22. Project the vector $b = (1, 1)$ onto the lines through $a_1 = (1, 0)$ and $a_2 = (1, 2)$. Draw the projections p_1 and p_2 and add $p_1 + p_2$. The projections do not add to b because the a's are not orthogonal.

23. In Problem 22, the projection of b onto the *plane* of a_1 and a_2 will equal b. Find $P = A(A^TA)^{-1}A^T$ for $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$.

24. Project $b = (1, 0, 0)$ onto the lines through a_1 and a_2 in Problem 21 and also onto $a_3 = (2, -1, 2)$. Add the three projections $p_1 + p_2 + p_3$.

25. Project $a_1 = (1, 0)$ onto $a_2 = (1, 2)$. Then project the result back onto a_1. Draw these projections and multiply the projection matrices P_1P_2: Is this a projection?

26. Continuing Problems 21, 24 find the projection matrix P_3 onto $a_3 = (2, -1, 2)$. Verify that $P_1 + P_2 + P_3 = I$. The basis a_1, a_2, a_3 is orthogonal!

3.3 PROJECTIONS AND LEAST SQUARES

Up to this point, $Ax = b$ 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	$2x = b_1$
than unknowns—	$3x = b_2$
no solution?	$4x = b_3$.

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 unsolvability, inconsistent equations arise all the time in practice. They have to be solved! 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:

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

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

Least-squares solution $$\widehat{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{a^Tb}{a^Ta}.$$

You recognize a^Tb in the numerator and a^Ta in the denominator.

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 $\widehat{x}$, 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, $\widehat{x} = (a_1b_1 + \cdots + a_mb_m)/(a_1^2 + \cdots + a_m^2)$, that geometry did earlier:

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

You see that we keep coming back to the geometrical interpretation of a least-squares problem—to minimize a distance. By setting the derivative of E^2 to zero, calculus confirms the geometry of the previous section. *The error vector e connecting b to p must be perpendicular to a*:

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

As a side remark, notice the degenerate case $a = 0$. All multiples of a are zero, and the line is only a point. Therefore $p = 0$ is the only candidate for the projection. But the formula for $\widehat{x}$ becomes a meaningless $0/0$, and correctly reflects the fact that $\widehat{x}$ is completely undetermined. All values of x give the same error $E = \|0x - b\|$, so E^2 is a horizontal line instead of a parabola. The "pseudoinverse" assigns the definite value $\widehat{x} = 0$, which is a more "symmetric" choice than any other number.

Least-Squares Problems with Several Variables

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.

We may use geometry or calculus to determine $\widehat{x}$. In n dimensions, we prefer the appeal of geometry; p must be the "projection of b onto the column space." ***The error vector $e = b - A\widehat{x}$ must be perpendicular to that space*** (Figure 3.8). Finding $\widehat{x}$ and the projection $p = A\widehat{x}$ 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 $e = b - A\widehat{x}$ must be in the nullspace of A^{T}:

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