any $\epsilon > 0$ there exists an integer $n \in \mathbb{N}$ and a polynomial $p \in \mathbb{R}_n[x]$ such that $||f - p|| < \epsilon$.

Positive definite matrices

A symmetric n square real matrix A is said to be *positive semi-definite* if and only if

$$\forall x \in \mathbb{R}^n, \quad x^T A x \geq 0.$$

A symmetric n square real matrix A is said to be *positive definite* if and only if

$$\forall x \in \mathbb{R}^n, x \neq 0, \quad x^T A x > 0.$$

Proposition 6.7

In this proposition, matrices A and B are assumed to be real symmetric n square matrices.

- *Matrix A is positive semi-definite if and only if all its eigenvalues are positive.*
  Matrix A is positive definite if and only if all its eigenvalues are strictly positive.
- *If A is positive semi-definite and invertible, then A is positive definite.*
- *If A is positive definite, then A^{-1} is positive definite.*
- *Matrix A is positive definite if and only if there exists an invertible n square matrix G such that $A = G^T G$.*
- *For any real matrix H of size (p, n), the matrix $H^T H$ is (a n square) symmetric positive semi-definite.*
- *If A is positive definite, then αA is positive definite for any real $\alpha > 0$.*
- *If A and B are positive semi-definite and if one of the two matrices A or B is invertible, then $A + B$ is definite positive.*
- *From the Gerschgorin-Hadamard theorem we deduce immediately the two following results.*

  a) A symmetric diagonally dominant real matrix A with non negative diagonal entries is positive semi-definite.

  b) A symmetric strictly diagonally dominant real matrix A with non negative diagonal entries is positive definite.

Minimization : main result

We consider the problem of minimizing a positive definite quadratic form $q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$q(x) = x^T A\, x - 2\, b^T x + c \tag{1}$$

where A is a symmetric n square real positive definite matrix, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Proposition 6.8

The positive definite quadratic form (1) is strictly convex, which means that

$$\forall x, y \in \mathbb{R}^n,\ \forall t \in]0,1[, \quad q\big((1-t)\,x + t\,y\big) < (1-t)\,q(x) + t\,q(y)$$

and the minimization problem

$$\textit{find } \tilde{x} \in \mathbb{R}^n \quad \textit{such that} \quad q(\tilde{x}) = \min_{x \in \mathbb{R}^n} q(x)$$

admits a unique global solution $\bar{x}$ on $\mathbb{R}^n$ defined as the unique solution of the linear system

$$A\,x = b.$$

Problem

We are concerned here by the resolution of the linear system

$$A\,x = b, \quad A \in \mathcal{M}_{n,p}(\mathbb{R}),\ x \in \mathbb{R}^p,\ b \in \mathbb{R}^n,\ n > p.$$

Usually n is much greater than p.
In general, such a linear system does not admit an exact solution. We are therefore looking for an *approximated* solution. Precisely, we replace the resolution of this linear system by the following optimization problem

$$\min_{x \in \mathbb{R}^p} ||A\,x - b|| \tag{2}$$

where $||.||$ is the classical Euclidean norm of $\mathbb{R}^n$.

In the following, a vector of $\mathbb{R}^k$ is identified with the column matrix of its components in the canonical basis. As an example, the inner product of two vectors x and y is written in the matrix form $x^T\,y$.

Weighted inner product

Consider the vector space $\mathbb{R}^n$ equipped with the inner product

$$(u, v) \in \mathbb{R}^n \times \mathbb{R}^n \quad \mapsto \quad \langle u, v \rangle_\Omega = u^T \, \Omega \, v = \sum_{i=1}^{n} w_i \, u_i \, v_i$$

where Ω is the diagonal matrix

$$\Omega = \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} \quad \text{with } \; w_i > 0, \;\; i = 1, \ldots, n,$$

which induces the norm

$$u \in \mathbb{R}^n \quad \mapsto \quad \| \, u \, \|_\Omega = \left(\langle u, u \rangle_\Omega \right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} w_i \, u_i^2 \right)^{\frac{1}{2}}.$$

We now consider the following optimization problem

$$\min_{x \in \mathbb{R}^p} ||A\,x - b||_\Omega \tag{3}$$

which allows to weight separately equations of the initial linear system with weights w_i.

$\rightarrow$ Minimizing the norm $||A\,x - b||_\Omega$ is equivalent to minimize the squared norm $||A\,x - b||_\Omega^2$.

Least squares approximation

The following proposition is the main result about least squares approximation.

We will give two proofs of this result.
- An algebraic proof.
- A simpler proof, using projection in an Euclidean space and proposition 6.4

Proposition 6.9

If matrix A has maximal rank p (which means that its columns are linearly independent), then the optimization problem

$$\min_{x \in \mathbb{R}^p} ||A\,x - b||_\Omega^2$$

admits a unique solution x^* defined by

$$A^T\,\Omega\,A\,x^* = A^T\,\Omega\,b\,. \tag{4}$$

Equations (4) are called the normal equations.

Algebraic proof of proposition 6.9

We first prove that the symmetric matrix $\hat{A} = A^T \Omega A$ of order p is positive definite

$$\forall z \in \mathbb{R}^p,\ z \neq 0, \quad z^T \hat{A}\, z = z^T (A^T \Omega A)\, z = (Az)^T \Omega (A\, z) = ||A\, z||_{\Omega}^2\ > 0$$

since $\ker(A) = \{0\}$ as A is of rank p.

Finally, we just need to remark that the application $x \mapsto ||A\, x - b||_{\Omega}^2$ is a positive definite quadratic form. Precisely, for $x \in \mathbb{R}^p$ we have

$$\begin{aligned} \| A\, x - b \|_{\Omega}^2 &= (A\, x - b)^T \Omega\, (A\, x - b) \\ &= x^T \left(A^T \Omega A\right) x - 2\, (A^T \Omega\, b)^T x + b^T \Omega\, b \\ &= x^T \hat{A}\, x - 2\, v^T x + c \end{aligned}$$

with $v = A^T \Omega b \in \mathbb{R}^p$ and $c = b^T \Omega\, b \in \mathbb{R}$, which concludes the proof by proposition 6.8.

Proof using projections in the Euclidean space $(\mathbb{R}^n, \langle . , . \rangle_\Omega)$

We introduce the vector subspace U defined by

$$U = \{A\,x,\ x \in \mathbb{R}^p\} \ = \ \mathrm{Im}\,(A).$$

so that, our optimization problem can be rewritten as follows

$$\min_{y \in U} ||y - b||_\Omega.$$

By proposition 6.4, the vector $\hat{y} \in U$ which minimizes the norm $||y-b||_\Omega$ is the orthogonal projection of b on U, and is characterized by

$$\begin{aligned}
\langle \hat{y} - b, y \rangle_\Omega &= 0, \quad \forall y \in U, \\
\langle A\hat{x} - b, Ax \rangle_\Omega &= 0, \quad \forall x \in \mathbb{R}^p, \qquad \text{with } \hat{y} = A\hat{x} \ \text{ and } \ y = Ax \\
(Ax)^T\, \Omega\, (A\hat{x} - b) &= 0, \quad \forall x \in \mathbb{R}^p, \\
x^T \left[A^T \Omega (A\hat{x} - b)\right] &= 0, \quad \forall x \in \mathbb{R}^p, \\
A^T\, \Omega\, A\, \hat{x} &= A^T\, \Omega\, b\,,
\end{aligned}$$

which shows that an optimal solution $\hat{x}$ of the optimization problem verify the normal equations (4).

The projection $\hat{y}$ of vector b on the subspace U is unique.
The unicity of the solution then depends on the rank of the matrix A, i.e., is a consequence of the injectivity of the linear map $x \mapsto A\,x$.

Objective

We consider the numerical resolution of the normal equations

$$A^T A\, x = A^T b\,,$$

with $A \in \mathcal{M}_{n,p}(\mathbb{R})$, $x \in \mathbb{R}^p$ and $b \in \mathbb{R}^n$, where $n \geq p$, which requires a specific treatment in order to avoid the propagation of numerical rounding errors.

If the matrix A has maximal rank p, the symmetric matrix $A^T A$ is positive definite, so that the normal equations $A^T A\, x = A^T b$ can be solved through the *Cholesky factorization* :

$$A^T A = L L^T$$

where L is a lower triangular matrix with positive diagonal.
However, such a factorization has the major drawback of propagating the rounding errors. For this reason, the QR factorization is preferred.

The QR factorization reduces the minimization of the norm

$$||A\, x - b||_2$$

to the resolution of a triangular linear system.

QR minimization

Proposition 6.10 (rectangular case)

Given a matrix $A \in \mathcal{M}_{n,p}(\mathbb{R})$, $n \geq p$, with maximal rank p, there exists an orthogonal matrix $Q \in \mathcal{M}_n(\mathbb{R})$ and a unique upper triangular matrix $R \in \mathcal{M}_{n,p}(\mathbb{R})$ with positive diagonal elements, such that

$$A = Q\,R.$$

We return to the optimization problem

$$\min_{x \in \mathbb{R}^p} \; ||A\,x - b||_2 \tag{5}$$

with $b \in \mathbb{R}^n$ and where $A \in \mathcal{M}_{n,p}(\mathbb{R})$, $n \geq p$, is of maximal rank p.

We consider the factorization $A = QR$ and we introduce the following notations :

$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, where R_1 is an upper p-square triangular matrix with positive diagonal,

$Q^T b = \begin{pmatrix} \left(Q^T b\right)_1 \\ \left(Q^T b\right)_2 \end{pmatrix}$, with $\left(Q^T b\right)_1 \in \mathbb{R}^p$ and $\left(Q^T b\right)_2 \in \mathbb{R}^{n-p}$,

$||.||_{2,r}$ is the usual Euclidean norm in $\mathbb{R}^r$ (by default $||.||_2 = ||.||_{2,n}$).

QR minimization

Then, for any vector $x \in \mathbb{R}^p$, we have

$$
\begin{aligned}
||A\,x - b||_2^2 &= ||QR\,x - b||_2^2 \\
&= ||Q^T\big(QR\,x - b\big)||_2^2 && \text{since } Q^T \text{ is orthogonal} \\
&= ||R\,x - Q^T b||_2^2 \\
&= ||\begin{pmatrix} R_1 x \\ 0 \end{pmatrix} - \begin{pmatrix} (Q^T b)_1 \\ (Q^T b)_2 \end{pmatrix}||_2^2 \\
&= ||R_1\,x - (Q^T b)_1||_{2,p}^2 + ||(Q^T b)_2||_{2,n-p}^2 && \text{since } R_1\,x - (Q^T b)_1 \perp (Q^T b)_2
\end{aligned}
$$

Finally, the norm $||A\,x - b||_2$, $x \in \mathbb{R}^p$, is minimal for $||R_1\,x - (Q^T b)_1||_{2,p} = 0$, from which we deduce the following proposition.

Proposition 6.11

With the previous hypotheses, the norm $||A\,x - b||_2$ is minimal for

$$\hat{x} = R_1^{-1}\,(Q^T b)_1$$

and the minimal value of the norm $||A\,x - b||_2$ is $||(Q^T b)_2||_{2,n-p}$.

Objective

Curve fitting is the process of constructing a curve (a mathematical function) that has the best fit to a series of data points.

Example :
We consider the problem of modeling the link between two variables X and Y for which we have a sample of n measurements

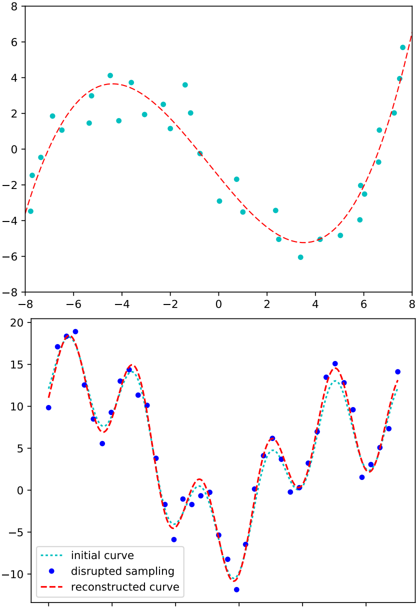
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Precisely, we desire to model the dependance between X and Y by the following relation

$$Y = a_1 f_1(X) + a_2 f_2(X) + \cdots + a_p f_p(X) \tag{6}$$

with p elementary functions f_k (e.g., x^α, $\ln x$, $\exp x$, $\sin x$, $\cos x$, $\ldots$) $(p < n)$ assumed to be linearly independent. Coefficients a_k are the unknown *parameters* of the model and will have to be estimated.

Objective



Problem modelling

Fitting the previous model (6)

$$Y = a_1 f_1(X) + a_2 f_2(X) + \cdots + a_p f_p(X)$$

to the measurement data leads to the n relations $(n > p)$

$$\left\{ \begin{array}{rcl} a_1 f_1(x_1) + a_2 f_2(x_1) + \cdots + a_p f_p(x_1) & = & y_1 + \epsilon_1 \\ a_1 f_1(x_2) + a_2 f_2(x_2) + \cdots + a_p f_p(x_2) & = & y_2 + \epsilon_2 \\ & \vdots & \\ a_1 f_1(x_i) + a_2 f_2(x_i) + \cdots + a_p f_p(x_i) & = & y_i + \epsilon_i \\ & \vdots & \\ a_1 f_1(x_n) + a_2 f_2(x_n) + \cdots + a_p f_p(x_n) & = & y_n + \epsilon_n \end{array} \right. \tag{7}$$

where each ϵ_i represents the error of the model on the measurement (x_i, y_i).
We then express these n linear equations in matrix form

$$\begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & & & \vdots \\ f_1(x_i) & f_2(x_i) & \cdots & f_p(x_i) \\ \vdots & & & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_p(x_n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \Leftrightarrow \mathbf{A}\,\mathbf{u} = \mathbf{b} + \boldsymbol{\epsilon} \tag{8}$$

Problem modelling

In order to estimate parameters a_j of the model (6) we introduce a *global error E* defined by

$$E(a_1,\dots,a_p) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \Big[a_1 f_1(x_i) + a_2 f_2(x_i) + \dots + a_p f_p(x_i) - y_i\Big]^2$$
$$= ||\epsilon||^2 = ||A\,u - b||^2,$$

where $||.||$ denotes the classical Euclidean norm.

Finally, we need to consider the minimization problem

$$\min_{a_1,\dots,a_p} E(a_1,\dots,a_p) \quad = \quad \min_u \; ||A\,u - b||^2$$

The function

$$E \;:\; \begin{array}{ccc} \mathbb{R}^p & \rightarrow & \mathbb{R} \\ (a_1,\dots,a_p) & \mapsto & E(a_1,\dots,a_p) \end{array}$$

is polynomial, quadratic and therefore of class C^2. Therefore, this minimization problem can be considered in two equivalent ways :

- as an over-determined linear system : minimization of a quadratic form
- as the minimization of a function of several variables (the coefficients a_k) with tools of section 4 of chapter on *Prerequisitesin Maths*. We will consider this last approach.

Minimization : $\min_{a_1,\ldots,a_p} E(a_1,\ldots,a_p)$

Determination of the *critical points* ⇒ system of p linear equations

$$\left\{\begin{array}{rcl} \dfrac{\partial E}{\partial a_1}(a_1,\ldots,a_p) & = & 2\displaystyle\sum_{i=1}^{n} f_1(x_i)\left[a_1 f_1(x_i)+\cdots+a_p f_p(x_i)-y_i\right]=0 \\ & \vdots & \\ \dfrac{\partial E}{\partial a_p}(a_1,\ldots,a_p) & = & 2\displaystyle\sum_{i=1}^{n} f_p(x_i)\left[a_1 f_1(x_i)+\cdots+a_p f_p(x_i)-y_i\right]=0 \end{array}\right.$$

$$\Leftrightarrow \begin{pmatrix} \sum_{i=1}^{n} f_1^2(x_i) & \sum_{i=1}^{n} f_1(x_i)f_2(x_i) & \cdots & \sum_{i=1}^{n} f_1(x_i)f_p(x_i) \\ \sum_{i=1}^{n} f_2(x_i)f_1(x_i) & \sum_{i=1}^{n} f_2^2(x_i) & \cdots & \sum_{i=1}^{n} f_2(x_i)f_p(x_i) \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} f_p(x_i)f_1(x_i) & \sum_{i=1}^{n} f_p(x_i)f_2(x_i) & \cdots & \sum_{i=1}^{n} f_p^2(x_i) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i f_1(x_i) \\ \sum_{i=1}^{n} y_i f_2(x_i) \\ \vdots \\ \sum_{i=1}^{n} y_i f_p(x_i) \end{pmatrix}$$

$$\Leftrightarrow \qquad \left(A^T A\right) u = A^T b$$

Minimization

We thus get the normal equations

$$\left(A^T A\right) u = A^T b$$

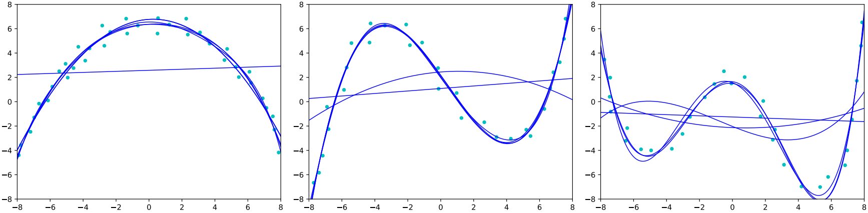
- If the rank of matrix A is maximum, that is equal to p, this linear system possesses a unique solution : the critical point $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_p)$, which defines a strict global minimum of the error function E.
- For the critical point $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_p)$, the global error $E(\hat{a}_1, \ldots, \hat{a}_p)$ is called *residual error* and the value

$$\left[\frac{1}{n}E(\hat{a})\right]^{\frac{1}{2}} = \left[\frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2\right]^{\frac{1}{2}}$$

is named the *residual standard deviation*.
Two regression models can be compared for a same data set by means of their residual standard deviations.

Example : residual standard deviation

Given a set of data points, we apply the least squares method so as to determine the best polynomial approximation for degree $d = 1, 2, \dots, 6$.



In each case, we evaluate the residual standard deviation $RSD(d)$.
Of course, the function $RSD(d)$ is decreasing with the degree, but we can see a gap for a certain degree in each case.

degree 1 RSD = 3.259601566	degree 1 RSD = 3.803313672	degree 1 RSD = 3.381303542
degree 2 RSD = 0.639429602	degree 2 RSD = 3.669523335	degree 2 RSD = 3.305089256
degree 3 RSD = 0.639135969	degree 3 RSD = 0.774789466	degree 3 RSD = 3.189389063
degree 4 RSD = 0.567980553	degree 4 RSD = 0.768156302	degree 4 RSD = 0.914385591
degree 5 RSD = 0.563194887	degree 5 RSD = 0.764840727	degree 5 RSD = 0.812359858
degree 6 RSD = 0.545339539	degree 6 RSD = 0.741644575	degree 6 RSD = 0.804901415

Fitting a straight line

We desire to fit a straight line (the model) with equation $Y = \alpha\, X + \beta$ to the data (x_i, y_i), which leads to the linear system

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \qquad \Leftrightarrow \qquad A\, u = b + \epsilon.$$

The normal equations $A^T A\, u = A^T b$ admits a unique solution, $\hat{\alpha}$, $\hat{\beta}$ given by

$$\hat{\alpha} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta} = \bar{y} - \hat{\alpha}\,\bar{x} \quad \text{with} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ \text{ and } \ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The straight line $Y = \hat{\alpha}\, X + \hat{\beta}$ is the *line of linear regression* of (or associated with) data points (x_i, y_i).

Note that the line of linear regression goes through the barycenter $(\bar{x}, \bar{y})$ of data points (x_i, y_i)

Linear correlation

With data points (x_i, y_i) we define

$$\mathrm{var(X)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{the empirical variance of } X$$

$$\mathrm{var(Y)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \text{the empirical variance of } Y$$

$$\mathrm{cov(X, Y)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad \text{the empirical covariance between } X \text{ and } Y$$

so that the line of linear regression $Y = \hat{\alpha}\, X + \hat{\beta}$ is defined by

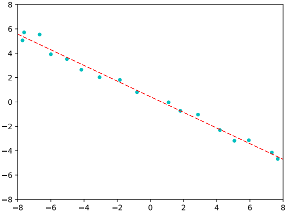
$$\hat{\alpha} = \frac{\mathrm{cov(X, Y)}}{\mathrm{var(X)}} \qquad \text{and} \qquad \hat{\beta} = \bar{y} - \hat{\alpha}\,\bar{x}.$$

Pearson's correlation coefficient :

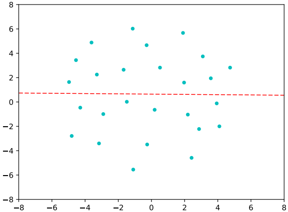
$$r = \mathrm{corr(X, Y)} = \frac{\mathrm{cov(X, Y)}}{\sqrt{\mathrm{var(X)}}\,\sqrt{\mathrm{var(Y)}}}\,, \qquad -1 \leq \mathrm{r} \leq 1$$

→ informally, *correlation* is synonymous with dependence
→ sensitive only to a *linear* relationship between two variables

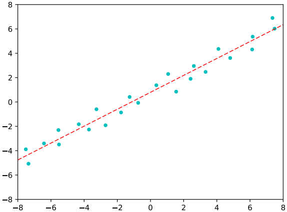
Linear correlation



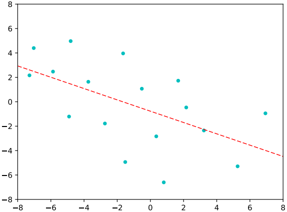
(a) $r = -0.99$



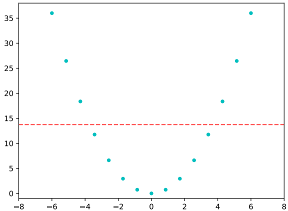
(b) $r = -0.011$



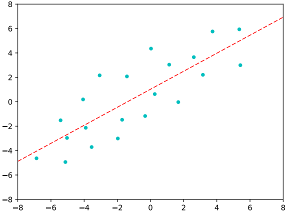
(c) $r = 0.98$



(d) $r = -0.56$



(e) $r = -10^{-16}$



(f) $r = 0.82$

Objective

Given a sequence of points $M_i = (x_i, y_i)$, $1 \leq i \leq N$, we look for a parametric polynomial curve of prescribed degree which approximates these data points, that is which passes as close as possible to each point (x_i, y_i) for some prescribed parameter t_i.

Precisely, we look for a parametric polynomial curve

$$s \; : \; \begin{array}{ccc} [a,b] \in \mathbb{R} & \longrightarrow & \mathbb{R}^2 \\ t & \longmapsto & s(t) = \begin{pmatrix} s_x(t) \\ s_y(t) \end{pmatrix} \end{array}$$

such that

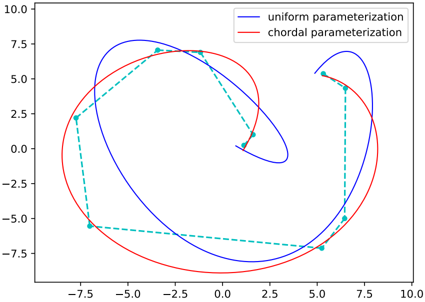
$$s(t_i) \simeq M_i \quad \Leftrightarrow \quad \left\{ \begin{array}{l} s_x(t_i) \simeq x_i \\ s_y(t_i) \simeq y_i \end{array} \right. , \quad 1 \leq i \leq N$$

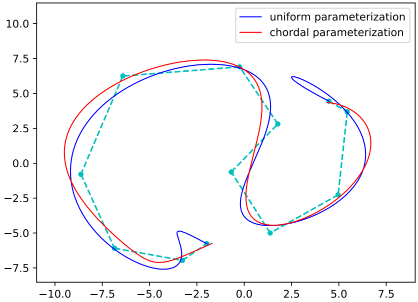
with a sequence of nodes $a = t_1 < t_2 < \cdots < t_N = b$, and where $s_x(t)$ and $s_y(t)$ are polynomials of prescribed degree.

$\rightarrow$ we are therefore reduced to solve two separate least squares approximation problems.

$\rightarrow$ we consider the *uniform* and the *chordal* parameterizations.

Examples





Least squares approximation of a polygon by parametric polynomial curves of degree 5 and 8, with uniform and chordal parameterization.