Imperial College London

Maths for Signals and Systems

Linear Algebra in Engineering

Lecture 6, Friday 21st October 2016

DR TANIA STATHAKI

READER (ASSOCIATE PROFFESOR) IN SIGNAL PROCESSING
IMPERIAL COLLEGE LONDON

In this set of lectures we will talk about:

- Spaces other than vector spaces.
- Orthogonal subspaces.
- Projections on to spaces.
- How to solve the problem $Ax = b$ when b does not belong in the column space of A.
- Lease Squares Minimization or Linear Regression.

Generalization of the concept of space

- Consider a set of entities (objects) that are not necessarily $1-D$ vectors.
- Assume that multiplication with a scalar and addition can be defined for these entities.
- A new space can be defined from all linear combinations of such entities.

Example: Space of matrices

- Consider a space M which contains all 3×3 matrices.
- Addition and multiplication with a scalar can be applied to matrices, e.g., if A and B are matrices, then $A+B$ and cA are matrices too.
- Examples of subspaces of M:
  - All upper triangular matrices.
  - All lower triangular matrices.
  - All symmetric matrices.
  - All diagonal matrices.

Imperial College London

Mathematics for Signals and Systems-Space of Matrices

- Consider the space M of all 3×3 matrices.
  - ➢ The dimension of this space is 9, i.e., $\dim(M) = 9$.
  - ➢ The simplest basis of this space is:

$$\begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 1\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\dots\begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}$$

  - ➢ Any 3×3 matrix can be written as a linear combination of the above matrices.

- Consider the subspace S of all 3×3 **symmetric** matrices. **(Keep in mind that symmetry implies square matrices only)**.
  - ➢ The dimension of this subspace is 6, i.e., $\dim(S) = 6$.
  - ➢ The simplest basis of this subspace is:

$$\begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\1 & 0 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 1\\0 & 0 & 0\\1 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\0 & 0 & 1\\0 & 1 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}$$

  - ➢ Any 3×3 symmetric matrix can be written as a linear combination of the above matrices.

Space of Matrices cont.

- Consider the subspace U of all 3×3 upper triangular matrices.
  - The dimension of this subspace is 6, i.e., $\dim(U) = 6$.
  - The simplest basis of this subspace is:

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&1\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$$

- Consider the subspace of the intersection of symmetric and upper triangular matrices $S \cap U$. **This subspace consists of diagonal matrices.**
  - The dimension of this subspace is 3, i.e., $\dim(S \cap U) = 3$.
  - A basis is:

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$$

- **Note that the union $S \cup U$ is not a subspace.**

Space of Matrices cont.

- As mentioned the union $S \cup U$ is not a subspace.

- Lets consider $S + U$.

  Question:

  What matrices can I get from $S + U$ where both S and U are of dimension $n \times n$?

  Answer:

  I can basically get any $n \times n$ matrix. **Can you possibly prove why?**

- We notice that for $n = 3$ we have:

$$\dim(S + U) = 9$$
$$\dim(S) = 6$$
$$\dim(U) = 6$$

- In general we can prove that:

$$\dim(S) + \dim(U) = \dim(S + U) + \dim(S \cap U)$$

Rank One Matrices

- An example of a 2×3 matrix of rank equal to 1 is:

$$A=\begin{bmatrix}1 & 4 & 5\\ 2 & 8 & 10\end{bmatrix}$$

- A basis of the row space of the above matrix is the vector $\begin{bmatrix}1 & 4 & 5\end{bmatrix}$.

- A basis of the column space of the above matrix is the vector $\begin{bmatrix}1\\ 2\end{bmatrix}$.

- Note that the well known property holds $\dim C(A)=r=\dim C(A^T)$.

- In this example $r=1$.

Rank One Matrices cont.

- We can write the previous matrix A as:

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}.$$

- Every rank one matrix can be decomposed into the product of a column vector and a row vector $A = v\,w^T$, with v and w column vectors.

- Furthermore, if v and w are column vectors, the matrix $v\,w^T$ is always of rank 1.

- All matrices can be written as a combination of rank one matrices.

- Rank one matrices are the building blocks for all matrices!

Imperial College London

Orthogonal subspaces: Definition and a couple of questions

- Suppose that a subspace S is orthogonal to a subspace T. **What does this mean?** It actually means that every vector in S is orthogonal to every vector in T.

- Assume we are in R^3. Consider two 2-D planes which are perpendicular to each other (for example a wall and the floor). Are these planes orthogonal?
  (The answer is NO: the line which is their intersection belongs to both)

- Assume we are in R^2.
  1. When is a line through the origin orthogonal to the entire plane?
  2. When is a line through the origin orthogonal to the 0 subspace?
  3. When is a line through the origin orthogonal to another line through the origin?

  (answers: 1. never, 2. always, 3. when they form a 90^o angle)

Orthogonal subspaces: more questions

- The row and null space of a matrix A of size $m \times n$ are orthogonal. **Why?**
  Vectors which satisfy $Ax = \mathbf{0}$ are orthogonal to all rows of A.

- Assume we are in R^3: consider two orthogonal lines. Could their subspaces be the row space and the null space of a matrix?
  (The answer is NO: dim$C(A^T)$+dim$N(A) = r + (3 - r) = 3$)

- Row space and null space are orthogonal and furthermore:
  - Their dimensions add up to the size of rows (or number of columns). This is basically the size of the maximum space that the rows can form.
  - Row space and null space are orthogonal complements in R^n.

Solve a system when there is no exact solution

- "Solve" $Ax = b$ when there is no exact solution, i.e., $b \notin C(A)$.

- We will realize, in subsequent sections, that there is a matrix which will play an important role for the solution to this problem: This is $A^T A$. It has the following properties:
  - It is square of size $n \times n$ (A is of size $m \times n$).
  - It is symmetric because $(A^T A)^T = A^T (A^T)^T = A^T A$.
  - It is invertible if A has n independent columns, i.e., the null space of A is zero (proof follows later).

Example 1: $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}$. Show that $A^T A$ is invertible.

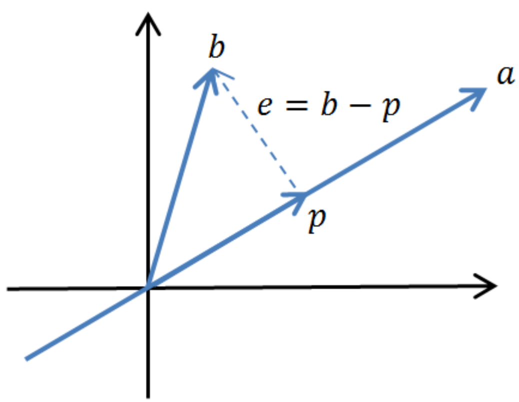
Example 2: $A = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}$. Show that $A^T A$ is NOT invertible.

Imperial College London

Projection matrix

PROBLEM: Find the projection of a vector b onto a line a. We need this to solve the problem $Ax = b$, $b \notin C(A)$. **(All vectors are assumed column vectors.)**

- This is a point p on the straight line formed by vector a, such that $p = xa$, where x is a scalar.
- p is defined such as the vector $e = b - p$ is orthogonal to the line formed by vector a.
- Inner product $a^T e = a^T(b - p) = a^T(b - xa) = 0$ or $x = \frac{a^T b}{a^T a}$.
- Based on the above $p = \frac{a^T b}{a^T a} a = \frac{aa^T}{a^T a} b = Pb$ with $P = \frac{aa^T}{a^T a}$ **(observe the form of P)**.
- Note that $a^T a$ is a scalar and aa^T is a square matrix.
- P is called the **projection matrix**.



Projection matrix properties

- **What is the column space of P?**

  It is the subspace formed by vector a. The reason is that if $a = [a_1 \quad a_2]^T$ then $P = \frac{1}{\|a\|^2}\begin{bmatrix} a_1{}^2 & a_1a_2 \\ a_1a_2 & a_2{}^2 \end{bmatrix}$ and therefore, $C(P)$: $c_1\begin{bmatrix} a_1{}^2 \\ a_1a_2 \end{bmatrix} + c_2\begin{bmatrix} a_1a_2 \\ a_2{}^2 \end{bmatrix} = (c_1a_1 + c_2a_2)\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = c\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$.

- **What is the rank of P?**

  Obviously 1.

- **What happens if I do the projection twice?**

  Nothing should happen. Therefore, the condition $P^2 = P$ must hold. Verify this using the above 2×2 matrix.

- **If we double b what happens to the projection?**

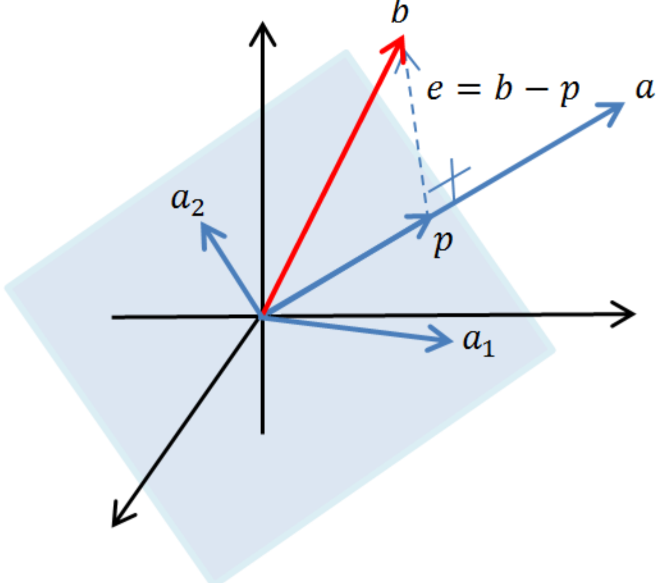
  The projection is doubled too.

- **If we double a what happens to the projection?**

  Nothing should happen to the projection in that case.

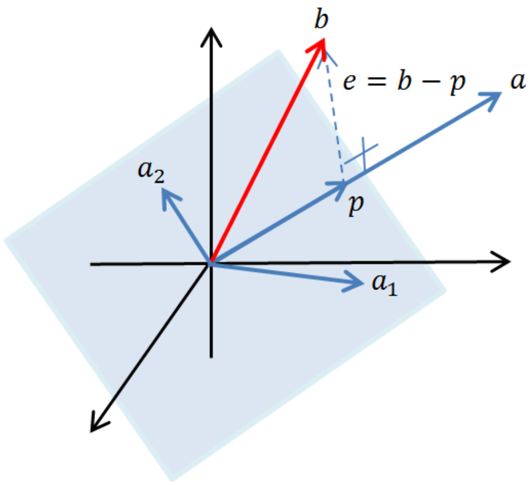
Solving Ax=b using projections

- Consider again the scenario where the system $Ax = b$ doesn't have solution.
- Goal: Solve $A\hat{x} = p$ instead, where p is the projection of b onto the column space of A.
- As already mentioned, the projection p of b onto the column space of A is found by forcing the error $e = b - p$ to be perpendicular to the column space of A.
- This scenario is depicted in the figure below for R^3.
- The quantities shown are column vectors and $A = [a_1 \quad a_2]$.



Projection error, solution and projection matrix

- The error e is perpendicular to a_1 and a_2.



$a_1{}^T(b - A\hat{x}) = 0$ and $a_2{}^T(b - A\hat{x}) = 0$

$$\begin{bmatrix} a_1{}^T \\ a_2{}^T \end{bmatrix}(b - A\hat{x}) = \mathbf{0} \Rightarrow A^T(b - A\hat{x}) = \mathbf{0}$$

Question:

Consider $A^T(b - A\hat{x}) = \mathbf{0}$.

In which subspace does $e = (b - A\hat{x})$ belong?

Answer:

Since $A^T(b - A\hat{x}) = \mathbf{0}$, we observe that $e \in N(A^T)$ (left nullspace). Therefore e is perpendicular to the column space of A.

- Solution of the "new" system: $A^T A\hat{x} = A^T b$
- If $A^T A$ is invertible then: $\hat{x} = (A^T A)^{-1} A^T b$
- Projection: $p = A\hat{x} = A(A^T A)^{-1} A^T b = Pb$
- Projection matrix: $P = A(A^T A)^{-1} A^T$

Projection matrix

- Projection matrix: $P = A(A^T A)^{-1} A^T$

- In general, A is not square (it is rectangular) and therefore, we cannot use the property $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$.

Question:

If A **was** a square and invertible matrix of size $n \times n$ what would P be?

Answer:

In that case the column space of A would be the entire R^n and therefore the vector b would belong to $C(A)$. Therefore, the projection of any vector b on $C(A)$ would be the vector itself. This can be also verified by:

$$P = A(A^T A)^{-1} A^T = AA^{-1}(A^T)^{-1} A^T = I \cdot I = I$$

Imperial College London

Projection matrix: Properties

- Properties of P.

  - Symmetric
    $A^T A$ symmetric and therefore $(A^T A)^{-1}$ is symmetric
    [to prove this we use the property $(A^{-1})^T = (A^T)^{-1}$]
    $P^T = [A(A^T A)^{-1} A^T]^T = (A^T)^T [(A^T A)^{-1}]^T A^T = A\,(A^T A)^{-1} A^T = P$

  - $P^2 = P$
    $P^2 = A\,(A^T A)^{-1} A^T A\,(A^T A)^{-1} A^T = A\,(A^T A)^{-1} [A^T A\,(A^T A)^{-1}] A^T =$
    $= A\,(A^T A)^{-1} I A^T = A\,(A^T A)^{-1} A^T = P$

Imperial College London

Least Squares Minimization

Problem:

Show that the proposed approach $A\hat{x} = p$ yields the solution which can be also obtained if we look for an $\hat{x}$ that minimizes the function:

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

Solution:

We are looking to minimize the function $\|A\hat{x} - b\|^2 = \|e\|^2 = (A\hat{x} - b)^T(A\hat{x} - b)$. This is a function of $\hat{x}$ only and therefore, we can formulate the problem:

$$\min f(\hat{x}) = \min[(A\hat{x} - b)^T(A\hat{x} - b)]$$

A function is minimized at points for which its first derivative is zero. Therefore, we are looking for the $\hat{x}$s that satisfy the equation:

$$\frac{\partial f(\hat{x})}{\partial \hat{x}} = \mathbf{0} \Rightarrow \frac{\partial}{\partial \hat{x}}(A\hat{x} - b)^T(A\hat{x} - b) = \mathbf{0}$$

$$(A\hat{x} - b)^T(A\hat{x} - b) = [(A\hat{x})^T - b^T](A\hat{x} - b) = (\hat{x}^T A^T - b^T)(A\hat{x} - b)$$

$$= \hat{x}^T A^T A\hat{x} - \hat{x}^T A^T b - b^T\ A\hat{x} + b^T b$$

$$\frac{\partial}{\partial \hat{x}}(\hat{x}^T A^T A\hat{x} - \hat{x}^T A^T b - b^T\ A\hat{x} + b^T b) = 2A^T A\hat{x} - A^T b - A^T\ b + \mathbf{0} = 2A^T A\hat{x} - 2A^T b$$

Therefore, the derivative is zero if $A^T A\hat{x} = A^T b$. This is equal to the previous solution. **This approach is called Least Squares Minimization.**