Numerical Analysis
FMN011

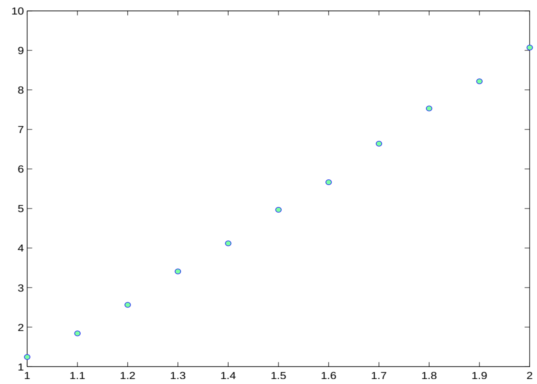
Carmen Arévalo

Lund University

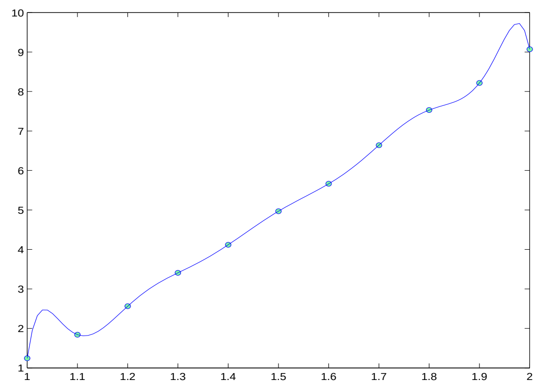
carmen@maths.lth.se

Least Squares

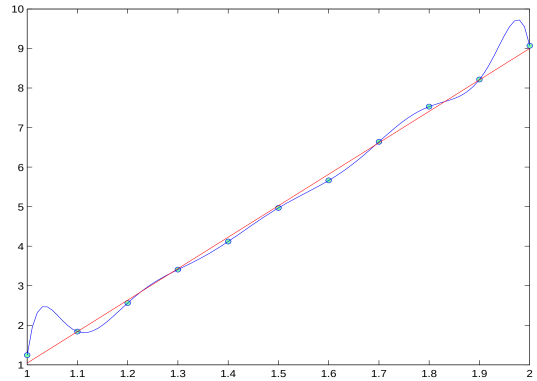
Curve Fitting



Data points too numerous, inaccurately collected.



Interpolating polynomial will have a high degree or a large error.



Underlying mathematical model $y = f(x)$ should go near, not necessarily through the data points

Inconsistent System of Equations

A small number of unknowns should satisfy a larger system of equations but there are slight errors in the coefficients ("noisy data"), so the equations are not exactly satisfied. There is no solution to the over-determined system.

$$\begin{aligned} 1.1\, w_1 + 0.9\, w_2 &= 1.02 \\ x_1 - w_2 &= 0.08 \\ 2.1\, w_1 + 2\, w_2 &= 0.95 \\ 3.2\, w_1 - 2.9\, w_2 &= 0.25 \end{aligned}$$

What is the closest possible "solution"?

Geometric interpretation

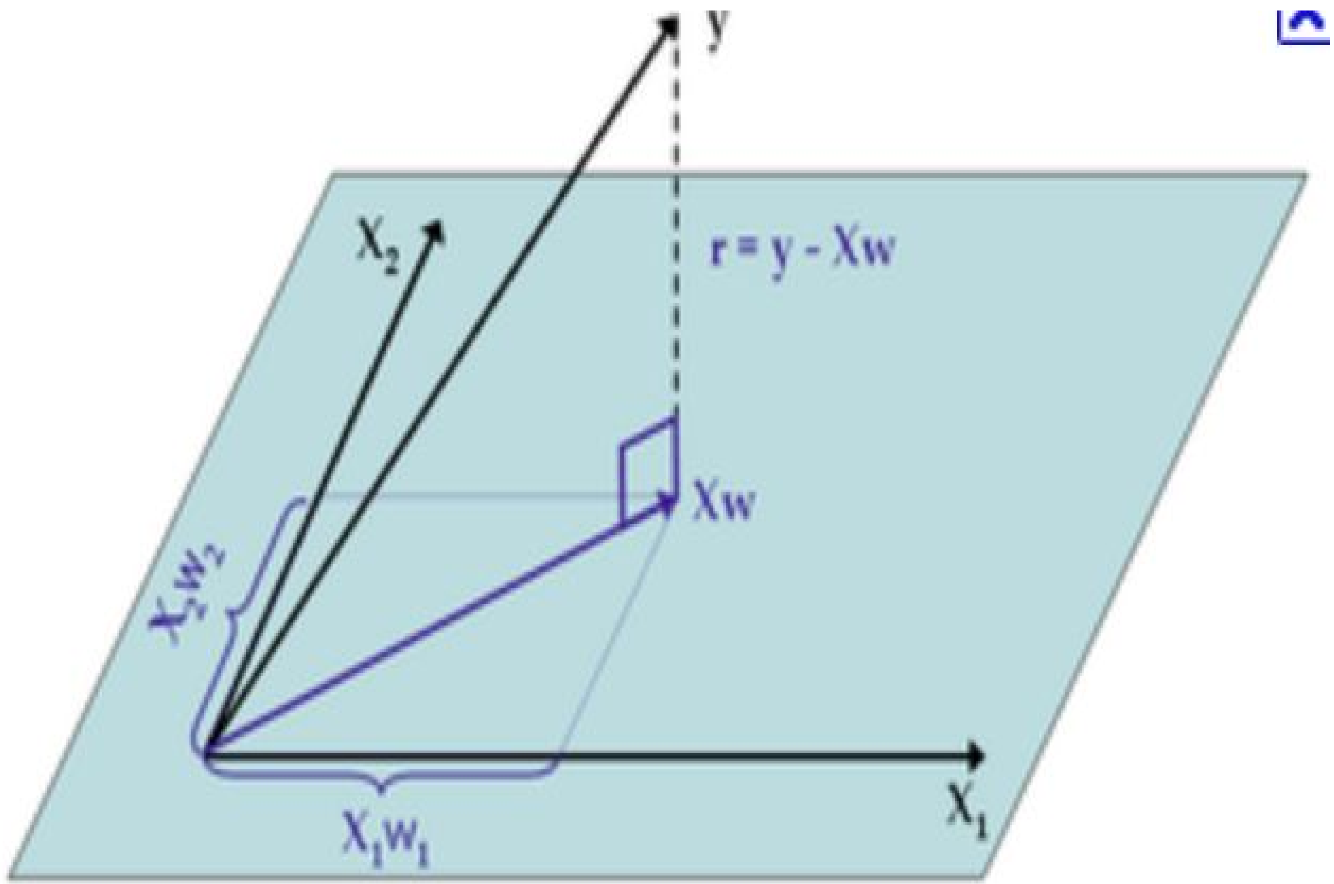
$$w_1 \begin{bmatrix} 1.1 \\ 1 \\ 2.1 \\ 3.2 \end{bmatrix} + w_2 \begin{bmatrix} 0.9 \\ -1 \\ 2 \\ -2.9 \end{bmatrix} = \begin{bmatrix} 1.02 \\ 0.08 \\ 0.95 \\ 0.25 \end{bmatrix}$$

There is no solution because the right-hand side vector $y = [1.02 \quad 0.08 \quad 0.95 \quad 0.25]^T$ is not a linear combination of the two vectors $X_1 = [1.1 \quad 1 \quad 2.1 \quad 3.2]^T$ and $X_2 = [0.9 \quad -1 \quad 2 \quad -2.9]^T$.

The vector on the plane generated by X_1 and X_2 that is closest to y is the perpendicular projection of y onto the plane.

Solution: Project y and solve the system with the new right-hand side.

Residual vector $r = y - Xw$ is perpendicular to the plane.


y
X_2
r = y - Xw
Xw
X_2w_2
X_1w_1
X_1

Fitting Data with Linear Least Squares

Linear least squares:
We look for the best model to fit m data points (t_i, y_i)

$$y(t) = f(t,x) = x_1\Phi_1(t) + x_2\Phi_2(t) + \cdots + x_n\Phi_n(t)$$

where $m > n$. The (yet unknown) parameter vector is $x = [x_1, \ldots, x_n]^T$

E.g., f could be a polynomial

$$f(t,x) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

We are looking for x such that $f(t_i, x) = y_i, \quad i = 1, \ldots, m$. We have an over-determined system with m equations and n unknowns $x_1, \cdots, x_n$.

In this situation there is (usually) no unique solution.

Data Fitting

Best fit: find the parameters x that minimize $\|b - f(t,x)\|_2$

Suppose there are 5 data points and we want f to be a quadratic polynomial.

$$b - f(t,x) = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} - \begin{bmatrix} x_1 + x_2 t_1 + x_3 t_1^2 \\ x_1 + x_2 t_2 + x_3 t_2^2 \\ x_1 + x_2 t_3 + x_3 t_3^2 \\ x_1 + x_2 t_4 + x_3 t_4^2 \\ x_1 + x_2 t_5 + x_3 t_5^2 \end{bmatrix}$$

Problem: Find the values of x_1, x_2, x_3 that make the 2-norm of this vector as small as possible.

$$\left\| \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} - \begin{bmatrix} t_1 + t_2 x_1 + t_3 x_1^2 \\ t_1 + t_2 x_2 + t_3 x_2^2 \\ x_1 + x_2 t_3 + x_3 t_3^2 \\ x_1 + x_2 t_4 + x_3 t_4^2 \\ x_1 + x_2 t_5 + x_3 t_5^2 \end{bmatrix} \right\|_2 \approx 0$$

$$\begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \approx \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \quad \Leftrightarrow \quad Ax \approx b$$

We arrive at an over-determined system with a Vandermonde matrix.

The Normal Equations for $Ax = b$

$b - A\hat{x} \perp Ax \Rightarrow (Ax)^T(b - A\hat{x}) = 0$ for all x

$x^T A^T (b - A\hat{x}) = 0$ for all $x \Rightarrow A^T(b - A\hat{x}) = 0$

$$\boxed{A^T A\hat{x} = A^T b}$$

$\hat{x}$ is the least squares solution of the over-determined system $Ax = b$.

$r = b - A\hat{x}$ is the residual vector of the least squares solution.

The least squares solution minimizes the 2-norm of the residual $\|b - Ax\|_2$.

For m data points, $\boxed{\|r\|_2/\sqrt{m}}$ is the root-mean-square error (RMSE).

Example

Fitting a straight line to a set of data points is called linear regression.

Fit a straight line through $(0,1),(1,2),(3,3)$.

$f(t,x) = x_1 + x_2 t$

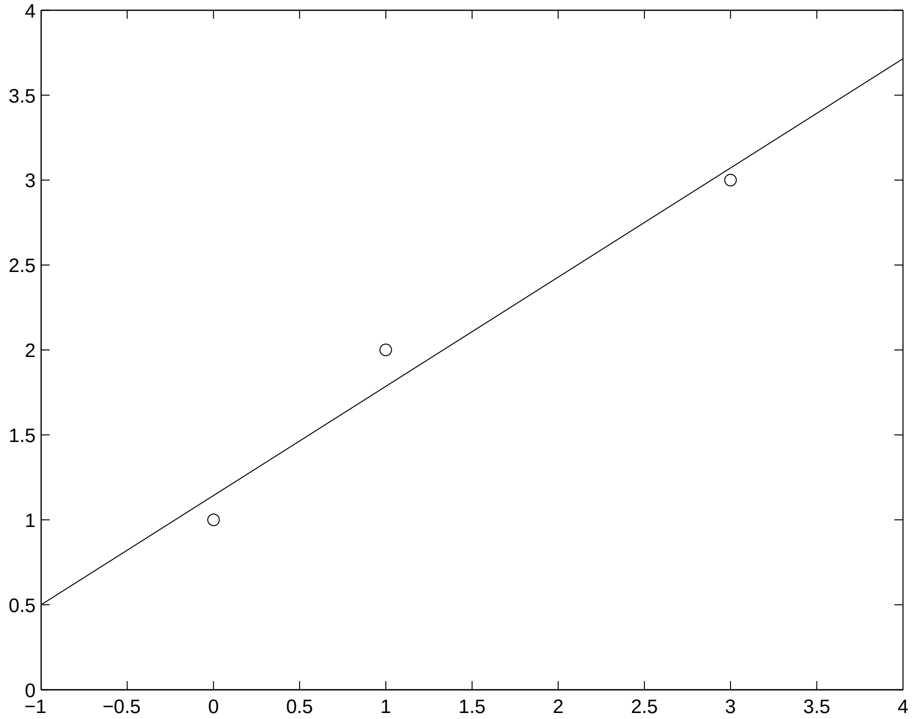
$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 3 & 4 \\ 4 & 10 \end{bmatrix}; \quad A^T b = \begin{bmatrix} 6 \\ 11 \end{bmatrix}$$

$$x_1 = \frac{8}{7}, \quad x_2 = \frac{9}{14} \Rightarrow \quad y = \frac{8}{7} + \frac{9}{14} t$$

Plot of data points and regression line

$$y = \frac{8}{7} + \frac{9}{14}t$$



Least Squares Fitting

Given m data points

1. Choose model (with unknown parameters x)

2. Substitute data into model (construct system $Ax = b$)

3. Solve normal equations ($A^T Ax = A^T b$). If A has full rank (all columns are linearly independent), then $A^T A$ is non-singular (there is a unique solution).

Example

The world oil production is shown below.

year	bbl/day $\times 10^6$	year	bbl/day $\times 10^6$
1994	67.052	1999	72.063
1995	68.008	2000	74.669
1996	69.803	2001	74.487
1997	72.024	2002	74.065
1998	73.400	2003	76.777

Find the best least squares
(a) line,
(b) parabola, and
(c) cubic curve,
and the RMSE of the fit. Estimate the 2010 production level.

Solution to part (a)

1. Model: a straight line

$$y = x_0 + x_1(2003 - t)$$

y is the oil production,

t is the year,

x_0 and x_1 are the parameters of the model.

2. Substitute data into model

$$\begin{pmatrix} 9 & 1 \\ 8 & 1 \\ 7 & 1 \\ 6 & 1 \\ 5 & 1 \\ 4 & 1 \\ 3 & 1 \\ 2 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} 67.052 \\ 68.008 \\ 69.803 \\ 72.024 \\ 73.400 \\ 72.063 \\ 74.669 \\ 74.487 \\ 74.065 \\ 76.777 \end{pmatrix}$$

3. Solve normal equations

$$A^T A x = A^T b$$

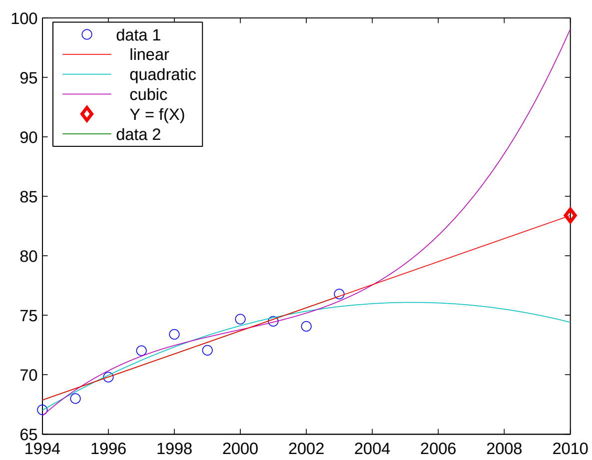
$$\begin{pmatrix} 285 & 45 \\ 45 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} 3170.6 \\ 722.3 \end{pmatrix}$$

$$x_1 = -0.9693, \quad x_0 = 76.5969$$

$$y = 76.5969 - 0.9693(2003 - t)$$

Estimation at $t = 2010$

$$y = 76.5969 - 0.9693(2003 - 2010) = 83.382$$



Root mean squared error

$$\text{RMSE} = \frac{\|r\|_2}{\sqrt{m}}$$

$$m = 10, \quad r = b - A\hat{x}$$

$\text{RMSE} = 3.0874/\sqrt{10} = 0.9763$

Accuracy of solving the normal equations

Note that the matrix in the undetermined system $Ax = b$ is a (rectangular) Vandermonde matrix. We know that this matrix is ill-conditioned.

What about $A^T A$? This matrix inherits the ill-conditioning of A, and in fact, it is even more ill-conditioned!

$$A = \begin{pmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{pmatrix} \qquad A^T A = \begin{pmatrix} 1+\varepsilon^2 & 1 & 1 \\ 1 & 1+\varepsilon^2 & 1 \\ 1 & 1 & 1+\varepsilon^2 \end{pmatrix}$$

For $\varepsilon = 0.1$, $\kappa(A) = 17.35$, $\kappa(A^T A) = 301 \approx \kappa(A)^2$.

Periodic data

A function g has period P if

$$g(x+P) = g(x)$$

Model: Trigonometric polynomial of order M

$$\boxed{T_M(x) = a_0 + \sum_{j=1}^{M} \left(a_j \cos(\frac{2\pi}{P} jx) + b_j \sin(\frac{2\pi}{P} jx) \right)}$$

Example

The Sturup weather station showed the following maximum temperature on the 6th of each month in 2007-2013:

month	1	2	3	4	5	6	7	8	9	10	11	12
T(C) 2007	8	1	7	8	19	23	16	24	16	15	8	9
T(C) 2008	2	8	8	8	20	23	24	20	21	13	10	5
T(C) 2009	1	4	3	12	11	13	24	24	18	14	8	6
T(C) 2010	-2	-1	1	9	8	21	19	24	17	13	8	1
T(C) 2011	6	8	13	21	24	26	26	24	23	22	13	9
T(C) 2012	9	12	16	18	25	23	27	29	24	17	10	7
T(C) 2013	9	5	12	17	24	27	29	28	22	17	11	9

Fit the data to an appropriate model.

Solution: $P = 12$,

$$T = a_0 + a_1 \cos m\pi/6 + b_1 \sin m\pi/6, \quad m = 1, 2, \ldots, 84$$

$$\begin{pmatrix} 1 & \cos\frac{\pi}{6} & \sin\frac{\pi}{6} \\ 1 & \cos\frac{2\pi}{6} & \sin\frac{2\pi}{6} \\ 1 & \cos\frac{3\pi}{6} & \sin\frac{3\pi}{6} \\ \vdots & \vdots & \vdots \\ 1 & \cos\frac{84\pi}{6} & \sin\frac{84\pi}{6} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} 8 \\ 1 \\ 7 \\ \vdots \\ 9 \end{pmatrix}$$

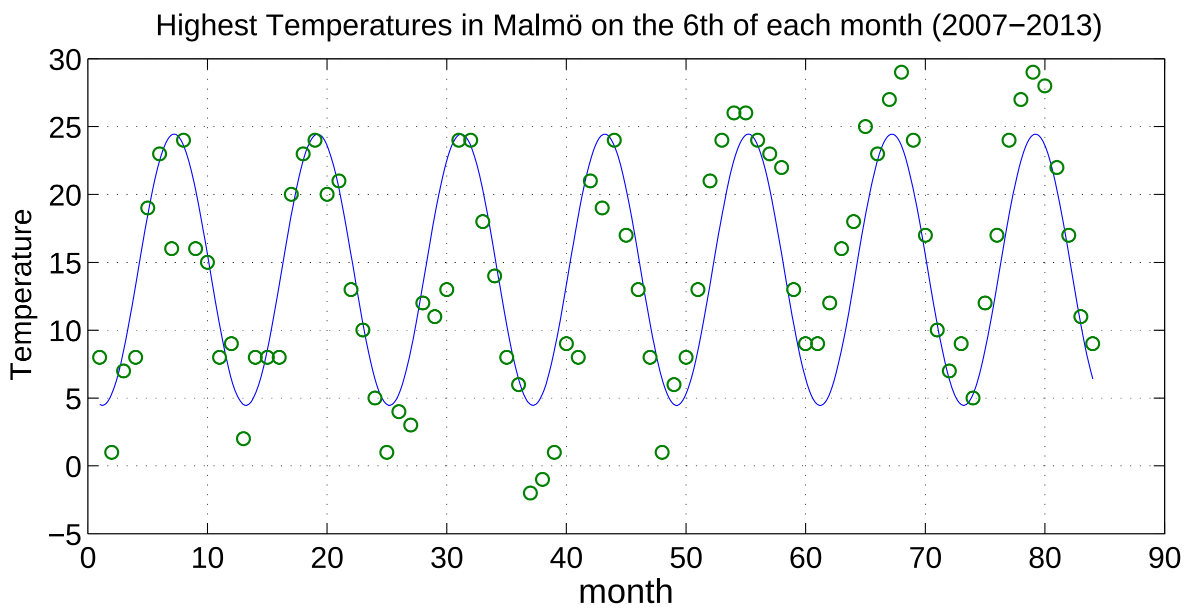
Solve $A^T A x = A^T b$ to get $a_0 = 14.4524, a_1 = -8.0446, b_1 = -5.9254$

$$T(m) = 14.4524 - 8.0446 \cos m\pi/6 - 5.9254 \sin m\pi/6$$

$$T(88) = 13.3431$$

Plot of order 2 trigonometric polynomial

$$
\begin{aligned}
T(m) &= 11.917 - 7.2863\cos m\pi/6 - 6.7863\sin m\pi/6 \\
&\quad +0.16667\cos m\pi/3 + 0.43301\sin m\pi/3
\end{aligned}
$$



Orthogonal matrices vs Vandermonde

We know that solving the normal equations is ill-conditioned, because $A^T A$ inherits and even worsens the large condition number of the Vandermonde matrix A.

An orthogonal matrix is a matrix with real entries such that $Q^{-1} = Q^T$.

$$\kappa_2(Q) = \|Q\|_2 \, \|Q^{-1}\|_2 = \sqrt{\rho(Q^T Q)} \, \sqrt{\rho(Q Q^T)} = \sqrt{\rho(I)}\sqrt{\rho(I)} = 1$$

As orthogonal matrices have 2-norm condition number equal to 1, we will develop a method to solve the least squares problem using orthogonal matrices.

Gram-Schmidt Orthogonalization

Compute an **orthogonal** basis for the space spanned by k given linearly independent vectors, $\{v_1, v_2, \ldots, v_k\}$.

Define the unit vector in the direction of v_1,

$$q_1 = \frac{v_1}{\|v_1\|_2}$$

The projection of v_2 onto q_1 is $(q_1^T v_2)q_1$, so vector $y_2 = v_2 - (q_1^T v_2)q_1$ is perpendicular to v_1 and

$$q_2 = \frac{y_2}{\|y_2\|_2}$$

is a unit vector that is perpendicular to v_1 and such that $\{q_1, q_2\}$ span the same subspace as $\{v_1, v_2\}$.

Gram-Schmidt algorithm

Define

1. $y_1 = v_1, \quad q_1 = \dfrac{v_1}{\|v_1\|_2}$

2. $y_2 = v_2 - q_1(q_1^T v_2), \quad q_2 = \dfrac{y_2}{\|y_2\|_2}$

$\vdots$

k. $y_k = v_k - q_1(q_1^T v_k) - \cdots - q_{k-1}(q_{k-1}^T v_k), \quad q_k = \dfrac{y_k}{\|y_k\|_2}$

Note that $q_j \perp q_i$

Example: Gram-Schmidt on the columns of a matrix

$$A = \begin{pmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{pmatrix} = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} \end{pmatrix}$$

$$\|y_1\|_2 = \sqrt{(-4)^2 + (-2)^2 + 4^2} = 6 \Rightarrow q_1 = \begin{pmatrix} -4/6 \\ -2/6 \\ 4/6 \end{pmatrix}$$

$$y_2 = \begin{pmatrix} -4 \\ 7 \\ -5 \end{pmatrix} + 3 \begin{pmatrix} -4/6 \\ -2/6 \\ 4/6 \end{pmatrix} = \begin{pmatrix} -6 \\ 6 \\ -3 \end{pmatrix} \Rightarrow q_2 = \begin{pmatrix} -6/9 \\ 6/9 \\ -3/9 \end{pmatrix}$$

$\begin{pmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{pmatrix}$ and $\begin{pmatrix} -4/6 & -6/9 \\ -2/6 & 6/9 \\ 4/6 & -3/9 \end{pmatrix}$ span the same plane.

Orthogonalization of a matrix

Let $r_{ii} = \|y_i\|_2$, $r_{ji} = q_j^T v_i$.

We can write $v_i = r_{ii} q_i + r_{1i} q_1 + \cdots + r_{i-1,i} q_{i-1}$, so

$$A = \begin{pmatrix} \mathbf{q_1} & \cdots & \mathbf{q_k} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ & & & r_{kk} \end{pmatrix}$$

If $\mathbf{v_1}, \ldots, \mathbf{v_k}$ are linearly independent, all r_{ii} are nonzero.

If $\mathbf{v_i}$ is spanned by $\mathbf{v_1}, \ldots, \mathbf{v_{i-1}}$, then $r_{ii} = 0$ and the Gram-Schmidt method terminates.

The QR-factorization

Once we have the Gram-Schmidt orthogonalization of an $n \times k$ matrix A, we can complete the orthonormal basis by adding vectors $\mathbf{q_{k+1}}, \ldots, \mathbf{q_n}$,

$$A = \begin{pmatrix} \mathbf{q_1} & \cdots & \mathbf{q_k} & \mathbf{q_{k+1}} & \cdots & \mathbf{q_n} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ & & & r_{kk} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

$Q = (\mathbf{q_1} \cdots \mathbf{q_n})$ is an orthogonal $n \times n$ matrix and R is an $n \times k$ upper triangular matrix.

Orthogonal matrices and the QR-factorization

A square matrix Q is **orthogonal** if $Q^T = Q^{-1}$.

A square matrix Q is orthogonal if its columns are pairwise orthogonal unit vectors ($\mathbf{q_i}^{\mathbf{T}}\mathbf{q_j} = 0$ and $\|\mathbf{q_i}\|_2 = 1$).

If Q is orthogonal, then $\|Qx\|_2 = \|x\|_2$.

In $A = QR$, Q is an orthogonal matrix and R is upper triangular.

Example: QR-factorization by Gram-Schmidt

To $\begin{pmatrix} -4 \\ -2 \\ 4 \end{pmatrix}$, $\begin{pmatrix} -4 \\ 7 \\ -5 \end{pmatrix}$ we add vector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ to complete the space $\mathbb{R}^3$.

$$y_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{4}{6}\begin{pmatrix} -4/6 \\ -2/6 \\ 4/6 \end{pmatrix} + \frac{6}{9}\begin{pmatrix} -6/9 \\ 6/9 \\ -3/9 \end{pmatrix} = \begin{pmatrix} 1/9 \\ 2/9 \\ 2/9 \end{pmatrix}, q_3 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}$$

$r_{11} = \|y_1\|_2 = 6,\ r_{12} = q_1^T v_2 = -3,\ r_{22} = \|y_2\|_2 = 9$

The QR-factorization of A is

$$\begin{pmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{pmatrix} = \begin{pmatrix} -2/3 & -2/3 & 1/3 \\ -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 6 & -3 \\ 0 & 9 \\ 0 & 0 \end{pmatrix}$$

Least Squares and QR-factorization

The least squares solution minimizes

$$\|b - Ax\|_2 = \|b - QRx\|_2 = \|Q^T(b - QRx)\|_2 = \|Q^Tb - Rx\|_2$$

Let $d = Q^Tb$. We can find x so that

$$\left(\begin{array}{c} d_1 \\ \vdots \\ d_k \\ \hline d_{k+1} \\ \vdots \\ d_n \end{array}\right) - \left(\begin{array}{cccc} r_{11} & r_{12} & \cdots & r_{1k} \\ & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ & & & r_{kk} \\ \hline 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{array}\right) \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline d_{k+1} \\ \vdots \\ d_n \end{array}\right)$$

Steps for least squares solving by QR-factorization

$$A = QR \Rightarrow Ax = QRx = b \Rightarrow Rx = Q^T b$$

Given the $m \times n, m > n$ system $Ax = b$,

1. Find Q and R such that $A = QR$

2. Set $\hat{R} =$ upper $n \times n$ submatrix of R

3. Set $\hat{d} =$ upper n entries of $d = Q^T b$

4. Solve $\hat{R}x = \hat{d}$

Example

Solve the least squares problem $\begin{pmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 9 \\ 0 \end{pmatrix}$

$$\begin{pmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{pmatrix} = \begin{pmatrix} -2/3 & -2/3 & 1/9 \\ -1/3 & 2/3 & 2/9 \\ 2/3 & -1/3 & 2/9 \end{pmatrix} \begin{pmatrix} 6 & -3 \\ 0 & 9 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 6 & -3 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -5 \\ 4 \end{pmatrix}$$