Chapter 8

Linear Approximation

§22. What Is Linear Approximation?

Before we can start solving a real-life problem using mathematics, we often need to collect numerical data. This is not always an easy task. For example, how can we measure the happiness of a person? The height (stature) of a person is considered a less controversial quantity, but precise measurements reveal that it is not constant during the same day even for an adult person. How about the speed of light? Since it is a physical constant, should not all observations by a skilled observer using the same tools and doing best to eliminate the sources of variation give exactly the same answer? Not at all! Even the most careful experiments produce variable results.

Even counting the passengers in an airplane sometimes gives discrepancies that may delay your flight. But airlines would like to know numbers of passengers not only in the present but also on future flights! (Cf. Exercise 4 in §24.)

A typical approach for a scientist is to observe the quantity several times and then take an average. An average is a single value that summarizes or represents a set of values. It is always between the minimal and maximal values. The averages used most often are the mean, median, and midrange. The differences between the observations and the selected average are called *residuals, discrepancies, vertical deviations,* or *error terms.* The term *vertical deviations* comes from a figure in which the average is represented by a horizontal line.

Now we define these averages and explain in what sense they are optimal. We consider m observations (numbers) $a_1, \ldots, a_m$. The arithmetic *mean* is

$$(a_1 + \cdots + a_m)/m.$$

To define other averages, it is convenient to order the observations in increasing order: $a_1 \leq \cdots \leq a_m$. In particular, $a_1 = \min(a_i)$ and $a_m = \max(a_i)$. Then the *midrange* is

$$(a_1 + a_m)/2.$$

When m is odd, the *median* is defined as $a_{(k+1)/2}$. When m is even, a *median* is any number x such that $a_{m/2} \leq x \leq a_{m/2+1}$. In some textbooks, it is defined to be $(a_{m/2} + a_{m/2+1})/2$ to make it unique. We will call the last number the *sample median* or the *central value.*

Example. For numbers 2, 5, 5, 7, the mean is $19/4 = 4.75$, the midrange is $9/2 = 4.5$, and the median is 5.

Remark. To compute the midrange and the interval of medians, it is not necessary to order the given numbers. It takes $m-1$ comparisons to find $\min(a_i)$. Then it takes $m-2$ comparisons to find $\max(a_i)$. So it takes $2m-3$ comparisons and 2 arithmetic operations to compute the midrange. Also, the medians can be found in time linear in m (see the Appendix). On the other hand, it takes at least $\log_2 m! \geq m(\log m - 1)$ comparisons to order m numbers. ∎

One reason that these three averages are used so often is the fact that they are optimal (the best fit to the given numbers) in the following three senses.

Theorem 22.1. The mean x_2 is the optimal solution for the following optimization problem:

$$\sum_{i=1}^{m} (a_i - x)^2 \to \min.$$

Proof. We can write the objective function as

$$\sum_{i=1}^{m} (a_i - x)^2 = m(x - x_2)^2 + C$$

with a constant C. Now it is obvious that $\min = C$ at $x = x_2$ and that this optimal solution is unique. ∎

So the mean is *the least squares fit,* or the best l^2-fit.

Theorem 22.2. The midrange x_∞ is the optimal solution for the following optimization problem:

$$\max_i |a_i - x| \to \min.$$

Proof. We order numbers as before. Then the objective function becomes

$$\max(a_1 - x, x - a_1, a_m - x, x - a_m)$$
$$= \begin{cases} a_m - x & \text{if } x \le x_\infty \\ x - a_1 & \text{otherwise} \end{cases} = |x - x_\infty| + (a_m - a_1)/2.$$

Now it is obvious that

$$\min = (a_m - a_1)/2 \text{ at } x = x_\infty$$

and that this optimal solution is unique. ■

So the midrange can be called the best l^∞-fit.

Theorem 22.3. A number x_1 is a median if and only if it is an optimal solution for the following optimization problem:

$$\sum_{i=1}^{m} |a_i - x| \to \min.$$

Proof. We order numbers as before. Then the objective function is affine on each of the following two rays and $m-1$ intervals:

$$x \le a_1; a_i \le x \le a_{i+1} (i = 1, 2, \ldots, m-1); x \ge a_m.$$

Its slope is $-m; -m + 2i$ (if $a_i \ne a_{i+1}$); m, respectively. Now it is obvious that the set of optimal solutions is the interval $a_{m/2} \le x_1 \le a_{m/2+1}$ when m is even and $a_{m/2} \ne a_{m/2+1}$. Otherwise, there is exactly one optimal solution which is

$$x_1 = \begin{cases} a_{m/2} = a_{m/2+1} & \text{when } m \text{ is even and } a_{m/2} = a_{m/2+1} \\ a_{(k+1)/2} & \text{when } m \text{ is odd.} \end{cases}$$

This agrees with the definition of medians. ■

Using the integer part function $\lfloor\ \rfloor$ the definition of medians x_1 can be written by one formula:

$$a_{\lfloor (k+1)/2 \rfloor} \le x_1 \le a_{\lfloor (k+2)/2 \rfloor}.$$

Theorem 22.3 tells us that the medians are the best l^1-fits.

Remark. Similarly, we can define best l^p-fit, but the values $p = 2, 1, \infty$ are most common. One reason for this is that those fits are easiest to compute. ■

To introduce bivariate models, we consider an example. Are you overweight? Underweight? Just right? A possible answer is, "It is my own business and I do not want to discuss it." Some health experts warn against excessive or insufficient weight (body mass). But what is the normal weight?

There are different points of view on this controversial issue. Some say that a person's ideal (or the setpoint) weight is a matter of genotype, the number of fat cells, health, lifestyle, and personal taste and has nothing to do with the weight of others.

There are some situations, however, when one's weight relative to the average in a group is important. For example, someone hoping to play on the offensive line of a football team may want to be heavier than other players in the game. A sumo wrestler may want to be heavier than his competition. You may want extra weight if you live in cold climates or compete in endurance tests such as the popular ' TV show "Survivor". On the other hand, a horse racing jockey may want to be the lightest among his competitors. Some runners may also strive to minimize their weight.

A simple-minded and probably politically incorrect way to judge your own weight is to compare it with an average weight of other persons (your peers). Depending on what average and which peers you use, the answer can be different. However, experts suggest a more sophisticated approach: Compare your weight with your own parameters such as your height.

For example, the Web site

http://health.yahoo.com/health

advises the following method:

> An easy way to determine your own desirable body weight is to use the following formula:
>
> Women: 100 pounds for the first 5 feet of height, 5 pounds for each additional inch; using this formula, the desirable body weight can be calculated.
>
> Men: 106 pounds of body weight for the first 5 feet of height, 6 pounds for each additional inch.

Writing w for the weight and h for the height, this recipe can be written $w = 5h - 200$ for women and $w = 6h - 254$ for men when $h \geq 60$ (in). Ever wonder where those coefficients 5, -200 and 6, -254 come from? Did a great scientist in an ivory tower compute them using basic laws of nature?

Or did somebody conduct a statistical analysis of real-life data? In the latter case, if you use these formulas, you compare implicitly your weight with the weights of other persons. A more explicit way of comparison is as follows: Assume that the ideal weight is a constant b (so the formula is $w = b$), and evaluate this constant as an average over a group of peers. Recall that we have considered three different concepts of an average.

Although you might like to be the heaviest in your group to make the football team, your doctor might be more concerned about the relation between your weight and your height. So we discuss now your weight relative to your height. This leads to more complicated mathematics. Our goal is to show how the coefficients a, b in the model $w = a + bh$ can be determined. We assume that you are a student in a class of 49 students on linear programming and all agreed to disclose their vital statistics. Consider the heights h_i and weights w_i in the class, $i = 1, 2, \ldots, 49$. We can plot the points (h_i, w_i) in the plane and look for a pattern in this scatterplot. It may happen that points cluster around a straight line in which case we want to find the line $w = a + bh$ that fits our data best.

It is not likely that a line $w = a+bh$ passes through all 49 points. In other words, it is unlikely that the system of 49 linear equations $w_i = a + bh_i$ $(i = 1, 2, \ldots, 49)$ for two unknowns a, b has a solution. So we are looking for an approximate "solution." Obviously, we want the best approximation. But how can we compare two different approximations and decide which is better?

In other words, we have 49 objective functions $|w_i - a - bh_i|$ of two variables a, b to minimize and we want to combine them into one objective function so that our optimization problem would make sense.

There are many ways to do this. The three most common ways are

$$e_1^2 + e_2^2 + \cdots + e_{49}^2 \to \min, \tag{22.4}$$

$$|e_1| + |e_2| + \cdots + |e_{49}| \to \min, \tag{22.5}$$

$$\max(|e_i|) \to \min, \tag{22.6}$$

where $e_i = w_i - a - bh_i$ are called *residuals, vertical deviations, or error terms.* Taking (22.4), (22.5), (22.6) as objective functions, we obtain the best l^p-fits for $p = 2, 1, \infty$ respectively.

Remark. Here is the reason why the objective function $\max |e_i|$ is referred to as the $p = \infty$ case:

$$\|e\|_p = (\sum_{i=1}^{m} |e_i|^p)^{1/p} \to \max(|e_i|)$$

as $p \to \infty$. ■

A more complicated way to relate the weight and height is $w = a + bh + ch^2 + dh^3$. Now we set $e_i = w_i - (a + bh + ch^2 + dh^3)$ and have one of three objective functions (22.4), (22.5), (22.6) to minimize. Besides height, other parameters can be brought into model. For example, the U.S. Navy uses a circumference method involving measurements of height, neck, and abdomen for men and height, abdomen, neck, and hip for women.

A simple model to relate h and w is used by CDC (the Centers for Disease Control and Prevention, the lead federal agency for protecting the health and safety of people), NIH (the National Institutes of Health, another federal agency), and AHA (the American Heart Association): $w = ch^2$. When the height h is measured in meters and weight w is in kilograms, the ratio w/h^2, measured in kg/m^2, is known as the body mass index (BMI).

By opinion of the CDC, NIH, and AHA, the BMI value is more useful for predicting health risks than weight alone. A BMI between 19 and 25 was considered to be "healthy" by AHA. These numbers were changed in 2001 to 18.5 to 24 (see AHA's Web site http://www.americanheart.org for updates; other Web sites give similar but different numbers that are changing with time). In a recent study (1996), researchers determined that 49% of women in the United States and 59% of men have a BMI of over 25, which would classify more than half of Americans as overweight. Of people between the ages of 50 and 60, 64% of women and 73% of men were identified as overweight.

Here is how the CDC answers the question "How does BMI relate to health among adults?": A healthy BMI for adults is between 18.5 and 24.9. BMI ranges are based on the effect body weight has on disease and death. In 1998 the NIH adapted the same range for "normal weight."

BMI has its limitations (e.g., for body builders), which are pointed out in the NIH guidelines (1998), where it is also suggested to use waist circumference for BMI between 25 and 34.9 kg/m in addition to BMI.

Thus, we plot points (h_i^2, w_i) and try to approximate them by a straight line passing through the origin. Once we choose (22.4), (22.5), or (22.6) as the objective function, with $e_i = w_i - ah_i^2$, we have an optimization problem in one variable c.

Remark. You should not make decisions about your health based solely on college textbooks. ■

Problem 22.7. Find the best l^p-fit $w = ch^2$ for $p = 1, 2, \infty$ given the following data:

i	1	2	3
Height h in m	1.6	1.5	1.7
Weight w in kg	65	60	70

Compare the optimal values for c with those for the best fits of the form $w/h^2 = c$ with the same p. Compare the minimums with those for the best fits of the form $w = b$ with the same p.

Solution. *Case* $p = 1$. We could convert this problem to a linear program with four variables and then solve it by the simplex method (see §23 below). But we can just consider the nonlinear problem with the objective function

$$f(c) = |65 - 1.6^2c| + |60 - 1.5^2c| + |70 - 1.7^2c|$$

to be minimized and no constraints. The function $f(c)$ is piecewise affine and convex, with the slope changing at

$$c = 70/1.7^2 \approx 24, c = 65/1.6^2 \approx 25, \text{ and } c = 60/1.5^2 \approx 27.$$

The slopes of $f(c)$ are

$$-1.6^2 - 1.5^2 - 1.7^2 < 0 \text{ for } c \le 70/1.7^2,$$
$$-1.6^2 - 1.5^2 + 1.7^2 \approx -2 \text{ for } 70/1.7^2 \le c \le 65/1.6^2,$$
$$1.6^2 - 1.5^2 + 1.7^2 \approx 3 \text{ for } 65/1.6^2 \le c \le 60/1.5^2,$$

and

$$1.6^2 + 1.5^2 + 1.7^2 > 0 \text{ for } c \ge 60/1.5^2.$$

Now it is clear that $f(c)$ is minimized at

$$c = x_1 = 65/1.6^2 \approx 25.39.$$

This value equals the median of the three observed BMIs

$$65/1.6^2, 60/1.5^2, 70/1.7^2.$$

The optimal value is

$$\min = |65 - c_1 1.6^2| + |60 - c_1 1.5^2| + |70 - c_1 1.7^2| = 6.25.$$

To compare this with the best l^1-fit for the model $w = b$, we compute the median $b = x_1 = 65$ and the corresponding optimal values:

$$\min = |65 - x_1| + |60 - x_1| + |70 - x_1| = 10.$$

So the model $w = ch^2$ is better than $w = b$ for our data with the l^1-approach.

Case $p = 2$. Our optimization problem can be reduced to solving a linear equation for c (see §23 below). Here we solve the problem using calculus, taking the advantage of the fact that our objective function

$$f(c) = (65 - 1.6^2c)^2 + (60 - 1.5^2c)^2 + (70 - 1.7^2c)^2$$

is differentiable. We set $f'(c) = 0$, which gives, after division by -2,

$$1.6^2(65 - 1.6^2c) + 1.5^2(60 - 1.5^2c) + 1.7^2(70 - 1.7^2c) = 0,$$

hence the optimal solution is

$$c = x_2 = 2518500/99841 \approx 25.225.$$

This x_2 is not the mean of the observed BMIs, which is about 25.426. The optimal value is $\min \approx 19$.

The mean of w_i is 65, and the corresponding minimal value is $5^2 + 0^2 + 5^2 = 50$. So again the model $w = ch^2$ is better than $w = b$.

Case $p = \infty$. We could reduce this problem to a linear program with two variables and then solve it by graphical method or simplex method (see §23). But we can do a graphical method with one variable. The objective function to minimize now is

$$f(c) = \max(|65 - 1.6^2c|, |60 - 1.5^2c|, |70 - 1.7^2c|).$$

This objective function $f(c)$ is piecewise affine and convex. We plot the function $f(c)$:

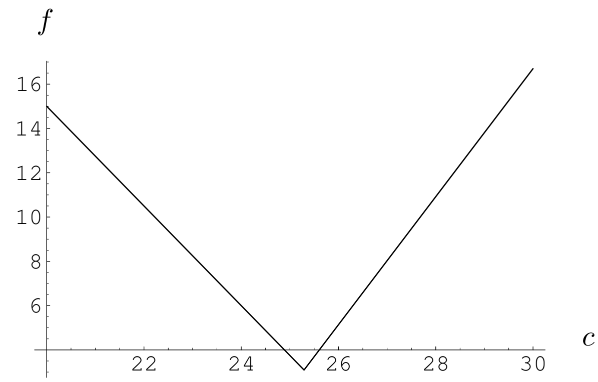

Figure 22.8. The objective function for l^∞-fit

The figure shows the optimal solution $a \approx 25$. Around this point,

$$65 - 1.6^2 c \approx 1, 60 - 1.5^2 c \approx 3.75, 70 - 1.7^2 c \approx -2.25;$$

hence

$$f(c) = \max(60 - 1.5^2 c, -70 + 1.7^2 c).$$

So the exact optimal solution satisfies $60 - 1.5^2 c = -70 + 1.7^2 c$; hence the optimal solution is

$$c_\infty = 6500/257 \approx 25.29.$$

It differs from the midrange of the BMIs, which is about 25.44. The optimal value is ≈ 3.

On the other hand, the midrange of the weights w_i is 65, which gives min $= 5$ for the model $w = b$ with the best l^∞-fit. So again the model $w = ch^2$ is better than $w = b$.

Remark 22.9. The optimal solutions for the values $p = 1, 2, \infty$ are all different in this example. It does not make sense to compare the corresponding optimal values M_p unless we normalize them:

$$M_p \mapsto (M_p/m)^{1/p} \text{ for } p \neq \infty, M_\infty \mapsto M_\infty,$$

where m is the number of observations ($m = 3$ in Problem 22.7). This transformation converts the l^p-norm of the column of residues to an average for the absolute values of the residues. But even after the transformation, the comparison of the averages with different p is difficult to justify. ∎

General Setup for Linear Approximation
In general, we may have data consisting of a column w of m entries and an m by n matrix A. We want to find a column X with n entries such that the column $e = (e_i) = w - AX$ of residuals is smallest in the sense of one of the following norms:

$$\|e\|_2 = (\sum e_i^2)^{1/2},\ \|e\|_1 = \sum |e_i|,\ \|e\|_\infty = \max(|e_i|). \qquad \blacksquare$$

So we have three optimization problems with m variables and nonlinear objective functions. These kinds of problems are typical in statistics. They also arise in other areas of mathematics and in computer science.

Remark 22.10. Geometrically, we want to approximate a vector w by a vector in the column space of A (by definition, the column space consists of all linear combinations of columns of A). In the next section, we will show that the first minimization problem is in fact about solving a system of linear equations, while the other two can be reduced to linear programs.

In the preceding theorems and examples, we considered the case when the matrix A consists of one column (i.e., $n = 1$). In fact, in the theorems A is the column with m ones, and this case was completely solved in the theorems. Finding optimal fits in the case $n = 1$ with general column A are optimization problems with one variable and no constraints. Solving them without a computer may present a challenge, as Problem 22.7 shows.

Examples like $w = a + bh + ch^2 + dh^3$ [so the residuals are $e_i = w_i - (a + bh + ch^2 + dh^3)$] are covered by the general setup for linear approximation with $n = 4$ because the residuals are linear functions of unknowns X. Functions used by scientists for data fitting could be any functions they know (polynomial, rational, trigonometric, exponential, etc.). But to find coefficients we either solve a system of linear equations (for least squares approximations) or use linear programming.

Remark 22.11. *Connection with statistics.* Linear approximation, especially with the least squares approach, appears in statistics. In regression analysis, traditionally, the first column of our matrix A consists of m ones. In simple regression analysis, the matrix A consists of $n = 2$ columns. In multiple regression, $n \geq 3$. In time series analysis, the second column of A is an arithmetic progression representing time; typically, this column is $[1, 2, ..., m]^T$.

Deep probability tools are used based on assumptions of normal distributions for residuals, which are considered as random noise masking a systematic pattern. These assumptions justify the least squares approach (l^2-fit), which is the method of choice in statistics. In the case when $n = 1$, this assumption gives preference to the mean over the other averages, but the sample median is also widely used.

Even though the assumptions of multiple regression cannot be tested explicitly, gross violations should be dealt with appropriately. In particular, outliers (i.e., extreme cases) can seriously bias the results by "pulling" or "pushing" the regression line in a particular direction, thereby leading to biased regression coefficients. Often, excluding just a single extreme case can yield a completely different set of results. The general purpose of multiple regression (the term was first used by Pearson, 1908) is to learn more about the relationship among several independent or predictor variables and a dependent or criterion variable. From the statistical point of view, the number m of observations should be much larger than the number n of variables to get reliable estimates for the n unknown coefficients in the column X.

Exercises

1–4 Compute the mean, the median, and the midrange of the following numbers.

1. 2, −7, 0, 2, 1.

2. 23, 56, −6, 0, 8, 0, 67.

3. 2, −7, 0, 2, 1, 0, 0, −1, 8.

4. 2, 3, 5, 6, −6, 0, 8, 0, 6, 7.

5. Construct examples when the mean x_2, the median x_1, and the midrange x_∞ of given numbers satisfy:

(a) $x_2 < x_1 < x_\infty$,
(b) $x_1 < x_2 < x_\infty$,
(c) $x_2 < x_\infty < x_1$,
(d) $x_\infty < x_2 < x_1$,
(e) $x_\infty < x_1 < x_2$,
(f) $x_1 < x_\infty < x_2$.

6. Note that a simple-minded computation of the median of BMIs in Problem 22.7 gives the same result $a_1 = 65/1.6^2 \approx 25.39$. Is it always the case (which would make the computation of the best l^1-fits for all models of the form $w = ax$ much easier) or a coincidence? *Hints*: Try other examples. Since the objective function is piecewise linear, nonconstant, and nonnegative, to find an optimal solution we need to look only at the points where the slope changes—that is, at the given BMIs w_i/h_i^2.

7. Using the data of Problem 22.7, find the best l^p-fit of the form $w = ah$ for $p = 1, 2, \infty$. Compare the results with the best fits in Problem 22.7.

8. Using the data of Problem 22.7, find the best l^p-fit of the form $w = ah^3$ for $p = 1, 2, \infty$. Compare the results with the best fits in Problem 22.7.

Remark. This model was suggested in literature. The quantity w/h^3 was named *normalized body mass* (NBM) and suggested as an alternative to BMI w/h^2. BMI was introduced by the Belgian statistician and anthropologist Lambert-Adolphe-Jacques Quetelet (1796-1874). The same metric units are used for both BMI and NBM. Also, the *ponderal index* $h/w^{1/3} = \text{NBM}^{-1/3}$ was suggested in literature. Another suggested alternative is w/h^4.

9. Find the best l^p-fit $w = ah^2$ for $p = 1, 2, \infty$ given the following data:

i	1	2	3	4
Height h in m	1.6	1.5	1.7	1.8
Weight w in kg	65	60	70	80

Compare the results with the best fits (with the same p) for the model $w = b$.

10. Find the best l^p-fit $w = ah$ for $p = 1, 2, \infty$ given the data in Exercise 9.

11. Find the best l^p-fit $w = ah^3$ for $p = 1, 2, \infty$ given the data in Exercise 9.

12. Here is the list of the first 100 primes p_n: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 521, 523, 541. Find the best l^p-fit $p_n = cn$ for $p = 1, 2, \infty$.

13. Using the data in Exercise 12, find the best l^p-fit $p_n = cn\log(n)$ for $p = 1, 2, \infty$.

14. The Fibonacci sequence F_t is defined by the recurrence

$$F_t = F_{t-1} + F_{t-2}, \;\; F_0 = F_1 = 1.$$

Using the first 50 Fibonacci numbers, compute the best l^p-fit $F_t = 2^{ct}$ for $p = 1, 2, \infty$.

§23. Linear Programming and Linear Approximation

The best l^2-fit is well known as the least squares fit. It is widely used for the following two reasons: It can be justified by some probability assumptions on the residuals and it can be found relatively easily. We remind now how to find it by solving a system of linear equations (which can be considered as a particular case of linear programming).

The Best l^2-Fit

The Euclidean norm $\|e\|_2 = \sum e_i^2$ is the most common way to measure the size of a vector and is used in Euclidean geometry. To find a vector in the column space, we drop a perpendicular from w onto the column space (see Remark 22.10). In other words, we want the vector $w - AX$ to be orthogonal to all columns of A—that is, $A^T(w - AX) = 0$. This gives a system of n linear equations $A^T AX = A^T w$ for n unknowns in the column X. The system always has a solution. Moreover, the best fit AX is the same for all solutions X. In the case when w belongs to the column space, the best fit is w. When the columns of A are linearly independent, X is unique. ■

Example 23.1. In the general setup, let $n = 1$, and let all entries of the column A be ones. Then we want to find a number $X = a$, such that $\sum(w_i - a)^2 \to \min$. In other words, we want to approximate m given numbers w_i by one number a. The equation $A^T AX = A^T w$ becomes $na = \sum w_i$; hence $a = X = \sum w_i/n$ is the arithmetic mean of the given numbers. This agrees with Theorem 22.1. ■

Problem 23.2. Find the best l^2-fit (up to two decimal points) of the form $w = a + bh$ to the following data:

i	1	2	3	4	5
h	1.5	1.6	1.7	1.7	1.8
w	60	65	70	75	80

Solution. In terms of the general setup, $X = [a, b]^T$,

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.5 & 1.6 & 1.7 & 1.7 & 1.8 \end{bmatrix},$$

and $w^T = [60, 65, 70, 75, 80]$. The system of linear equations $A^T AX = A^T w$ takes the form

$$\begin{bmatrix} 5 & 8.3 \\ 8.3 & 13.83 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 350 \\ 584.5 \end{bmatrix}.$$

Solving this system, we find $a \approx -41.7, b \approx 67.3$.

Problem 23.3. Using the data in Problem 23.2, find the best l^2-fit (up to two decimal points) of the form $w = a + ch^2$.

Solution. In terms of the general setup, $X = [a, c]^T$,

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.5^2 & 1.6^2 & 1.7^2 & 1.7^2 & 1.8^2 \end{bmatrix}$$

and $w^T = [60, 65, 70, 75, 80]$. The system of linear equations $A^T AX = A^T w$ takes the form

$$\begin{bmatrix} 5 & 13.83 \\ 13.83 & 38.82 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} \approx \begin{bmatrix} 350 \\ 979.65 \end{bmatrix}.$$

Solving this system, we find $a \approx 13.37, c \approx 20.47$.

The Best l^1-Fit

We reduce the optimization problem with the objective function $\|e\|_1 \to \min$, where $e = (e_i) = w - AX$, to a linear program using m additional variables u_i such that $|e_i| \le u_i$ for all i. We obtain the following linear program with $m + n$ variables a_j, u_i and $2m$ linear constraints:

$$\sum u_i \to \min, \ \ -u_i \le w_i - A_i X \le u_i \text{ for } i = 1, \dots, m,$$

where A_i is the i^{th} row of the given matrix A.

Problem 23.4. Using the data from Problem 23.2, find the best l^1-fit of the form $w = bh$.

Solution. Here $A = [1.5, 1.6, 1.7, 1.7, 1.8]^T$. As previously, we can convert this problem to a linear program with six variables and then solve it by the simplex method. But we can just consider the nonlinear problem with the objective function $f = \sum |w_i - bh_i| \to \min$ in one variable a and no constraints and solve it graphically or as follows. First we compute $c_i = w_i/h_i$ and obtain

$$c_1 = 40, c_2 = 65/1.6 \approx 40.6, c_3 = 70/1.7 \approx 41.2,$$

$$c_4 = 75/1.7 \approx 44.1, c_5 = 80/1.8 \approx 44.4.$$

If $b \le c_1 = 40 = \min(c_i)$, then $f = \sum(w_i - ah_i)$ and its slope is

$$-\sum h_i = -h_1 - h_2 - h_3 - h_4 - h_5 = -8.3.$$

On the next interval, $c_1 \le b \le c_2$, the slope of f is

$$h_1 - h_2 - h_3 - h_4 - h_5 = -5.3.$$

On the next interval, $c_2 \le b \le c_3$, the slope of f is

$$h_1 + h_2 - h_3 - h_4 - h_5 = -2.1.$$

On the next interval, $c_3 \le b \le c_4$, the slope of f is

$$h_1 + h_2 + h_3 - h_4 - h_5 = 1.3.$$

For a bigger b, the slope is larger. So it is clear that the slope changes sign at $b = c_3$, hence the minimum is obtained at $b = c_3 = 70/1.7 \approx 41.2$. So our answer is $w = 70h/1.7$.

Remark. The first l^1-approximation problems appeared in connection with data on star movements. Bosovitch (about 1756), Laplace (1789), Gauss (1809), and Fourier (about 1822) proposed methods of solving those problems. In fact, Fourier considered also l^∞-approxima-tion, and he suggested a method of finding feasible solutions for an arbitrary system of linear constraints. Strangely enough, works on l^2-approximation, where finding the best fit reduces to solving a system of linear equations, appeared only in the nineteenth century (Legendre, Gauss).

The best l^∞-fit is also known as the least-absolute-deviation fit and the Chebyshev approximation.

The Best l^∞-fit

We reduce the optimization problem with the objective function $\|e\|_\infty \to \min$, where $e = (e_i) = w - AX$ to a linear program using an additional variable u such that $|e_i| \le u$ for all i. A similar trick was used when we reduced solving matrix games to linear programming. We obtain the following linear program with $n+1$ variables X, u and $2m$ linear constraints:

$$u \to \min, \quad -u \le w_i - A_iX \le u \text{ for } i = 1, \dots, m,$$

where A_i is the i-th row of the given matrix A. ■

Problem 23.5 Find the best l^∞-fit of the form $w = ah$ using the data from Problem 23.2.

Solution. We can reduce our problem to a linear program with two variables a, u and ten linear constraints and then solve this problem graphically or by the simplex method. Or we can plot the objective function $f = \max(w_i - ah_i)$ (Figure 23.6).

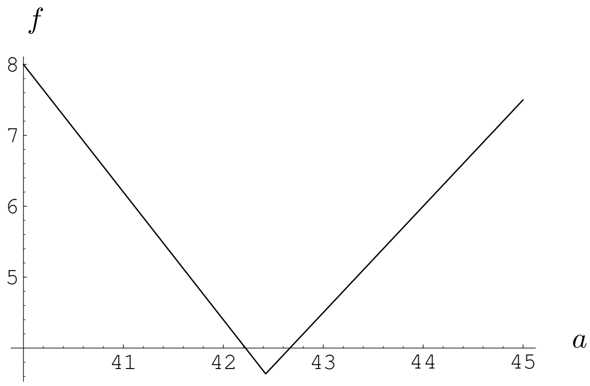

Figure 23.6. $f = \max(w_i - ah_i)$

We see that the optimal solution is $a \approx 42.4$. For a more precise answer, we can plot all five functions $|w_i - ah_i|$ near $a = 42.4$, f being the maximum of these five functions (Figure 23.7):

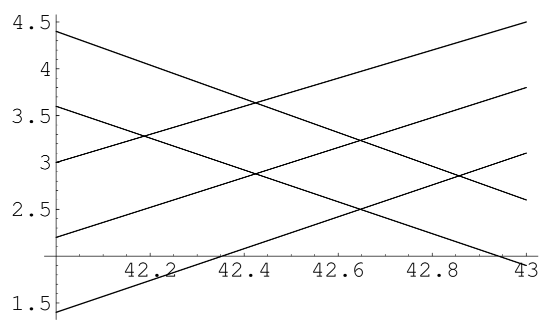

Figure 23.7.

We see that near $a = 42.4$ our objective function f is

$$\max(1.5a - 60, 80 - 1.8a),$$

so the optimal solution is

$$a = 140/3.3 \approx 42.42.$$

Problem 23.8. Find the half-life of a radioactive isotope using the following radiation measurements made every hour:

time in hours t	0	1	2	3	4
radiation level r	100	76	58	45	34.

Solution. We want to find the half-life λ in hours. Without any computations it is clear that $2 < \lambda < 3$. The radiation r in t hours should be $c2^{-t/\lambda}$. The given numbers should form a geometric progression up to measurement errors and round-offs. But the model $r = c2^{-t/\lambda}$ is not linear with respect to λ. Taking log, we can make it linear with respect to $w = \log_2 r, a = \log_2 c$ and $b = 1/\lambda$:

$$w = a - tb. \tag{23.9}$$

There are several options to proceed with. The simplest one is to compute an average of r_t/r_{t-1}. Thus, we obtain the following estimates for 2^b : the mean 1.30964, the midrange 1.30079, the central value 1.31966. They give the following estimates for λ: 2.56958, 2.6358, 2.49896.

Now we will work with the linear model (23.9). The best l^2-fit is obtained by solving the linear system $A^T AX = A^T w$ with

$$w = [\log_2 100, \log_2 76, \log_2 58, \log_2 45, \log_2 34]^T, \tag{23.10}$$

$$X = \begin{bmatrix} a \\ -b \end{bmatrix}, \; A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}.$$

So the system is

$$\begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix} \begin{bmatrix} a \\ -b \end{bmatrix} = \begin{bmatrix} \log_2 674424000 \\ \log_2 31133130272352000 \end{bmatrix} \approx \begin{bmatrix} 29.3291 \\ 54.7893 \end{bmatrix}.$$

Solving this, we obtain $b \approx 0.387$; hence $\lambda \approx 2.58$.

Let us now find the best l^∞-fit for the model (23.9). The corresponding linear program is

$$u \to \min, -u \le w_t - a + bt \le u \text{ for } t = 0, 1, 2, 3, 4,$$

where w_t are the entries of the column w in (23.10). This problem has three unknowns a, b, u and ten linear constraints. The optimal value for b is $b \approx 0.385259$ which gives $\lambda \approx 2.596$.

Let us now find the best l^1-fit for the model (23.9). The corresponding linear program is

$$u_0 + u_1 + u_2 + u_3 + u_4 \to \min,$$

$$-u_t \le w_t - a + bt \le u_t \text{ for } t = 0, 1, 2, 3, 4$$

where w_t are the entries of the column w in (23.10).

This problem has seven unknowns $a, b, u_0, u_1, u_2, u_3, u_4$ and ten linear constraints. The optimal value for b is $b \approx 0.389098$, which gives $\lambda \approx 2.570$.

Unless we know more about how the data were produced, it is hard to decide which method is better. The answer $\lambda = 2.55 \pm 0.05$ looks reasonable.

Problem 23.11. A radiation counter was calibrated using three samples: 1 mg of isotope A, 1 mg of isotope B, 1 mg of isotope C:

time in hours t	0	1	2	3	4
1 mg of isotope A	4100	510	64	8	1
1 mg of isotope B	1300	320	80	20	5
1 mg of isotope C	160	80	40	20	10.

After this, the counter was used to find contents of the isotopes in different samples. A sample S, consisting of the isotopes A, B, C and nonradioactive components, gave the following readings:

time in hours t	0	1	2	3	4
radiation level r	100	76	58	45	34.

Find the weights (in mg) of the isotopes A, B, C in the sample.

Solution. Let a, b, c be weights of A,B,C in the sample.

Unless we know something about nuclear interactions between isotopes, we assume that the radiation level is

$100 = 4100a + 1300b + 160c + e_0$ at $t = 0$,
$76 = 510a + 320b + 80c + e_1$ at $t = 1$,
$58 = 64a + 80b + 40c + e_2$ at $t = 2$,
$45 = 8a + 20b + 20c + e_3$ at $t = 3$,
$34 = a + 5b + 10c + e_4$ at $t = 4$

with small errors e_0, e_1, e_2, e_3, e_4.

The best l^p-fits are

$a \approx 0.12, b \approx -0.65, c \approx 2.74$ for $p = 2$,
$a \approx 0.13, b \approx -0.70, c \approx 2.89$ for $p = 1$,
$a \approx 0.16, b \approx -0.78, c \approx 2.96$ for $p = \infty$.

We see that the answer depends on the criterion. But the negative value for b in all three solutions is not acceptable. We could prevent this, imposing the sign restrictions on the unknowns. The problems with l^1- and l^∞-criteria would stay linear programs, but the additional constraints would take the l^2-problem from linear algebra to nonlinear programming.

But since the negative values for b are not so close to 0, this is a strong indication that something is wrong with our solutions, our data, or our assumptions. Speaking about the assumptions, could it be that an unexpected isotope is present in the sample?

Speaking about solutions, if a computer was used, did we introduce the data correctly according to the software specifications? Does the software have a bug that resulted in a wrong solution? Did the computer make a random mistake?

We describe now two different ways to find the best l^2-fit with the software package *Mathematica*. The first way uses the command "FindMinimum." We introduce data (with e0 = e_0, etc.):

e0 = -100 + 4100a + 1300b + 160c;
e1 = -76 + 510a + 320b + 80c;
e2 = -58 + 64a + 80b + 40c;
e3 = -45 + 8a + 20b + 20c;
e4 = -34 + a + 5b + 10c;

[the semicolons are used to suppress printing data back, and they allow us to enter the data in one block]. Then we type and enter

FindMinimum[e0ˆ 2 + e1ˆ 2 + e2ˆ 2 + e3ˆ 2 + e4ˆ 2,
$\{a, 1\}, \{b, 1\}, \{c, 1\}$]

which results in the following response:

{158.785, {a − > 0.124638, b − > −0.653399, c − > 2.74108}}

The first number in the response, 158.785, is the optimal value (± 0.005). The numbers 1 in

$\{a, 1\}, \{b, 1\}, \{c, 1\}$

indicate an initial point in an iterative procedure for searching for optimal solutions.

The second way is to reduce finding the least squares fit to solving a system of linear equations $A^T AX = A^T B$ where $X = [a, b, c]^T$ and

$$[A|B] = \begin{bmatrix} 4100 & 1300 & 160 & |100 \\ 510 & 320 & 80 & |76 \\ 64 & 80 & 160 & |58 \\ 8 & 20 & 20 & |45 \\ 1 & 5 & 10 & |34 \end{bmatrix}.$$

We input A, X, and B into *Mathematica* as follows:

X={a,b,c}; B={100, 76, 58, 45, 34}; A={{4100, 1300, 160}, {510, 320, 80}, {64, 80, 40}, {8, 20, 20}, {1, 5, 10}};

In *Mathematica*, the system $A^T AX = A^T B$ is

Transpose[A].A.X==Transpose[A].B

To solve it we use the command "Solve":

Solve[Transpose[A].A.X==Transpose[A].B,{a,b,c}]

The output is

$$\{\{\text{a}->\frac{69369996}{556572001},\text{b}->-\frac{10909908649}{16697160030},\text{c}->\frac{91536610621}{33394320060}\}\}$$

To get this in decimals, we input N[%] and get

{{ a −− > 0.124638, b −− > -0.653399, c −− > 2.74108}}.

This agrees with what we found with "FindMinimum." Note that *Mathematica* also has other commands to compute the least squares fits.

For additional checking we solve the problem with *Maple*. Here is how we input matrices into *Maple*:

```
A := array( [[4100, 1300, 160], [510, 320, 80],
[64, 80, 40], [8, 20, 20], [1, 5, 10]] );
B := array( [100, 76, 58, 45, 34] );
```

Here is how to compute the least squares fit:

```
with(linalg);
leastsqrs(A, B); .
```

Here is how the output looks:

```
[   69369996         -10909908649        91536610621   ]
[  - - - - -    ,    - - - - - - - -  ,  - - - - - - -  ]
[  556572001          16697160030        33394320060   ]
```

This agrees with the answer by *Mathematica*. So probably something is wrong with our data or our assumptions.

Problem 23.12. Here are data about SAT (the Scholastic Aptitude Test) and GPA (the Grade Point Average) of ten students at Oxbridge University:

SAT1	x	750	720	710	780	700	730	760	770	720	720
SAT2	y	740	730	710	770	720	740	770	760	710	730
GPA	z	3.5	3.4	3.6	3.7	3.2	3.2	3.8	3.7	3.5	3.4.

Find the best l^p-fit of the form $z = ax + by$ for $p = 1, 2, \infty$.

Solution. We set

$$A = \begin{bmatrix} 750 & 720 & 710 & 780 & 700 & 730 & 760 & 770 & 720 & 720 \\ 740 & 730 & 710 & 770 & 720 & 740 & 770 & 760 & 710 & 730 \end{bmatrix}^T,$$

$$b = [\,3.5, 3.4, 3.6, 3.7, 3.2, 3.2, 3.8, 3.7, 3.5, 3.4\,]^T\,,$$

$$e = A \begin{bmatrix} x \\ y \end{bmatrix} - b.$$

Our three optimization problems are

$$\|e\|_p \to \min \text{ for } p = 1, 2, \infty.$$

The three optimal solutions are:

$x \approx 0.008, y \approx -0.004$ for $p = 2$;
$x \approx 0.015, y \approx -0.010$ for $p = \infty$;
$x \approx 0.008, y \approx -0.003$ for $p = 1$.

The negative value here is not as impossible as in the previous example, but the conclusion that the high score in Part 2 of the SAT inhibits GPA seems to be questionable. So we can say that our model is not good or our data are not sufficient to come to any conclusion. There are zillions of models and computations that ended up in the trash rather than in publications. Some published models and computations also belong to trash.

In general, interpretation of results of computations is a very important part of applied mathematics. Details about collecting and handling data, computing with data, and interpreting results of computations can be found in textbooks on statistics.

Exercises

1. Find the least squares fit $w = a + bh$ for the data in Problem 22.7. Compare the optimal value for this fit with those for the least squares fit of the form $w = ah^2$.

2. Find the least squares solution of the linear system in Example 6.10. Recall that the system has no solutions. In general, the least squares solution $\hat{x}$ of a system $Ax = b$ is not a solution (i.e., $A\hat{x} \neq b$ in general). Rather, $A\hat{x}$ is as close to b as possible.

3. Find the least squares solution of the linear system

$$\begin{array}{c} \begin{array}{cc} a & \;\; b \end{array} \\ \begin{bmatrix} 3 & 4 \\ -1 & 5 \\ 3 & 0 \\ 1 & -7 \end{bmatrix} \end{array} \begin{array}{l} \\ = 1 \\ = 2 \\ = 4 \\ = 5 \end{array}$$

4. Find the least squares fit $w = a + bh^2$ for the data in Exercise 9 of §22.

5. Find the best l^1-fit of the form $w = ah^2$ to the data in Problem 23.2.

6. Find the best l^1-fit of the form $w = ah^3$ to the data in Problem 23.2.

7. Find the best l^1-fit of the form $w = ah+b$ to the data in Problem 23.2.

8. Compare the minimal value in Problem 23.2 with those in Exercises 5 and 6 and find which model gives us the best fit for our data.

9. Find the best l^∞-fit of the form $w = ah^2$ to the data in Problem 23.2.

10. Find the best l^∞-fit of the form $w = ah^3$ to the data in Problem 23.2.

11. Find the best l^∞-fit of the form $w = ah + b$ to the data in Problem 23.2.

12. Find the best l^2-fit of the form $p_n = an + b$ to the data in Exercise 12 in §22. Compare the result with the fit $p_n = n\log(n)$, where log means the natural logarithm (which has no parameters).

Remark. It is known that $p_n/(n\log(n)) \to 1$ as $n \to \infty$.

13. Find the best l^2-fit of the form $F_n = a2^{bn}$ to the data in Exercise 13 in §22. Compare the result with the fit $F_n = \alpha^{n+1}/\sqrt{5}$ where $\alpha = (\sqrt{5}+1)/2$, the golden section ratio.

Remark. It is known that $F_t = (\alpha^{t+1} - (-1/\alpha)^{t+1})/\sqrt{5}$ for all t.

14. Rewrite as a linear program:

$$|e_1| + 2|e_2| + \max[|e_3|, |e_4|] \to \min$$

subject to

$e_1 = 2x_1 + 3x_2 - 1,$
$e_2 = x_1 - 2x_2 - 2,$
$e_3 = -x_1 + x_2 + 3,$
$e_4 = x_1 - x_2 - 4,$
$e_1 \le |e_3| \le 4.$

Solve the program.

§24. More Examples

In the following examples we discuss what model to use and how to interpret the results of computations, and we pay little attention to how to solve the corresponding optimization problem, which can be reduced to linear programs as explained in §23. The data are not made up. They either come from Web sites or from mathematical research problems.

Example 24.1. *Time series.* Suppose you are interested in per capita chocolate consumption w (in grams) in Japan in 1995, but you know only what it was in 10 preceding years:

1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
1253	1313	1394	1535	1590	1535	1648	1626	1585	1499

How can you use these numbers (instead of more traditional things like tarot cards, which you would not expect in this book) to predict the number for 1995? Anybody who could find a good way to answer this kind of question could make a lot of money playing the stock market.

We will try a simple model, $w = ah + b$, where h is the year and w is the per capita chocolate consumption in grams. Maybe w depends in fact on the weather, per capita income, chocolate price, and/or health concerns, but suppose that we do not have any data about these factors. We decide to use the best l^p-fit for $p = 1, 2, \infty$ rather than other fits. Here are answers obtained by a computer together with the prediction $w_{11} = 1995a + b$ for 1995:

$a \approx 46.4, b \approx -90802.8, w_{11} \approx 1765.2$ for $p = 1$,

$a \approx 33.7, b \approx -65548.4, w_{11} \approx 1683.2$ for $p = 2$,

$a \approx 27.333, b \approx -52888.2, w_{11} \approx 1641.83$ for $p = \infty$.

It is up to the reader to pass judgment on whether those fits predicted sufficiently well the value 1566 g for 1995. All data were taken from the Web site

http://202.167.121.158/ebooks/jetro/November.html#01.

Figure 24.2 plots $w - 1253$ versus $h - 1984$ for 11 years together with the best l^∞-fit.

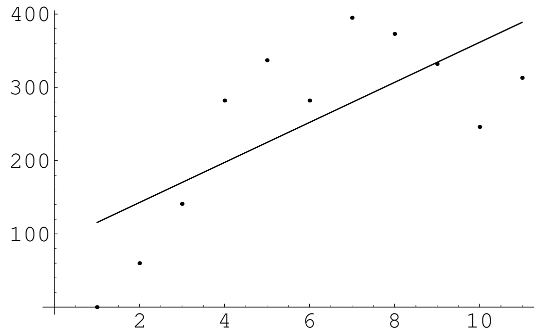

Figure 24.2. Per capita chocolate consumption w (in grams, reduced by 1253) in Japan in 1985–1995

Looking at the figure (which is usually a good thing to do before any computations) reveals that the model $w = ah + b$ does not work well. It seems that the trend in consumption is not a gradual growth with a constant rate a, but the rate $w_i - w_{i-1}$ of growth goes down and becomes negative in 1992. Since the trend does appear to be monotonous, a better model could be $w_i - w_{i-1} = a'h + b'$—that is, $w = ah^2 + bh + c$, where $a = a'/2, b = b' + a'/2$.

Since the model $w = ah^2+bh+c$ is more general than $w = bh+c$, the fits for 1985–1994 must be better or the same, but the predictions for 1995 need not be better. Here they are: $w_{11} \approx 1487.2$ for $p = 1$, $w_{11} \approx 1453.1$ for $p = 2$, $w_{11} \approx 1641.8$ for $p = \infty$.

It is always possible to find a model that predicts exactly the answer you want. But we would like a simple model which predicts an answer we do not know. See

http://202.167.121.158/ebooks/jetro/November.html#01

for the data interpretation.

Example 24.3. A statistician is interested in how often the digit 1 occurs in the number $\pi = 3.14159\ldots$. She computed 2 billion digits and recorded the positions after the decimal point where 1 occurs:

$$2, 4, 38, 41, 50, 69, 95, 96, 104, 111, 139, 149, 154, 155, 156, \ldots.$$

She asks us to approximate the position w_i as ai with unknown constant a. Her conjecture is that a must be close to 10.

What can we do with the given 15 numbers using the tools we learned? First we can compute the three averages of the numbers w_i/i $(1 \le i \le 15)$.

The mean is

$$2793047/270270 \approx 10.3343;$$

the midrange is

$$109/14 \approx 7.78571;$$

and the central value is

$$415/36 \approx 11.5278.$$

Thus, we found the best l^p-fits of the form $w_i/i = a$ for $p = 1, 2, \infty$. The mean is closest to 10.

Now we compute the best l^∞-fits for the model $w_i = ai$. They are $a \approx 11.56$ for $p = 1$, $a = 11.5$ for $p = 2$, $a = 11$ for $p = \infty$.

Should we try now to repeat these computations for 2 billion digits? We will have to wait a few years for more powerful computers.

Remark. How can we try to confirm or refute this conjecture? Notice that if you change or drop finitely many members of an infinite sequence, you do not change the limit (if it exists).

Example 24.4. *One-sided fits.*

Your monthly paycheck of \$5K is deposited electronically (available on first day of the month) to your money market account (MMA) where you get 3% interest. The interest is computed monthly on the minimal balance and credited to your account on December 31. You also have a checking account (CA) at the same bank with no interest paid. You use the CA to pay all your bills by mail and never use cash, except cash for post stamps, which you withdraw from your MMA. So you have to go to the bank often to transfer money from the MMA to the CA to pay for the mortgage, telephone, cable TV, cellular phone, Internet access, car insurance, credit cards, and other bills every month, totaling 10 checks, \$3K. On top of this, you pay five bills every quarter (in March, June, September, December) totaling \$2K per quarter and tax bills, \$2K and \$5K in April and \$4K in August. Finally, each year, on December 31 you write your last seven checks of the year for the rest of your annual income, \$5K minus postage total, and send them to your favorite mutual funds and charities. We assume that the initial balance at your MMA is sufficiently large so you need not worry about overdrawing.

Now your bank offers to change your routine: They will pay all your bills from your CA, without any fee. So you do not need to write 150 checks and addresses every year and you save on postage and checks. On the top of this, the bank offers to set up an automatic monthly transfer (on the first day of each month) from your MMA to your CA and one automatic yearly transfer (on January 1) at no cost to you. If you accept, you do not need to go to the bank or post office ever again. You have to decide about the amounts a and b of your monthly and annual transfers.

If you set the transfer amount to be all your salary, \$5K (that is, $a = 5, b = 0$) everything works well accept that you do not get any interest from your MMA account. For small b you get fines and other troubles for an insufficient balance in your CA. So what is the optimal solution? To see this we put all data in a table:

Month t	To pay in \$K	Transfer in \$K	CA balance e_t in \$K
1	3	$a+b$	$a+b-3$
2	3	a	$2a+b-6$
3	$3+2$	a	$3a+b-11$
4	$3+7$	a	$4a+b-21$
5	3	a	$5a+b-24$
6	$3+2$	a	$6a+b-29$
7	3	a	$7a+b-32$
8	$3+4$	a	$8a+b-39$
9	$3+2$	a	$9a+b-44$
10	3	a	$10a+b-47$
11	3	a	$11a+b-50$
12	$3+2+5$	a	$12a+b-60$

Table 24.5. Data and variables for Example 24.4

You are ready to state your optimization problem. You have two variables, a and b, subject to the conditions

$$a \geq 0, b \geq 0, 12a + b = 60.$$

In addition, you have 12 constraints $e_i \geq 0$ (see Table 24.5). You can state your objective without the MMA balances since maximizing interest in your MMA is equivalent to minimizing the total balance $e_1 + \cdots + e_{12}$ in your CA, where you get no interest.

We leave solving this particular problem to the reader (see Exercise 3 on the next page). We observe that the problem is similar to finding the best l^1-fit $\sum |e_i| \to \min$ with $e_i = ah_i + b - w_i$ but with additional constraints $e_i \geq 0$ and $a, b \geq 0$. These sign restrictions $a, b \geq 0$, $e_i \geq 0$ help to rewrite the optimization problem as a linear program in canonical form—namely, $\sum_i ah_i + b - w_i \to \min$, $ah_i + b - w_i \geq 0$ for all i.

Exercises

1. Here is the U.S. fresh strawberry production w (in millions of pounds) for 9 years. Predict the production in 1993 using the model $w = ah + b$ and the best l^p-fits for $p = 1, 2, \infty$. The data are from

http://www.nalusda.gov/pgdic/Strawberry/ers/ers.htm.

Compare your predictions with actual production 987.6. *Hint*: Some computer software does not like big numbers. Replace the year h by $h - 1988$ and the production w by $w - x_2$, where x_2 is its mean over 9 years.

1984	1985	1986	1987	1988	1989	1990	1991	1992
748.2	754.1	734.8	780.4	855.5	861.6	864.2	971.5	980.3

2. A student is interested in the number w of integer points $[x, y]$ in the disc $x^2 + y^2 \leq r^2$ of radius r. He computed w for some r:

r	1	2	3	4	5	6	7	8	9
w	5	13	29	45	81	113	149	197	253

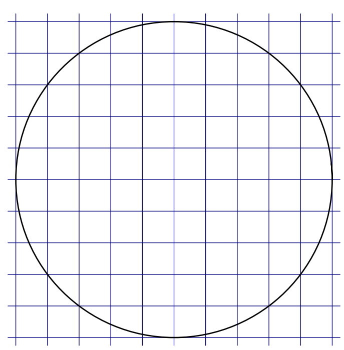

Figure 24.6. 81 integer points in the disc of radius 5

The student wants to approximate w by a simple formula $w = ar + b$ with constants a, b. But you feel that the area of the disc, πr^2 would be a better approximation, and hence the best l^2-fit of the form $w = ar^2$ should work even better for the numbers above.

Compute the best l^2-fit (the least squares fit) for both models, $w = ar + b$ and $w = ar^2$ and find which is better. Also compare both optimal values with

$$\sum_{i=1}^{9}(w_i - \pi i^2)^2.$$

3. Solve the optimization problem in Example 24.4. *Hint*: Use the graphical method.

4. Some passengers with confirmed reservations were denied boarding ("bumped") from their flights because the flights were oversold. The airlines oversell because they cannot be sure how many passengers will show up. Here are the year t and the numbers x and y of boarded (the second row, in millions) and bumped (the third row, in thousands) passengers for the domestic nonstop scheduled flights by the 10 largest U.S. air carriers. Data are taken from

http://www.bts.gov/btsprod/nts/Ch1_web/1-55.htm:

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
421	429	445	449	457	460	481	503	514	523
628	646	764	683	824	843	957	1072	1126	1070

Find the best l^p-fits with $p = 1, 2, \infty$ of the form

$$y = at + bx + c$$

for the data for 1990–1998 and use these fits to predict the number 1070 in 1999.