COMP 7180
Quantitative Methods for Data Analytics and Artificial Intelligence

Lecture 2 – Linear Independence, Rank, and Orthogonality

What is Linear Algebra

- Linear algebra is the study of vectors, matrices, and linear functions/equations.
- Vectors (columns of numbers) and matrices (2D arrays of numbers) are the language of data.
- **Linear algebra is a branch of mathematics, but the truth of it is that linear algebra is the mathematics of data.**

Open Questions

- $\mathbf{AB} \neq \mathbf{BA}$ (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.

$$\mathbf{Ax} = \mathbf{b} \quad \text{(multiply both sides by } \mathbf{A}^{-1}\text{)}$$

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \quad (\mathbf{A}^{-1}\mathbf{A} = \mathbf{I})$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

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

1. Which square matrices are invertible?
2. What if A is square matrix but not invertible?
3. What if 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:

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

- 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:

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

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

Properties of Transpose

- The transpose of A^{-1} is $(A^{-1})^T = (A^T)^{-1}$
- Proof:
- 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., $(A^{-1})^T = (A^T)^{-1}$.

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

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: $a_{ij} = a_{ji}$.

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

  - Property: $R^T R$ and RR^T are symmetric matrices.
  - Proof: $(R^T R)^T = R^T (R^T)^T = R^T R$ (since $(R^T)^T = R$)

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

Small Exercise

- If A is not a zero matrix, is $A^2 = 0$ possible? is $A^T A = 0$ possible?

Solution to Small Exercise

- If A is not a zero matrix, is $A^2 = 0$ possible? is $A^TA = 0$ possible?
- Solution: $A^2 = 0$ is possible. For example, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$
- On the other hand, $A^TA = 0$ is not possible, which can be proved by exploring its diagonal.

Outline of Today’s Content

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

Outline of Today's 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 $c_1 v_1 + \cdots + c_k v_k = 0$ only happens when $c_1 = \cdots = c_k = 0$. Then the vectors $v_1, \ldots, v_k$ 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

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

- <u>Example 1.</u> 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 $w1$, $w2$, $w3$ are independent vectors, let

$$v1 = w2 - w3,$$

$$v2 = w1 - w3,$$

$$v3 = w1 - w2.$$

- Are $v1$, $v2$, and $v3$ independent?

Solution to In-Class Exercise 1

- If w1, w2, w3 are independent vectors, let

$$v1 = w2 - w3,$$
$$v2 = w1 - w3,$$
$$v3 = w1 - w2.$$

- Are v1, v2, and v3 independent?

- Solution: No, because v1 – v2 + v3 = 0.

In-Class Exercise 2

- If w1, w2, w3 are independent vectors, let

$$v1 = w2 + w3,$$

$$v2 = w1 + w3,$$

$$v3 = w1 + w2.$$

- Are v1, v2, and v3 independent?

Solution to In-Class Exercise 2

- If $w1, w2, w3$ are independent vectors, let

$$v1 = w2 + w3,$$

$$v2 = w1 + w3,$$

$$v3 = w1 + w2.$$

- Are $v1, v2$, and $v3$ independent?
- Solution: Yes. Assume that $c1*v1+c2*v2+c3*v3 = 0$. Then we have $c1*(w2+w3)+c2*(w1+w3)+c3*(w1+w2) = 0$. Then we have $(c2+c3)*w1 + (c1+c3)*w2 + (c1+c2)*w3 = 0$. We know that $w1, w2, w3$ are independent, so we have $c2+c3=0$, $c1+c3=0$, and $c1+c2=0$, which gives us $c1=c2=c3=0$. So $v1, v2$, and $v3$ 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 $w_1, \ldots, w_l$, then these vectors **<u>span the space</u>**. Every vector v in $\mathbf{V}$ is some combination of the w's:
- Every v comes from w's: $v = c_1 w_1 + \cdots + c_l w_l$ for some coefficients c_i.

Basis for a Vector Space

- The crucial idea of a **basis**:
  - A basis for $\mathbf{V}$ 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 $\mathbf{V}$ (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 $\mathbf{V}$ 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 $\mathbf{V}$.

Dimension of a Vector Space

- If $v_1, \dots, v_m$ and $w_1, \dots, w_n$ are both bases for the same vector space, then $m = n$. The number of vectors is the same.
- Proof:
  ① Suppose there are more w's than v's ($n > m$).
  ② Since the v's form a basis, they must span the space. Every w_j can be written as a combination of the v's: If $w_1 = a_{11}v_1 + \dots + a_{m1}v_m$, this is the first column of a matrix multiplication VA.

$$W = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \begin{bmatrix} a_{11} & & \\ \vdots & & \\ a_{m1} & & \end{bmatrix} = VA$$

  ③ So A is a short, wide matrix, since $n > m$. **There is a nonzero solution to** $Ax = 0$. Then $VAx = 0$ which is $Wx = 0$. **Then w's could not be a basis**. **CONTRADICTION!**
  ④ So we cannot have $n > m$.

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 $v_1, v_2, \ldots, v_6$ 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 $\mathbf{R}^4$.

Solution to In-Class Exercise 3

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

(a) Those vectors (do)(do not)$\boxed{\text{(might not)}}$ span $\mathbf{R}^4$.

(b) Those vectors (are)$\boxed{\text{(are not)}}$(might be) linearly independent.

(c) Any four of those vectors (are)(are not)$\boxed{\text{(might be)}}$ a basis for $\mathbf{R}^4$.

Outline of Today's 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**.
- Proof:
  - Suppose $c_1 v_1 + \cdots + c_k v_k = 0$.
  - Orthogonality of the v's leaves only one term:

$$v_1^{\mathrm{T}}(c_1 v_1 + \cdots + c_k v_k) = c_1 v_1^{\mathrm{T}} v_1 = 0.$$

  - 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**!

Subspaces

- **Definition:** A **subspace** of a vector space is a nonempty subset that satisfies the requirements for a vector space: **Linear combinations stay in the subspace**.
  I. If we add any vectors x and y in the subspace, $x + y$ is in the subspace.
  II. If we multiply any vector x in the subspace by any scalar c, cx is in the subspace.
- **Notice: The zero vector will belong to every subspace.**
- The smallest subspace $\mathbf{Z}$ contains only one vector, the zero vector.
- The largest subspace is the whole of the original space.

Orthogonal Subspaces

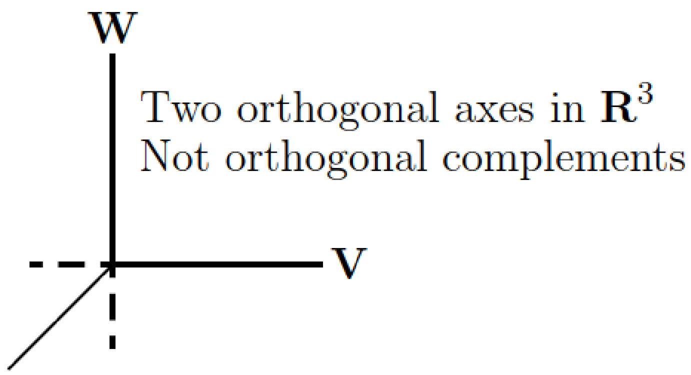
- The orthogonality of two subspaces:
  - **Two subspaces** $\mathbf{V}$ **and** $\mathbf{W}$ **of the same space** R^n **are orthogonal if every vector** v **in** $\mathbf{V}$ **is orthogonal to every vector** w **in** $\mathbf{W}$: $v^T w = 0$ **for all** v **and** w.
  - Example:
    - Suppose $\mathbf{V}$ is the plane spanned by $v_1 = (1,0,0)$ and $v_2 = (0,1,0)$.
    - If $\mathbf{W}$ is the line spanned by $w = (0,0,1)$, then w is orthogonal to both v's. The line $\mathbf{W}$ will be orthogonal to the whole plane $\mathbf{V}$.

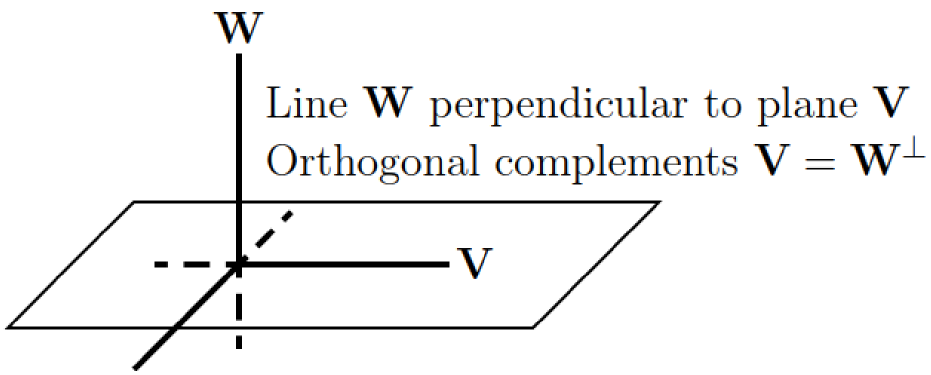
Orthogonal Complement

- **Definition**:
  – Given a subspace $\boldsymbol{V}$ of $\mathbf{R}^n$, the space of **all** vectors orthogonal to V is called the **orthogonal complement** of $\boldsymbol{V}$. It is denoted by $\boldsymbol{V}^{\perp} = \text{"}\mathbf{V}\ \text{perp."}$.

Splitting $\mathbf{R}^n$ into Orthogonal Parts

- Orthogonal complements in $\mathbf{R}^3$:
  - The dimensions of $\mathbf{V}$ and $\mathbf{W}$ are right, and the whole space $\mathbf{R}^3$ is being decomposed into two perpendicular parts.





- Splitting $\mathbf{R}^n$ into orthogonal parts will split every vector into $x = v + w$.
  - The vector v is the projection onto the subspace $\boldsymbol{V}$.
  - The orthogonal component w is the projection of x onto $\mathbf{W}$.

Example

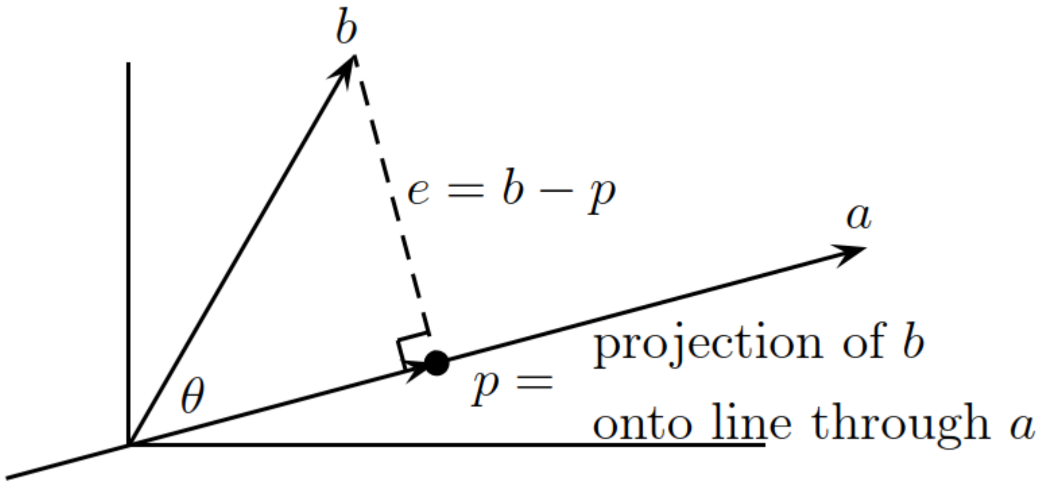
- Assume the whole space is $\mathbf{R}^2$
- If V is the subspace spanned by [1,0], then W is the subspace spanned by [0,1].
- Splitting $\mathbf{R}^2$ into orthogonal parts will split every vector into $x = v + w$. For example, x = [2,3] = [2,0] + [0,3]
  - v = [2,0] is the projection onto the subspace $\boldsymbol{V}$.
  - The orthogonal component w =[0,3] is the projection of x onto $\mathbf{W}$.

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 perpendicular to** a.

– Given a plane (or any subspace $\boldsymbol{S}$) instead of a line, again the problem is to find the point p on that subspace that is closest to b.

➢ **This point** p **is the projection of** b **onto the subspace**.



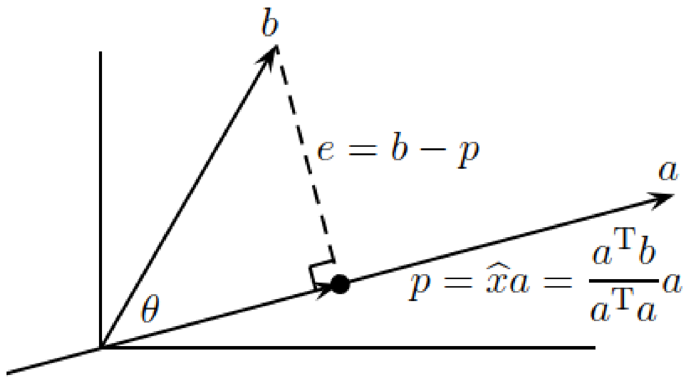
Projection onto a Line

- 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:

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

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

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



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:

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

Example

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

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

  - P is a symmetric matrix.
  - Its square is itself: $P^2 = P$.
    - ➢ Can you prove that?
    - ➢ What does it mean?

Projections and Least Squares

- $Ax = b$ either has solution(s) or not.

More equations than unknowns— no solution?

$$\begin{aligned} 2x &= b_1 \\ 3x &= b_2 \\ 4x &= b_3. \end{aligned}$$

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

$$\begin{aligned} 2x &= b_1 \\ 3x &= b_2 \\ 4x &= b_3. \end{aligned}$$

  – 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 (抛物線). <u>**The minimum error is at the lowest point, where the derivative is zero**</u>:

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

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

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

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:

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

- 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 $\widehat{x} = \dfrac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}$.

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

Orthogonality of a and e $\quad a^{\mathrm{T}}(b - \widehat{x}a) = a^{\mathrm{T}}b - \dfrac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}a^{\mathrm{T}}a = 0.$

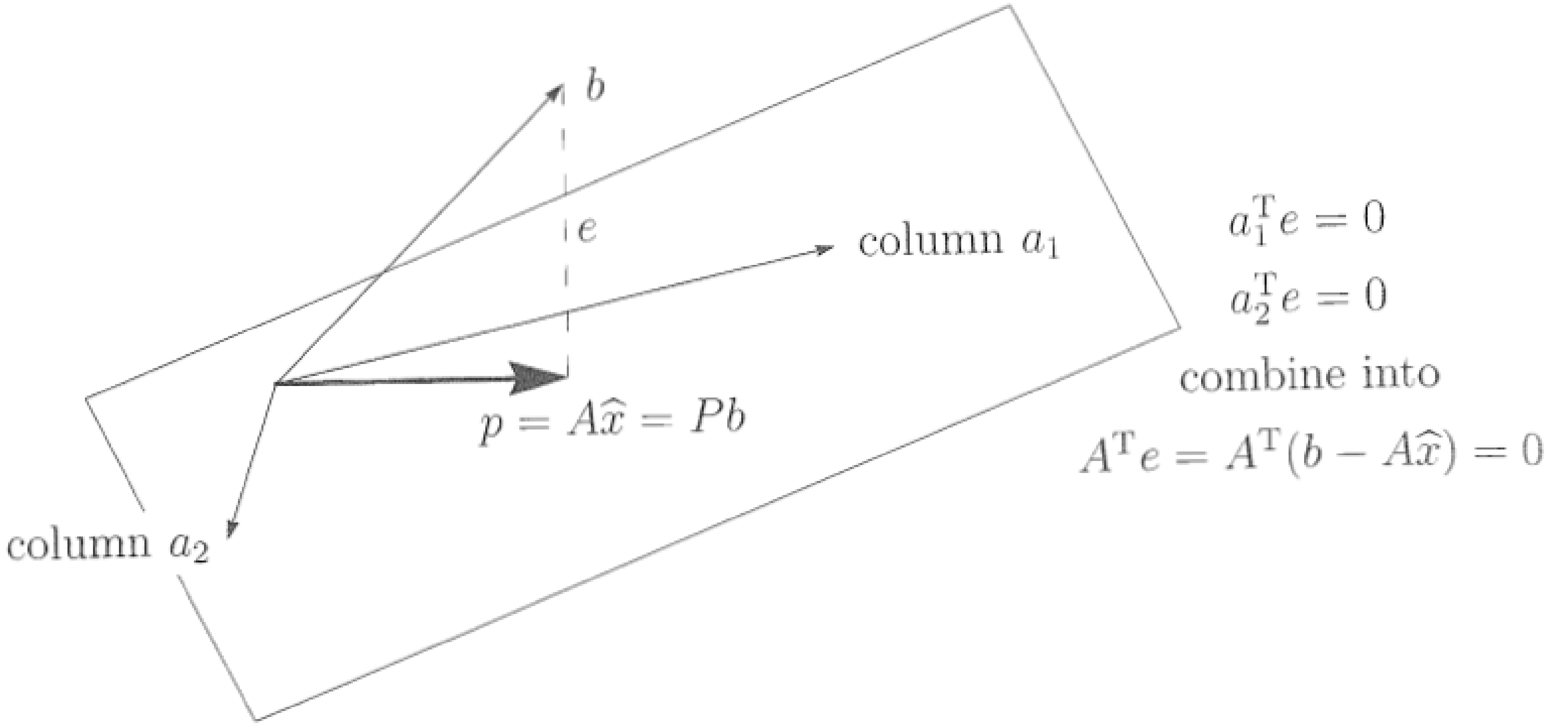
Least Squares Problems with Several Variables

- Now, projecting b onto a **subspace** 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:

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



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 $\|Ax - b\|^2$:

$$\textbf{Normal equations} \qquad A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b.$$

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

$$\textbf{Best estimate } \widehat{x} \qquad \widehat{x} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

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

$$\textbf{Projection} \qquad p = A\widehat{x} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

The Cross-Product Matrix $A^T A$

- The matrix $A^T A$ is certainly **symmetric**.
  - $(A^T A)^T = A^T A^{TT} = A^T A$

- $Ax = 0 \Rightarrow A^T A x = 0$?
- $A^T A x = 0 \Rightarrow Ax = 0$?

- ***A*** **has independent columns == $\boldsymbol{A^T A}$ is invertible.**

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:

$$\textbf{Projection matrix} \qquad P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}.$$

  – $I - P$ **is also a projection matrix**! It projects b onto the orthogonal complement, and the projection is $b - Pb$.

Properties of Projection Matrices

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

– Proof:

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

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

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

$$P^{\mathrm{T}} = (A^{\mathrm{T}})^{\mathrm{T}}\left((A^{\mathrm{T}}A)^{-1}\right)^{\mathrm{T}}A^{\mathrm{T}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = P.$$

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** $b = C + Dt$. For example:

  The cost of producing t **books is nearly linear**, $b = C + Dt$, with editing and typesetting in C and then printing and binding in D. C is the set-up cost and D is the cost for each additional book.

Least-Squares Fitting of Data

- How to compute C and D?

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

- This is an **overdetermined** system, with m equations and only 2 unknowns.
  - If there is **no experimental error**, then two measurements of b will determine the line $b = C + Dt$.
  - If **errors are presenting**, it will have no solution.
- The best solution is $(\hat{C}, \widehat{D})$ that minimizes the squared error E^2:

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

Example

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

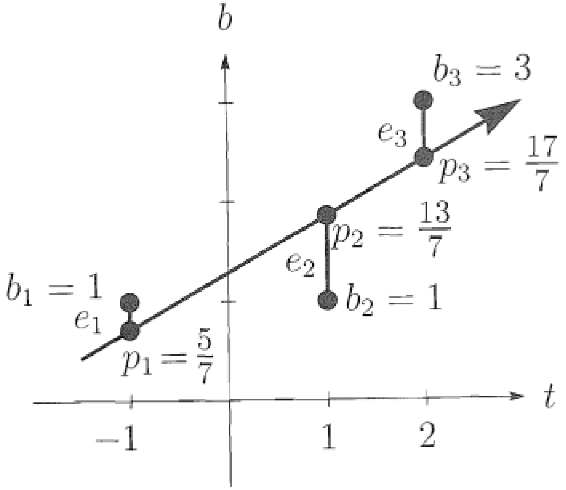
$$b = 1 \quad \text{at} \quad t = -1, \qquad b = 1 \quad \text{at} \quad t = 1, \qquad b = 3 \quad \text{at} \quad t = 2.$$

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

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

  – They are solved by least squares:

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



  – The best solution is $\hat{C} = \frac{9}{7}, \widehat{D} = \frac{4}{7}$ 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 $q_1, \ldots, q_n$ are **orthonormal** if:

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

- **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 $e_1, \ldots, e_n$ again consists of the columns of $Q = I$:

Standard basis

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

Orthogonal Matrices

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

Orthonormal columns

$$\begin{bmatrix} — & q_1^{\mathrm{T}} & — \\ — & q_2^{\mathrm{T}} & — \\ & \vdots & \\ — & q_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \end{bmatrix} = I.$$

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

Example

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

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

- Q rotates every vector through the angle θ, and Q^T rotates it back through $-\theta$.
- The columns of Q and Q^T are **orthonormal** because $\sin^2\theta + \cos^2\theta = 1$.
- They are orthonormal matrices.

Properties of Orthogonal Matrix

- **Multiplication by any Q preserves lengths**, because $(Qx)^T(Qx) = x^TQ^TQx = x^Tx$.

Lengths unchanged $\quad ||Qx|| = ||x|| \quad$ for every vector x.

- It also preserves inner products and angles, since $(Qx)^T(Qy) = x^TQ^TQy = x^Ty$.

Coefficients of the Basis Vectors

- If we have an orthonormal basis, then any vector is a combination of the basis vectors. The problem is **to find the coefficients of the basis vectors**:

Write b ***as a combination*** $b = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n.$

① The method:

- Compute x_1: Multiply both sides of the equation by q_1^T, we are left with:

$$q_1^{\mathrm{T}} b = x_1 q_1^{\mathrm{T}} q_1. \xrightarrow{q_1^{\mathrm{T}} q_1 = 1} x_1 = q_1^{\mathrm{T}} b.$$

- Similarly, the second coefficient is $x_2 = q_2^T b$.
- Each piece of b has a simple formula, and recombining the pieces gives back b:

Every vector b ***is equal to*** $(q_1^{\mathrm{T}} b) q_1 + (q_2^{\mathrm{T}} b) q_2 + \cdots + (q_n^{\mathrm{T}} b) q_n.$

Rectangular Matrices with Orthogonal Columns

- When the columns are orthonormal, the "cross-product matrix" $A^T A$ becomes $Q^T Q = I$.
- We emphasize that **<u>those projections do not reconstruct</u>** b. In the square case $m = n$, they did. In the rectangular case $m > n$, they don't.
  - They give the projection p and not the original vector b and the q's are no longer a basis.
- The projection matrix is usually $A(A^T A)^{-1} A^T$, and here it simplifies to:

$$P = Q(Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}} \qquad \text{or} \qquad P = QQ^{\mathrm{T}}.$$

  - Notice that $Q^T Q$ is the n by n identity matrix, but QQ^T is an m by m projection P.

The End