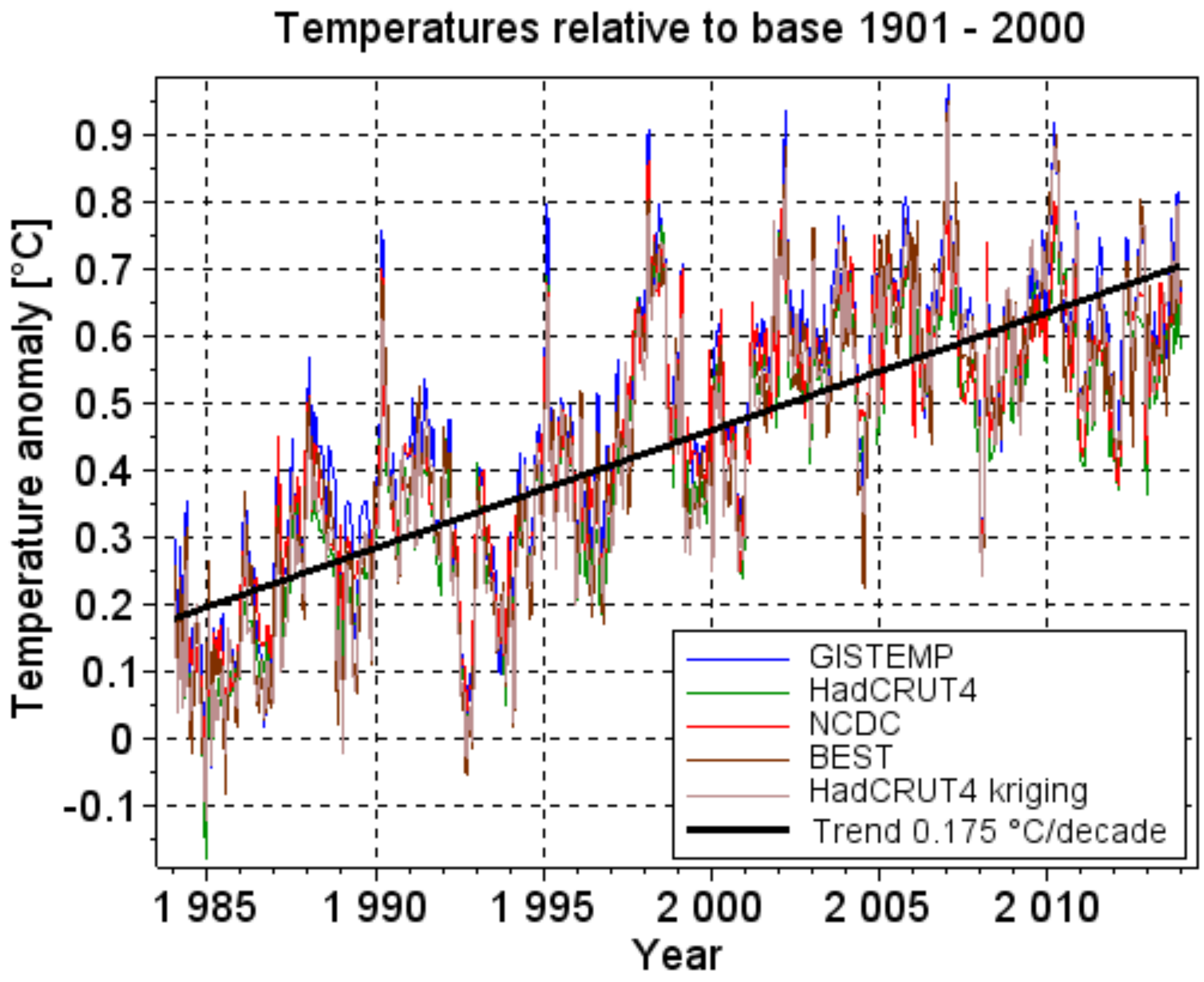
Linear Regression



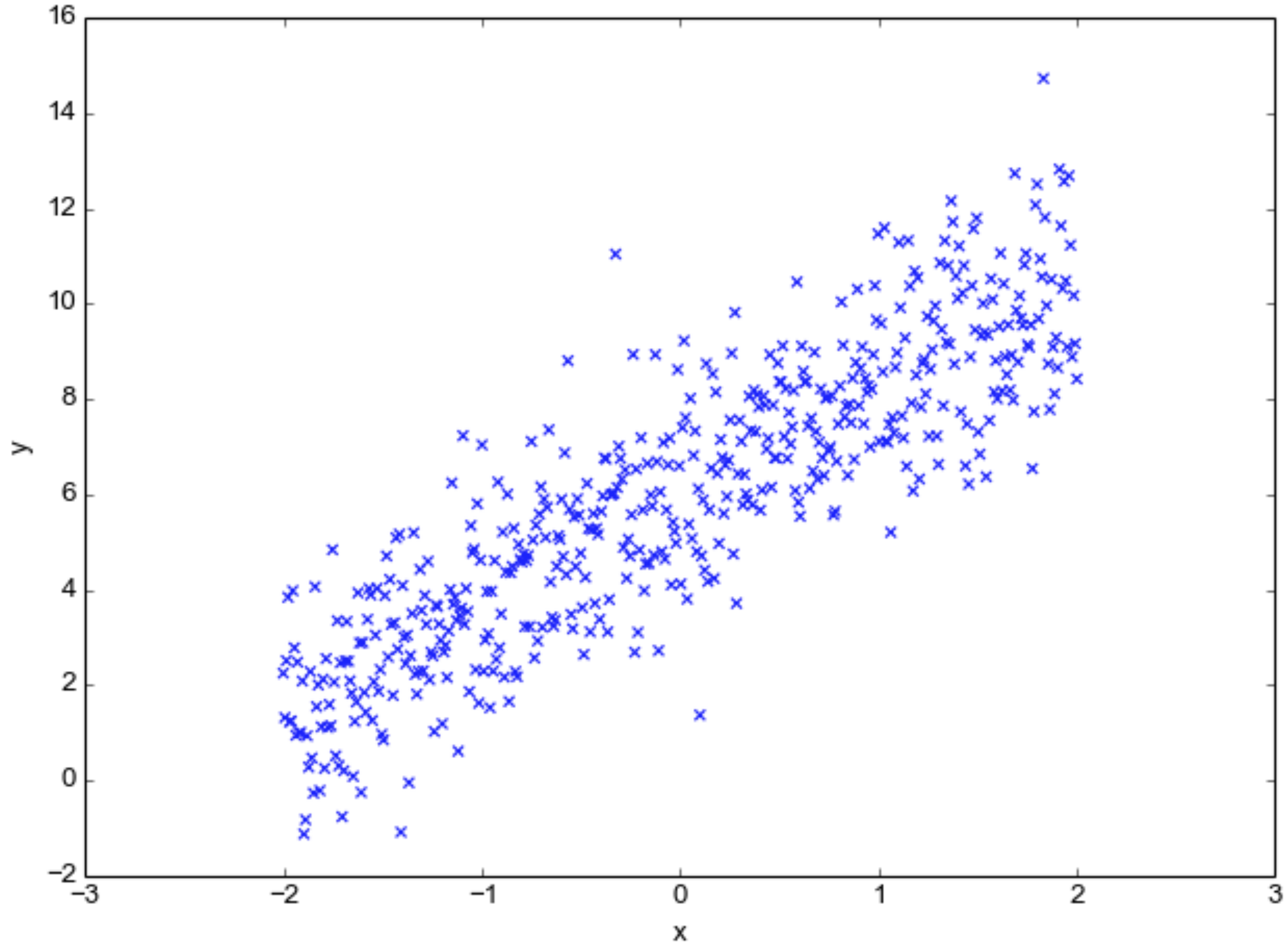
Mean global temperature compared to the average 1901-2000

Linear Regression
Modeling Data with Linear Functions

- Many scientific problems require prediction
- Examples are temperature prediction, economic forecasting, etc.

- We want a rule that predicts y from x

Example



- Datapoints (x_i, y_i) are given

- We assume a linear relation

$$y(x) = a + b * x$$

- But there are errors

$$y(x) = a + b * x + \epsilon(x)$$

First modeling idea

- Large system of equations

$$y_1 = a + b \cdot x_1$$
$$y_2 = a + b \cdot x_2$$
$$\vdots$$
$$y_N = a + b \cdot x_N$$

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$$y = X\beta, \quad X \in \mathbb{R}^{N \times 2}$$

- Usually, $N \gg 2$, the linear system is overdetermined and we can't solve it

First modeling idea

- There is an error in every equation

$$y_1 + \epsilon_1 = a + b \cdot x_1$$
$$y_2 + \epsilon_2 = a + b \cdot x_2$$
$$\vdots$$
$$y_N + \epsilon_N = a + b \cdot x_N$$

$$y = X\beta + \epsilon, \quad X \in \mathbb{R}^{N \times 2}$$

- Distorted linear system
- We cannot solve the overdetermined problem
- But, we can find the solution that minimizes the error

$$\epsilon = y - X\beta \to \min$$

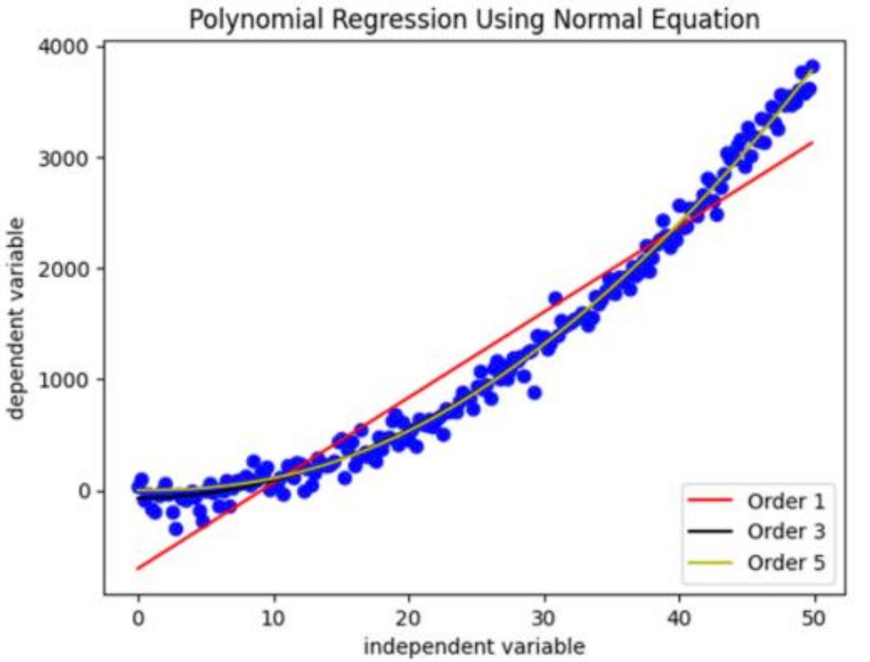
Least Squares Problem

$$\epsilon = y - X\beta \rightarrow \min$$

- Minimization of the error in the $l_2 - \mathrm{Norm}$

$$\min_{\beta} \|\epsilon\|_2^2 = \min_{\beta} \|y - X\beta\|_2^2$$

Generalization



- The relation between x and y must not be linear
- general polynomial model

$$
\begin{aligned}
y_1 + \epsilon_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots + \beta_p x_1^p \\
y_2 + \epsilon_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \cdots + \beta_p x_2^p \\
&\vdots \\
y_N + \epsilon_N &= \beta_0 + \beta_1 x_N + \beta_2 x_N^2 + \cdots + \beta_p x_N^p
\end{aligned}
$$

- We still get a linear system

$$
y = X\beta + \epsilon \qquad X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^p \end{pmatrix} \in \mathbb{R}^{N \times p+1}
$$

Solving the Least Squares Problem

Theorem Let $X \in \mathbb{R}^{Nxp}$ with $N > p$ and **full rank** $\boldsymbol{p}$. Then, the unique solution $\beta \in \mathbb{R}^p$ to the Least Squares Problem

$$\min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2$$

is given as solution to the **Normal equation**

$$X^T X\beta = X^T y$$

Proof

(1) Show that $X^T X$ is sym. pos. def.

a) $(X^T X)^T = X^T (X^T)^T = X^T X$

b) $\langle X^T X \beta, \beta \rangle = \langle X\beta, X\beta \rangle = \| X\beta \|_2^2$

$X \in \mathbb{R}^{N \times p}$ with $N > p$ has full rank p

$\Rightarrow X\beta = 0$ only if $\beta = 0$

(II) Let $\beta \in \mathbb{R}^p$ be the solution normal equation
$X^T X \beta = X^T y$. Then β is the minimum

For any $\alpha \in \mathbb{R}^p$

$$\|y - A(\alpha+\beta)\|_2^2 = \langle y - A(\alpha+\beta), y - A(\alpha+\beta)\rangle$$

$$= \|y - A\beta\|_2^2 + \underbrace{\|A\alpha\|_2^2}_{\geq 0} - \underbrace{2\langle y - A\beta, A\alpha\rangle}_{= -2\langle \underbrace{A^T y - A^T A\beta}_{=0}, \alpha\rangle}$$

$$= \|y - A\beta\|_2^2 + \|A\alpha\|_2^2$$

$$\geq \|y - A\beta\|_2^2$$

(iii) Assume $\|y - A\beta\|_2^2 \leq \|y - A(\beta + \alpha)\|_2^2 \quad \forall \alpha \in \mathbb{R}^p$

$$= \|y - A\beta\|_2^2 + \|A\alpha\|_2^2 - 2\langle y - A\beta, A\alpha\rangle \quad \forall \alpha \in \mathbb{R}^p$$

$$\Leftrightarrow$$

$$2\langle y - A\beta, A\alpha\rangle \leq \|A\alpha\|_2^2$$

Let $\alpha = s \cdot e_i$ where $s \in \mathbb{R}$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{th pos.}$$

$\Rightarrow$

$$2\langle A^T y - A^T A\beta, s e_i\rangle \leq s^2 \|Ae_i\|_2^2$$

$$\leq s^2 \|A\|_2^2$$

$$2\langle A^T y - A^T A\beta, s e_i \rangle \le s^2 \|Ae_i\|_2^2$$
$$\le s^2 \|A\|_2^2$$

$\Leftrightarrow$

$$2(A^T y - A^T A\beta)_i \cdot s \le |s|^2 \|A\|_2^2$$

$\forall s \in \mathbb{R}$
$\forall i = 1, \ldots, N$

$\Rightarrow$

$$2(A^T y - A^T A\beta)_i \le |s| \cdot \|A\|_2^2 \qquad s > 0$$
$$2(A^T y - A^T A\beta)_i \ge -|s| \cdot \|A\|_2^2 \qquad s < 0$$

$\xrightarrow[|s| \to 0]{} \quad (A^T y - A^T A\beta)_i = 0 \Leftrightarrow A^T y - A^T A\beta = 0$
$\Leftrightarrow A^T A\beta = A^T y \quad \square$

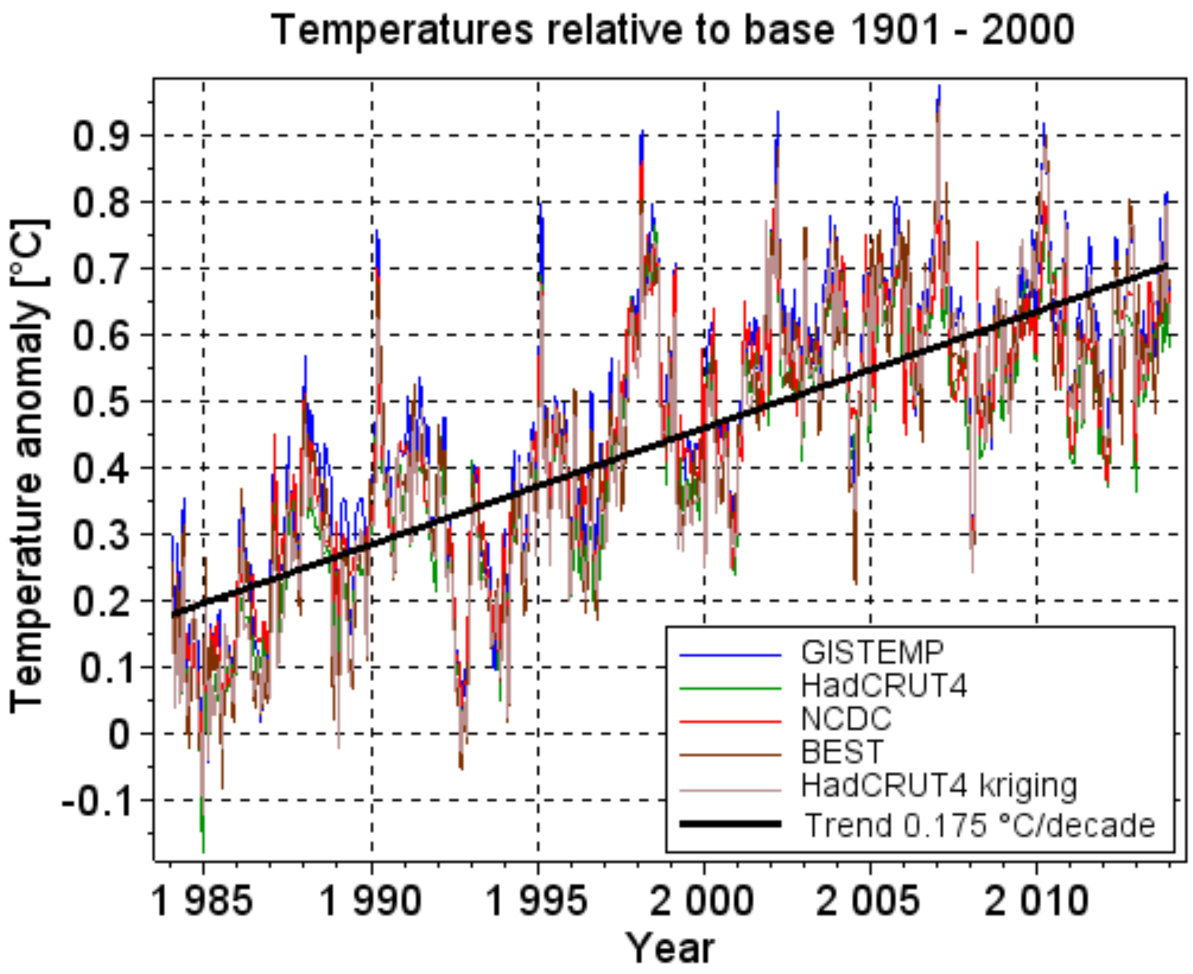
Example: Linear Regression

- We consider the linear model

$$y_i + \epsilon_i = \beta_0 + \beta_1 x_i$$

- The matrix of the **normal equation** is given as

$$A^T A = \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$$



Properties of the Normal Equation

Remark

Let $X \in \mathbb{R}^{Nxp}$ with $N > p$ and **full rank $\boldsymbol{p}$**.

- The matrix $X^T X$ is always symmetric positive definite.
- The matrix can be very ill-conditioned

Condition number?

Theorem 3.5 (Perturbation Theorem for Linear Systems) *Let $A \in \mathbb{R}^{n\times n}$ be a regular matrix, $b \in \mathbb{R}^n$ the right-hand side. Furthermore, let $x \in \mathbb{R}^n$ be the solution of the linear system $Ax = b$. For the solution $\tilde{x} \in \mathbb{R}^n$ of the perturbed system $\tilde{A}\tilde{x} = \tilde{b}$ with perturbations $\delta b = \tilde{b} - b$ and $\delta A = \tilde{A} - A$, under the condition*

$$\|\delta A\| < \frac{1}{\|A^{-1}\|},$$

the following estimate holds

$$\frac{\|\delta x\|}{\|x\|} \le \frac{\mathrm{cond}(A)}{1 - \mathrm{cond}(A)\|\delta A\|/\|A\|}\left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right),$$

where the condition number *is defined as*

$$\mathrm{cond}(A) = \|A\|\,\|A^{-1}\|.$$

If A is sym. pos. def.,

$$\mathrm{cond}_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$$

Fraction of Eigenvalues

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \cdots \\ \frac{1}{3} & \frac{1}{4} & \vdots & \cdots \\ \frac{1}{4} & \vdots & \vdots & \end{pmatrix}$$

Hilbert Matrix

$$\operatorname{cond}(A) \approx 28000$$

Condition number?

Remark

Let $X \in \mathbb{R}^{Nxp}$ with $N > p$ and **full rank** $\boldsymbol{p}$.

- The matrix $X^T X$ is always symmetric positive definite.
- The **matrix can be very ill-conditioned**

$$\begin{aligned} \operatorname{cond}(X^T X) &= \|X^T X\| \cdot \|(X^T X)^{-1}\| \leq \|X^T\| \cdot \|X\| \cdot \|X^{-T}\| \cdot \|X^{-1}\| \\ &\leq \|X\|^2 \cdot \|X^{-1}\|^2 = \operatorname{cond}(X)^2 \end{aligned}$$

Orthogonal Matrices

Definition 4.1 (Orthogonal Matrix) A matrix $Q \in \mathbb{R}^{n\times n}$ is called *orthogonal*, if its row and column vectors form an orthonormal basis of $\mathbb{R}^n$. It holds $Q^TQ = I$, thus $Q^{-1} = Q^T$.

Theorem 4.2 (Orthogonal Matrix) *Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. Then Q is regular, and it holds*

$$Q^{-1} = Q^T, \quad Q^TQ = I, \quad \|Q\|_2 = 1, \quad \mathrm{cond}_2(Q) = 1.$$

$\|Q^{-1}\|_2 = \|Q^T\|_2 = \|Q\|_2 = 1$

$$\underbrace{\begin{pmatrix} \cdots & q_1 & \cdots \\ \cdots & q_2 & \cdots \\ & \vdots & \\ \cdots & q_n & \cdots \end{pmatrix}}_{Q^T} \underbrace{\begin{pmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & & \vdots \end{pmatrix}}_{Q} = I$$

$Q \iff \langle q_i, q_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Further,

1. *It holds* $\det(Q) = 1$ *or* $\det(Q) = -1$.
2. *For two orthogonal matrices $Q_1, Q_2 \in \mathbb{R}^{n\times n}$, the product Q_1Q_2 is also an orthogonal matrix. For any matrix $A \in \mathbb{R}^{n\times n}$ it holds $\|QA\|_2 = \|A\|_2$.*
3. *For arbitrary vectors $x, y \in \mathbb{R}^n$ it holds*

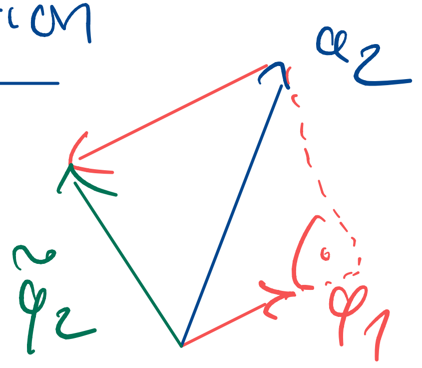
$$\|Qx\|_2 = \|x\|_2, \quad (Qx, Qy)_2 = (x, y)_2.$$

The QR-Decomposition

Theorem 4.4 (QR Factorization) *Let $A \in \mathbb{R}^{n\times n}$ be a regular matrix. Then there exists a factorization $A = QR$ into an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and an upper right triangular matrix $R \in \mathbb{R}^{n\times n}$.*

regular

Gram-Schmidt Orthogonalization



Given $a_1, \dots, a_p$ linearly independent

$$q_1 := \frac{a_1}{\|a_1\|}$$

$i > 1$

$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} \langle a_i, q_j \rangle \cdot q_j \quad , \quad q_i := \frac{\tilde{q}_i}{\|\tilde{q}_i\|_2}$$

$$Q := \begin{pmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

$$Q^T Q = I$$

$$\langle q_i, q_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and

$$\langle q_i, a_j \rangle = \begin{cases} 0 & i > j \\ ? & i \leq j \end{cases}$$

$$\Rightarrow \quad Q^T \cdot \begin{pmatrix} \vdots & \vdots & & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} * & \cdots & \cdots & * \\ 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & * \end{pmatrix}$$

QR for rectangular matrices

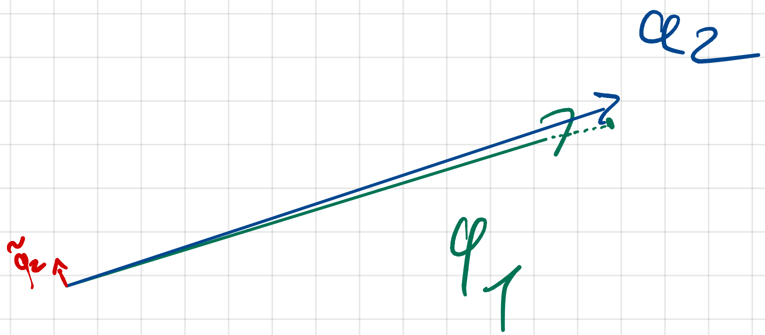
$$\begin{pmatrix} \vdots & & \vdots \\ a_1 & \cdots & a_p \\ \vdots & & \vdots \end{pmatrix} \in \mathbb{R}^{N \times p} \quad \text{with} \quad N > p$$

rank p $\Rightarrow$ $a_1, \dots, a_p$ are linearly independent

Gram-Schmidt $\quad q_1, \dots, q_p \in \mathbb{R}^N$

$$\underbrace{\begin{pmatrix} \cdots & q_1 & \cdots \\ & \vdots & \\ \cdots & q_p & \cdots \end{pmatrix}}_{Q^T} \begin{pmatrix} \vdots & & \vdots \\ a_1 & & a_p \\ \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ 0 & & * \end{pmatrix}$$

Stability of Gram-Schmidt



$$\begin{array}{r} 1234{,}1251 \\ 1234{,}1249 \\ \hline 0{,}0002 \end{array}$$

rel. err. 10^{-8}

$$\begin{array}{r} 1234{,}1252 \\ 1234{,}1248 \\ \hline 0{,}0004 \end{array}$$

$\leadsto$ rel. err 1

Using the QR-Decomposition for Least Squares

- We solve the normal equation

$$X^T X \beta = X^T y$$

- Without even computing the matrix $X^T X$

$$\begin{aligned}
\mathrm{cond}_2(R) &= \mathrm{cond}(Q^T X) \\
&= \|Q^T X\| \cdot \|(Q^T X)^{-1}\| \\
&\leq \|Q\| \, \|Q^{-1}\| \cdot \|X\| \cdot \|X^{-1}\| \\
&= \underbrace{\mathrm{cond}_2(Q)}_{=1} \cdot \mathrm{cond}_2(X)
\end{aligned}$$

1. Compute the QR-Decomposition of *X*

$$X = QR, \quad Q \in \mathbb{R}^{N \times p}, \quad R \in \mathbb{R}^{p \times p}$$

2. Insert the QR-Decomposition into the normal equation

$$(QR)^T QR\beta = (QR)^T y \quad \Leftrightarrow \quad R^T \underbrace{Q^T Q}_{=I} R\beta = R^T Q^T y \quad \Leftrightarrow \quad R\beta = Q^T y$$

Orthogonalization

Implementation 4.8: Gram-Schmidt Orthogonalization

```python
def gram_schmidt(A):
    n = A.shape[0]
    Q = np.zeros_like(A)

    for i in range(n):
        Q[:, i] = A[:, i]
        for j in range(i):
            Q[:, i] -= np.inner(A[:, i], Q[:, j]) * Q[:, j]
        Q[:, i] /= np.linalg.norm(Q[:, i])
    return Q
```

$$q_1 := \frac{a_1}{\|a_1\|}, \quad \tilde{q}_i = a_i - \sum_{j=1}^{i-1} (a_i, q_j) q_j, \; q_i := \frac{\tilde{q}_i}{\|\tilde{q}_i\|} \quad i = 2, 3, \ldots, m.$$

Conditioning of Gram-Schmidt Orthogonalization

$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} (a_i, q_j)_2 q_j$$

- Each step is a projection

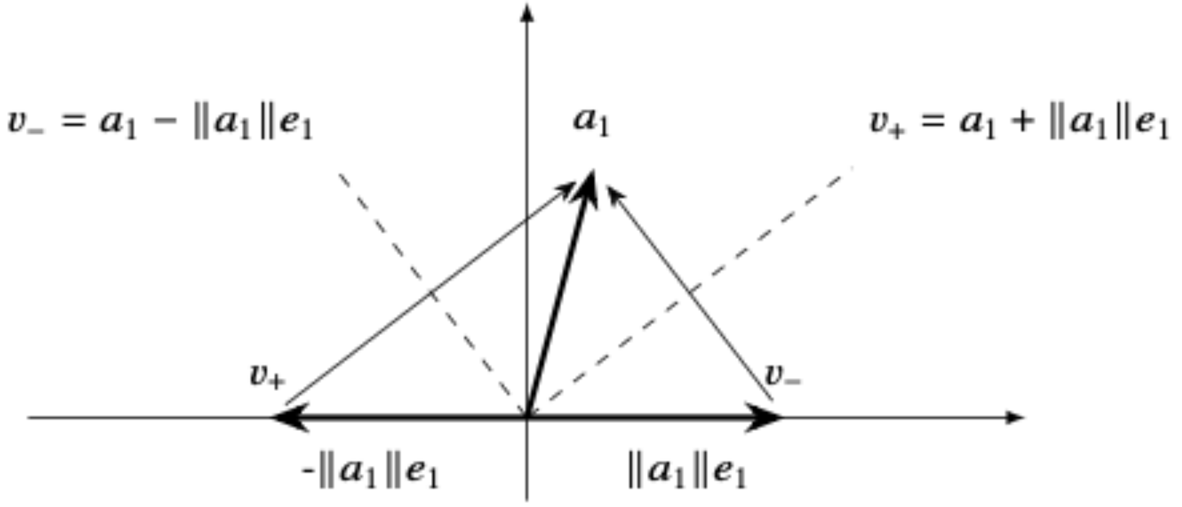
$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} q_j q_j^T a_i = \left(I - \sum_{j=1}^{i-1} q_j q_j^T \right) a_i$$

- Ill-conditioned, if new vector is nearly orthogonal

$$q_j^T a_i \approx \begin{cases} 1 & i = j \\ 0 & i \neq 0 \end{cases} \quad \Rightarrow \quad \tilde{q}_i \approx a_i - a_i = 0$$

Better Option: Householder transformation

- In Step *i* find reflection S_i such that $q_i = S_i a_i = \alpha e_i$



$$A := \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & \ddots & & \vdots \\ a_{31} & a_{32} & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{pmatrix} \quad \Rightarrow \quad A^{(1)} := S^{(1)} A = \left(\begin{array}{c|cccc} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1n}^{(1)} \\ \hline 0 & a_{22}^{(1)} & \ddots & & \vdots \\ 0 & a_{32}^{(1)} & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \cdots & \cdots & a_{nn}^{(1)} \end{array}\right)$$

Householder transformation

Definition 4.16 (Householder Transformation) For a vector $v \in \mathbb{R}^n$ with $\|v\|_2 = 1$, the *dyadic product* is defined by vv^T and the matrix

$$S := I - 2vv^T \in \mathbb{R}^{n \times n}$$

is called *Householder transformation*.

Theorem 4.17 (Householder Transformation) *Every Householder transformation $S = I - 2vv^T$ with $\|v\|_2 = 1$ is symmetric and orthogonal. The product of two Householder transformations $S_1 S_2$ is again an orthogonal matrix.*

Householder transformation

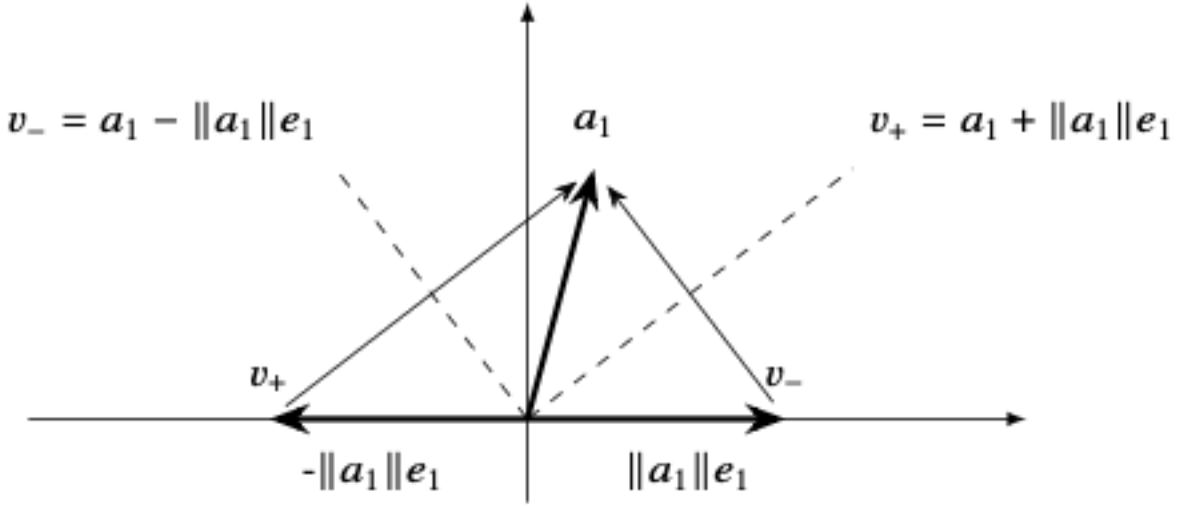
Remark 4.18 (Householder Transformation as Reflection) Let $v \in \mathbb{R}^n$ be an arbitrary normalized vector with $\|v\|_2 = 1$. Further, let $x \in \mathbb{R}^n$ be given with $x = \alpha v + w^\perp$, where $w^\perp \in \mathbb{R}^n$ is a vector with $v^T w^\perp = 0$ in the orthogonal complement to v. Then, for the Householder transformation $S = I - 2vv^T$, it holds:

$$S(\alpha v + w^\perp) = [I - 2vv^T](\alpha v + w^\perp) = \alpha(v - 2v \underbrace{v^T v}_{=1}) + w^\perp - 2v \underbrace{v^T w^\perp}_{=0} = -\alpha v + w^\perp.$$

That is, the Householder transformation describes a reflection on the plane perpendicular to v. ♦

Householder transformation

$$v^{(1)} := \frac{a_1 + \operatorname{sign}(a_{11})\|a_1\|e_1}{\|a_1 + \operatorname{sign}(a_{11})\|a_1\|e_1\|},$$



- The sign is chosen for stability

- The matrix gets smaller and smaller

$$\tilde{v}^{(2)} := \frac{\tilde{a}^{(1)} + \operatorname{sign}(a_{22}^{(1)})\|\tilde{a}^{(1)}\|e_1}{\|\tilde{a}^{(1)} + \operatorname{sign}(a_{22}^{(1)})\|\tilde{a}^{(1)}\|e_1\|}, \quad S^{(2)} = \left(\begin{array}{c|c} 1 & 0 \cdots\cdots \quad 0 \\ \hline 0 & \\ 0 & I - 2\tilde{v}^{(2)}(v^{(2)})^T \\ \vdots & \\ 0 & \end{array}\right) \in \mathbb{R}^{n\times n}.$$

QR with Householder

Theorem 4.19 (QR Factorization With Householder Transformations) *Let $A \in \mathbb{R}^{n\times n}$ be a regular matrix. Then the QR factorization of A according to Householder can be performed in*

$$\frac{2}{3}n^3 + O(n^2)$$

numerically stable elementary operations (i.e., each a multiplication or division as well as an addition).

- QR with Householder is efficient and numerically stable