

COMP 7180
***Quantitative Methods for Data
Analytics and Artificial
Intelligence***

Lecture 2 – Linear Independence, Rank, and
Orthogonality

Open Questions

- **$AB \neq BA$** (the commutative “law” is usually broken)
 1. Are there any relations between **AB** and **BA** ?
 2. If so, what are they?

Open Questions

- Solve $\mathbf{Ax} = \mathbf{b}$
- Suppose \mathbf{A} is a square matrix and \mathbf{A} is invertible.
 $\mathbf{Ax} = \mathbf{b}$ (multiply both sides by \mathbf{A}^{-1})
 $\mathbf{A}^{-1} \mathbf{Ax} = \mathbf{A}^{-1} \mathbf{b}$ ($\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$)
 $\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$

Questions (We will answer them in the following lectures):

1. Which square matrices are invertible?
2. What if \mathbf{A} is square matrix but not invertible?
3. What if \mathbf{A} is not a square matrix?

Transpose Matrix

- The transpose of A is denoted by A^T . Its columns are taken directly from the rows of A — the i -th row of A becomes the i -th column of A^T :

Transpose If $A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}$ then $A^T = \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 4 & 3 \end{bmatrix}$

- If A is an m by n matrix, then A^T is n by m . The final effect is to flip the matrix across its main diagonal, and the entry:

Entries of A^T $(A^T)_{ij} = A_{ji}$.

Properties of Transpose

- $(A+B)^T = A^T + B^T$
- The transpose of AB is $(AB)^T = B^T A^T$

Start from

$$AB = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 5 & 5 & 5 \end{bmatrix}$$

Transpose to

$$B^T A^T = \begin{bmatrix} 3 & 2 \\ 3 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 3 & 5 \\ 3 & 5 \end{bmatrix}.$$

Properties of Transpose

- The transpose of A^{-1} is $(A^{-1})^T = (A^T)^{-1}$
- Why?
- To establish the formula for $(A^{-1})^T$, start from $AA^{-1} = I$. Take transpose, we have $(AA^{-1})^T = I^T = I$.
- From the previous slides, we know $(AA^{-1})^T = (A^{-1})^T A^T$.
- So we have $(A^{-1})^T A^T = I$. This indicates that $(A^{-1})^T$ is the inverse of A^T , i.e., $(A^{-1})^T = (A^T)^{-1}$.

$$\text{Inverse of } A^T = \text{Transpose of } A^{-1}$$

Symmetric Matrix

- A symmetric matrix is a matrix that equals its own transpose:
 $A^T = A$.
 - Each entry on one side of the diagonal equals its “mirror image” on the other side: $a_{ij} = a_{ji}$.

Symmetric matrices $A = \begin{bmatrix} 1 & 2 \\ 2 & 8 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$

- Property: $R^T R$ and $R R^T$ are symmetric matrices.
- Why:

$$(R^T R)^T = R^T (R^T)^T = R^T R \text{ (since } (R^T)^T = R)$$

$$(R R^T)^T = (R^T)^T R^T = R R^T$$

Small Exercise

- If A is not a zero matrix, is $A^2 = 0$ possible? is $A^T A = 0$ possible?
- Solution: $A^2 = 0$ is possible.
- On the other hand, $A^T A = 0$ is not possible, which can be proved by exploring its diagonal.

More Advanced Content

- Linear Independence, Basis, Rank, and Dimension
- Orthogonality

More Advanced Content

- **Linear Independence, Basis, Rank, and Dimension**
- Orthogonality

Linear Independence and Dependence

- Rank r :
 - Physical meaning: the amount of useful and non-redundant information in the data (matrix)
 - Formal definition: The number of independent rows/columns in the matrix A .
 - $m = n = r \Rightarrow$ invertible
- Definition of linear independence and dependence:
 - Suppose $c_1 v_1 + \dots + c_k v_k = 0$ only happens when $c_1 = \dots = c_k = 0$. Then the vectors v_1, \dots, v_k are **linearly independent**.
 - If any c 's are(is) nonzero, the v 's are **linearly dependent**. One vector is a combination of the others.

Linear Independence and Dependence

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}$$

- Example 1. The columns of the above matrix are linearly dependent, since the second column is three times the first.

Linear Independence and Dependence

- Example 2. The columns of the following triangular matrix are linearly independent.

No zeros on the diagonal $A = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 2 \end{bmatrix}$

Solve $Ac = 0$ $c_1 \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

c_1, c_2, c_3 are all forced to be zero.

Linear Independence and Dependence

- A set of n vectors in \mathbf{R}^m must be linearly dependent if $n > m$.
- Example 3. The three columns of the following matrix cannot be independent

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix}$$

In-Class Exercise 1

- If w_1, w_2, w_3 are independent vectors, let

$$v_1 = w_2 - w_3,$$

$$v_2 = w_1 - w_3,$$

$$v_3 = w_1 - w_2.$$

- Are v_1, v_2 , and v_3 independent?

- Solution: No, because $v_1 - v_2 + v_3 = 0$.

In-Class Exercise 2

- If w_1, w_2, w_3 are independent vectors, let

$$v_1 = w_2 + w_3,$$

$$v_2 = w_1 + w_3,$$

$$v_3 = w_1 + w_2.$$

- Are v_1, v_2 , and v_3 independent?
- Solution: Yes. Assume that $c_1*v_1+c_2*v_2+c_3*v_3 = 0$. Then we have $c_1*(w_2+w_3)+c_2*(w_1+w_3)+c_3*(w_1+w_2) = 0$. Then we have $(c_2+c_3)*w_1 + (c_1+c_3)*w_2 + (c_1+c_2)*w_3 = 0$. We know that w_1, w_2, w_3 are independent, so we have $c_2+c_3=0$, $c_1+c_3=0$, and $c_1+c_2=0$, which gives us $c_1=c_2=c_3=0$. So v_1, v_2 , and v_3 are independent.

Vector Spaces

- Most important spaces:
 1. \mathbf{R}^1 : Line (One-dimensional space)
 2. \mathbf{R}^2 : Represented by usual x - y plane (Two-dimensional space)
 3. \mathbf{R}^3 : Represented by usual x - y - z space (Three-dimensional space)
- **A real vector space is a set of vectors together with rules for vector addition and multiplication by real numbers**
 - Linear combinations stay in the vector space
 - If we add any vectors x and y in the vector space, $x+y$ is in the vector space.
 - If we multiply any vector x in the vector space by any scalar c , cx is in the vector space.

Spanning a Vector Space

- If a vector space \mathbf{V} consists of all linear combinations of w_1, \dots, w_l , then these vectors **span the space**. Every vector v in \mathbf{V} is some combination of the w 's:
- Every v comes from w 's: $v = c_1w_1 + \dots + c_lw_l$ for some coefficients c_i .

Basis for a Vector Space

- The crucial idea of a **basis**:
 - A basis for V is a sequence of vectors having two properties at once:
 1. The vectors are linearly independent (not too many vectors).
 2. They span the space V (not too few vectors).
 - A vector space has **infinitely many different bases**.
 - Whenever a square matrix is invertible, its columns are independent—and they are a basis for \mathbf{R}^n .

Dimension of a Vector Space

- The number of basis vectors is a property of the space itself:
 - Any two bases for a vector space V contain the same number of vectors.
 - This number, which is shared by all bases and expresses the number of “degrees of freedom” of the space, is the **dimension** of V .

Dimension of a Vector Space

- **In a subspace of dimension k , no set of more than k vectors can be independent, and no set of less than k vectors can span the space.**
 - Any linearly independent set in V can be extended to a basis, by adding more vectors if necessary. (A basis is a **maximal independent set**. It cannot be made larger without losing independence.)
 - Any spanning set in V can be reduced to a basis, by discarding vectors if necessary. (A basis is also a **minimal spanning set**. It cannot be made smaller and still span the space.)

In-Class Exercise 3

Suppose v_1, v_2, \dots, v_6 are six vectors in \mathbf{R}^4 .

- (a) Those vectors (do)(do not)(might not) span \mathbf{R}^4 .
- (b) Those vectors (are)(are not)(might be) linearly independent.
- (c) Any four of those vectors (are)(are not)(might be) a basis for \mathbf{R}^4 .

More Advanced Content

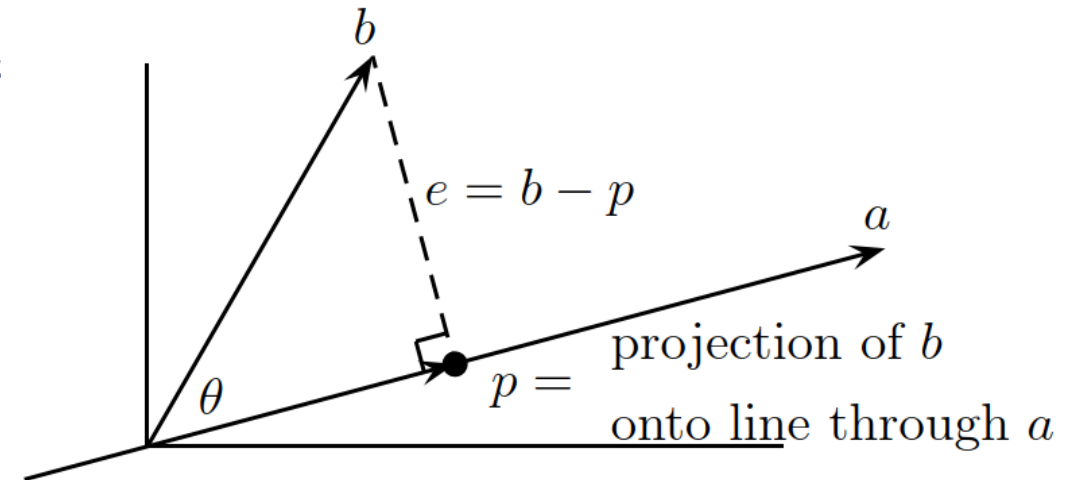
- Linear Independence, Basis, Rank, and Dimension
- **Orthogonality**

Orthogonality and Independence

- Useful fact: If nonzero vectors v_1, \dots, v_k are mutually orthogonal, then those vectors are linearly independent.
- Why?
 - Suppose $c_1 v_1 + \dots + c_k v_k = 0$.
 - Orthogonality of the v 's leaves only one term:
$$v_1^T (c_1 v_1 + \dots + c_k v_k) = c_1 v_1^T v_1 = 0.$$
 - The vectors are nonzero, so $v_1^T v_1 \neq 0$ and therefore $c_1 = 0$.
 - The same is true of every c_i . So the only combination of the v 's producing zero has all $c_i = 0$: **independence!**

Projections

- Suppose we want to find the distance from a point b to the line in the direction of the vector a .
 - The dotted line connecting b to p is orthogonal/perpendicular to a .
- Given a plane (or any vector space S) instead of a line, again the problem is to find the point p on that space that is closest to b .
 - This point p is the projection of b onto the vector space.



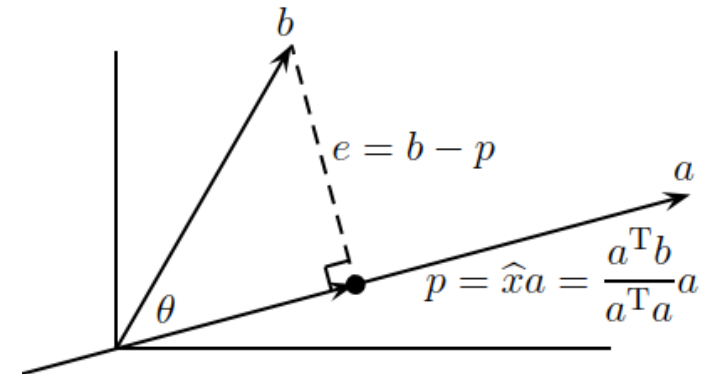
Projection onto a Line

- Find the projection point p :
 - All we need is the geometrical fact that the line from b to the closest point $p = \hat{a} = \hat{x}a$ is orthogonal to the vector a :

$$(b - \hat{a}) \perp a, \quad \text{or} \quad a^T(b - \hat{a}) = 0, \quad \text{or} \quad \hat{x} = \frac{a^T b}{a^T a}.$$

- The projection of the vector b onto the line in the direction of a is $p = \hat{x}a$:

Projection onto a line $p = \hat{x}a = \frac{a^T b}{a^T a}a.$



Projection Matrix of Rank 1

- The projection of b onto the line through a lies at $p = a(a^T b / a^T a)$.
- Projection onto a line is carried out by a **projection matrix** P . P is the matrix that multiplies b and produces p :

$$p = a \frac{a^T b}{a^T a} \quad \text{so the projection matrix is} \quad P = \frac{a a^T}{a^T a}.$$

Example

- The matrix that projects onto the line through $a = (1,1,1)^T$ is:

$$P = \frac{aa^T}{a^T a} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

- P is a symmetric matrix.
- Its square is itself: $P^2 = P$.
 - Why?
 - What does it mean?

Projections and Least Squares

- $Ax = b$ either has solution(s) or not. **More equations than unknowns—no solution?**
 $2x = b_1$
 $3x = b_2$
 $4x = b_3.$
- In spite of their unsolvability, **inconsistent equations arise all the time in practice.** They have to be solved!
 - Rather than expecting no error in some equations and large errors in the others, **it is much better to choose the x that minimizes an average error E in the m equations.**

Projections and Least Squares

- The most convenient “average” comes from the sum of squares:

Squared error $E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2.$

$2x = b_1$
 $3x = b_2$
 $4x = b_3.$

- If there is an exact solution, the minimum error is $E = 0$.
- In the more likely case that b is not proportional to a , the graph of E^2 will be a parabola (拋物線). **The minimum error is at the lowest point, where the derivative is zero:**

$$\frac{dE^2}{dx} = 2[(2x - b_1)2 + (3x - b_2)3 + (4x - b_3)4] = 0.$$

- Solving for x , the least-squares solution of this model system $ax = b$ is denoted by \hat{x} :

Leastsquares solution $\tilde{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{a^T b}{a^T a}.$

Projections and Least Squares

- The **general case** is the same, We “solve” $ax = b$ by minimizing:

$$E^2 = \|ax - b\|^2 = (a_1x - b_1)^2 + \cdots + (a_mx - b_m)^2.$$

- The derivative of E^2 is zero at the point \hat{x} , if:

$$(a_1\hat{x} - b_1)a_1 + \cdots + (a_m\hat{x} - b_m)a_m = 0.$$

- We are **minimizing the distance from** b **to the line through** a , and calculus gives the same answer, $\hat{x} = (a_1b_1 + \cdots + a_mb_m)/(a_1^2 + \cdots + a_m^2)$, that geometry did earlier:

The least-squares solution to a problem $ax = b$ in one unknown is $\hat{x} = \frac{a^T b}{a^T a}$.

- **The error vector** e **connecting** b **to** p **must be perpendicular to** a :

Orthogonality of a and e $a^T(b - \hat{x}a) = a^Tb - \frac{a^Tb}{a^Ta}a^Ta = 0.$

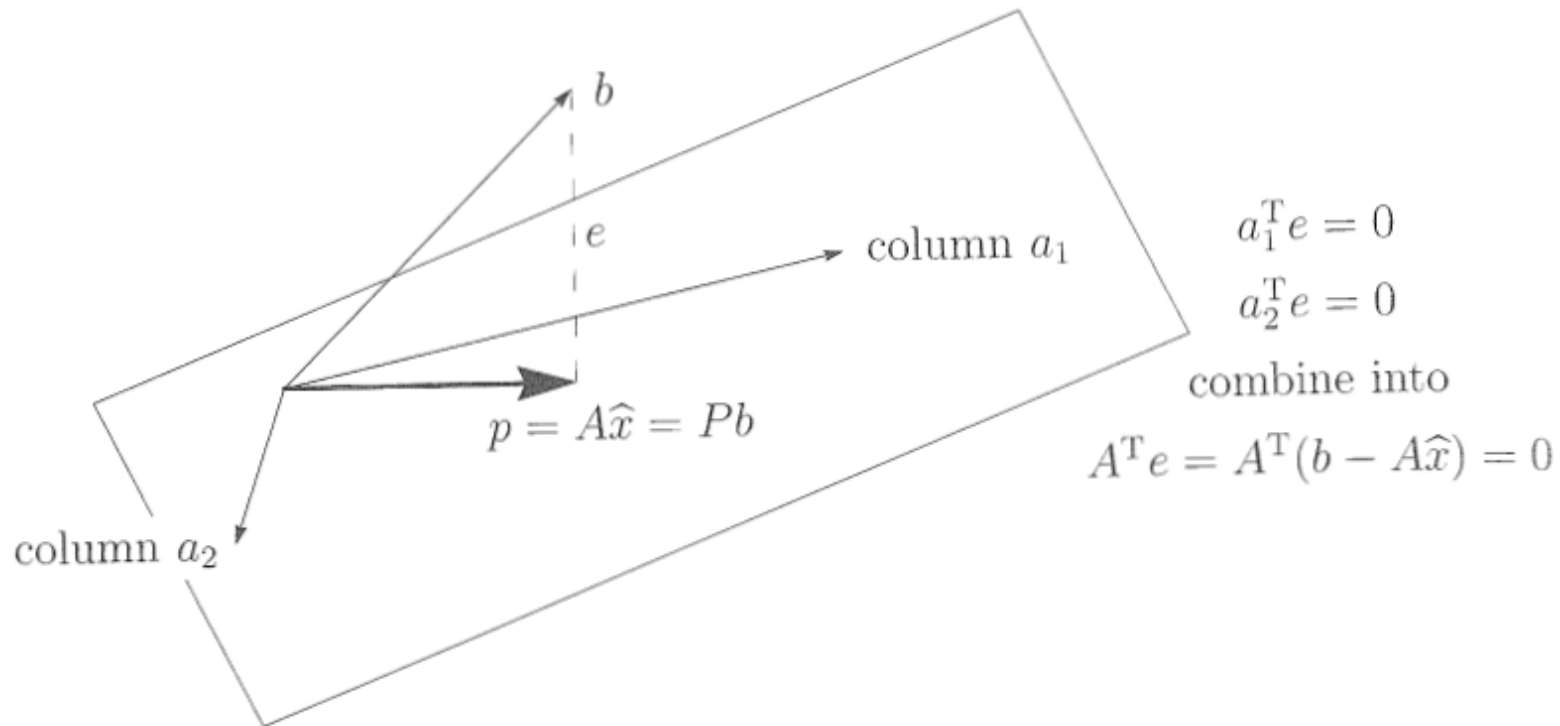
Least Squares Problems with Several Variables

- Now, projecting b onto a **vector space** rather than just onto a line. This problem arises from $Ax = b$ when A is an m by n matrix.
 - The number m of equations is still larger than the number n of unknowns, so it must be expected that $Ax = b$ will be inconsistent. **Probably, there will not exist a choice of x that perfectly fits the data b .**
- Again, the problem is to choose \hat{x} so as to minimize the error, and again this minimization will be done in the least-squares sense:
 - The error is $E = \|Ax - b\|$.
 - Searching for the least-squares solution \hat{x} , which minimizes E , **is the same as locating the point $p = A\hat{x}$ that is closer to b than any other point in the column space.**

Least Squares Problems with Several Variables

- We find \hat{x} and the projection $p = A\hat{x}$ as follows:

$$A^T(b - A\hat{x}) = 0 \quad \text{or} \quad A^T A\hat{x} = A^T b.$$



Normal Equations

- The equations $A^T A \hat{x} = A^T b$ are known in statistics as the **normal equations**.

- When $Ax = b$ is inconsistent, its least-squares solution minimizes $\|Ax - b\|^2$:

$$\text{Normal equations} \quad A^T A \hat{x} = A^T b.$$

- $A^T A$ **is invertible exactly when the columns of A are linearly independent**! Then:

$$\text{Best estimate } \hat{x} \quad \hat{x} = (A^T A)^{-1} A^T b.$$

- The projection of b onto the column space is the nearest point $A\hat{x}$:

$$\text{Projection} \quad p = A\hat{x} = A(A^T A)^{-1} A^T b.$$

Projection Matrices

- The closest point to b is $p = A(A^T A)^{-1} A^T b$. The matrix that gives p is a **projection matrix**, denoted by P :

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

Properties of Projection Matrices

- The projection matrix $P = A(A^T A)^{-1} A^T$ has two basic properties:
 1. It equals its square: $P^2 = P$.
 2. It equals its transpose: $P^T = P$.

– Why?

1. If we start with any b , then Pb lies in the subspace we are projecting onto. **When we project again nothing is changed.** $P(Pb)$ is still Pb . Two or three or fifty projections give the same point p as the first projection:

$$P^2 = A(A^T A)^{-1} A^T A (A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P.$$

2. Multiply the transposes in reverse order, and use symmetry of $(A^T A)^{-1}$, to come back to P :

$$P^T = (A^T)^T ((A^T A)^{-1})^T A^T = A(A^T A)^{-1} A^T = P.$$

Least-Squares Fitting of Data

- Suppose we do a series of experiments, and expect the output b to be a linear function of the input t .
- We look for a **straight line** $b = C + Dt$.
- How to compute C and D ?

$$\begin{array}{rcl} C + Dt_1 & = & b_1 \\ C + Dt_2 & = & b_2 \\ & \vdots & \\ C + Dt_m & = & b_m. \end{array} \quad \longrightarrow \quad \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad \text{or} \quad Ax = b.$$

- The best solution is (\hat{C}, \hat{D}) that minimizes the squared error E^2 :

$$\textbf{Minimize} \quad E^2 = \|b - Ax\|^2 = (b_1 - C - Dt_1)^2 + \cdots + (b_m - C - Dt_m)^2.$$

Example

- Supposing three measurements b_1, b_2, b_3 are marked:

$$b = 1 \quad \text{at} \quad t = -1, \quad b = 1 \quad \text{at} \quad t = 1, \quad b = 3 \quad \text{at} \quad t = 2.$$

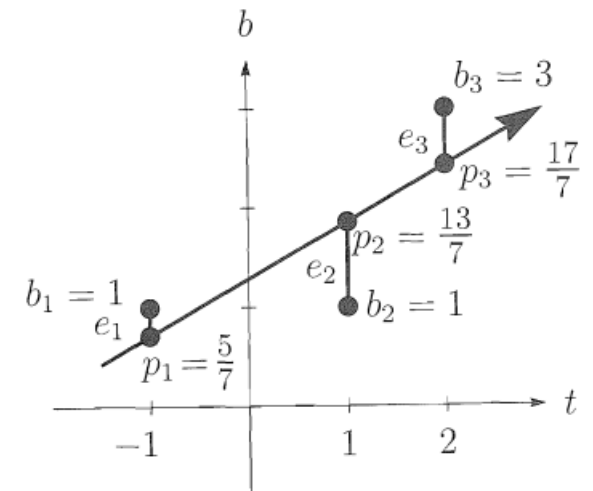
- The first step is **to write the equations**:

$$Ax = b \quad \text{is} \quad \begin{array}{rcl} C - D & = & 1 \\ C + D & = & 1 \\ C + 2D & = & 3 \end{array} \quad \text{or} \quad \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

- They are solved by least squares:

$$A^T A \hat{x} = A^T b \quad \text{is} \quad \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

- The best solution is $\hat{C} = \frac{9}{7}, \hat{D} = \frac{4}{7}$ and the best line is $\frac{9}{7} + \frac{4}{7}t$.



Orthonormal Basis

- In an **orthogonal basis**, every vector is perpendicular to every other vector.
- Divide each vector by its length, to make it a unit vector. That changes an orthogonal basis into an **orthonormal basis** of q 's.
 - The vectors q_1, \dots, q_n are **orthonormal** if:

$$q_i^T q_j = \begin{cases} 0 & \text{whenever } i \neq j, \\ 1 & \text{whenever } i = j, \end{cases} \quad \begin{array}{l} \text{giving the orthogonality;} \\ \text{giving the normalization.} \end{array}$$

- **A matrix with orthonormal columns will be called Q .**

Standard Orthonormal Basis

- The most important example of Q is the **standard basis**:
 - For the x - y plane, the best-known axes $e_1 = (1,0)$ and $e_2 = (0,1)$.
 - In n dimensions the standard basis e_1, \dots, e_n again consists of the columns of $Q = I$:

**Standard
basis**

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Orthogonal Matrices

- If Q has orthonormal columns, then $Q^T Q = I$:

$$\begin{array}{c} \text{Orthonormal} \\ \text{columns} \end{array} \begin{bmatrix} \text{---} & q_1^T & \text{---} \\ \text{---} & q_2^T & \text{---} \\ & \vdots & \\ \text{---} & q_n^T & \text{---} \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \end{bmatrix} = I.$$

- An orthogonal matrix is a square matrix with orthonormal columns. Its transpose is the inverse $Q^T = Q^{-1}$.

References and Acknowledgement

- Strang G. Linear algebra and its applications[J]. 2000.

The End