Orthogonality

3.2 Projections onto Lines

Projection onto a Line

Now we want to find the projection point p. This point must be some multiple $p = \widehat{x}a$ of the given vector a—every point on the line is a multiple of a. The problem is to compute

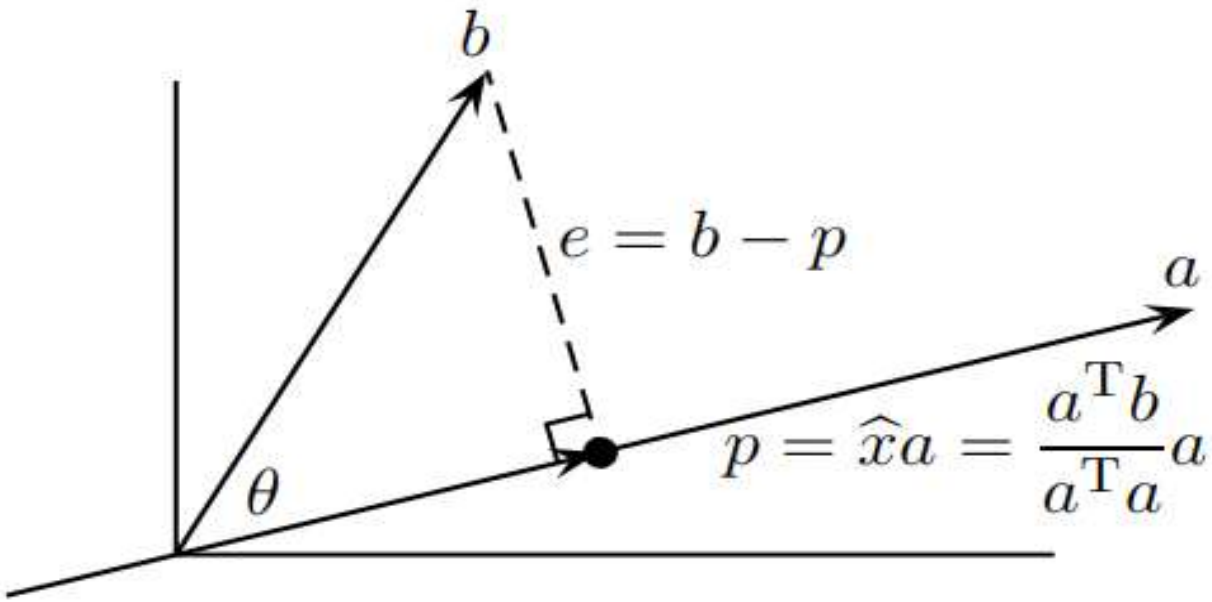


Figure 3.7: The projection p of b onto a,

the coefficient $\widehat{x}$. All we need is the geometrical fact that ***the line from*** b ***to the closest point*** $p = \widehat{x}a$ ***is perpendicular to the vector*** a:

$$(b - \widehat{a}) \perp a, \quad \text{or} \quad a^{\mathrm{T}}(b - \widehat{a}) = 0, \quad \text{or} \quad \widehat{x} = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}. \tag{4}$$

That gives the formula for the number $\widehat{x}$ and the projection p:

3H The projection of the vector b onto the line in the direction of a is $p = \widehat{x}a$:

$$\textbf{Projection onto a line} \qquad p = \widehat{x}a = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}a. \tag{5}$$

Example 1. Project $b = (1,2,3)$ onto the line through $a = (1,1,1)$ to get $\widehat{x}$ and p:

$$\widehat{x} = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a} = \frac{6}{3} = 2.$$

The projection is $p = \widehat{x}a = (2,2,2)$.

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

Projection Matrix of Rank 1

P is the matrix that multiplies b and produces p:

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

Projection Matrix of Rank 1

The projection of b onto the line through a lies at $p = a(a^{\mathrm{T}}b/a^{\mathrm{T}}a)$. That is our formula $p = \widehat{x}a$, but it is written with a slight twist: The vector a is put before the number $\widehat{x} = a^{\mathrm{T}}b/a^{\mathrm{T}}a$. There is a reason behind that apparently trivial change. Projection onto a line is carried out by a ***projection matrix*** P, and written in this new order we can see what it is. *P is the matrix that multiplies b and produces p*:

$$P = a\frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a} \quad \text{so the projection matrix is} \quad P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}. \tag{7}$$

That is a column times a row—a square matrix—divided by the number $a^{\mathrm{T}}a$.

Example 2. The matrix that projects onto the line through $a = (1,1,1)$ is

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

This matrix has two properties that we will see as typical of projections:

1. P is a symmetric matrix.
2. Its square is itself: $P^2 = P$.

P^2b is the projection of Pb—and Pb is already on the line! So $P^2b = Pb$. This matrix P also gives a great example of the four fundamental subspaces:

The column space consists of the line through $a = (1,1,1)$.

The nullspace consists of the plane perpendicular to a.

The rank is $r = 1$.

Every column is a multiple of a, and so is $Pb = \widehat{x}a$. The vectors that project to $p = 0$ are especially important. They satisfy $a^{\mathrm{T}}b = 0$—they are perpendicular to a and their component along the line is zero. They lie in the nullspace = perpendicular plane.

Actually that example is too perfect. It has the nullspace orthogonal to the column space, which is haywire. The nullspace should be orthogonal to the *row space*. But because P is symmetric, its row and column spaces are the same.

Remark on scaling The projection matrix $aa^{\mathrm{T}}/a^{\mathrm{T}}a$ is the same if a is doubled:

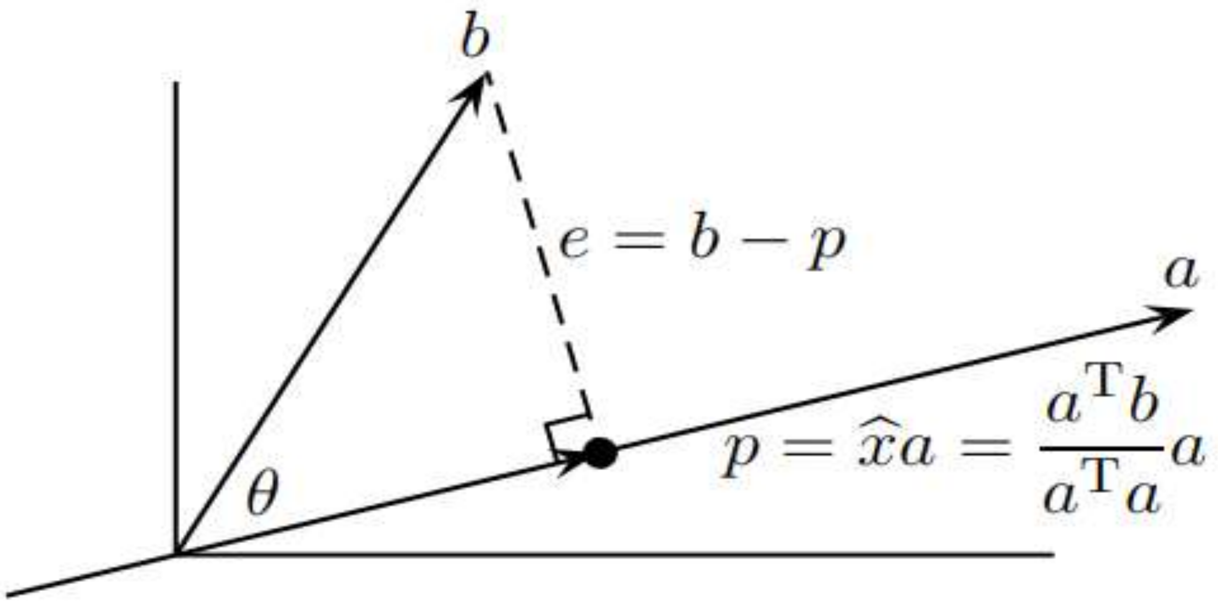
$$a = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \quad \text{gives} \quad P = \frac{1}{12}\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 2 & 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad \text{as before.}$$

The line through a is the same, and that's all the projection matrix cares about. If a has unit length, the denominator is $a^{\mathrm{T}}a = 1$ and the matrix is just $P = aa^{\mathrm{T}}$.

We emphasize that it produces the projection p:

To project b onto a, multiply by the projection matrix P: $p = Pb$.

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



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

Now we are ready for the serious step, ***to project*** b ***onto a subspace***—rather than just onto a line. This problem arises from $Ax = b$ when 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 $\widehat{x}$ 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 $\widehat{x}$, which minimizes E, is the same as locating the point $p = A\widehat{x}$ that is closer to b than any other point in the column space.

The error vector $e = b - A\hat{x}$ must be perpendicular to that space

p must be the "projection of b onto the column space."

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:

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

Remember That :

3C Fundamental theorem of orthogonality The row space is orthogonal to the nullspace (in $\mathbf{R}^n$). The column space is orthogonal to the left nullspace (in $\mathbf{R}^m$).

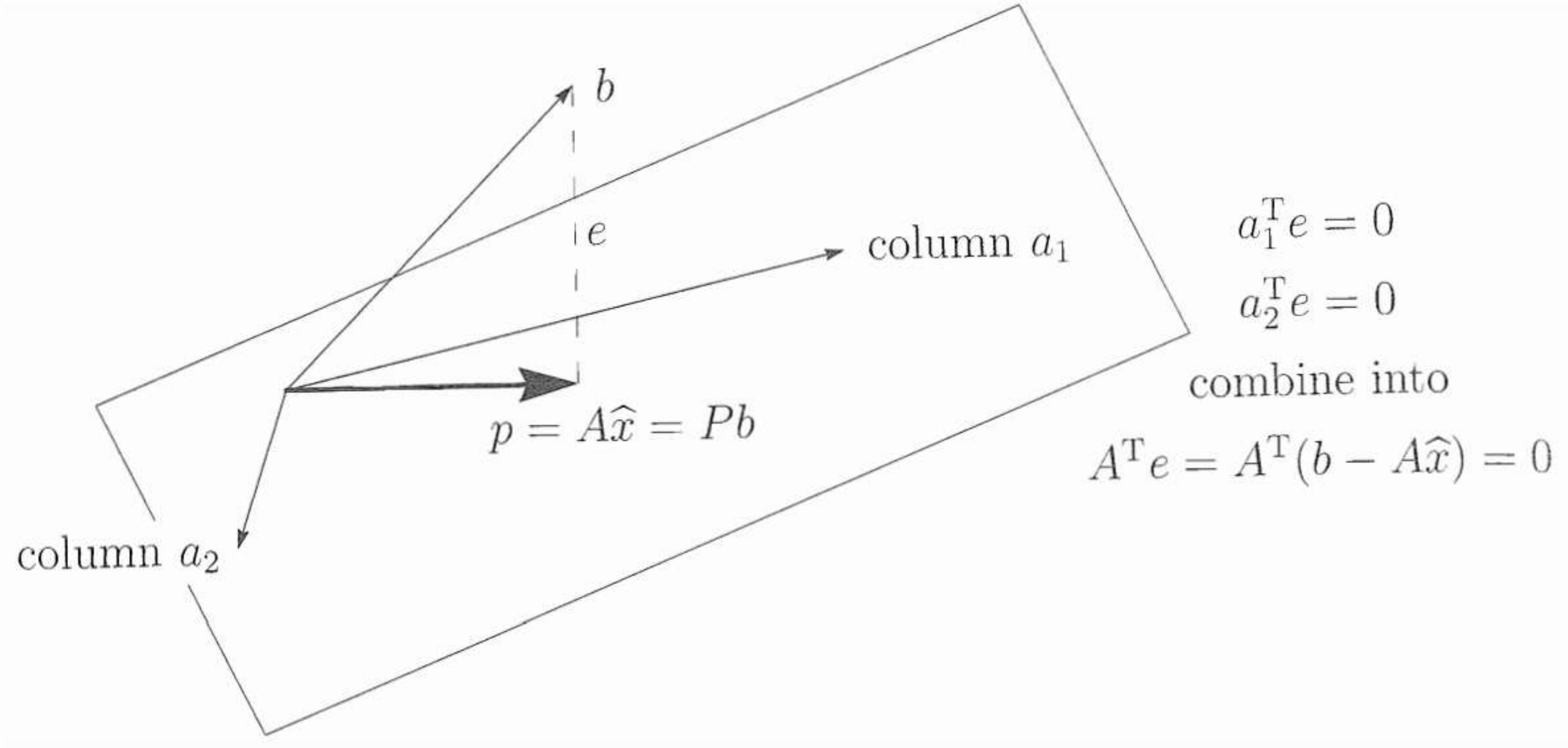


Figure 3.8: Projection onto the column space of a 3 by 2 matrix.

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 $A^{\mathrm{T}}(b - A\widehat{x}) = 0$ and $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$,

3L When $Ax = b$ is inconsistent,

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

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

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

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

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

We choose an example in which our intuition is as good as the formulas:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \qquad \begin{array}{ll} Ax = b & \text{has no solution} \\ A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b & \text{gives the best } x. \end{array}$$

Both columns end with a zero, so $C(A)$ is the x-y plane within three-dimensional space The projection of $b=(4,5,6)$ is $p=(4,5,0)$—the x and y components stay the same but $z=6$ will disappear. That is confirmed by solving the normal equations:

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

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

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

***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:

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

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:

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

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

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

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.

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

Projection Matrices

We have shown that the closest point to b is $p = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b$. *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:

$$\textbf{Projection matrix} \qquad P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}. \tag{4}$$

This matrix projects any vector b onto the column space of A.[1] 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 $N(A^{\mathrm{T}})$—which is orthogonal to the column space.

These projection matrices can be understood geometrically and algebraically.

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

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

(ii) It equals its transpose: $P^{\mathrm{T}} = P$.

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

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

Orthogonality

3.4 Orthogonal Bases and Gram-Schmidt

3.4 Orthogonal Bases and Gram-Schmidt

In an orthogonal basis, every vector is perpendicular to every other vector. The coordinate axes are mutually orthogonal. That is just about optimal, and the one possible improvement is easy: 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.

Orthogonal Matrices

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

Orthonormal columns

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

An orthogonal matrix is a square matrix with orthonormal columns.[2] 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 $q_j^{\mathrm{T}}q_j = 0$. On the diagonal where $i = j$, we have $q_i^{\mathrm{T}}q_i = 1$. That is the normalization to unit vectors of length 1.

Note that $Q^{\mathrm{T}}Q = I$ even if Q is rectangular. But then Q^{T} is only a left-inverse.

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.

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

the special property $Q^{\mathrm{T}} = Q^{-1}$.

to find the coefficients of the basis vectors:

$$\boxed{\textbf{Write } b \textbf{ as a combination } \; b = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n.}$$

To compute x_1 there is a neat trick. *Multiplv both sides of the eauation by* q_1^{T}.

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

Since $q_1^{\mathrm{T}} q_1 = 1$, we have found $x_1 = q_1^{\mathrm{T}} b$. Similarly the second coefficient is $x_2 = q_2^{\mathrm{T}} b$;

$$\textbf{\textit{Every vector } } b \textbf{\textit{ is equal to }} (q_1^{\mathrm{T}} b) q_1 + (q_2^{\mathrm{T}} b) q_2 + \cdots + (q_n^{\mathrm{T}} b) q_n. \tag{4}$$

$x_1 q_1 + \cdots + x_n q_n = b$ is identical to $Qx = b$.

Its solution is $x = Q^{-1} b$. But since $Q^{-1} = Q^{\mathrm{T}}$-

enters—the solution is also $x = Q^{\mathrm{T}} b$:

$$x = Q^{\mathrm{T}} b = \begin{bmatrix} - & q_1^{\mathrm{T}} & - \\ & \vdots & \\ - & q_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \begin{bmatrix} q_1^{\mathrm{T}} b \\ \vdots \\ q_n^{\mathrm{T}} b \end{bmatrix} \tag{5}$$

The components of x are the inner products $q_i^{\mathrm{T}} b$, as in equation (4).

Expressing b as a combination $x_1 a_1 + \cdots + x_n a_n$ is the same as solving $Ax = b$. The basis vectors go into the columns of A. In that case we need A^{-1}, which takes work. In the orthonormal case we only need Q^{T}.

The Gram-Schmidt Process

Suppose you are given three independent vectors a, b, c. If they are orthonormal, life is easy. To project a vector v onto the first one, you compute $(a^{\mathrm{T}}v)a$. To project the same vector v onto the plane of the first two, you just add $(a^{\mathrm{T}}v)a+(b^{\mathrm{T}}v)b$. To project onto the span of a, b, c, you add three projections. All calculations require only the inner products $a^{\mathrm{T}}v$, $b^{\mathrm{T}}v$, and $c^{\mathrm{T}}v$. But to make this true, we are forced to say, "***If*** they are orthonormal." Now we propose to find a way to ***make*** them orthonormal.

The method is simple. 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 $q_1 = a/\|a\|$ is a unit vector. The real problem begins with q_2—which 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***:

$$\textbf{Second vector} \qquad B = b-(q_1^{\mathrm{T}}b)q_1 \quad \text{and} \quad q_2 = B/\|B\|. \tag{9}$$

B is orthogonal to q_1. It is the part of b that goes in a new direction, and not in the a. In Figure 3.10, B is perpendicular to q_1. It sets the direction for q_2.

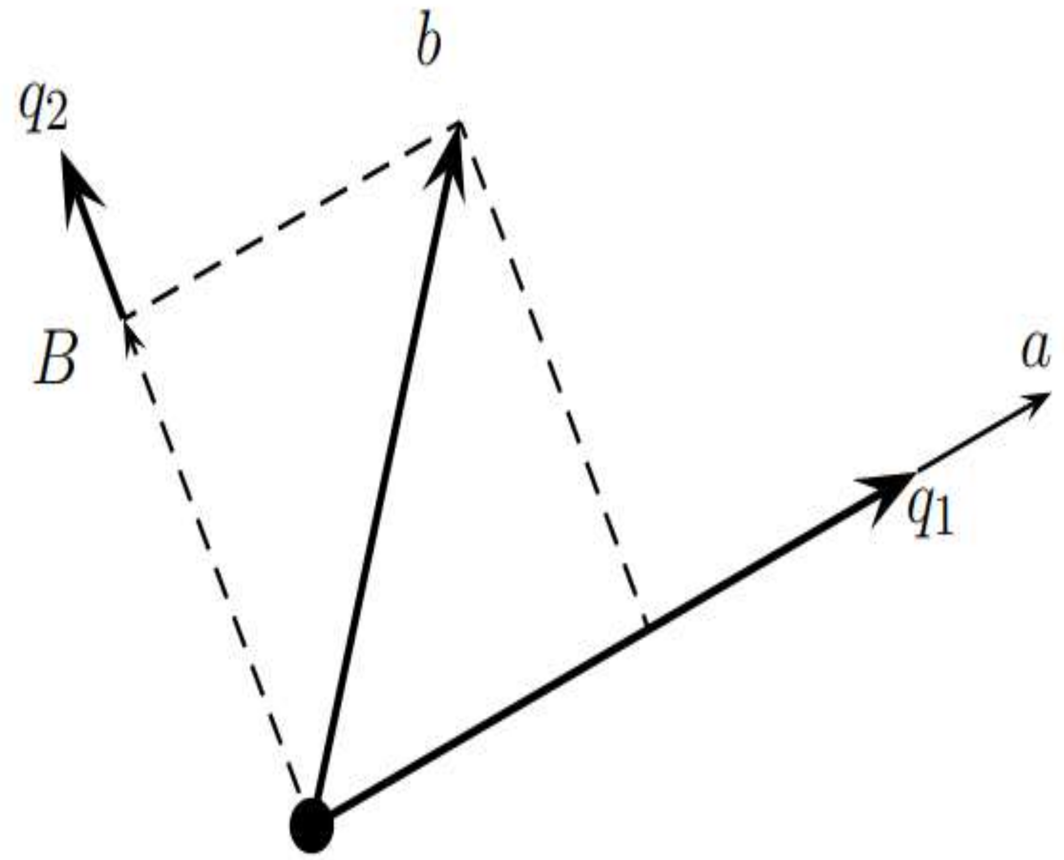


Figure 3.10: The q_i component of b is removed; a and B normalized to q_1 and q_2.

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. (If the result is $C = 0$, this signals that a, b, c were not independent in the first place) What is left is the component C we want, the part that is in a new direction perpendicular to the plane:

$$\textbf{Third vector} \qquad C = c - (q_1^{\mathrm{T}} c) q_1 - (q_2^{\mathrm{T}} c) q_2 \quad \text{and} \quad q_3 = C/\|C\|. \tag{10}$$

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***. That idea is used over and over again.[3] When there is a fourth vector, we subtract away its components in the directions of q_1, q_2, q_3.

Example 5. Gram-Schmidt 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. I went to desperate lengths to cut down the number of square roots (the painful part of Gram-Schmidt). The result is a set of orthonormal vectors q_1, q_2, q_3, which go into the columns of an orthogonal matrix Q:

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

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:

$$A_j = a_j - (q_1^{\mathrm{T}} a_j) q_1 - \cdots - (q_{j-1}^{\mathrm{T}} a_j) q_{j-1}. \tag{11}$$

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

Remark on the calculations I think it is easier to compute the orthogonal a, B, C, without forcing their lengths to equal one. Then square roots enter only at the end, when dividing by those lengths. The example above would have the same B and C, without using square roots. Notice the $\frac{1}{2}$ from $a^{\mathrm{T}} b / a^{\mathrm{T}} a$ instead of $\frac{1}{\sqrt{2}}$ from $q^{\mathrm{T}} b$:

$$B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and then} \quad C = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}.$$

The Factorization $A = QR$

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 in Figure 3.10 is a combination of the orthonormal q_1 and q_2, and we know what combination it 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*** $A = QR$:

$$\textbf{\textit{QR} factors} \qquad A = \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}}a & q_1^{\mathrm{T}}b & q_1^{\mathrm{T}}c \\ & q_2^{\mathrm{T}}b & q_2^{\mathrm{T}}c \\ & & q_3^{\mathrm{T}}c \end{bmatrix} = QR \qquad (12)$$

Notice the zeros in the last matrix! R is *upper triangular* because of the way Gram-Schmidt was done. The first vectors a and q_1 fell on the same line. Then q_1, q_2 were in the same plane as a, b. The third vectors c and q_3 were not involved until step 3.

The QR factorization is like $A = LU$, except that the first factor Q has orthonormal columns. The second factor is called R, because the nonzeros are to the *right* of the diagonal (and the letter U is already taken). The off-diagonal entries of R are the numbers $q_1^{\mathrm{T}}b = 1/\sqrt{2}$ and $q_1^{\mathrm{T}}c = q_2^{\mathrm{T}}c = \sqrt{2}$, found above. The whole factorization is

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ & 1/\sqrt{2} & \sqrt{2} \\ & & 1 \end{bmatrix} = QR.$$

You see the lengths of a, B, C on the diagonal of R. The orthonormal vectors q_1, q_2, q_3, which are the whole object of orthogonalization, are in the first factor Q.

Problem Set 3.4

Orthogonal Bases and Gram-Schmidt

2. Project $b = (0,3,0)$ onto each of the orthonormal vectors $a_1 = (\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$ and $a_2 = (-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, and then find its projection p onto the plane of a_1 and a_2.

3. Find also the projection of $b = (0,3,0)$ onto $a_3 = (\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$, and add the three projections. Why is $P = a_1a_1^{\mathrm{T}} + a_2a_2^{\mathrm{T}} + a_3a_3^{\mathrm{T}}$ equal to I?

9. If the vectors q_1, q_2, q_3 are orthonormal, what combination of q_1 and q_2 is closest to q_3?

12. What multiple of $a_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ should be subtracted from $a2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ to make the result orthogonal to a_1? Factor $\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$ into QR with orthonormal vectors in Q.

13. Apply the Gram-Schmidt process to

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

and write the result in the form $A = QR$.