```
      Acol = A(:,col);
      [t, row] = max(-Acol./(b+mu*b0));
   else
      mu = mu_row;
      Arow = A(row,:);
      [s, col] = max(-Arow./(c+mu*c0));
   end
```

Finally, as part of every pivot we have to update b0 and c0:

```
   brow = b0(row);
   b0 = b0 - brow*Acol/a;
   b0(row) = -brow/a;

   ccol = c0(col);
   c0 = c0 - ccol*Arow/a;
   c0(col) = ccol/a;
```

The code was run 1000 times. Figure 12.4 shows the number of pivots plotted against the sum $m + n$. Just as we saw with the primal simplex method in Chapter 4, $m + n$ does not seem to be a good measure of problem size as many problems of a given size solve much more quickly than the more typical cases. Hence, there are a number of "outliers." Overlayed on the scatter plot are the L^1 and L^2 regression lines. While neither regression line follows what appears to the an upper line of points that seems to dominate the results, the L^1 is closer to that than is the L^2 line.

The result of the $L1$-regression is:

$$T \approx e^{-0.722} e^{1.12 \log(m+n)} = 0.486(m+n)^{1.12}.$$

The result of the $L2$-regression is:

$$T \approx e^{-0.606} e^{1.05 \log(m+n)} = 0.546(m+n)^{1.05}.$$

Finally, as in Chapter 4, $\min(m, n)$ is a better measure of problem size for these randomly generated problems. Figure 12.5 shows the same data plotted against $\min(m, n)$. In this case, both regression lines are about the same:

$$T \approx e^{-0.2} e^{1.46 \log(\min(m,n))} = 0.8 \min(m, n)^{1.46}.$$

Exercises

12.1 Find the L^2-regression line for the data shown in Figure 12.6.

12.2 Find the L^1-regression line for the data shown in Figure 12.6.

12.3 *Midrange*. Given a sorted set of real numbers, $\{b_1, b_2, \ldots, b_m\}$, show that the midrange, $\tilde{x} = (b_1 + b_m)/2$, minimizes the maximum deviation from the set of observations. That is,

$$\frac{1}{2}(b_1 + b_m) = \text{argmin}_{x \in \mathbb{R}} \max_i |x - b_i|.$$

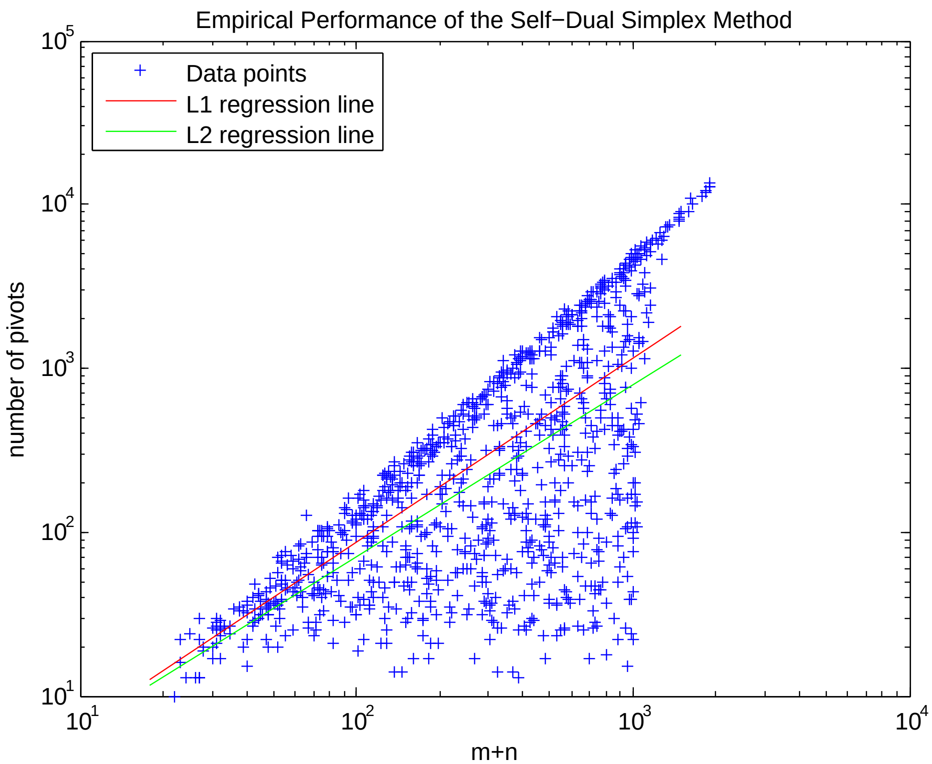

FIGURE 12.4. The parametric self-dual simplex method was used to solve 1000 problems known to have an optimal solution. Shown here is a log-log plot showing the number of pivots required to reach optimality plotted against $m+n$. Also shown are the L^1 and L^2 regression lines.

12.4 *Centroid*. Given a set of points $\{b_1, b_2, \ldots, b_m\}$ on the plane $\mathbb{R}^2$, show that the centroid

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} b_i$$

minimizes the sum of the squares of the distance to each point in the set. That is, $\bar{x}$ solves the following optimization problem:

$$\text{minimize} \sum_{i=1}^{m} \|x - b_i\|_2^2$$

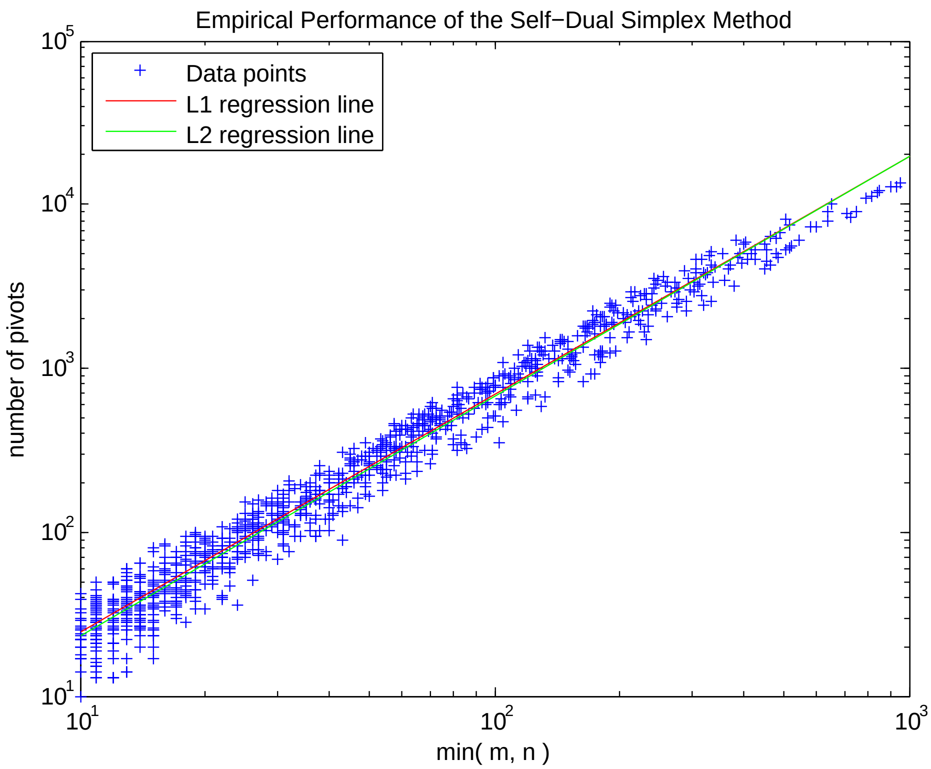

FIGURE 12.5. The parametric self-dual simplex method was used to solve 1000 problems known to have an optimal solution. Shown here is a log-log plot showing the number of pivots required to reach optimality plotted against $\min(m, n)$. In this case, the L^1 and L^2 regression lines are almost exactly on top of each other.

Note: Each data point b_i is a vector in $\mathbb{R}^2$ whose components are denoted, say, by b_{i1} and b_{i2}, and, as usual, the subscript 2 on the norm denotes the Euclidean norm. Hence,

$$\|x - b_i\|_2 = \sqrt{(x_1 - b_{i1})^2 + (x_2 - b_{i2})^2}.$$

12.5 *Facility Location.* A common problem is to determine where to locate a facility so that the distance from its customers is minimized. That is, given a set of points $\{b_1, b_2, \ldots, b_m\}$ on the plane $\mathbb{R}^2$, the problem is to find $\hat{x} =$

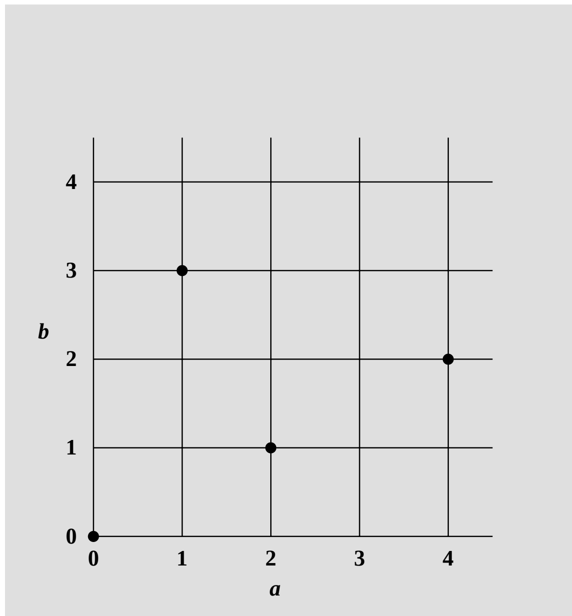

FIGURE 12.6. Four data points for a linear regression.

$(\hat{x}_1, \hat{x}_2)$ that solves the following optimization problem:

$$\text{minimize} \sum_{i=1}^{m} \|x - b_i\|_2.$$

As for L^1-regression, there is no explicit formula for $\hat{x}$, but an iterative scheme can be derived along the same lines as in Section 12.5. Derive an explicit formula for this iteration scheme.

12.6 *A Simple Steiner Tree*. Suppose there are only three customers in the facility location problem of the previous exercise. Suppose that the triangle formed by b_1, b_2, and b_3 has no angles greater than 120 degrees. Show that the solution $\hat{x}$ to the facility location problem is the unique point in the triangle from whose perspective the three customers are 120 degrees apart from each other. What is the solution if one of the angles, say, at vertex b_1, is more than 120 degrees?

12.7 *Sales Force Planning*. A distributor of office equipment finds that the business has seasonal peaks and valleys. The company uses two types of sales persons: (a) regular employees who are employed year-round and cost the

Jan	390	May	310	Sep	550
Feb	420	Jun	590	Oct	360
Mar	340	Jul	340	Nov	420
Apr	320	Aug	580	Dec	600 .

TABLE 12.2. Projected labor hours by month.

company \$17.50/hr (fully loaded for benefits and taxes) and (b) temporary employees supplied by an outside agency at a cost of \$25/hr. Projections for the number of hours of labor by month for the following year are shown in Table 12.2. Let a_i denote the number of hours of labor needed for month i and let x denote the number of hours per month of labor that will be handled by regular employees. To minimize total labor costs, one needs to solve the following optimization problem:

$$\text{minimize} \sum_i (25 \max(a_i - x, 0) + 17.50x).$$

(a) Show how to reformulate this problem as a linear programming problem.
(b) Solve the problem for the specific data given above.
(c) Use calculus to find a formula giving the optimal value for x.

12.8 *Acceleration Due to Gravity*. The law of gravity from classical physics says that an object dropped from a tall building will, in the absence of air resistance, have a constant rate of acceleration g so that the height x, as a function of time t, is given by

$$x(t) = -\frac{1}{2}gt^2.$$

Unfortunately, the effects of air resistance cannot be ignored. To include them, we assume that the object experiences a retarding force that is directly proportional to its speed. Letting $v(t)$ denote the velocity of the object at time t, the equations that describe the motion are then given by

$$\begin{aligned} x'(t) &= v(t), \qquad & t > 0, \qquad & x(0) = 0, \\ v'(t) &= -g - fv(t), \qquad & t > 0, \qquad & v(0) = 0 \end{aligned}$$

(f is the unknown constant of proportionality from the air resistance). These equations can be solved explicitly for x as a function of t:

$$x(t) = -\frac{g}{f^2}\left(e^{-ft} - 1 + ft\right)$$
$$v(t) = -\frac{g}{f}\left(1 - e^{-ft}\right).$$

It is clear from the equation for the velocity that the *terminal velocity* is g/f. It would be nice to be able to compute g by measuring this velocity, but this is not possible, since the terminal velocity involves both f and g. However, we can use the formula for $x(t)$ to get a two-parameter model from which we can compute both f and g. Indeed, if we assume that all measurements are taken after terminal velocity has been "reached" (i.e., when e^{-ft} is much smaller than 1), then we can write a simple linear expression relating position to time:

$$x = \frac{g}{f^2} - \frac{g}{f}t.$$

Now, in our experiments we shall set values of x (corresponding to specific positions below the drop point) and measure the time at which the object passes these positions. Since we prefer to write regression models with the observed variable expressed as a linear function of the control variables, let us rearrange the above expression so that t appears as a function of x:

$$t = \frac{1}{f} - \frac{f}{g}x.$$

Using this regression model and the data shown in Table 12.3, do an L^2-regression to compute estimates for $1/f$ and $-f/g$. From these estimates derive an estimate for g. If you have access to linear programming software, solve the problem using an L^1-regression and compare your answers.

12.9 *Iteratively Reweighted Least Squares*. Show that the sequence of iterates in the iteratively reweighted least squares algorithm produces a monotonically decreasing sequence of objective function values by filling in the details in the following outline. First, recall that the objective function for L^1-regression is given by

$$f(x) = \|b - Ax\|_1 = \sum_{i=1}^{m} \epsilon_i(x),$$

where

$$\epsilon_i(x) = \left| b_i - \sum_{j=1}^{n} a_{ij}x_j \right|.$$

Obs. Number	Position (meters)	Time (secs)
1	-10	3.72
2	-20	7.06
3	-30	10.46
4	-10	3.71
5	-20	7.00
6	-30	10.48
7	-10	3.67
8	-20	7.08
9	-30	10.33

TABLE 12.3. Time at which a falling object passes certain points.

Also, the function that defines the iterative scheme is given by

$$T(x) = \left(A^T E_x^{-1} A\right)^{-1} A^T E_x^{-1} b,$$

where E_x denotes the diagonal matrix with the vector $\epsilon(x)$ on its diagonal. Our aim is to show that

$$f(T(x)) < f(x).$$

In order to prove this inequality, let

$$g_x(z) = \sum_{i=1}^{m} \frac{\epsilon_i^2(z)}{\epsilon_i(x)} = \|E_x^{-1/2}(b - Az)\|_2^2.$$

(a) Use calculus to show that, for each x, $T(x)$ is a global minimum of g_x.
(b) Show that $g_x(x) = f(x)$.
(c) By writing

$$\epsilon_i(T(x)) = \epsilon_i(x) + (\epsilon_i(T(x)) - \epsilon_i(x))$$

and then substituting the right-hand expression into the definition of $g_x(T(x))$, show that

$$g_x(T(x)) \geq 2f(T(x)) - f(x).$$

(d) Combine the three steps above to finish the proof.

12.10 In our study of means and medians, we showed that the median of a collection of numbers, $b_1, b_2, \ldots, b_n$, is the number $\hat{x}$ that minimizes $\sum_j |x - b_j|$. Let μ be a real parameter.

(a) Give a statistical interpretation to the following optimization problem:

$$\text{minimize} \sum_j \left(|x - b_j| + \mu(x - b_j)\right).$$

Hint: the special cases $\mu = 0, \pm 1/2, \pm 1$ *might help clarify the general situation.*

(b) Express the above problem as a linear programming problem.

(c) The parametric simplex method can be used to solve families of linear programming problems indexed by a parameter μ (such as we have here). Starting at $\mu = \infty$ and proceeding to $\mu = -\infty$ one solves all of the linear programs with just a finite number of pivots. Use the parametric simplex method to solve the problems of the previous part in the case where $n = 4$ and $b_1 = 1$, $b_2 = 2$, $b_3 = 4$, and $b_4 = 8$.

(d) Now consider the general case. Write down the dictionary that appears in the k-th iteration and show by induction that it is correct.

12.11 Show that the L^∞-norm is just the maximum of the absolute values. That is,

$$\lim_{p \to \infty} \|x\|_p = \max_i |x_i|.$$

Notes

Gonin & Money (1989) and Dodge (1987) are two references on regression that include discussion of both L^2 and L^1 regression. The standard reference on L^1 regression is Bloomfield & Steiger (1983).

Several researchers, including Smale (1983), Borgwardt (1982), Borgwardt (1987*a*), Adler & Megiddo (1985), and Todd (1986), have studied the average number of iterations of the simplex method as a function of m and/or n. The model discussed in this chapter is similar to the sign-invariant model introduced by Adler & Berenguer (1981).