Projections and Least Squares

Orthogonal Bases and Gram-Schmidt

Projections and Least Squares

- $Ax = \ b$ either has a solution or not.
- If b is not in the column space C(A), the system is inconsistent and Gaussian elimination fails.
- This failure is almost certain when there are several equations and only one unknown:
- More equations than unknowns— no solution?

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

- This is solvable when b_1, b_2, b_3 are in the ratio $2{:}\,3{:}\,4$.
- The solution x will exist only if b is on the same line as the column $a = (2,3,4)$.

- In spite of their insolvability, inconsistent equations arise all the time in practice.
- One possibility is to determine x from part of the system, and ignore the rest; this is hard to justify if all m equations come from the same source.
- 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.
- The most convenient "average" comes from the sum of squares:

Squared error:

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

- 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.
- The minimum error is at the lowest point, where the derivative is zero:

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

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

Least-squares solution $$\hat{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}.$$

- 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, $\hat{x} = \frac{a_1b_1 + \cdots + a_mb_m}{a_1^2 + \cdots + a_m^2}$:

3K The least-squares solution to a problem $ax = b$ in one unknown is $\widehat{x} = \dfrac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}$.

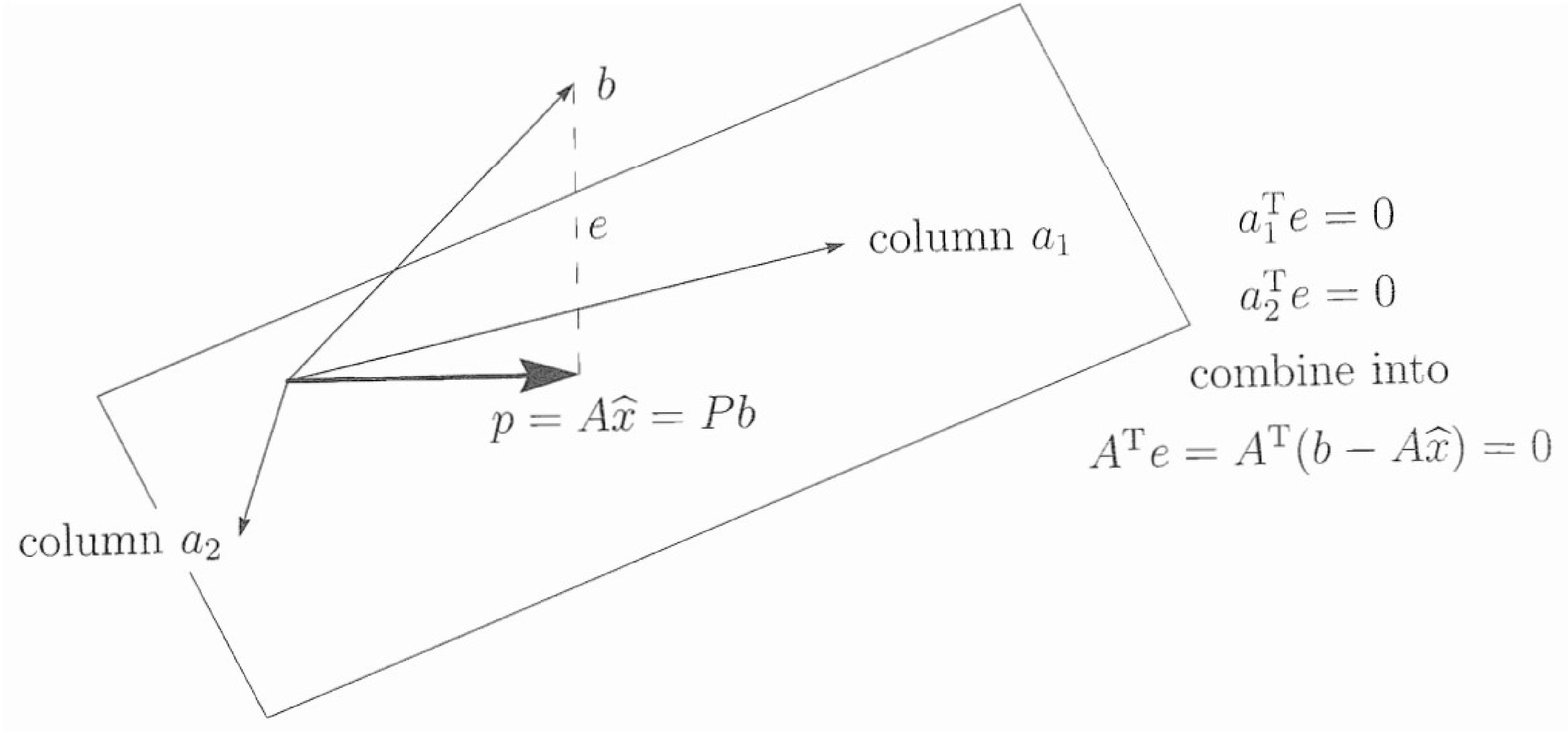
- The error vector e connecting b to p must be perpendicular to a:

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

Least Squares Problems with Several Variables

- **Goal:** To project b onto a subspace—rather than just onto a line.
- $Ax = b$ where A is an m by n matrix.
- Instead of one column and one unknown x, the matrix now has n columns.
- The number m of observations 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. In other words, the vector b probably will not be a combination of the columns of A; it will be outside the column space.

- 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 $E = \|Ax - b\|$, and this is exactly the distance from b to the point Ax in the column space.
- 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.
- In n dimensions, p must be the “projection of b onto the column space.”
- The error vector $e = b - A\hat{x}$ must be perpendicular to that space.



Projection onto the column space of a 3 by 2 matrix.

- Finding $\hat{x}$ and the projection $p = A\hat{x}$ is so fundamental that we do it in two ways:
    1. All vectors perpendicular to the column space lie in the left nullspace. Thus the error vector $e = b - A\hat{x}$ must be in the nullspace of A^T:

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

    2. The error vector must be perpendicular to each column $a_1, \ldots, a_n$ of A:

$$\begin{matrix} a_1^{\mathrm{T}}(b - A\widehat{x}) = 0 \\ \vdots \\ a_n^{\mathrm{T}}(b - A\widehat{x}) = 0 \end{matrix} \qquad \text{or} \qquad \begin{bmatrix} a_1^{\mathrm{T}} \\ \vdots \\ a_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \\ b - A\widehat{x} \\ \\ \end{bmatrix} = 0.$$

- This is again $A^T(b - A\hat{x}) = 0$ and $A^TA\hat{x} = A^Tb$
- The equations $A^TA\hat{x} = A^Tb$ are known in statistics as the normal equations.

3L When $Ax = b$ is inconsistent, its least-squares solution minimizes $\|Ax - b\|^2$:

$$\textbf{Normal equations} \qquad A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b. \tag{1}$$

$A^{\mathrm{T}}A$ is invertible exactly when the columns of A are linearly independent! Then,

$$\textbf{Best estimate } \widehat{x} \qquad \widehat{x} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b. \tag{2}$$

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

$$\textbf{Projection} \qquad p = A\widehat{x} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b. \tag{3}$$

Example 1

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

- $Ax = b$ has no solution. $A^T A\,\hat{x} = \mathrm{A^T b}$ gives the best x.

$$\mathrm{A^T}A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}.$$

$$\hat{x} = (\mathrm{A^T}A)^{-1}A^{\mathrm{T}}b = \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Projection $$p = A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

- In this special case, the best we can do is to solve the first two equations of $Ax = b$. Then $\hat{x}_1 = 2$ and $\hat{x}_2 = 1$. The error in the equation $0x_1 + 0x_2 = 6$ is sure to be 6.

Remark **4.** Suppose b is actually in the column space of A—it is a combination $b = Ax$ of the columns. Then the projection of b is still b:

$$\textbf{b in column space} \qquad p = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}Ax = Ax = b.$$

The closest point p is just b itself—which is obvious.

Remark **5.** At the other extreme, suppose b is *perpendicular* to every column, so $A^{\mathrm{T}}b = 0$. In this case b projects to the zero vector:

$$\textbf{b in left nullspace} \qquad p = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b = A(A^{\mathrm{T}}A)^{-1}0 = 0.$$

Remark **6.** When A is square and invertible, the column space is the whole space. Every vector projects to itself, p equals b, and $\widehat{x} = x$:

$$\textbf{If A is invertible} \qquad p = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b = AA^{-1}(A^{\mathrm{T}})^{-1}A^{\mathrm{T}}b = b.$$

This is the only case when we can take apart $(A^{\mathrm{T}}A)^{-1}$, *and write it as* $A^{-1}(A^{\mathrm{T}})^{-1}$. When A is rectangular that is not possible.

Remark **7.** Suppose A has only one column, containing a. Then the matrix $A^{\mathrm{T}}A$ is the number $a^{\mathrm{T}}a$ and $\widehat{x}$ is $a^{\mathrm{T}}b/a^{\mathrm{T}}a$. We return to the earlier formula.

The Cross-Product Matrix $A^T A$

- The matrix $A^T A$ is certainly symmetric.
- $A^T A$ has the same null space as A.
- Certainly if $Ax = 0$ then $A^T Ax = 0$.
- Vectors x in the null space of A are also in the null space of $A^T A$. To go in the other direction, start by supposing that $A^T Ax = 0$, and take the inner product with x to show that $Ax = 0$:

$$x^{\mathrm{T}}A^{\mathrm{T}}Ax = 0, \quad \text{or} \quad \|Ax\|^2 = 0, \quad \text{or} \quad Ax = 0.$$

- The two null spaces are identical. In particular, if A has independent columns (and only $x = 0$ is in its null space), then the same is true for $A^T A$:

3M If A has independent columns, then $A^{\mathrm{T}}A$ is *square*, *symmetric*, and *invertible*.

Projection Matrices

- The closest point to b is $p = A(A^T A)^{-1} A^T b$. This formula expresses in matrix terms the construction of a perpendicular line from b to the column space of A.
- The matrix that gives p is a projection matrix, denoted by P:

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

- This matrix projects any vector b onto the column space of A.
- In other words, $p = Pb$ is the component of b in the column space, and the error $e = b - Pb$ is the component in the orthogonal complement.
- $I - P$ is also a projection matrix! It projects b onto the orthogonal complement, and the projection is $b - Pb$.
- In short, we have a matrix formula for splitting any b into two perpendicular components. Pb is in the column space C(A), and the other component $(I - P)b$ is in the left nullspace $N(A^T)$—which is orthogonal to the column space.

3N The projection matrix $P = A(A^TA)^{-1}A^T$ has two basic properties:

(i) It equals its square: $P^2 = P$.

(ii) It equals its transpose: $P^T = P$.

Conversely, any symmetric matrix with $P^2 = P$ represents a projection.

Proof:

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

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

- For the converse, we have to deduce from $P^2 = P$ and $P^T = P$ that Pb is the projection of b onto the column space of P.
- The error vector $b - Pb$ is orthogonal to the space. For any vector Pc in the space, the inner product is zero:

$$(b - Pb)^TPc = b^T(I - P)^TPc = b^T(P - P^2)c = 0.$$

- Thus $b - Pb$ is orthogonal to the space, and Pb is the projection onto the column space.

Example 2

- Suppose A is actually invertible. If it is $4\ by\ 4$, then its four columns are independent and its column space is all of $\boldsymbol{R}^4$. What is the projection onto the whole space?

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

- It is the identity matrix.
- The identity matrix is symmetric, $I^2 = I$, and the error $b - Ib$ is zero.

Least-Squares Fitting of Data

- Consider an overdetermined system, with m equations and only two unknowns.

$$\begin{aligned} C &+ Dt_1 &= b_1 \\ C &+ Dt_2 &= b_2 \\ &\vdots& \\ C &+ Dt_m &= b_m. \end{aligned}$$

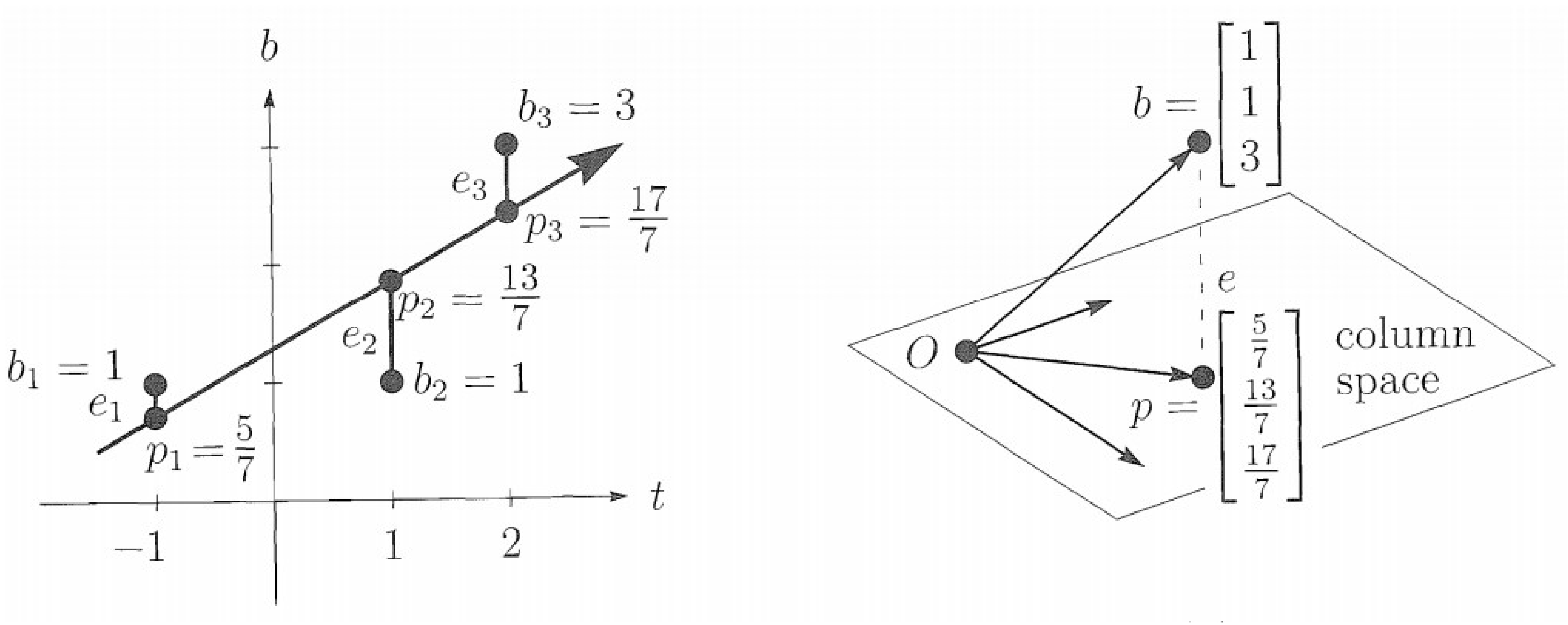
- If errors are present, it will have no solution. A has two columns, and $x = (C, D)$:

$$\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \qquad \text{or} \qquad Ax = b.$$

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

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

- The vector $p = A\hat{x}$ is as close as possible to b. Of all straight lines $b = C + Dt$, we are choosing the one that best fits the data. On the graph, the errors are the vertical distances $b - C - Dt$ to the straight line (not perpendicular distances!).
- It is the vertical distances that are squared, summed, and minimized.



- Straight-line approximation matches the projection p of b.

- Three measurements b_1, b_2, b_3 are marked on the figure:
- $b = 1$ at $t = -1$, $b = 1$ at $t = 1$, $b = 3$ at $t = 2$.
- Note that the values $t = -1, 1, 2$ are not required to be equally spaced.
- The first step is to write the equations that would hold if a line could go through all three points. Then every $C + Dt$ would agree exactly with b:

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

- If those equations $Ax = b$ could be solved, there would be no errors. They can't be solved because the points are not on a line. Therefore 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 $\frac{9}{7} + \frac{4}{7}t$.

3O The measurements $b_1, \ldots, b_m$ are given at distinct points $t_1, \ldots, t_m$. Then the straight line $\widehat{C} + \widehat{D}t$ which minimizes E^2 comes from least squares:

$$A^{\mathrm{T}}A \begin{bmatrix} \widehat{C} \\ \widehat{D} \end{bmatrix} = A^{\mathrm{T}}b \qquad \text{or} \qquad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \widehat{C} \\ \widehat{D} \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Orthogonal Bases and Gram-Schmidt

- In an orthogonal basis, every vector is perpendicular to every other vector. The coordinate axes are mutually orthogonal.
- **Orthonormal basis:** Divide each vector by its length, to make it a unit vector. That changes an orthogonal basis into an orthonormal basis of q's:

3P 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.

Example of Orthonormal Basis

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

3Q If Q (square or rectangular) has orthonormal columns, then $Q^{\mathrm{T}}Q = I$:

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.
- Then Q^T is Q^{-1} . **For square orthogonal matrices, the transpose is the inverse.**
- When row i of Q^T multiplies column j of Q, the result is $q_j^T q_j = 0$. On the diagonal where $i = j$, we have $q_i^T q_i = 1$. That is the normalization to unit vectors of length 1.

Example 1

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

- Q rotates every vector through the angle θ, and Q^T rotates it back through $-\theta$. The columns are clearly orthogonal, and they are orthonormal because $\sin^2\theta + \cos^2\theta = 1$. The matrix Q^T is just as much an orthogonal matrix as Q.

Example 2

- Any permutation matrix P is an orthogonal matrix. The columns are certainly unit vectors and certainly orthogonal—because the 1 appears in a different place in each column: The transpose is the inverse.

$$\text{If} \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{then} \quad P^{-1} = P^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

- Projections reduce the length of a vector, whereas orthogonal matrices have a property that is the most important and most characteristic of all:

3R Multiplication by any Q preserves lengths:

$$\textbf{Lengths unchanged} \qquad \|Qx\| = \|x\| \quad \text{for every vector } x.$$

It also preserves inner products and angles, since

$$(Qx)^{\mathrm{T}}(Qy) = x^{\mathrm{T}}Q^{\mathrm{T}}Qy = x^{\mathrm{T}}y.$$

- The preservation of lengths comes directly from $Q^TQ = I$:

  $$\|Qx\|^2 = \|x\|^2 \quad \text{because} \quad (Qx)^{\mathrm{T}}(Qx) = x^{\mathrm{T}}Q^{\mathrm{T}}Qx = x^{\mathrm{T}}x.$$

- All inner products and lengths are preserved, when the space is rotated or reflected.

- We come now to the calculation that uses the special property $Q^T = Q^{-1}$.
- If we have a 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 $b = x_1q_1 + x_2q_2 + \cdots + x_nq_n$.

- To compute x_1, multiply both sides of the equation by q_1^T.
- On the left-hand side is $q_1^T b$. On the right-hand side all terms disappear (because $q_1^T\, q_j = 0$) except the first term.
- We are left with $q_1^T\, b = x_1 q_1^T\, q_1$.
- Since $q_1^T\, q_1 = 1$, we have found $x_1 = q_1^T b$. Similarly the second coefficient is $x_2 = q_2^T b$

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

- Putting this orthonormal basis into a square matrix Q, the vector equation $x_1q_1 + \cdots + x_nq_n = b$ is identical to $Qx = b$.
- The columns of Q multiply the components of x.
- Its solution is $x = Q^{-1}b$. But since $Q^{-1} = Q^T$—this is where orthonormality enters—the solution is also $x = Q^T b$:

$$x = Q^{\mathrm{T}}b = \begin{bmatrix} - & q_1^{\mathrm{T}} & - \\ & \vdots & \\ - & q_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \begin{bmatrix} q_1^{\mathrm{T}}b \\ \vdots \\ q_n^{\mathrm{T}}b \end{bmatrix}$$

- The components of x are the inner products $q_i^T b$

- **The rows of a square matrix are orthonormal whenever the columns are.**
- **Example:**

Orthonormal columns
Orthonormal rows

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}.$$

Rectangular Matrices with Orthogonal Columns

- The n orthonormal vectors q_i in the columns of Q have m > n components. Then Q is an m by n matrix and we cannot expect to solve $Qx = b$ exactly. We solve it by least squares.
- The key is to notice that we still have $Q^TQ = I$. So Q^T is still the left-inverse of Q.
- The normal equations are $Q^TQx = Q^Tb$. But Q^TQ is the identity matrix! Therefore $\hat{x} = Q^Tb$, whether Q is square and $\hat{x}$ is an exact solution, or Q is rectangular and we need least squares.

3S If Q has orthonormal columns, the least-squares problem becomes easy: rectangular system with no solution for most b.

$Qx = b$		rectangular system with no solution for most b.
$Q^{\mathrm{T}}Q\widehat{x} = Q^{\mathrm{T}}b$		normal equation for the best $\widehat{x}$—in which $Q^{\mathrm{T}}Q = I$.
$\widehat{x} = Q^{\mathrm{T}}b$		$\widehat{x}_i$ is $q_i^{\mathrm{T}}b$.
$p = Q\widehat{x}$		the projection of b is $(q_1^{\mathrm{T}}b)q_1 + \cdots + (q_n^{\mathrm{T}}b)q_n$.
$p = QQ^{\mathrm{T}}b$		the projection matrix is $P = QQ^{\mathrm{T}}$.

- The projection matrix is usually $A(A^T A)^{-1}A^T$, and here it simplifies to

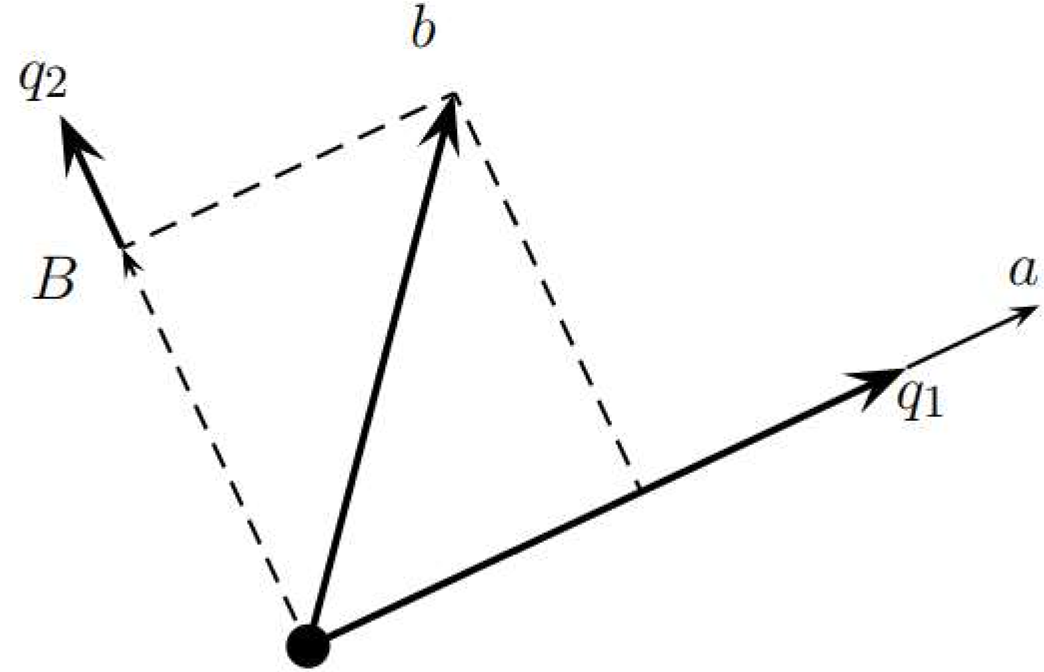
$$P = Q(Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}} \qquad \text{or} \qquad P = QQ^{\mathrm{T}}.$$

The Gram-Schmidt Process

- This method is used to make the vectors orthonormal to each other.
- We are given a, b, c and we want q_1, q_2, q_3.
- There is no problem with q_1: it can go in the direction of a.
- We divide by the length, so that $q_1 = \frac{a}{\|a\|}$ is a unit vector.
- Now the second vector q_2—has to be orthogonal to q_1. If the second vector b has any component in the direction of q_1 (which is the direction of a), that component has to be subtracted:

Second vector $\quad B = b - (q_1^{\mathrm{T}} b) q_1 \quad \text{and} \quad q_2 = B / \|B\|.$

- B is orthogonal to q_1. It is the part of b that goes in a new direction, and not in the a. B is perpendicular to q_1. It sets the direction for q_2.



The q_i component of b is removed; a and B normalized to q_1 and q_2.

- At this point q_1 and q_2 are set.
- The third orthogonal direction starts with c.
- It will not be in the plane of q_1 and q_2, which is the plane of a and b.
- However, it may have a component in that plane, and that has to be subtracted.
- What is left is the component C we want, the part that is in a new direction perpendicular to the plane:

Third vector $\quad C = c - (q_1^{\mathrm{T}} c) q_1 - (q_2^{\mathrm{T}} c) q_2 \quad \text{and} \quad q_3 = C / \|C\|.$

- This is the one idea of the whole Gram-Schmidt process, **to subtract from every new vector its components in the directions that are already settled.**

Example

- Suppose the independent vectors are a, b, c:

$$a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

- To find q_1, make the first vector into a unit vector: $q_1 = a/\sqrt{2}$. To find q_2, subtract from the second vector its component in the first direction:

$$B = b - (q_1^{\mathrm{T}} b) q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

- The normalized q_2 is B divided by its length, to produce a unit vector:

$$q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}.$$

- To find q_3, subtract from c its components along q_1 and q_2:

$$\begin{aligned} C &= c - (q_1^{\mathrm{T}} c) q_1 - (q_2^{\mathrm{T}} c) q_2 \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned}$$

- This is already a unit vector, so it is q_3.

- **Final Answer:**

Orthonormal basis $$Q = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}.$$

3T The Gram-Schmidt process starts with independent vectors $a_1, \ldots, a_n$ and ends with orthonormal vectors $q_1, \ldots, q_n$. At step j it subtracts from a_j its components in the directions $q_1, \ldots, q_{j-1}$ that are already settled:

$$A_j = a_j - (q_1^{\mathrm{T}} a_j) q_1 - \cdots - (q_{j-1}^{\mathrm{T}} a_j) q_{j-1}.$$

Then q_j is the unit vector $A_j / \|A_j\|$.

The Factorization $\boldsymbol{A = QR}$

- We started with a matrix A, whose columns were a, b, c.
- We ended with a matrix Q, whose columns are q_1, q_2, q_3.
- What is the relation between those matrices?
- The matrices A and Q are m by n when the n vectors are in m-dimensional space, and there has to be a third matrix that connects them.

- The idea is to write the a's as combinations of the q's. The vector b is a combination of the orthonormal q_1 and q_2, and the combination is:

$$b = (q_1^{\mathrm{T}} b) q_1 + (q_2^{\mathrm{T}} b) q_2.$$

- Every vector in the plane is the sum of its q_1 and q_2 components. Similarly c is the sum of its q_1, q_2, q_3 components: $c = (q_1^{\mathrm{T}} c) q_1 + (q_2^{\mathrm{T}} c) q_2 + (q_3^{\mathrm{T}} c) q_3.$
- If we express that in matrix form we have the new factorization $A = QR$:

QR factors

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

Where R is upper triangular

3U Every m by n matrix with independent columns can be factored into $A = QR$. The columns of Q are orthonormal, and R is upper triangular and invertible. When $m = n$ and all matrices are square, Q becomes an orthogonal matrix.

- Orthogonalization simplifies the least-squares problem $Ax = b$. The normal equations are still correct, but $A^T A$ becomes easier:

$$A^{\mathrm{T}}A = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}R.$$

- The fundamental equation $A^T A\,\hat{x} = A^T b$ simplifies to a triangular system:

$$R^{\mathrm{T}}R\widehat{x} = R^{\mathrm{T}}Q^{\mathrm{T}}b \qquad \text{or} \qquad R\widehat{x} = Q^{\mathrm{T}}b.$$

- Instead of solving $QRx = b$, which can't be done, we solve $R\hat{x} = Q^T b$ which is just back-substitution because R is triangular.
- The real cost is the mn^2 operations of Gram Schmidt, which are needed to find Q and R in the first place.