45. Suppose an n by n matrix is invertible: $AA^{-1} = I$. Then the first column of A^{-1} is orthogonal to the space spanned by which rows of A?

46. Find $A^{\mathrm{T}}A$ if the columns of A are unit vectors, all mutually perpendicular.

47. Construct a 3 by 3 matrix A with no zero entries whose columns are mutually perpendicular. Compute $A^{\mathrm{T}}A$. Why is it a diagonal matrix?

48. The lines $3x+y=b_1$ and $6x+2y=b_2$ are ____. They are the same line if ____. In that case (b_1,b_2) is perpendicular to the vector ____. The nullspace of the matrix is the line $3x+y=$ ____. One particular vector in that nullspace is ____.

49. Why is each of these statements false?

(a) $(1,1,1)$ is perpendicular to $(1,1,-2)$, so the planes $x+y+z=0$ and $x+y-2z=0$ are orthogonal subspaces.

(b) The subspace spanned by $(1,1,0,0,0)$ and $(0,0,0,1,1)$ is the orthogonal complement of the subspace spanned by $(1,-1,0,0,0)$ and $(2,-2,3,4,-4)$.

(c) Two subspaces that meet only in the zero vector are orthogonal.

50. Find a matrix with $v=(1,2,3)$ in the row space and column space. Find another matrix with v in the nullspace and column space. Which pairs of subspaces can v *not* be in?

51. Suppose A is 3 by 4, B is 4 by 5, and $AB=0$. Prove $\operatorname{rank}(A)+\operatorname{rank}(B)\le 4$.

52. The command `N = null(A)` will produce a basis for the nullspace of A. Then the command `B = null(N')` will produce a basis for the ____ of A.

3.2 Cosines and Projections onto Lines

Vectors with $x^{\mathrm{T}}y=0$ are orthogonal. Now we allow inner products that are ***not zero***, and angles that are ***not right angles***. We want to connect inner products to angles, and also to transposes. In Chapter 1 the transpose was constructed by flipping over a matrix as if it were some kind of pancake. We have to do better than that.

One fact is unavoidable: *The orthogonal case is the most important.* Suppose we want to find the distance from a point b to the line in the direction of the vector a. We are looking along that line for the point p closest to b. The key is in the geometry: ***The line connecting*** b ***to*** p (the dotted line in Figure 3.5) ***is perpendicular to*** a. *This fact will allow us to find the projection* p. Even though a and b are not orthogonal, the distance problem automatically brings in orthogonality.

The situation is the same when we are given a plane (or any subspace **S**) instead of a line. Again the problem is to find the point p on that subspace that is closest to b. ***This***

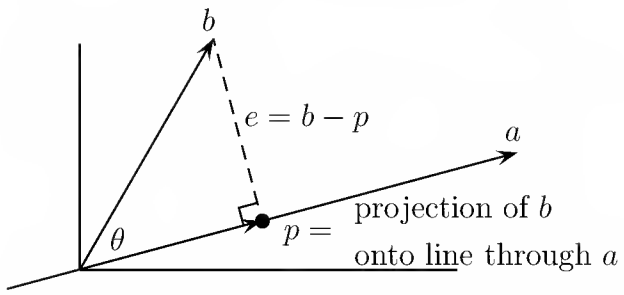

Figure 3.5: The projection p is the point (on the line through a) closest to b.

point p is the projection of b onto the subspace. A perpendicular line from b to **S** meets the subspace at p. Geometrically, that gives the distance between points b and subspaces **S**. But there are two questions that need to be asked:

1. Does this projection actually arise in practical applications?

2. If we have a basis for the subspace **S**, is there a formula for the projection p?

The answers are certainly yes. This is exactly the problem of the ***least-squares solution to an overdetermined system***. The vector b represents the data from experiments or questionnaires, and it contains too many errors to be found in the subspace **S**. When we try to write b as a combination of the basis vectors for **S**, it cannot be done—the equations are inconsistent, and $Ax = b$ has no solution.

The least-squares method selects p as the best choice to replace b. There can be no doubt of the importance of this application. In economics and statistics, least squares enters *regression analysis*. In geodesy, the U.S. mapping survey tackled 2.5 million equations in 400,000 unknowns.

A formula for p is easy when the subspace is a line. We will project b onto a in several different ways, and relate the projection p to inner products and angles. Projection onto a higher dimensional subspace is by far the most important case; it corresponds to a least-squares problem with several parameters, and it is solved in Section 3.3. The formulas are even simpler when we produce an orthogonal basis for **S**.

inner products and cosines

We pick up the discussion of inner products and angles. You will soon see that it is not the angle, but ***the cosine of the angle***, that is directly related to inner products. We look back to trigonometry in the two-dimensional case to find that relationship. Suppose the vectors a and b make angles α and β with the x-axis (Figure 3.6). The length $\|a\|$ is the hypotenuse in the triangle OaQ. So the sine and cosine of α are

$$\sin\alpha = \frac{a_2}{\|a\|}, \qquad \cos\alpha = \frac{a_1}{\|a\|}.$$

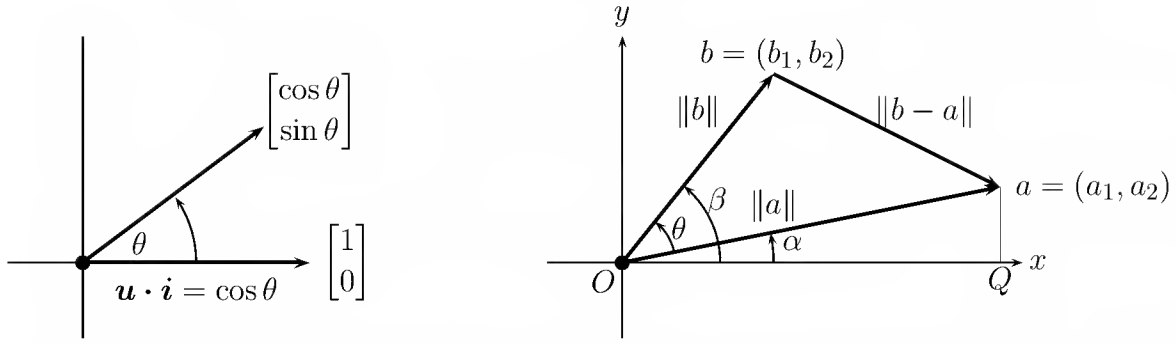

Figure 3.6: The cosine of the angle $\theta = \beta - \alpha$ using inner products.

For the angle β, the sine is $b_2/\|b\|$ and the cosine is $b_1/\|b\|$. The cosine of $\theta = \beta - \alpha$ comes from an identity that no one could forget:

$$\textbf{Cosine formula} \qquad \cos\theta = \cos\beta\cos\alpha + \sin\beta\sin\alpha = \frac{a_1b_1 + a_2b_2}{\|a\|\|b\|}. \tag{1}$$

The numerator in this formula is exactly the inner product of a and b. It gives the relationship between $a^{\mathrm{T}}b$ and $\cos\theta$:

3G The cosine of the angle between any nonzero vectors a and b is

$$\textbf{Cosine of } \theta \qquad \cos\theta = \frac{a^{\mathrm{T}}b}{\|a\|\|b\|}. \tag{2}$$

This formula is dimensionally correct; if we double the length of b, then both numerator and denominator are doubled, and the cosine is unchanged. Reversing the sign of b, on the other hand, reverses the sign of $\cos\theta$—and changes the angle by 180°.

There is another law of trigonometry that leads directly to the same result. It is not so unforgettable as the formula in equation (1), but it relates the lengths of the sides of any triangle:

$$\textbf{Law of Cosines} \qquad \|b-a\|^2 = \|b\|^2 + \|a\|^2 - 2\|b\|\|a\|\cos\theta. \tag{3}$$

When θ is a right angle, we are back to Pythagoras: $\|b-a\|^2 = \|b\|^2 + \|a\|^2$. For any angle θ, the expression $\|b-a\|^2$ is $(b-a)^{\mathrm{T}}(b-a)$, and equation (3) becomes

$$b^{\mathrm{T}}b - 2a^{\mathrm{T}}b + a^{\mathrm{T}}a = b^{\mathrm{T}}b + a^{\mathrm{T}}a - 2\|b\|\|a\|\cos\theta.$$

Canceling $b^{\mathrm{T}}b$ and $a^{\mathrm{T}}a$ on both sides of this equation, you recognize formula (2) for the cosine: $a^{\mathrm{T}}b = \|a\|\|b\|\cos\theta$. In fact, this proves the cosine formula in n dimensions, since we only have to worry about the plane triangle Oab.

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

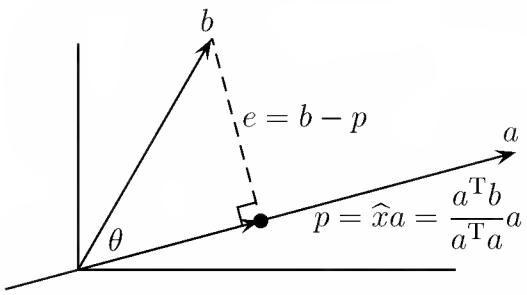

Figure 3.7: The projection p of b onto a, with $\cos\theta = \dfrac{Op}{Ob} = \dfrac{a^{\mathrm{T}}b}{\|a\|\|b\|}$.

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 $p = \widehat{x}a$:

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

This allows us to redraw Figure 3.5 with a correct formula for p (Figure 3.7).

This leads to the Schwarz inequality in equation (6), which is the most important inequality in mathematics. A special case is the fact that arithmetic means $\frac{1}{2}(x+y)$ are larger than geometric means $\sqrt{xy}$. (It is also equivalent—see Problem 1 at the end of this section—to the triangle inequality for vectors.) The Schwarz inequality seems to come almost accidentally from the statement that $\|e\|^2 = \|b - p\|^2$ in Figure 3.7 cannot be negative:

$$\left\| b - \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}a \right\|^2 = b^{\mathrm{T}}b - 2\frac{(a^{\mathrm{T}}b)^2}{a^{\mathrm{T}}a} + \left(\frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}\right)^2 a^{\mathrm{T}}a = \frac{(b^{\mathrm{T}}b)(a^{\mathrm{T}}a) - (a^{\mathrm{T}}b)^2}{(a^{\mathrm{T}}a)} \geq 0.$$

This tells us that $(b^{\mathrm{T}}b)(a^{\mathrm{T}}a) \geq (a^{\mathrm{T}}b)^2$—and then we take square roots:

3I All vectors a and b satisfy the ***Schwarz inequality***, which is $|\cos\theta| \leq 1$ in $\mathbf{R}^n$:

$$\textbf{Schwarz inequality} \qquad |a^{\mathrm{T}}b| \leq \|a\|\|b\|. \tag{6}$$

According to formula (2), the ratio between $a^{\mathrm{T}}b$ and $\|a\|\|b\|$ is exactly $|\cos\theta|$. Since all cosines lie in the interval $-1 \leq \cos\theta \leq 1$, this gives another proof of equation (6): *the Schwarz inequality is the same as* $|\cos\theta| \leq 1$. In some ways that is a more easily understood proof, because cosines are so familiar. Either proof is all right in $\mathbf{R}^n$, but