Imperial College London

Maths for Signals and Systems

Linear Algebra in Engineering

Lectures 7 – 8, Tuesday 25th October 2016

DR TANIA STATHAKI

READER (ASSOCIATE PROFFESOR) IN SIGNAL PROCESSING

IMPERIAL COLLEGE LONDON

Mathematics for Signals and Systems

In this set of lectures we will talk about:

- An application of Least Squares method
- Semi-orthogonal matrices
- Rotation, Permutation and Householder Reflection matrices
- Gram-Schmidt Orthogonalisation
- QR Decomposition

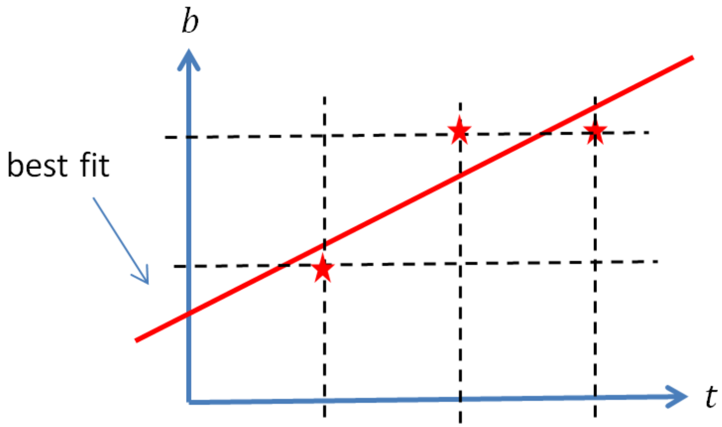
Application: Least squares method. Fitting by a line.

Problem:

I am given the three points shown with stars in the figure below. I want to fit them on the "best" possible straight line.

- The given points are (1,1), (2,2), (3,2).
- The required line is described by an equation of the form $b = C + Dt$, with C and D unknowns.
- The three given points must satisfy the line equation:

$$C + D = 1$$
$$C + 2D = 2$$
$$C + 3D = 2$$



Application: Least squares method. Fitting by a line cont.

Problem:

The previous problem is translated to solving the system

$$Ax = b = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

The system is not solvable because $b \notin C(A)$ (show that).

Solution:

Solve $A\hat{x} = p$ instead, where p is the projection of b onto $C(A)$.

For a random b we write $b - p = e \Rightarrow b = p + e$.

p is in the column space of A and e is perpendicular to the column space of A.

As already mentioned, projection eliminates e and keeps p.

Imperial College London

Solution of the specific example

- As proven, the proposed approach $A\hat{x} = p$ yields the same solution which can be obtained if we look for an $\hat{x}$ that minimizes the function:

$$\|A\hat{x} - b\|^2 = \|e\|^2$$

- The above function is the square of the magnitude of the error vector.
- Apart from **Least Squares Minimizations**, this method is also called **Linear Regression**.
- For the particular problem we have:

$$A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \quad A^T b = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

- We can use the inverse $(A^T A)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix}$.
- Or we can solve directly the equations

$$3C + 6D = 5$$

$$6C + 14D = 11.$$

- Final solution is $D = \frac{1}{2}, C = \frac{2}{3}$. The "best" line is $b = \frac{2}{3} + \frac{1}{2}t$ shown in the previous figure in **red**.

Imperial College London

Solution of the specific example

- As mentioned, an alternative approach is to find the unknowns that minimize the error function:

$$(C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2$$

- We must take the partial derivatives with respect to the two unknowns and set them to zero.
- By implementing the above we get the same solution as previously.
- The vector p is obtained by:

$$p_1 = C + D = \frac{1}{2} + \frac{2}{3} = \frac{7}{6}$$

$$p_2 = C + 2D = \frac{2}{3} + 1 = \frac{5}{3}$$

$$p_3 = C + 3D = \frac{2}{3} + \frac{3}{2} = \frac{13}{6}$$

Solution of the specific example

$$e_1 = b_1 - p_1 = 1 - \frac{7}{6} = -\frac{1}{6}$$

$$e_2 = b_2 - p_2 = 2 - \frac{5}{3} = \frac{1}{3}$$

$$5$$

$$e_3 = b_3 - p_3 = 2 - \frac{13}{6} = -\frac{1}{6}$$

$$p^T \cdot e = \begin{bmatrix} \frac{7}{6} & \frac{5}{3} & \frac{13}{6} \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix} = -\frac{7}{36} + \frac{20}{36} - \frac{13}{36} = 0$$

p and e are perpendicular to each other as expected.

The matrix $A^T A$

Problem:

We mentioned previously that if a matrix A of dimension $m \times n$ has independent columns $a_1, a_2, \dots, a_n$ then $A^T A$ is invertible. The proof of this statement is given below.

Solution:

- I must prove that $A^T A x = \mathbf{0}$ implies $x = \mathbf{0}$ where x is a column vector $[x_1 \quad x_2 \quad \cdots \quad x_n]^T$. I assume that $A^T A x = \mathbf{0}$.
- The above implies $x^T A^T A x = 0 \Rightarrow (Ax)^T A x = 0$.
- I define $Ax = y$ and therefore $y^T y = 0 \Rightarrow \|y\|^2 = 0$.
- We know that $y^T y = \|y\|^2$ is the sum of the squares of the elements of a vector.
- Note that for complex vectors $\|y\|^2$ is defined as $y^{*T} y$.
- Therefore, $\|y\|^2 = 0$ implies that $y = \mathbf{0}$.
- In the above case $y = Ax = \mathbf{0}$ implies $\sum_{i=1}^{n} x_i a_i = \mathbf{0}$. This condition holds only if $x = \mathbf{0}$ if A has independent columns.

The matrix AA^T

Problem:

If A has independent rows then AA^T is invertible. The proof of this follows in a similar fashion as in the previous slide.

Solution:

- I must prove that $AA^T x = \mathbf{0}$ implies $x = \mathbf{0}$ where x is a column vector. I assume that $AA^T x = \mathbf{0}$.
- The above implies $x^T AA^T x = 0 \Rightarrow (A^T x)^T A^T x = 0$.
- I define $A^T x = y$ and therefore $y^T y = 0 \Rightarrow ||y||^2 = 0$.
- $||y||^2 = 0$ implies that $y = \mathbf{0}$.
- In the above case $y = A^T x = \mathbf{0}$ implies that $x = \mathbf{0}$ if A^T has independent columns (see previous slide) or, equivalently, if A has independent rows.

Orthogonal and Orthonormal vectors revision

- Lets recall that:

  - The column vectors $q_1, \ldots, q_n$ are **orthogonal** if ${q_i}^T \cdot q_j = 0$ for $i \neq j$.

  - If their lengths are all 1, then the vectors are called **orthonormal**.

$$ {q_i}^T \cdot q_j = \begin{cases} 0 & \text{when} \quad i \neq j \quad (\textbf{orthogonal} \text{ vectors}) \\ 1 & \text{when} \quad i = j \quad (\textbf{unit} \text{ vectors: } \|q_i\| = 1) \end{cases} $$

- In order for a set of n vectors to satisfy the above, their dimension m must be at least n, i.e., $m \geq n$. This is because the maximum number of $m-$ dimensional vectors that can be orthogonal is m.

Imperial College London

Semi-orthogonal matrices with more rows than columns

- I assign to a matrix with n orthonormal m −dimensional columns q_i the special letter Q.
- I wish to deal first with the case where Q is strictly non-square (it is rectangular), and therefore, $m > n$.
- The matrix Q is called **semi-orthogonal**.

Problem:

Consider a semi-orthogonal matrix Q with real entries, of dimension $m \times n$, $m > n$. The columns are orthonormal vectors. Prove that $Q^T Q = I_{n\times n}$.

Solution:

$$Q^T Q = \begin{bmatrix} {q_1}^T \\ {q_2}^T \\ \vdots \\ {q_n}^T \end{bmatrix} [q_1 \quad q_2 \quad \cdots \quad q_n] = I_{n\times n}.$$

- We see that Q^T is an **inverse from the left.**
- This is because there isn't a matrix Q' for which $QQ' = I_{m\times m}$. This would imply that we could find m independent vectors of dimension n, with $m > n$. **This is not possible.**

Semi-orthogonal matrices with more columns than rows

- I again assign to a matrix with m orthonormal n −dimensional rows r_i the special letter Q.
- I wish to deal with the case where Q is strictly non-square (it is rectangular), and therefore, $m < n$.
- Obviously the matrix Q is defined now is also **semi-orthogonal**.

Problem:

Consider a semi-orthogonal matrix Q with real entries, of dimension $m \times n$, $m < n$. The columns are orthonormal vectors. Prove that $QQ^T = I_{m\times m}$.

Solution:

$$QQ^T = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \begin{bmatrix} r_1{}^T & r_2{}^T & \dots & r_n{}^T \end{bmatrix} = I_{m\times m}.$$

- We see now that Q^T is an **inverse from the right.**

Semi-orthogonal matrices: Generalization

- In linear algebra, a **semi-orthogonal matrix** is a non-square matrix with real entries where: if the number of rows exceeds the number of columns, then the columns are orthonormal vectors; but if the number of columns exceeds the number of rows, then the rows are orthonormal vectors.

- Equivalently, a rectangular matrix of dimension $m \times n$ is semi-orthogonal if

$$Q^T Q = I_{n \times n}, m > n \text{ or } QQ^T = I_{m \times m}, n > m$$

- The above formula yields the terms **left-invertible** or **right-invertible** matrix.

- In the above cases, the left or right inverse is the transpose of the matrix. For that reason, a rectangular orthogonal matrix is called **semi-unitary**. (To remind you: a **unitary** matrix is the one with an inverse being its transpose.)

Semi-orthogonal matrices: Generalization

Problem 1:

Show that for left-invertible, semi-orthogonal matrices of dimension $m \times n$, $m > n$ $\|Qx\| = \|x\|$ for every $n-$ dimensional vector x.

Solution:

$\|Qx\|^2 = (Qx)^T(Qx) = x^TQ^TQx = x^TIx = x^Tx \Rightarrow \|Qx\|^2 = \|x\|^2 \Rightarrow \|Qx\| = \|x\|$.

Problem 2:

Show that for right-invertible, semi-orthogonal matrices of dimension $m \times n$, $m < n$, $\|Q^Tx\| = \|x\|$ for every $m-$ dimensional vector x.

Solution:

$\|Q^Tx\|^2 = (Q^Tx)^T(Q^Tx) = x^TQQ^Tx = x^TIx = x^Tx \Rightarrow \|Q^Tx\|^2 = \|x\|^2 \Rightarrow \|Q^Tx\| = \|x\|$.

Orthogonal matrices

Problem 1:

Extend the relationship $Q^T Q = I_{n \times n}$ for the case when Q is a square matrix of dimension $n \times n$ and has orthogonal columns.

Solution:

$Q^T Q = I_{n \times n} \Rightarrow Q^{-1} = Q^T$. **The inverse is the transpose.**

Problem 2:

Prove that $QQ^T = I_{n \times n}$.

Solution:

Since Q is a full rank matrix we can find Q' such that $QQ' = I_{n \times n}$. This gives:

$$Q^T QQ' = Q^T I_{n \times n} \Rightarrow I_{n \times n} Q' = Q^T \Rightarrow Q' = Q^T$$

Therefore, we see that Q^T is the **two-sided inverse** of Q.

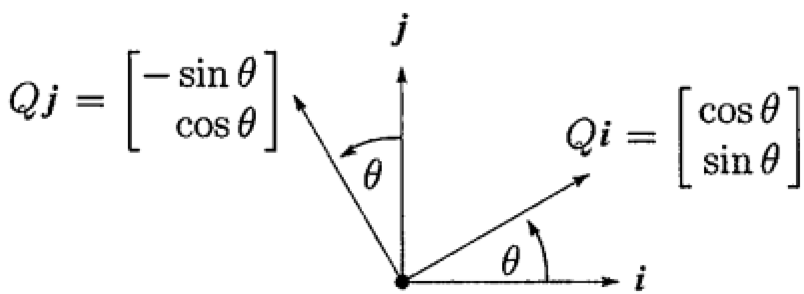
Examples of elementary orthogonal matrices. Rotation matrices.

- The **rotation matrix** of size 2×2 is defined as:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ and } Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Problems:

- The columns (and rows) of Q are orthogonal (straightforward to prove).
- The columns (and rows) of Q are vectors of magnitude 1 (also straightforward).
- Explain the effect that the rotation matrix has on vectors $j = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, when it multiplies them from the left.
- The matrix causes rotation of the vectors.



Examples of elementary orthogonal matrices
Permutation Matrices

- **Permutation** matrices reorder the rows of identity matrices. Examples are:

$Q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ and $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. $Q^T = Q^{-1}$ in both cases.

Problems:

- The columns of Q are orthogonal (straightforward).
- The columns of Q are unit vectors (straightforward).
- Explain the effect that the permutation matrices have on a random vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ or $\begin{bmatrix} x \\ y \end{bmatrix}$ when they multiply the vector from the left.
- Obviously the matrices cause re-ordering of the elements of these vectors.

Examples of elementary orthogonal matrices Householder Reflection Matrices

- **Householder reflection** matrices are defined as:

$Q = I - 2uu^T$ with u any vector that satisfies the condition $u^T u = 1$ (unit vector).
$Q^T = I^T - (2uu^T)^T = I - 2uu^T = Q$
$Q^T Q = Q^2 = (I - 2uu^T)(I - 2uu^T) = I - 4uu^T + 4uu^T\ uu^T$ and $u^T\ u = 1$ and therefore, $Q^T Q = Q^2 = I - 4uu^T + 4uu^T = I$

Problems:

- For $u_1 = [1 \quad 0]^T$ and $u_2 = \left[1/\sqrt{2} \quad -1/\sqrt{2}\right]^T$ find $Q_i = I - 2u_i {u_i}^T$, $i = 1,2$.
- Explain the effect that matrix Q_1 has on the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ when it multiplies the vector from the left.
- Explain the effect that matrix Q_2 has on the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ when it multiplies the vector from the left.

- A generalized definition is $Q = I - 2\frac{vv^T}{\|v\|^2}$ with v any column vector.

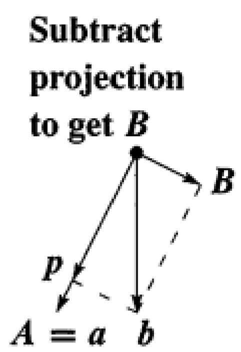
The Gram-Schmidt process

- The goal here is to start with three independent vectors a, b, c and construct three orthogonal vectors A, B, C and finally three orthonormal vectors.

$$q_1 = A/\|A\|,\ q_2 = B/\|B\|,\ q_3 = C/\|C\|$$

- We begin by choosing $A = a$. This first direction is accepted.
- The next direction B must be perpendicular to A. We start with b and subtract its projection along A. This leaves the part of b which we call vector B (what we knew before as the error of projection), defined as:

$$B = b - \frac{AA^T}{A^T A} b$$



The Gram-Schmidt process

Problem:

Show that A and B are orthogonal. Note that $A = a$.

Solution:

$$A^T B = A^T b - A^T \frac{AA^T}{A^T A} b = = A^T b - \frac{A^T AA^T}{A^T A} b = A^T b - A^T b = 0$$

The inner product between A and B is 0 and therefore, A and B are orthogonal.

Problem:

Show that if a and b are independent then B is not zero.

$$B = b - \frac{AA^T}{A^T A} b$$

Solution:

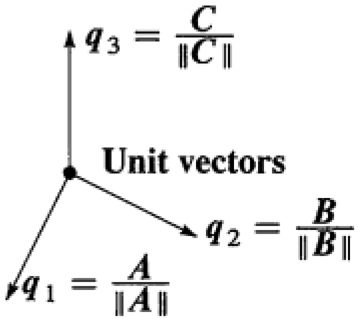
The vector $\frac{AA^T}{A^T A} b$ is the projection of vector b onto vector a. In order for B to be zero the projection of vector b onto vector a must be equal to b itself. This happens only when a and b are dependent.

The Gram-Schmidt process

- The third direction starts with c. This is not a combination of A and B.
- Most likely c is not already perpendicular to A and B .
- Therefore, we subtract the projections of c along A and B to get C:

$$C = c - \frac{AA^T}{A^TA}c - \frac{BB^T}{B^TB}c$$

- **In general we subtract from every new vector its projections in the directions already set.**
- If we had a fourth vector d, we would subtract three projections onto A, B, C to get D.
- We make the resulting vectors orthonormal.
- This is done by dividing the vectors with their magnitudes.



The factorization A=QR known as QR decomposition

- Assume matrix A whose columns are a, b, c.
- Assume matrix Q whose columns are q_1, q_2, q_3 defined previously.
- We are looking for a matrix R such that $A = QR$. Since Q is an orthogonal matrix we have that $R = Q^T A$.

$$R = Q^T A = \begin{bmatrix} q_1{}^T \\ q_2{}^T \\ q_3{}^T \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1{}^T a & q_1{}^T b & q_1{}^T c \\ q_2{}^T a & q_2{}^T b & q_2{}^T c \\ q_3{}^T a & q_3{}^T b & q_3{}^T c \end{bmatrix}$$

- We know that from the method that was used to construct q_i we have

$$q_2{}^T a = 0,\ \ q_3{}^T a = 0,\ \ q_3{}^T b = 0$$ **(see Appendix)**

and therefore,

$$R = \begin{bmatrix} q_1{}^T a & q_1{}^T b & q_1{}^T c \\ 0 & q_2{}^T b & q_2{}^T c \\ 0 & 0 & q_3{}^T c \end{bmatrix}$$

- QR decomposition can facilitate the solution of the system $Ax = b$, since $Ax = b \Rightarrow QRx = b \Rightarrow Rx = Q^T b$. The later system is easy to solve due to the upper triangular form of R.
- **So far you have learnt two types of decompositions: the LU and the QR.**

Imperial College London

Generalization of QR decomposition

- **Any matrix A of dimension $m \times n$ with independent columns can be factored into QR.**

- The $m \times n$ matrix Q has orthonormal columns.

- The square matrix R is upper triangular with positive diagonal.

- $A^T A = R^T Q^T Q R = R^T R$

- With the use of QR decomposition the least squares solution of the system of equations $Ax = b$ becomes:
$$A^T A\hat{x} = A^T b \Rightarrow R^T R\hat{x} = R^T Q^T b \Rightarrow R\hat{x} = Q^T b$$

- The system of equations $R\hat{x} = Q^T b$ can be easily solved with back-substitution.

Imperial College London

Appendix QR

- ${q_2}^T a = 0$. The proof of this is straightforward.

- ${q_3}^T a = 0$. In order to prove this we can prove that $C^T a = 0$.

$C^T = c^T - c^T \frac{AA^T}{A^T A} - c^T \frac{BB^T}{B^T B} \Rightarrow C^T a = c^T a - c^T \frac{AA^T}{A^T A} a - c^T \frac{BB^T}{B^T B} a \Rightarrow$

$C^T a = c^T a - c^T \frac{AA^T}{A^T A} a = 0$ since $A = a$.

- ${q_3}^T b = 0$. In order to prove this we can prove that $C^T b = 0$.

$C^T = c^T - c^T \frac{AA^T}{A^T A} - c^T \frac{BB^T}{B^T B} \Rightarrow C^T b = c^T b - c^T \frac{AA^T}{A^T A} b - c^T \frac{BB^T}{B^T B} b \Rightarrow$

$C^T b = c^T b + c^T (B - b) - c^T \frac{BB^T}{B^T B} b \Rightarrow C^T b = c^T B - c^T \frac{BB^T}{B^T B} b.$

$\frac{BB^T}{B^T B} b = \frac{BB^T}{B^T B} (B + \frac{AA^T}{A^T A} b) = \frac{BB^T}{B^T B} B + \frac{BB^T AA^T}{B^T B A^T A} b$ but $B^T A = 0$. Therefore $\frac{BB^T}{B^T B} b = B.$

Thus, $C^T b = c^T B - c^T \frac{BB^T}{B^T B} b = c^T B - c^T B = 0$.