# Section 7.8 Best Approximation and Least Squares

## Minimum Distance Problems

We will be concerned here with the following problem.

**The Minimum Distance Problem in $R^n$** Given a subspace $W$ and a vector $\mathbf{b}$ in $R^n$, find a vector $\hat{\mathbf{w}}$ in $W$ that is closest to $\mathbf{b}$ in the sense that $\|\mathbf{b} - \hat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$ for every vector $\mathbf{w}$ in $W$ that is distinct from $\hat{\mathbf{w}}$. Such a vector $\hat{\mathbf{w}}$, if it exists, is called a ***best approximation to* b *from* W** (Figure 7.8.1).

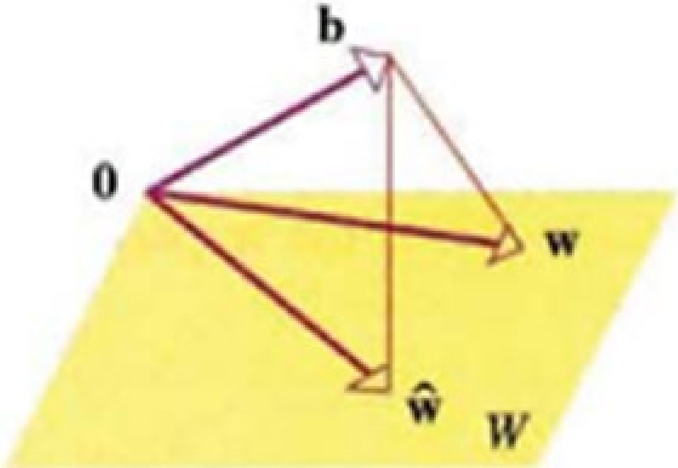


Figure 7.8.1

$\hat{\mathbf{w}}$ is closer to $\mathbf{b}$ than any other vector $\mathbf{w}$ in $W$.

To motivate a method for solving the minimum distance problem, let us focus on $R^3$. We know from geometry that if $\mathbf{b}$ is a point in $R^3$ and $W$ is a plane through the origin, then the point $\hat{\mathbf{w}}$ in $W$ that is closest to $\mathbf{b}$ is obtained by dropping a perpendicular from $\mathbf{b}$ to $W$; that is, $\hat{\mathbf{w}} = \text{proj}_W \mathbf{b}$.

It follows from this that the distance from $\mathbf{b}$ to $W$ is $d = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|$, or equivalently, $d = \|\text{proj}_{W^\perp} \mathbf{b}\|$, where $W^\perp$ is the line through the origin that is perpendicular to $W$ (Figure 7.8.2).

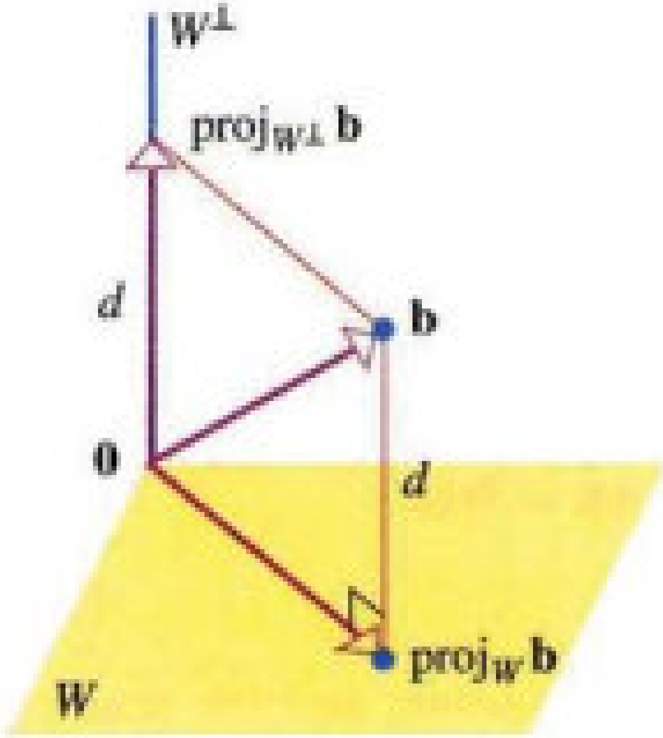


Figure 7.8.2

**EXAMPLE 1** Distance from a Point to a Plane in $R^3$

Use an appropriate orthogonal projection to find a formula for the distance $d$ from the point $(x_0, y_0, z_0)$ to the plane $ax + by + cz = 0$.

***Solution*** Let $\mathbf{b} = (x_0, y_0, z_0)$, let $W$ be the given plane, and let $l$ be the line through the origin that is perpendicular to $W$ (i.e., $l$ is $W^{\perp}$). The line $l$ is spanned by the normal $\mathbf{n} = (a, b, c)$ and hence it follows from Formula (14) of Section 7.7 that

$$d = \|\text{proj}_{W^{\perp}}\mathbf{b}\| = \|\text{proj}_{\mathbf{n}}\mathbf{b}\| = \frac{|\mathbf{n} \cdot \mathbf{b}|}{\|\mathbf{n}\|}$$

Substituting the components for $\mathbf{n}$ and $\mathbf{b}$ into this formula yields

$$d = \frac{|(a, b, c) \cdot (x_0, y_0, z_0)|}{\|(a, b, c)\|} = \frac{|ax_0 + by_0 + cz_0|}{\sqrt{a^2 + b^2 + c^2}} \tag{1}$$

Thus, for example, the distance from the point $(-1, 5, 4)$ to the plane $x - 2y + 3z = 0$ is

$$d = \frac{|ax_0 + by_0 + cz_0|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|(1)(-1) + (-2)(5) + (3)(4)|}{\sqrt{1^2 + (-2)^2 + 3^2}} = \frac{1}{\sqrt{14}}$$ ■

$$\|\text{proj}_{\mathbf{a}}\mathbf{x}\| = \frac{|\mathbf{x} \cdot \mathbf{a}|}{\|\mathbf{a}\|} \tag{14}$$

In light of the preceding discussion, the following theorem should not be surprising.

**Theorem 7.8.1** (*Best Approximation Theorem*) *If $W$ is a subspace of $R^n$ and $\mathbf{b}$ is a point in $R^n$, then there is a unique best approximation to $\mathbf{b}$ from $W$, namely $\hat{\mathbf{w}} = \text{proj}_W \mathbf{b}$.*

there is a right angle between them.

*Proof* For every vector $\mathbf{w}$ in $W$ we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W \mathbf{b}) + (\text{proj}_W \mathbf{b} - \mathbf{w})$$

The two terms on the right side of this equation are orthogonal (since the first term is in $W^\perp$, and the second term, being a difference of vectors in $W$, is in $W$). Thus, we can apply the theorem of Pythagoras to write

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 + \|\text{proj}_W \mathbf{b} - \mathbf{w}\|^2$$

If $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, then the second term on the right side of this equation is positive and hence

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2$$

This implies that $\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$ if $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, which tells us that $\hat{\mathbf{w}} = \text{proj}_W \mathbf{b}$ is a best approximation to $\mathbf{b}$ from $W$; we leave the proof of the uniqueness as an exercise. ■

smallest distance

Motivated by Figure 7.8.2, we define the ***distance*** from a point $\mathbf{b}$ to a subspace $W$ in $R^n$ to be

$$d = \|\mathbf{b} - \text{proj}_W \mathbf{b}\| \qquad \textbf{[Distance from b to } W\textbf{]} \tag{2}$$

or equivalently,

$$d = \|\text{proj}_{W^\perp} \mathbf{b}\| \qquad \textbf{[Distance from b to } W\textbf{]} \tag{3}$$

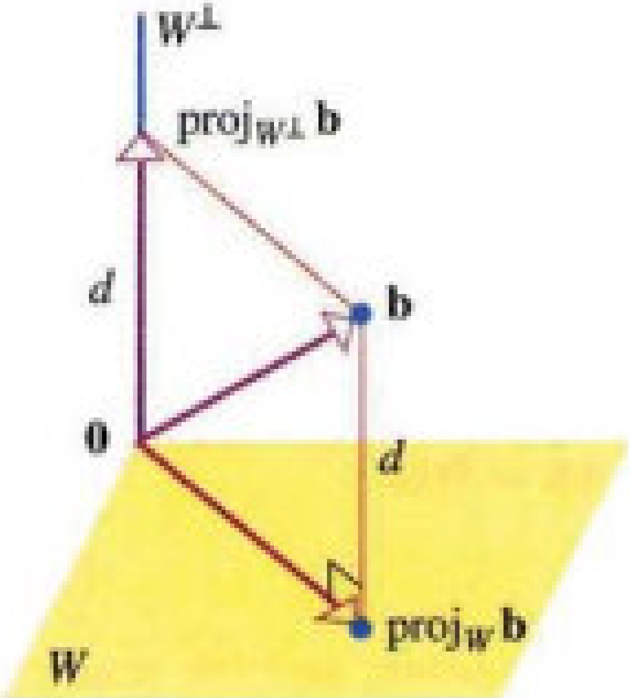


Figure 7.8.2

The following example extends the result in Example 1 to $R^n$.

## EXAMPLE 2 Distance from a Point to a Hyperplane

Find a formula for the distance $d$ from a point $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $R^n$ to the hyperplane $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$.

***Solution*** Denote the hyperplane by $W$. This hyperplane is the orthogonal complement of $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, so Theorem 7.4.6 implies that $W^\perp = \text{span}\{\mathbf{a}\}$. Thus, Formula (3) above and Formula (14) of Section 7.7 imply that

$$d = \|\text{proj}_{W^\perp}\mathbf{b}\| = \|\text{proj}_{\mathbf{a}}\mathbf{b}\| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{\|\mathbf{a}\|} = \frac{|a_1b_1 + a_2b_2 + \cdots + a_nb_n|}{\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}} \tag{4}$$

With the appropriate change in notation, this reduces to Formula (1) in $R^3$. ■

## LEAST SQUARES SOLUTIONS OF LINEAR SYSTEMS

There are many applications in which a linear system $A\mathbf{x} = \mathbf{b}$ should be consistent on theoretical grounds but fails to be so because of measurement errors in the entries of $A$ or $\mathbf{b}$.

In such cases, a common scientific procedure is to look for vectors that come as close as possible to being solutions in the sense that they minimize $\|\mathbf{b} - A\mathbf{x}\|$. Accordingly, we make the following definition.

**Definition 7.8.2** If $A$ is an $m \times n$ matrix and $\mathbf{b}$ is a vector in $R^m$, then a vector $\hat{\mathbf{x}}$ in $R^n$ is called a ***best approximate solution*** or a ***least squares solution*** of $A\mathbf{x} = \mathbf{b}$ if

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\| \tag{5}$$

for all $\mathbf{x}$ in $R^n$. The vector $\mathbf{b} - A\hat{\mathbf{x}}$ is called the ***least squares error vector***, and the scalar $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the ***least squares error***.

$e_m$

**REMARK** To understand the terminology in this definition, let $\mathbf{b} - A\mathbf{x} = (e_1, e_2, \ldots, e_n)$. The components of this vector can be interpreted as the errors that result in the individual components when $\mathbf{x}$ is used as an approximate solution. Since a best approximate solution minimizes

$$\|\mathbf{b} - A\mathbf{x}\| = \left(e_1^2 + e_2^2 + \cdots + e_n^2\right)^{1/2} \tag{6}$$

$e_m^2$

this solution also minimizes $e_1^2 + e_2^2 + \cdots + e_n^2$, which is the sum of the squares of the errors in the components, and hence the term *least squares solution*.

→ Be careful: $e_1$, $e_2$, ..., $e_m$ above are components of **b** – A**x** ;
they are not components of the standard unit vectors.

## FINDING LEAST SQUARES SOLUTIONS OF LINEAR SYSTEMS

Our next objective is to develop a method for finding least squares solutions of a linear system $A\mathbf{x} = \mathbf{b}$ of $m$ equations in $n$ unknowns. To start, observe that $A\mathbf{x}$ is in the column space of $A$ for all $\mathbf{x}$ in $R^n$, so $\|\mathbf{b} - A\mathbf{x}\|$ is minimized when

$$A\mathbf{x} = \text{proj}_{\text{col}(A)}\mathbf{b} \tag{7}$$

(Figure 7.8.3).

Since $\text{proj}_{\text{col}(A)}\mathbf{b}$ is a vector in the column space of $A$, system (7) is consistent and its solutions are the least squares solutions of $A\mathbf{x} = \mathbf{b}$. Thus, we are guaranteed that *every linear system $A\mathbf{x} = \mathbf{b}$ has at least one least squares solution.*

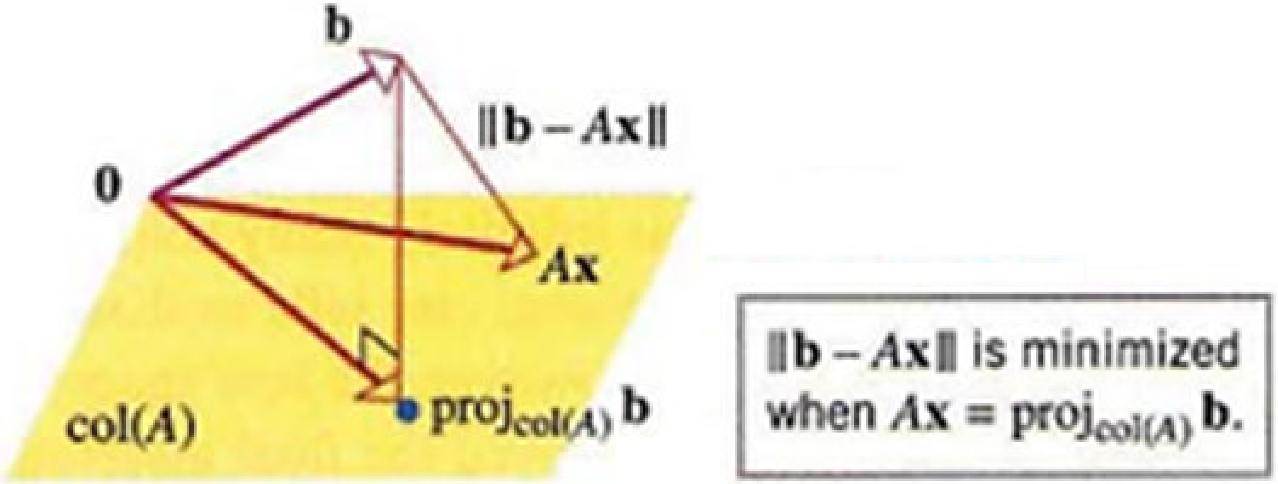


Figure 7.8.3

$$A\mathbf{x} = \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b} \tag{7}$$

As a practical matter, least squares solutions are rarely obtained by solving (7), since this equation can be rewritten in an alternative form that eliminates the need to calculate the orthogonal projection. To see how this can be accomplished, rewrite (7) as

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b} \tag{8}$$

and multiply both sides of this equation by $A^T$ to obtain

$$A^T(\mathbf{b} - A\mathbf{x}) = A^T(\mathbf{b} - \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b}) \tag{9}$$

Since the orthogonal complement of $\mathrm{col}(A)$ is $\mathrm{null}(A^T)$, it follows from Formula (24) of Section 7.7 that

$$\mathbf{b} - \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b} = \mathrm{proj}_{\mathrm{null}(A^T)}\mathbf{b}$$

This implies that $\mathbf{b} - \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b}$ is in the null space of $A^T$, and hence that

$$A^T(\mathbf{b} - \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b}) = \mathbf{0}$$

$$\boxed{\mathbf{x} = \mathrm{proj}_W\mathbf{x} + \mathrm{proj}_{W^\perp}\mathbf{x} \qquad (24)}$$

Thus, (9) can be rewritten as $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$ or, alternatively, as

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{10}$$

This is called the ***normal equation*** or ***normal system*** associated with $A\mathbf{x} = \mathbf{b}$. The individual equations in (10) are called the ***normal equations*** associated with $A\mathbf{x} = \mathbf{b}$.

Using this terminology, the problem of finding least squares solutions of $A\mathbf{x} = \mathbf{b}$ has been reduced to solving the associated normal system exactly. The following theorem provides the basic facts about solutions of the normal equation.

**Theorem 7.8.3**

(*a*) *The least squares solutions of a linear system* $A\mathbf{x} = \mathbf{b}$ *are the exact solutions of the normal equation*

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{11}$$

(*b*) *If A has full column rank, the normal equation has a unique solution, namely*

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b} \tag{12}$$

(c) *If A does not have full column rank, then the normal equation has infinitely many solutions, but there is a unique solution in the row space of A. Moreover, among all solutions of the normal equation, the solution in the row space of A has the smallest norm.*

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{11}$$

***Proof (a)*** We have already established that every least squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies (11). Conversely, if $\mathbf{x}$ satisfies (11), then this vector also satisfies the equation

$$A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$$

This implies that $\mathbf{b} - A\mathbf{x}$ is orthogonal to the row vectors of $A^T$, and hence to the column vectors of $A$, and hence to the column space of $A$. It follows from this that the equation

$$\mathbf{b} = A\mathbf{x} + (\mathbf{b} - A\mathbf{x})$$

expresses $\mathbf{b}$ as the sum of a vector in $\operatorname{col}(A)$ and a vector orthogonal to $\operatorname{col}(A)$, which implies that $A\mathbf{x} = \operatorname{proj}_{\operatorname{col}(A)}\mathbf{b}$ by Theorem 7.7.4. Thus, $\mathbf{x}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$.

***Proof (b)*** If $A$ has full column rank, then Theorem 7.5.10 implies that $A^TA$ is invertible, so (12) is the unique solution of (11).

***Proof* (c)** If $A$ does not have full column rank, then $A^TA$ is not invertible (Theorem 7.5.10), so (11) is a consistent linear system whose coefficient matrix does not have full column rank. This being the case, it follows from part (*b*) of Theorem 7.7.7 that (11) has infinitely many solutions but has a unique solution in the row space of $A^TA$.

Moreover, that theorem also tells us that the solution in the row space of $A^TA$ is the solution with smallest norm. However, the row space of $A^TA$ is the same as the row space of $A$ (Theorem 7.5.8), so we have proved the final assertion of the theorem. ■

**Theorem 7.7.7** *Suppose that $A$ is an $m \times n$ matrix and $\mathbf{b}$ is in the column space of $A$.*

(*a*) *If $A$ has full column rank, then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution, and that solution is in the row space of $A$.*

(*b*) *If $A$ does not have full column rank, then the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions, but there is a unique solution in the row space of $A$. Moreover, among all solutions of the system, the solution in the row space of $A$ has the smallest norm.*

## EXAMPLE 3 Least Squares Solutions

Find the least squares solutions of the linear system

$$\begin{aligned} x_1 - \quad x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned}$$

***Solution*** The matrix form of the system is $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

Since the columns of $A$ are not scalar multiples of one another, the matrix has full column rank. Thus, it follows from Theorem 7.8.3 that there is a unique least squares solution given by Formula (12). We leave it for you to confirm that

$$A^TA = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}$$

$$(A^TA)^{-1} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{21}{285} & \frac{3}{285} \\ \frac{3}{285} & \frac{14}{285} \end{bmatrix}$$

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \begin{bmatrix} \frac{21}{285} & \frac{3}{285} \\ \frac{3}{285} & \frac{14}{285} \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{21}{285} & \frac{3}{285} \\ \frac{3}{285} & \frac{14}{285} \end{bmatrix} \begin{bmatrix} 1 \\ 10 \end{bmatrix} = \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix}$$

Thus, the least squares solution is

$$x_1 = \frac{17}{95}, \quad x_2 = \frac{143}{285} \qquad \blacksquare$$

## ORTHOGONALITY PROPERTY OF LEAST SQUARES ERROR VECTORS

Before considering another example, it will be helpful to develop some of the properties of least squares error vectors. For this purpose, consider a linear system $A\mathbf{x} = \mathbf{b}$, and recall from Formula (30) of Section 7.7 that $\mathbf{b}$ can be written as

$$\mathbf{b} = \operatorname{proj}_{\operatorname{col}(A)}\mathbf{b} + \operatorname{proj}_{\operatorname{null}(A^T)}\mathbf{b}$$

from which it follows that

$$\mathbf{b} - A\mathbf{x} = (\operatorname{proj}_{\operatorname{col}(A)}\mathbf{b} - A\mathbf{x}) + \operatorname{proj}_{\operatorname{null}(A^T)}\mathbf{b} \tag{13}$$

However, we know from (7) that $\mathbf{x}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if

$$\operatorname{proj}_{\operatorname{col}(A)}\mathbf{b} - A\mathbf{x} = \mathbf{0}$$

which, together with (13), implies that $\hat{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if

$$\mathbf{b} - A\hat{\mathbf{x}} = \operatorname{proj}_{\operatorname{null}(A^T)}\mathbf{b} \tag{14}$$

$$\boxed{\mathbf{b} = \mathbf{b}_{\operatorname{col}(A)} + \mathbf{b}_{\operatorname{null}(A^T)} \qquad (30)}$$

This shows that every least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ has the same error vector, namely

$$\text{least squares error vector} = \mathbf{b} - A\hat{\mathbf{x}} = \text{proj}_{\text{null}(A^T)}\mathbf{b} \tag{15}$$

Thus, the least squares error can be written as

$$\text{least squares error} = \|\mathbf{b} - A\hat{\mathbf{x}}\| = \|\text{proj}_{\text{null}(A^T)}\mathbf{b}\| \tag{16}$$

Moreover, since the least squares error vector lies in $\text{null}(A^T)$, and since this space is orthogonal to $\text{col}(A)$, we have also established the following result.

**Theorem 7.8.4** *A vector $\hat{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of $A$.*

## EXAMPLE 4 Least Squares Solutions and Their Error Vector

Find the least squares solutions and least squares error for the linear system

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &= 2 \\ x_1 - 4x_2 + 3x_3 &= -2 \\ x_1 + 10x_2 - 7x_3 &= 1 \end{aligned}$$

***Solution*** The matrix form of the system is $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

Since it is not evident by inspection whether $A$ has full column rank (in fact it does not), we will simply proceed by solving the associated normal system $A^TA\mathbf{x} = A^T\mathbf{b}$. We leave it for you to confirm that

$$A^TA = \begin{bmatrix} 11 & 12 & -7 \\ 12 & 120 & -84 \\ -7 & -84 & 59 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 22 \\ -15 \end{bmatrix}$$

Thus, the augmented matrix for the normal system is

$$\left[\begin{array}{rrr|r} 11 & 12 & -7 & 5 \\ 12 & 120 & -84 & 22 \\ -7 & -84 & 59 & -15 \end{array}\right]$$

We leave it for you to confirm that the reduced row echelon form of this matrix is

$$\left[\begin{array}{rrr|r} 1 & 0 & \frac{1}{7} & \frac{2}{7} \\ 0 & 1 & -\frac{5}{7} & \frac{13}{84} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Thus, there are infinitely many least squares solutions, and they are given by

$$x_1 = \tfrac{2}{7} - \tfrac{1}{7}t$$

$$x_2 = \tfrac{13}{84} + \tfrac{5}{7}t$$

$$x_3 = t$$

As a check, let us verify that all least squares solutions produce the same least squares error vector and the same least squares error. To see that this is so, we first compute

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \begin{bmatrix} \frac{2}{7} - \frac{1}{7}t \\ \frac{13}{84} + \frac{5}{7}t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ -\frac{1}{3} \\ \frac{11}{6} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ -\frac{5}{3} \\ -\frac{5}{6} \end{bmatrix}$$

Since $\mathbf{b} - A\mathbf{x}$ does not depend on $t$, all least squares solutions produce the same error vector. The resulting least squares error is

$$\|\mathbf{b} - A\mathbf{x}\| = \sqrt{\left(\frac{5}{6}\right)^2 + \left(-\frac{5}{3}\right)^2 + \left(-\frac{5}{6}\right)^2} = \frac{5}{6}\sqrt{6}$$

We leave it for you to confirm that the error vector is orthogonal to the column vectors of the matrix

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix}$$

in agreement with Theorem 7.8.4. ■

## FITTING A CURVE TO EXPERIMENTAL DATA

A common problem in experimental work is to obtain a mathematical relationship between two variables $x$ and $y$ by "fitting" a curve $y = f(x)$ of a specified form to a set of points in the plane that correspond to experimentally determined values of $x$ and $y$, say

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

The curve $y = f(x)$ is called a ***mathematical model*** for the data. The form of the function $f$ is sometimes determined by a physical theory and sometimes by the pattern of the data.

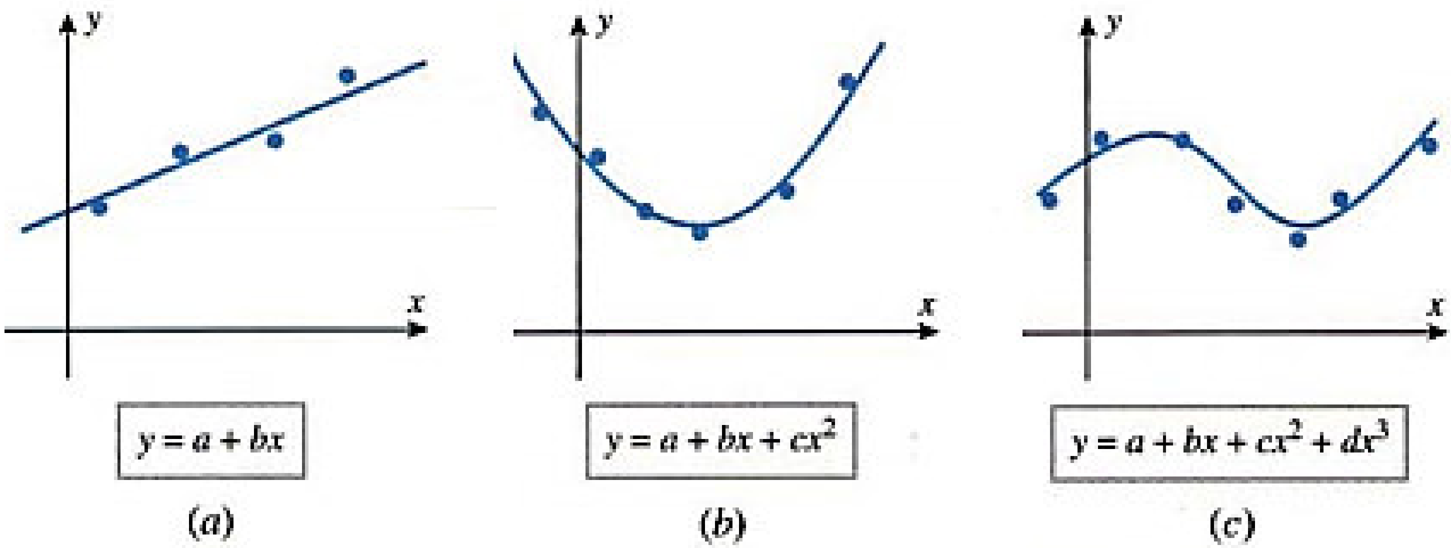


**Figure 7.8.5**

For example, Figure 7.8.5 shows some data patterns that suggest polynomial models. Once the form of the function has been decided on, the idea is to determine values of coefficients that make the graph of the function fit the data *as closely as possible.* In this section we will be concerned exclusively with polynomial models, but we will discuss some other kinds of mathematical models in the exercises.

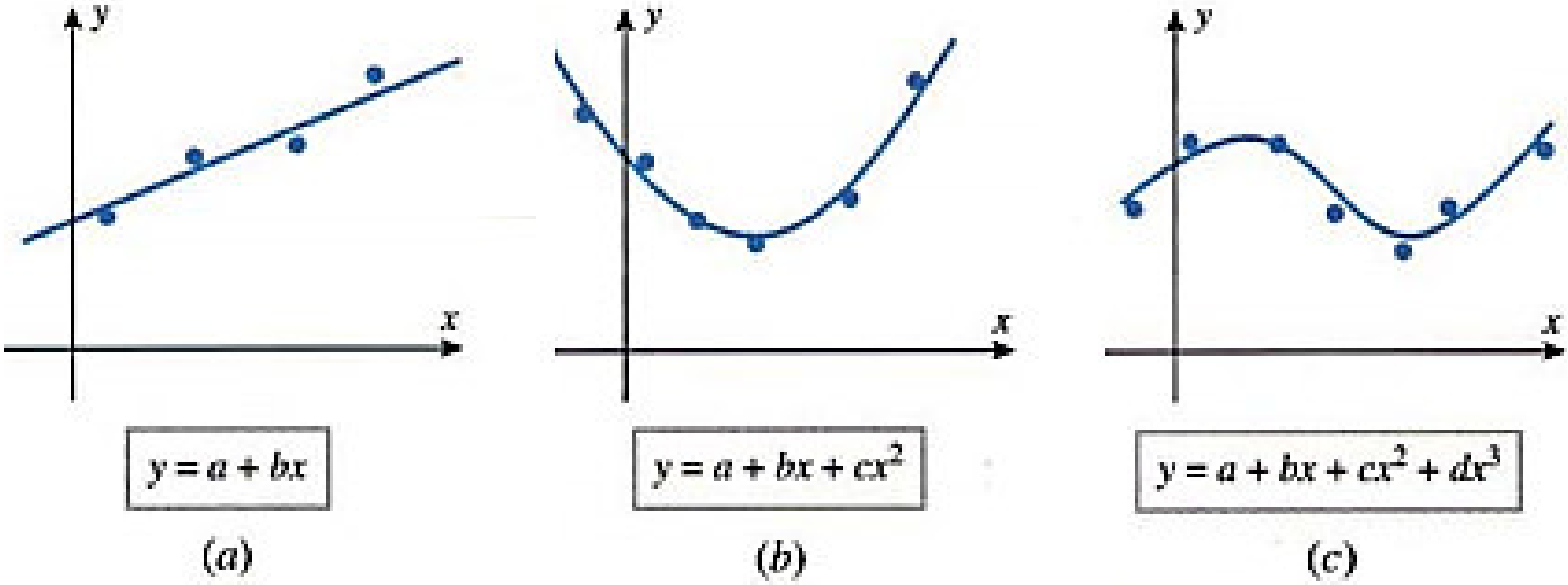


**Figure 7.8.5**

We will begin with linear models (polynomials of degree 1). For this purpose let $x$ and $y$ be given variables, and assume that there is evidence to suggest that the variables are related by a linear equation

$$y = a + bx \tag{17}$$

where $a$ and $b$ are to be determined from two or more data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

If the data happen to fall exactly on a line, then the coordinates of each point will satisfy (17), and the unknown coefficients will be a solution of the linear system

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned} \tag{18}$$

We can write this system in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{19}$$

or more compactly as

$$M\mathbf{v} = \mathbf{y} \tag{20}$$

where

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{21}$$

If there are measurement errors in the data, then the data points will typically not lie on a line, and (20) will be inconsistent. In this case we look for a least squares approximation to the values of $a$ and $b$ by solving the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y} \tag{22}$$

If the $x$-coordinates of the data points are not all the same, then $M$ will have rank 2 (full column rank), and the normal system will have a unique least squares solution

$$\mathbf{v} = (M^T M)^{-1} M^T\mathbf{y} \tag{23}$$

**Theorem 7.8.3**

(*a*) *The least squares solutions of a linear system $A\mathbf{x} = \mathbf{b}$ are the exact solutions of the normal equation*

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{11}$$

(*b*) *If $A$ has full column rank, the normal equation has a unique solution, namely*

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b} \tag{12}$$

The line $y = a + bx$ that results from this solution is called the ***least squares line of best fit*** to the data or, alternatively, the ***regression line***. Referring to the equations in (18), we see that this line minimizes

$$S = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2 \qquad (24)$$

The differences in Equation (24) are called ***residuals***, so, in words, *the least squares line of best fit minimizes the sum of the squares of the residuals* (Figure 7.8.6).

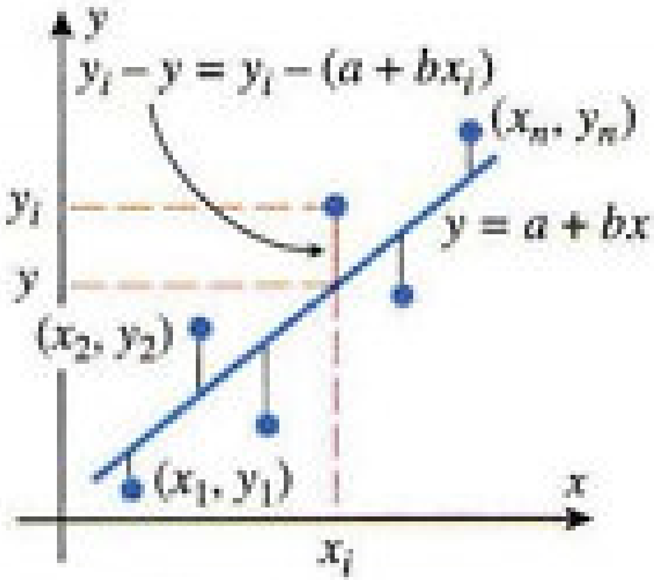


**Figure 7.8.6**

**EXAMPLE 5** Find the least squares line of best fit to the four points (0, 1), (1, 3), (2, 4), and (3, 4).

***Solution*** The first step is to use the data to build the matrices $M$ and $\mathbf{y}$ in (21). This yields

$$M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$$

Since the $x$-coordinates of the data points are not all the same, the normal system has a unique solution, and the coefficients for the least squares line of best fit can be obtained from Formula (23). We leave it for you to confirm that

$$M^T M = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix} \quad \text{and} \quad (M^T M)^{-1} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{2}{10} \end{bmatrix}$$

Thus, applying Formula (23) yields

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{2}{10} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

Thus, the approximate values of $a$ and $b$ are $a = 1.5$ and $b = 1$, and the least squares straight line fit to the data is $y = 1.5 + x$. This line and the data points are shown in Figure 7.8.7.

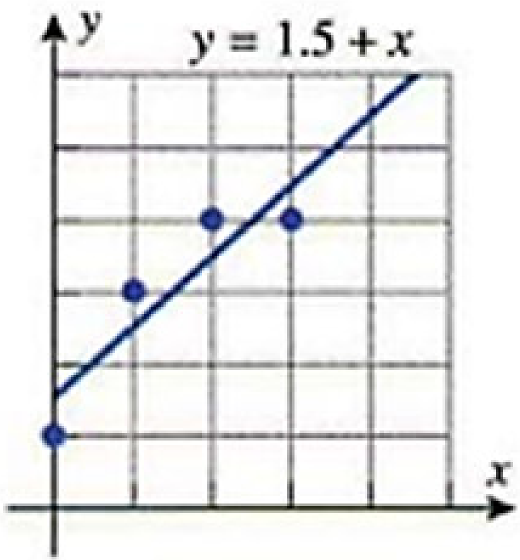


Figure 7.8.7