

CHAPTER 12

SOLVING LINEAR EQUATIONS

12.1 Least-Squares Analysis

Consider a system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \geq n$, and $\text{rank } \mathbf{A} = n$. Note that the number of unknowns, n , is no larger than the number of equations, m . If $\mathbf{b}$ does not belong to the range of $\mathbf{A}$, that is, if $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, then this system of equations is said to be *inconsistent* or *overdetermined*. In this case there is no solution to the above set of equations. Our goal then is to find the vector (or vectors) $\mathbf{x}$ minimizing $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$. This problem is a special case of the nonlinear least-squares problem discussed in Section 9.4.

Let $\mathbf{x}^*$ be a vector that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$; that is, for all $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \geq \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2.$$

We refer to the vector $\mathbf{x}^*$ as a *least-squares solution* to $\mathbf{A}\mathbf{x} = \mathbf{b}$. In the case where $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution, then the solution is a least-squares solution.

Otherwise, a least-squares solution minimizes the norm of the difference between the left- and right-hand sides of the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. To characterize least-squares solutions, we need the following lemma.

Lemma 12.1 *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$. Then, $\text{rank } \mathbf{A} = n$ if and only if $\text{rank } \mathbf{A}^\top \mathbf{A} = n$ (i.e., the square matrix $\mathbf{A}^\top \mathbf{A}$ is nonsingular).* $\square$

Proof. $\Rightarrow$: Suppose that $\text{rank } \mathbf{A} = n$. To show $\text{rank } \mathbf{A}^\top \mathbf{A} = n$, it is equivalent to show $\mathcal{N}(\mathbf{A}^\top \mathbf{A}) = \{\mathbf{0}\}$. To proceed, let $\mathbf{x} \in \mathcal{N}(\mathbf{A}^\top \mathbf{A})$; that is, $\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{0}$. Therefore,

$$\|\mathbf{A}\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} = 0,$$

which implies that $\mathbf{A}\mathbf{x} = \mathbf{0}$. Because $\text{rank } \mathbf{A} = n$, we have $\mathbf{x} = \mathbf{0}$.

$\Leftarrow$: Suppose that $\text{rank } \mathbf{A}^\top \mathbf{A} = n$; that is, $\mathcal{N}(\mathbf{A}^\top \mathbf{A}) = \{\mathbf{0}\}$. To show $\text{rank } \mathbf{A} = n$, it is equivalent to show that $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$. To proceed, let $\mathbf{x} \in \mathcal{N}(\mathbf{A})$; that is, $\mathbf{A}\mathbf{x} = \mathbf{0}$. Then, $\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{0}$, and hence $\mathbf{x} = \mathbf{0}$. $\blacksquare$

Recall that we assume throughout that $\text{rank } \mathbf{A} = n$. By Lemma 12.1 we conclude that $(\mathbf{A}^\top \mathbf{A})^{-1}$ exists. The following theorem characterizes the least-squares solution.

Theorem 12.1 *The unique vector $\mathbf{x}^*$ that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ is given by the solution to the equation $\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$; that is, $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.* $\square$

Proof. Let $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$. First observe that

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 &= \|\mathbf{A}(\mathbf{x} - \mathbf{x}^*) + (\mathbf{A}\mathbf{x}^* - \mathbf{b})\|^2 \\ &= (\mathbf{A}(\mathbf{x} - \mathbf{x}^*) + (\mathbf{A}\mathbf{x}^* - \mathbf{b}))^\top (\mathbf{A}(\mathbf{x} - \mathbf{x}^*) + (\mathbf{A}\mathbf{x}^* - \mathbf{b})) \\ &= \|\mathbf{A}(\mathbf{x} - \mathbf{x}^*)\|^2 + \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2 + 2[\mathbf{A}(\mathbf{x} - \mathbf{x}^*)]^\top (\mathbf{A}\mathbf{x}^* - \mathbf{b}). \end{aligned}$$

We now show that the last term in this equation is zero. Indeed, substituting the expression above for $\mathbf{x}^*$,

$$\begin{aligned} [\mathbf{A}(\mathbf{x} - \mathbf{x}^*)]^\top (\mathbf{A}\mathbf{x}^* - \mathbf{b}) &= (\mathbf{x} - \mathbf{x}^*)^\top \mathbf{A}^\top [\mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top - \mathbf{I}_n] \mathbf{b} \\ &= (\mathbf{x} - \mathbf{x}^*)^\top [(\mathbf{A}^\top \mathbf{A})(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top - \mathbf{A}^\top] \mathbf{b} \\ &= (\mathbf{x} - \mathbf{x}^*)^\top (\mathbf{A}^\top - \mathbf{A}^\top) \mathbf{b} \\ &= 0. \end{aligned}$$

Hence,

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \|\mathbf{A}(\mathbf{x} - \mathbf{x}^*)\|^2 + \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2.$$

If $\mathbf{x} \neq \mathbf{x}^*$, then $\|\mathbf{A}(\mathbf{x} - \mathbf{x}^*)\|^2 > 0$, because $\text{rank } \mathbf{A} = n$. Thus, if $\mathbf{x} \neq \mathbf{x}^*$, we have

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 > \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2.$$

Thus, $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$ is the unique minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$. $\blacksquare$

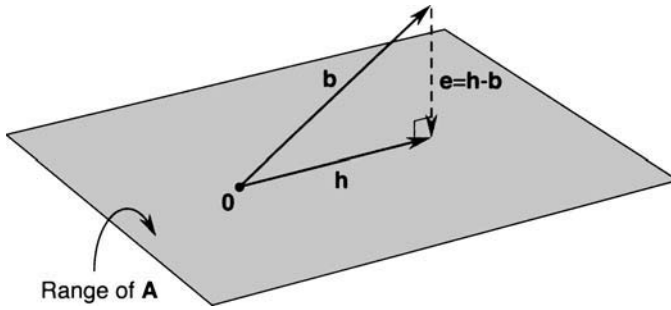


Figure 12.1 Orthogonal projection of $\mathbf{b}$ on the subspace $\mathcal{R}(\mathbf{A})$.

We now give a geometric interpretation of the Theorem 12.1. First note that the columns of $\mathbf{A}$ span the range $\mathcal{R}(\mathbf{A})$ of $\mathbf{A}$, which is an n -dimensional subspace of $\mathbb{R}^m$. The equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ lies in this n -dimensional subspace $\mathcal{R}(\mathbf{A})$. If $m = n$, then $\mathbf{b} \in \mathcal{R}(\mathbf{A})$ always, and the solution is $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$. Suppose now that $m > n$. Intuitively, we would expect the “likelihood” of $\mathbf{b} \in \mathcal{R}(\mathbf{A})$ to be small, because the subspace spanned by the columns of $\mathbf{A}$ is very “thin.” Therefore, let us suppose that $\mathbf{b}$ does not belong to $\mathcal{R}(\mathbf{A})$. We wish to find a point $\mathbf{h} \in \mathcal{R}(\mathbf{A})$ that is “closest” to $\mathbf{b}$. Geometrically, the point $\mathbf{h}$ should be such that the vector $\mathbf{e} = \mathbf{h} - \mathbf{b}$ is orthogonal to the subspace $\mathcal{R}(\mathbf{A})$ (see Figure 12.1). Recall that a vector $\mathbf{e} \in \mathbb{R}^m$ is said to be orthogonal to the subspace $\mathcal{R}(\mathbf{A})$ if it is orthogonal to every vector in this subspace. We call $\mathbf{h}$ the *orthogonal projection* of $\mathbf{b}$ onto the subspace $\mathcal{R}(\mathbf{A})$. It turns out that $\mathbf{h} = \mathbf{A}\mathbf{x}^* = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$. Hence, the vector $\mathbf{h} \in \mathcal{R}(\mathbf{A})$ minimizing $\|\mathbf{b} - \mathbf{h}\|$ is exactly the orthogonal projection of $\mathbf{b}$ onto $\mathcal{R}(\mathbf{A})$. In other words, the vector $\mathbf{x}^*$ minimizing $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ is exactly the vector that makes $\mathbf{A}\mathbf{x} - \mathbf{b}$ orthogonal to $\mathcal{R}(\mathbf{A})$.

To proceed further, we write $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, where $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of $\mathbf{A}$. The vector $\mathbf{e}$ is orthogonal to $\mathcal{R}(\mathbf{A})$ if and only if it is orthogonal to each of the columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ of $\mathbf{A}$. To see this, note that

$$\langle \mathbf{e}, \mathbf{a}_i \rangle = 0, \quad i = 1, \dots, n$$

if and only if for any set of scalars $\{x_1, x_2, \dots, x_n\}$, we also have

$$\langle \mathbf{e}, x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n \rangle = 0.$$

Any vector in $\mathcal{R}(\mathbf{A})$ has the form $x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$.

Proposition 12.1 *Let $\mathbf{h} \in \mathcal{R}(\mathbf{A})$ be such that $\mathbf{h} - \mathbf{b}$ is orthogonal to $\mathcal{R}(\mathbf{A})$. Then, $\mathbf{h} = \mathbf{A}\mathbf{x}^* = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$. $\square$*

Proof. Because $\mathbf{h} \in \mathcal{R}(\mathbf{A}) = \text{span}[\mathbf{a}_1, \dots, \mathbf{a}_n]$, it has the form $\mathbf{h} = x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$, where $x_1, \dots, x_n \in \mathbb{R}$. To find $x_1, \dots, x_n$, we use the assumption

that $\mathbf{e} = \mathbf{h} - \mathbf{b}$ is orthogonal to $\text{span}[\mathbf{a}_1, \dots, \mathbf{a}_n]$; that is, for all $i = 1, \dots, n$, we have

$$\langle \mathbf{h} - \mathbf{b}, \mathbf{a}_i \rangle = 0,$$

or, equivalently,

$$\langle \mathbf{h}, \mathbf{a}_i \rangle = \langle \mathbf{b}, \mathbf{a}_i \rangle.$$

Substituting $\mathbf{h}$ into the equations above, we obtain a set of n linear equations of the form

$$\langle \mathbf{a}_1, \mathbf{a}_i \rangle x_1 + \dots + \langle \mathbf{a}_n, \mathbf{a}_i \rangle x_n = \langle \mathbf{b}, \mathbf{a}_i \rangle, \quad i = 1, \dots, n.$$

In matrix notation this system of n equations can be represented as

$$\begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_1 \rangle \\ \vdots & & \vdots \\ \langle \mathbf{a}_1, \mathbf{a}_n \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_n \rangle \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \langle \mathbf{b}, \mathbf{a}_1 \rangle \\ \vdots \\ \langle \mathbf{b}, \mathbf{a}_n \rangle \end{bmatrix}.$$

Note that we can write

$$\begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_1 \rangle \\ \vdots & & \vdots \\ \langle \mathbf{a}_1, \mathbf{a}_n \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_n \rangle \end{bmatrix} = \mathbf{A}^\top \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} [\mathbf{a}_1 \cdots \mathbf{a}_n].$$

We also note that

$$\begin{bmatrix} \langle \mathbf{b}, \mathbf{a}_1 \rangle \\ \vdots \\ \langle \mathbf{b}, \mathbf{a}_n \rangle \end{bmatrix} = \mathbf{A}^\top \mathbf{b} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \mathbf{b}.$$

Because $\text{rank } \mathbf{A} = n$, $\mathbf{A}^\top \mathbf{A}$ is nonsingular, and thus we conclude that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} = \mathbf{x}^*.$$

■

Notice that the matrix

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_1 \rangle \\ \vdots & & \vdots \\ \langle \mathbf{a}_1, \mathbf{a}_n \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_n \rangle \end{bmatrix}$$

plays an important role in the least-squares solution. This matrix is often called the *Gram matrix* (or *Grammian*).

An alternative method of arriving at the least-squares solution is to proceed as follows. First, we write

$$\begin{aligned} f(\mathbf{x}) &= \|\mathbf{Ax} - \mathbf{b}\|^2 \\ &= (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b}) \\ &= \frac{1}{2} \mathbf{x}^\top (2\mathbf{A}^\top \mathbf{A}) \mathbf{x} - \mathbf{x}^\top (2\mathbf{A}^\top \mathbf{b}) + \mathbf{b}^\top \mathbf{b}. \end{aligned}$$

Therefore, f is a quadratic function. The quadratic term is positive definite because $\text{rank } \mathbf{A} = n$. Thus, the unique minimizer of f is obtained by solving the FONC (see Exercise 6.33); that is,

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^\top \mathbf{Ax} - 2\mathbf{A}^\top \mathbf{b} = \mathbf{0}.$$

The only solution to the equation $\nabla f(\mathbf{x}) = \mathbf{0}$ is $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.

Example 12.1 Suppose that you are given two different types of concrete. The first type contains 30% cement, 40% gravel, and 30% sand (all percentages of weight). The second type contains 10% cement, 20% gravel, and 70% sand. How many pounds of each type of concrete should you mix together so that you get a concrete mixture that has as close as possible to a total of 5 pounds of cement, 3 pounds of gravel, and 4 pounds of sand?

The problem can be formulated as a least-squares problem with

$$\mathbf{A} = \begin{bmatrix} 0.3 & 0.1 \\ 0.4 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix},$$

where the decision variable is $\mathbf{x} = [x_1, x_2]^\top$ and x_1 and x_2 are the amounts of concrete of the first and second types, respectively. After some algebra, we obtain the solution:

$$\begin{aligned} \mathbf{x}^* &= (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \\ &= \frac{1}{(0.34)(0.54) - (0.32)^2} \begin{bmatrix} 0.54 & -0.32 \\ -0.32 & 0.34 \end{bmatrix} \begin{bmatrix} 3.9 \\ 3.9 \end{bmatrix} \\ &= \begin{bmatrix} 10.6 \\ 0.961 \end{bmatrix}. \end{aligned}$$

(For a variation of this problem solved using a different method, see Example 15.7.) ■

We now give an example in which least-squares analysis is used to fit measurements by a straight line.

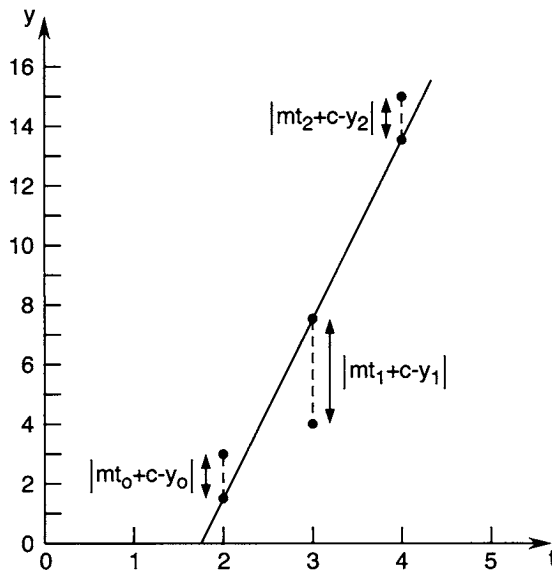
Table 12.1 Experimental Data for Example 12.2.

i	0	1	2
t_i	2	3	4
y_i	3	4	15

Example 12.2 *Line Fitting.* Suppose that a process has a single input $t \in \mathbb{R}$ and a single output $y \in \mathbb{R}$. Suppose that we perform an experiment on the process, resulting in a number of measurements, as displayed in Table 12.1. The i th measurement results in the input labeled t_i and the output labeled y_i . We would like to find a straight line given by

$$y = mt + c$$

that fits the experimental data. In other words, we wish to find two numbers, m and c , such that $y_i = mt_i + c$, $i = 0, 1, 2$. However, it is apparent that there is no choice of m and c that results in the requirement above; that is, there is no straight line that passes through all three points simultaneously. Therefore, we would like to find the values of m and c that best fit the data. A graphical illustration of our problem is shown in Figure 12.2.

**Figure 12.2** Fitting a straight line to experimental data.

We can represent our problem as a system of three linear equations of the form

$$\begin{aligned} 2m + c &= 3 \\ 3m + c &= 4 \\ 4m + c &= 15. \end{aligned}$$

We can write this system of equations as

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 4 \\ 15 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} m \\ c \end{bmatrix}.$$

Note that since

$$\text{rank } \mathbf{A} < \text{rank } [\mathbf{A}, \mathbf{b}],$$

the vector $\mathbf{b}$ does not belong to the range of $\mathbf{A}$. Thus, as we have noted before, the system of equations above is inconsistent.

The straight line of best fit is the one that minimizes

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \sum_{i=0}^2 (mt_i + c - y_i)^2.$$

Therefore, our problem lies in the class of least-squares problems. Note that the foregoing function of m and c is simply the total squared vertical distance (squared error) between the straight line defined by m and c and the experimental points. The solution to our least-squares problem is

$$\mathbf{x}^* = \begin{bmatrix} m^* \\ c^* \end{bmatrix} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} = \begin{bmatrix} 6 \\ -32/3 \end{bmatrix}.$$

Note that the error vector $\mathbf{e} = \mathbf{A}\mathbf{x}^* - \mathbf{b}$ is orthogonal to each column of $\mathbf{A}$. ■

Next, we give an example of the use of least-squares in wireless communications.

Example 12.3 Attenuation Estimation. A wireless transmitter sends a discrete-time signal $\{s_0, s_1, s_2\}$ (of duration 3) to a receiver, as shown in Figure 12.3. The real number s_i is the value of the signal at time i .

The transmitted signal takes two paths to the receiver: a direct path, with delay 10 and attenuation factor a_1 , and an indirect (reflected) path, with delay 12 and attenuation factor a_2 . The received signal is the sum of the signals from these two paths, with their respective delays and attenuation factors.

We can easily check our solution by computing $\mathbf{x}^{(2)}$ directly using the formula $\mathbf{x}^{(2)} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}^{(0)} \\ \mathbf{b}^{(1)} \\ \mathbf{b}^{(2)} \end{bmatrix}.$$

■

12.3 Solution to a Linear Equation with Minimum Norm

Consider now a system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \leq n$, and $\text{rank } \mathbf{A} = m$. Note that the number of equations is no larger than the number of unknowns. There may exist an infinite number of solutions to this system of equations. However, as we shall see, there is only one solution that is closest to the origin: the solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ whose norm $\|\mathbf{x}\|$ is minimal. Let $\mathbf{x}^*$ be this solution; that is, $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ and $\|\mathbf{x}^*\| \leq \|\mathbf{x}\|$ for any $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$. In other words, $\mathbf{x}^*$ is the solution to the problem

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{x}\| \\ & \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

In Part IV, we study problems of this type in more detail.

Theorem 12.2 *The unique solution $\mathbf{x}^*$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$ that minimizes the norm $\|\mathbf{x}\|$ is given by*

$$\mathbf{x}^* = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}.$$

□

Proof. Let $\mathbf{x}^* = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}$. Note that

$$\begin{aligned} \|\mathbf{x}\|^2 &= \|(\mathbf{x} - \mathbf{x}^*) + \mathbf{x}^*\|^2 \\ &= ((\mathbf{x} - \mathbf{x}^*) + \mathbf{x}^*)^\top ((\mathbf{x} - \mathbf{x}^*) + \mathbf{x}^*) \\ &= \|\mathbf{x} - \mathbf{x}^*\|^2 + \|\mathbf{x}^*\|^2 + 2\mathbf{x}^{*\top}(\mathbf{x} - \mathbf{x}^*). \end{aligned}$$

We now show that

$$\mathbf{x}^{*\top}(\mathbf{x} - \mathbf{x}^*) = 0.$$

Indeed,

$$\begin{aligned} \mathbf{x}^{*\top}(\mathbf{x} - \mathbf{x}^*) &= [\mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}]^\top [\mathbf{x} - \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}] \\ &= \mathbf{b}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} [\mathbf{A}\mathbf{x} - (\mathbf{A}\mathbf{A}^\top)(\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}] \\ &= \mathbf{b}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} [\mathbf{b} - \mathbf{b}] = 0. \end{aligned}$$

Therefore,

$$\|\mathbf{x}\|^2 = \|\mathbf{x}^*\|^2 + \|\mathbf{x} - \mathbf{x}^*\|^2.$$

Because $\|\mathbf{x} - \mathbf{x}^*\|^2 > 0$ for all $\mathbf{x} \neq \mathbf{x}^*$, it follows that for all $\mathbf{x} \neq \mathbf{x}^*$,

$$\|\mathbf{x}\|^2 > \|\mathbf{x}^*\|^2,$$

which implies that

$$\|\mathbf{x}\| > \|\mathbf{x}^*\|.$$

■

Example 12.7 Find the point closest to the origin of $\mathbb{R}^3$ on the line of intersection of the two planes defined by the following two equations:

$$x_1 + 2x_2 - x_3 = 1,$$

$$4x_1 + x_2 + 3x_3 = 0.$$

Note that this problem is equivalent to the problem

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\| \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}, \end{array}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ 4 & 1 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Thus, the solution to the problem is

$$\mathbf{x}^* = \mathbf{A}^\top (\mathbf{AA}^\top)^{-1} \mathbf{b} = \begin{bmatrix} 0.0952 \\ 0.3333 \\ -0.2381 \end{bmatrix}.$$

■

In the next section we discuss an iterative algorithm for solving $\mathbf{Ax} = \mathbf{b}$.

12.4 Kaczmarz's Algorithm

As in Section 12.3, let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \leq n$, and $\text{rank } \mathbf{A} = m$. We now discuss an iterative algorithm for solving $\mathbf{Ax} = \mathbf{b}$, originally analyzed by Kaczmarz in 1937 [70]. The algorithm converges to the vector $\mathbf{x}^* = \mathbf{A}^\top (\mathbf{AA}^\top)^{-1} \mathbf{b}$ without explicitly having to invert the matrix $\mathbf{AA}^\top$. This is important from a practical point of view, especially when $\mathbf{A}$ has many rows.

Let $\mathbf{a}_j^\top$ denote the j th row of $\mathbf{A}$, and b_j the j th component of $\mathbf{b}$, and μ a positive scalar, $0 < \mu < 2$. With this notation, Kaczmarz's algorithm is:

$$3. (\mathbf{A}\mathbf{A}^\dagger)^\top = \mathbf{A}\mathbf{A}^\dagger.$$

$$4. (\mathbf{A}^\dagger\mathbf{A})^\top = \mathbf{A}^\dagger\mathbf{A}.$$

The Penrose definition above is equivalent to Definition 12.1 (see Exercise 12.31). For more information on generalized inverses and their applications, we refer the reader to the books by Ben-Israel and Greville [6] and Campbell and Meyer [23].

EXERCISES

12.1 A rock is accelerated to 3, 5, and 6 m/s² by applying forces of 1, 2, and 3 N, respectively. Assuming that Newton's law $F = ma$ holds, where F is the force and a is the acceleration, estimate the mass m of the rock using the least-squares method.

12.2 A spring is stretched to lengths $L = 3, 4$, and 5 cm under applied forces $F = 1, 2$, and 4 N, respectively. Assuming that Hooke's law $L = a + bF$ holds, estimate the normal length a and spring constant b using the least-squares approach.

12.3 Suppose that we perform an experiment to calculate the gravitational constant g as follows. We drop a ball from a certain height and measure its distance from the original point at certain time instants. The results of the experiment are shown in the following table.

Time (seconds)	1.00	2.00	3.00
Distance (meters)	5.00	19.5	44.0

The equation relating the distance s and the time t at which s is measured is given by

$$s = \frac{1}{2}gt^2.$$

- Find a least-squares estimate of g using the experimental results from the table above.
- Suppose that we take an additional measurement at time 4.00 and obtain a distance of 78.5. Use the recursive least-squares algorithm to calculate an updated least-squares estimate of g .

12.4 Suppose that we have a speech signal, represented as a finite sequence of real numbers $x_1, x_2, \dots, x_n$. Suppose that we record this signal onto magnetic tape. The recorded speech signal is represented by another sequence of real numbers $y_1, y_2, \dots, y_n$.

Suppose that we model the recording process as a simple scaling of the original signal (i.e., we believe that a good model of the relationship between the recorded signal and the original signal is $y_i = \alpha x_i$ for some constant α that does not depend on i). Suppose that we know exactly the original signal $x_1, x_2, \dots, x_n$ as well as the recorded signal $y_1, y_2, \dots, y_n$. Use the least-squares method to find a formula for estimating the scale factor α given this data. (You may assume that at least one x_i is nonzero.)

12.5 Suppose that we wish to estimate the value of the resistance R of a resistor. Ohm's law states that if V is the voltage across the resistor and I is the current through the resistor, then $V = IR$. To estimate R , we apply a 1-ampere current through the resistor and measure the voltage across it. We perform the experiment on n voltage-measuring devices and obtain measurements of $V_1, \dots, V_n$. Show that the least-squares estimate of R is simply the average of $V_1, \dots, V_n$.

12.6 The table below shows the stock prices for three companies, X, Y, and Z, tabulated over three days:

	Day		
	1	2	3
X	6	4	5
Y	1	1	3
Z	2	1	2

Suppose that an investment analyst proposes a model for the predicting the stock price of X based on those of Y and Z:

$$p_X = ap_Y + bp_Z,$$

where p_X , p_Y , and p_Z are the stock prices of X, Y, and Z, respectively, and a and b are real-valued parameters. Calculate the least-squares estimate of parameters a and b based on the data in the table above.

12.7 We are given two mixtures, A and B. Mixture A contains 30% gold, 40% silver, and 30% platinum, whereas mixture B contains 10% gold, 20% silver, and 70% platinum (all percentages of weight). We wish to determine the ratio of the weight of mixture A to the weight of mixture B such that we have as close as possible to a total of 5 ounces of gold, 3 ounces of silver, and 4 ounces of platinum. Formulate and solve the problem using the linear least-squares method.

12.8 *Background:* If $A\mathbf{x} + \mathbf{w} = \mathbf{b}$, where $\mathbf{w}$ is a “white noise” vector, then define the *least-squares estimate* of $\mathbf{x}$ given $\mathbf{b}$ to be the solution to the problem

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\|^2.$$

This problem is related to Wiener filtering.

Application: Suppose that a given speech signal $\{u_k : k = 1, \dots, n\}$ (with $u_k \in \mathbb{R}$) is transmitted over a telephone cable with input-output behavior given by $y_k = ay_{k-1} + bu_k + v_k$, where, at each time k , $y_k \in \mathbb{R}$ is the output, $u_k \in \mathbb{R}$ is the input (speech signal value), and v_k represents white noise. The parameters a and b are fixed known constants, and the initial condition is $y_0 = 0$.

We can measure the signal $\{y_k\}$ at the output of the telephone cable, but we cannot directly measure the desired signal $\{u_k\}$ or the noise signal $\{v_k\}$. Derive a formula for the linear least-squares estimate of the signal $\{u_k : k = 1, \dots, n\}$ given the signal $\{y_k : k = 1, \dots, n\}$.

Note: Even though the vector $\mathbf{v} = [v_1, \dots, v_n]^\top$ is a white noise vector, the vector $\mathbf{D}\mathbf{v}$ (where $\mathbf{D}$ is a matrix) is, in general, not.

12.9 Line Fitting. Let $[x_1, y_1]^\top, \dots, [x_p, y_p]^\top$, $p \geq 2$, be points in $\mathbb{R}^2$. We wish to find the straight line of best fit through these points (“best” in the sense that the total squared error is minimized); that is, we wish to find $a^*, b^* \in \mathbb{R}$ to minimize

$$f(a, b) = \sum_{i=1}^p (ax_i + b - y_i)^2.$$

Assume that the x_i , $i = 1, \dots, p$, are not all equal. Show that there exist unique parameters a^* and b^* for the line of best fit, and find the parameters in terms of the following quantities:

$$\begin{aligned}\bar{X} &= \frac{1}{p} \sum_{i=1}^p x_i, \\ \bar{Y} &= \frac{1}{p} \sum_{i=1}^p y_i, \\ \overline{X^2} &= \frac{1}{p} \sum_{i=1}^p x_i^2, \\ \overline{Y^2} &= \frac{1}{p} \sum_{i=1}^p y_i^2, \\ \overline{XY} &= \frac{1}{p} \sum_{i=1}^p x_i y_i.\end{aligned}$$

12.10 Suppose that we take measurements of a sinusoidal signal $y(t) = \sin(\omega t + \theta)$ at times $t_1, \dots, t_p$, and obtain values $y_1, \dots, y_p$, where $-\pi/2 \leq \omega t_i + \theta \leq \pi/2$, $i = 1, \dots, p$, and the t_i are not all equal. We wish to determine the values of the frequency ω and phase θ .

- a. Express the problem as a system of linear equations.
- b. Find the least-squares estimate of ω and θ based on part a. Use the following notation:

$$\begin{aligned}\bar{T} &= \frac{1}{p} \sum_{i=1}^p t_i, \\ \overline{T^2} &= \frac{1}{p} \sum_{i=1}^p t_i^2, \\ \overline{TY} &= \frac{1}{p} \sum_{i=1}^p t_i \arcsin y_i, \\ \bar{Y} &= \frac{1}{p} \sum_{i=1}^p \arcsin y_i.\end{aligned}$$

12.11 We are given a point $[x_0, y_0]^\top \in \mathbb{R}^2$. Consider the straight line on the $\mathbb{R}^2$ plane given by the equation $y = mx$. Using a least-squares formulation, find the point on the straight line that is closest to the given point $[x_0, y_0]$, where the measure of closeness is in terms of the Euclidean norm on $\mathbb{R}^2$.

Hint: The given line can be expressed as the range of the matrix $\mathbf{A} = [1, m]^\top$.

12.12 Consider the affine function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ of the form $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + c$, where $\mathbf{a} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

- a. We are given a set of p pairs $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_p, y_p)$, where $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$, $i = 1, \dots, p$. We wish to find the affine function of best fit to these points, where “best” is in the sense of minimizing the total square error

$$\sum_{i=1}^p (f(\mathbf{x}_i) - y_i)^2.$$

Formulate the above as an optimization problem of the form: minimize $\|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2$ with respect to $\mathbf{z}$. Specify the dimensions of $\mathbf{A}$, $\mathbf{z}$, and $\mathbf{b}$.

- b. Suppose that the points satisfy

$$\mathbf{x}_1 + \cdots + \mathbf{x}_p = \mathbf{0}$$

and

$$y_1 \mathbf{x}_1 + \cdots + y_p \mathbf{x}_p = \mathbf{0}.$$

Find the affine function of best fit in this case, assuming that it exists and is unique.

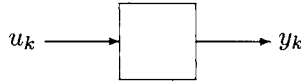


Figure 12.5 Input-output system in Exercise 12.13.

12.13 For the system shown in Figure 12.5, consider a set of input-output pairs $(u_1, y_1), \dots, (u_n, y_n)$, where $u_k \in \mathbb{R}$, $y_k \in \mathbb{R}$, $k = 1, \dots, n$.

- Suppose that we wish to find the best linear estimate of the system based on the input-output data above. In other words, we wish to find a $\hat{\theta}_n \in \mathbb{R}$ to fit the model $y_k = \hat{\theta}_n u_k$, $k = 1, \dots, n$. Using the least-squares approach, derive a formula for $\hat{\theta}_n$ based on $u_1, \dots, u_n$ and $y_1, \dots, y_n$.
- Suppose that the data in part a are generated by

$$y_k = \theta u_k + e_k,$$

where $\theta \in \mathbb{R}$ and $u_k = 1$ for all k . Show that the parameter $\hat{\theta}_n$ in part a converges to θ as $n \rightarrow \infty$ if and only if

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n e_k = 0.$$

12.14 Consider a discrete-time linear system $x_{k+1} = ax_k + bu_k$, where u_k is the input at time k , x_k is the output at time k , and $a, b \in \mathbb{R}$ are system parameters. Suppose that we apply a constant input $u_k = 1$ for all $k \geq 0$ and measure the first four values of the output to be $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, $x_3 = 8$. Find the least-squares estimate of a and b based on the data above.

12.15 Consider a discrete-time linear system $x_{k+1} = ax_k + bu_k$, where u_k is the input at time k , x_k is the output at time k , and $a, b \in \mathbb{R}$ are system parameters. Given the first $n + 1$ values of the impulse response $h_0, \dots, h_n$, find the least-squares estimate of a and b . You may assume that at least one h_k is nonzero.

Note: The *impulse response* is the output sequence resulting from an input of $u_0 = 1$, $u_k = 0$ for $k \neq 0$ and zero initial condition $x_0 = 0$.

12.16 Consider a discrete-time linear system $x_{k+1} = ax_k + bu_k$, where u_k is the input at time k , x_k is the output at time k , and $a, b \in \mathbb{R}$ are system parameters. Given the first $n + 1$ values of the step response $s_0, \dots, s_n$, where $n > 1$, find the least-squares estimate of a and b . You may assume that at least one s_k is nonzero.

Note: The *step response* is the output sequence resulting from an input of $u_k = 1$ for $k \geq 0$, and zero initial condition $x_0 = 0$ (i.e., $s_0 = x_0 = 0$).

12.17 Consider a known discrete-time signal on the time interval $\{1, \dots, n\}$, represented by the vector $\mathbf{x} \in \mathbb{R}^n$ (x_i is the value of the signal at time i). We transmit the signal $a\mathbf{x}$ over a communication channel, where $a \in \mathbb{R}$ represents the “amplitude” of the transmission, a quantity unknown to the receiver. The receiver receives a signal $\mathbf{y} \in \mathbb{R}^n$, which is a distorted version of the transmitted signal (so that $\mathbf{y}$ may not be equal to $a\mathbf{x}$ for any a). Formulate the problem of estimating the quantity a according to a least-squares criterion, and solve it (stating whatever appropriate assumptions are necessary, if any).

12.18 Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \geq n$, and $\text{rank } \mathbf{A} = n$. Consider the constrained optimization problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \mathbf{x}^\top \mathbf{x} - \mathbf{x}^\top \mathbf{b} \\ & \text{subject to} && \mathbf{x} \in \mathcal{R}(\mathbf{A}), \end{aligned}$$

where $\mathcal{R}(\mathbf{A})$ denotes the range of $\mathbf{A}$. Derive an expression for the global minimizer of this problem in terms of $\mathbf{A}$ and $\mathbf{b}$.

12.19 Solve the problem

$$\begin{aligned} & \text{minimize} && \|\mathbf{x} - \mathbf{x}_0\| \\ & \text{subject to} && \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \mathbf{x} = 1, \end{aligned}$$

where $\mathbf{x}_0 = [0, -3, 0]^\top$.

12.20 Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \leq n$, $\text{rank } \mathbf{A} = m$, and $\mathbf{x}_0 \in \mathbb{R}^n$. Consider the problem

$$\begin{aligned} & \text{minimize} && \|\mathbf{x} - \mathbf{x}_0\| \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

Show that this problem has a unique solution given by

$$\mathbf{x}^* = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b} + (\mathbf{I}_n - \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{A}) \mathbf{x}_0.$$

12.21 Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, $\text{rank } \mathbf{A} = n$, and $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^m$, consider the problem

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}_1\|^2 + \|\mathbf{A}\mathbf{x} - \mathbf{b}_2\|^2 + \dots + \|\mathbf{A}\mathbf{x} - \mathbf{b}_p\|^2.$$

Suppose that $\mathbf{x}_i^*$ is a solution to the problem

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}_i\|^2,$$

where $i = 1, \dots, p$. Write the solution to the problem in terms of $\mathbf{x}_1^*, \dots, \mathbf{x}_p^*$.

12.22 Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, $\text{rank } \mathbf{A} = n$, $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^m$, and $\alpha_1, \dots, \alpha_p \in \mathbb{R}$, consider the problem

$$\text{minimize } \alpha_1 \|\mathbf{Ax} - \mathbf{b}_1\|^2 + \alpha_2 \|\mathbf{Ax} - \mathbf{b}_2\|^2 + \cdots + \alpha_p \|\mathbf{Ax} - \mathbf{b}_p\|^2.$$

Suppose that $\mathbf{x}_i^*$ is the solution to the problem

$$\text{minimize } \|\mathbf{Ax} - \mathbf{b}_i\|^2,$$

where $i = 1, \dots, p$. Assuming that $\alpha_1 + \cdots + \alpha_p > 0$, derive a simple expression for the solution to this problem in terms of $\mathbf{x}_1^*, \dots, \mathbf{x}_p^*$ and $\alpha_1, \dots, \alpha_p$.

12.23 Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $m \leq n$, and $\text{rank } \mathbf{A} = m$. Show that $\mathbf{x}^* = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{b}$ is the only vector in $\mathcal{R}(\mathbf{A}^\top)$ satisfying $\mathbf{Ax}^* = \mathbf{b}$.

12.24 The purpose of this question is to derive a recursive least-squares algorithm where we *remove* (instead of add) a data point. To formulate the algorithm, suppose that we are given matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ such that

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{a}_1^\top \end{bmatrix},$$

where $\mathbf{a}_1 \in \mathbb{R}^n$. Similarly, suppose that vectors $\mathbf{b}^{(0)}$ and $\mathbf{b}^{(1)}$ satisfy

$$\mathbf{b}^{(0)} = \begin{bmatrix} \mathbf{b}^{(1)} \\ b_1 \end{bmatrix},$$

where $b_1 \in \mathbb{R}$. Let $\mathbf{x}^{(0)}$ be the least-squares solution associated with $(\mathbf{A}_0, \mathbf{b}^{(0)})$ and $\mathbf{x}^{(1)}$ the least-squares solution associated with $(\mathbf{A}_1, \mathbf{b}^{(1)})$. Our goal is to write $\mathbf{x}^{(1)}$ in terms of $\mathbf{x}^{(0)}$ and the data point “removed,” $(\mathbf{a}_1, b_1)$. As usual, let $\mathbf{G}_0$ and $\mathbf{G}_1$ be the Grammians associated with $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(1)}$, respectively.

- Write down expressions for the least-squares solutions $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(1)}$ in terms of $\mathbf{A}_0$, $\mathbf{b}^{(0)}$, $\mathbf{A}_1$, and $\mathbf{b}^{(1)}$.
- Derive a formula for $\mathbf{G}_1$ in terms of $\mathbf{G}_0$ and $\mathbf{a}_1$.
- Let $\mathbf{P}_0 = \mathbf{G}_0^{-1}$ and $\mathbf{P}_1 = \mathbf{G}_1^{-1}$. Derive a formula for $\mathbf{P}_1$ in terms of $\mathbf{P}_0$ and $\mathbf{a}_1$. (The formula must not contain any matrix inversions.)
- Derive a formula for $\mathbf{A}_0^\top \mathbf{b}^{(0)}$ in terms of $\mathbf{G}_1$, $\mathbf{x}^{(0)}$, and $\mathbf{a}_1$.
- Finally, derive a formula for $\mathbf{x}^{(1)}$ in terms of $\mathbf{x}^{(0)}$, $\mathbf{P}_1$, $\mathbf{a}_1$, and b_1 . Use this and part c to write a recursive algorithm associated with successive removals of rows from $(\mathbf{A}_k, \mathbf{b}^{(k)})$.

12.25 Show that in Kaczmarz's algorithm, if $\mathbf{x}^{(0)} = \mathbf{0}$, then $\mathbf{x}^{(k)} \in \mathcal{R}(\mathbf{A}^\top)$ for all k .

12.26 Consider Kaczmarz's algorithm with $\mathbf{x}^{(0)} \neq \mathbf{0}$.

- a. Show that there exists a unique point minimizing $\|\mathbf{x} - \mathbf{x}^{(0)}\|$ subject to $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$.
- b. Show that Kaczmarz's algorithm converges to the point in part a.

12.27 Consider Kaczmarz's algorithm with $\mathbf{x}^{(0)} = \mathbf{0}$, where $m = 1$; that is, $\mathbf{A} = [\mathbf{a}^\top] \in \mathbb{R}^{1 \times n}$ and $\mathbf{a} \neq \mathbf{0}$, and $0 < \mu < 2$. Show that there exists $0 \leq \gamma < 1$ such that $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq \gamma \|\mathbf{x}^{(k)} - \mathbf{x}^*\|$ for all $k \geq 0$.

12.28 Show that in Kaczmarz's algorithm, if $\mu = 1$, then $\mathbf{b}_{R(k)+1} - \mathbf{a}_{R(k)+1}^\top \mathbf{x}^{(k+1)} = 0$ for each k .

12.29 Consider the problem of minimizing $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ over $\mathbb{R}^n$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^*$ be a solution. Suppose that $\mathbf{A} = \mathbf{B}\mathbf{C}$ is a full-rank factorization of $\mathbf{A}$; that is, $\text{rank } \mathbf{A} = \text{rank } \mathbf{B} = \text{rank } \mathbf{C} = r$, and $\mathbf{B} \in \mathbb{R}^{m \times r}$, $\mathbf{C} \in \mathbb{R}^{r \times n}$. Show that the minimizer of $\|\mathbf{B}\mathbf{y} - \mathbf{b}\|$ over $\mathbb{R}^r$ is $\mathbf{C}\mathbf{x}^*$.

12.30 Prove the following properties of generalized inverses:

- a. $(\mathbf{A}^\top)^\dagger = (\mathbf{A}^\dagger)^\top$.
- b. $(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$.

12.31 Show that the Penrose definition of the generalized inverse is equivalent to Definition 12.1.

12.32 Construct matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ such that $(\mathbf{A}_1\mathbf{A}_2)^\dagger \neq \mathbf{A}_2^\dagger\mathbf{A}_1^\dagger$.