the equations using algorithm 11.2 ($3n^3$ versus $2n^3$), so algorithm 11.2 is the usual method of choice. While the matrix inverse appears in many formulas (such as the solution of a set of linear equations), it is *computed* far less often.

Sparse linear equations. Systems of linear equations with sparse coefficient matrix arise in many applications. By exploiting the sparsity of the coefficient matrix, these linear equations can be solved far more efficiently than by using the generic algorithm 11.2. One method is to use the same basic algorithm 11.2, replacing the QR factorization with a variant that handles sparse matrices (see page 190). The memory usage and complexity of these methods depends in a complicated way on the sparsity pattern of the coefficient matrix. In order, the memory usage is typically a modest multiple of $\mathbf{nnz}(A) + n$, the number of scalars required to specify the problem data A and b, which is typically much smaller than $n^2 + n$, the number of scalars required to store A and b if they are not sparse. The flop count for solving sparse linear equations is also typically closer in order to $\mathbf{nnz}(A)$ than n^3, the order when the matrix A is not sparse.

11.4 Examples

Polynomial interpolation. The 4-vector c gives the coefficients of a cubic polynomial,

$$p(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$$

(see pages 154 and 120). We seek the coefficients that satisfy

$$p(-1.1) = b_1, \qquad p(-0.4) = b_2, \qquad p(0.2) = b_3, \qquad p(0.8) = b_4.$$

We can express this as the system of 4 equations in 4 variables $Ac = b$, where

$$A = \begin{bmatrix} 1 & -1.1 & (-1.1)^2 & (-1.1)^3 \\ 1 & -0.4 & (-0.4)^2 & (-0.4)^3 \\ 1 & 0.2 & (0.2)^2 & (0.2)^3 \\ 1 & 0.8 & (0.8)^2 & (0.8)^3 \end{bmatrix},$$

which is a specific Vandermonde matrix (see (6.7)). The unique solution is $c = A^{-1}b$, where

$$A^{-1} = \begin{bmatrix} -0.5784 & 1.9841 & -2.1368 & 0.7310 \\ 0.3470 & 0.1984 & -1.4957 & 0.9503 \\ 0.1388 & -1.8651 & 1.6239 & 0.1023 \\ -0.0370 & 0.3492 & 0.7521 & -0.0643 \end{bmatrix}$$

(to 4 decimal places). This is illustrated in figure 11.1, which shows the two cubic polynomials that interpolate the two sets of points shown as filled circles and squares, respectively.

The columns of A^{-1} are interesting: They give the coefficients of a polynomial that evaluates to 0 at three of the points, and 1 at the other point. For example, the

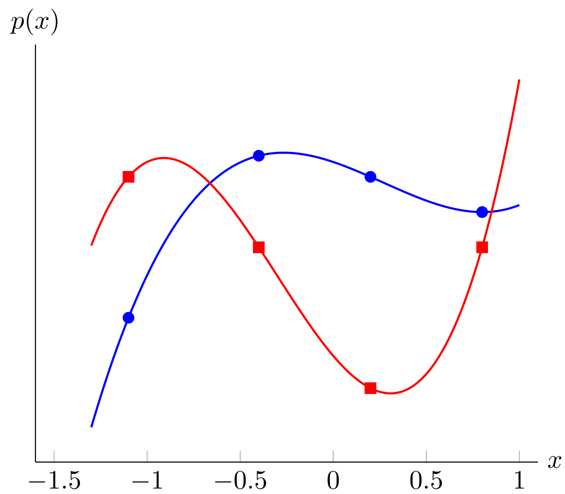

Figure 11.1 Cubic interpolants through two sets of points, shown as circles and squares.

first column of A^{-1}, which is $A^{-1}e_1$, gives the coefficients of the polynomial that has value 1 at -1.1, and value 0 at -0.4, 0.2, and 0.8. The four polynomials with coefficients given by the columns of A^{-1} are called the *Lagrange polynomials* associated with the points -1.1, -0.4, 0.2, 0.8. These are plotted in figure 11.2. (The Lagrange polynomials are named after the mathematician Joseph-Louis Lagrange, whose name will re-appear in several other contexts.)

The rows of A^{-1} are also interesting: The ith row shows how the values b_1, $\ldots$, b_4, the polynomial values at the points -1.1, -0.4, 0.2, 0.8, map into the ith coefficient of the polynomial, c_i. For example, we see that the coefficient c_4 is not very sensitive to the value of b_1 (since $(A^{-1})_{41}$ is small). We can also see that for each increase of one in b_4, the coefficient c_2 increases by around 0.95.

Balancing chemical reactions. (See page 154 for background.) We consider the problem of balancing the chemical reaction

$$a_1\mathrm{Cr_2O_7^{2-}} + a_2\mathrm{Fe^{2+}} + a_3\mathrm{H^+} \longrightarrow b_1\mathrm{Cr^{3+}} + b_2\mathrm{Fe^{3+}} + b_3\mathrm{H_2O},$$

where the superscript gives the charge of each reactant and product. There are 4 atoms (Cr, O, Fe, H) and charge to balance. The reactant and product matrices are (using the order just listed)

$$R = \begin{bmatrix} 2 & 0 & 0 \\ 7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 2 & 1 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 3 & 3 & 0 \end{bmatrix}.$$

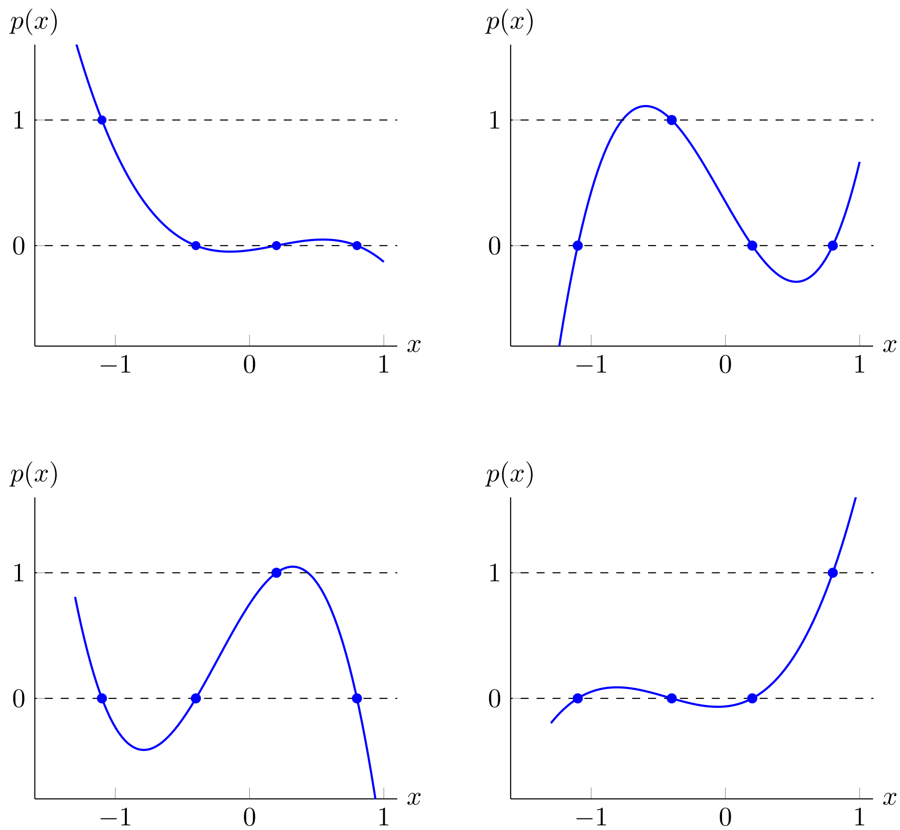

Figure 11.2 Lagrange polynomials associated with the points −1.1, −0.4, 0.2, 0.8.

Imposing the condition that $a_1 = 1$ we obtain a square set of 6 linear equations,

$$\begin{bmatrix} 2 & 0 & 0 & -1 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -2 \\ -2 & 2 & 1 & -3 & -3 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Solving these equations we obtain

$$a_1 = 1, \quad a_2 = 6, \quad a_3 = 14, \qquad b_1 = 2, \quad b_2 = 6, \quad b_3 = 7.$$

(Setting $a_1 = 1$ could have yielded fractional values for the other coefficients, but in this case, it did not.) The balanced reaction is

$$\mathrm{Cr_2O_7^{2-}} + 6\mathrm{Fe^{2+}} + 14\mathrm{H^+} \longrightarrow 2\mathrm{Cr^{3+}} + 6\mathrm{Fe^{3+}} + 7\mathrm{H_2O}.$$

Heat diffusion. We consider a diffusion system as described on page 155. Some of the nodes have fixed potential, *i.e.*, e_i is given; for the other nodes, the associated external source s_i is zero. This would model a thermal system in which some nodes are in contact with the outside world or a heat source, which maintains their temperatures (via external heat flows) at constant values; the other nodes are internal, and have no heat sources. This gives us a set of n additional equations:

$$e_i = e_i^{\mathrm{fix}}, \quad i \in \mathcal{P}, \qquad s_i = 0, \quad i \notin \mathcal{P},$$

where $\mathcal{P}$ is the set of indices of nodes with fixed potential. We can write these n equations in matrix-vector form as

$$Bs + Ce = d,$$

where B and C are the $n \times n$ diagonal matrices, and d is the n-vector given by

$$B_{ii} = \begin{cases} 0 & i \in \mathcal{P} \\ 1 & i \notin \mathcal{P}, \end{cases} \qquad C_{ii} = \begin{cases} 1 & i \in \mathcal{P} \\ 0 & i \notin \mathcal{P}, \end{cases} \qquad d_i = \begin{cases} e_i^{\mathrm{fix}} & i \in \mathcal{P} \\ 0 & i \notin \mathcal{P}. \end{cases}$$

We assemble the flow conservation, edge flow, and the boundary conditions into one set of $m + 2n$ equations in $m + 2n$ variables (f, s, e):

$$\begin{bmatrix} A & I & 0 \\ R & 0 & A^T \\ 0 & B & C \end{bmatrix} \begin{bmatrix} f \\ s \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ d \end{bmatrix}.$$

(The matrix A is the incidence matrix of the graph, and R is the resistance matrix; see page 155.) Assuming the coefficient matrix is invertible, we have

$$\begin{bmatrix} f \\ s \\ e \end{bmatrix} = \begin{bmatrix} A & I & 0 \\ R & 0 & A^T \\ 0 & B & C \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ d \end{bmatrix}.$$

This is illustrated with an example in figure 11.3. The graph is a 100×100 grid, with 10000 nodes, and edges connecting each node to its horizontal and vertical neighbors. The resistance on each edge is the same. The nodes at the top and bottom are held at zero temperature, and the three sets of nodes with rectilinear shapes are held at temperature one. All other nodes have zero source value.

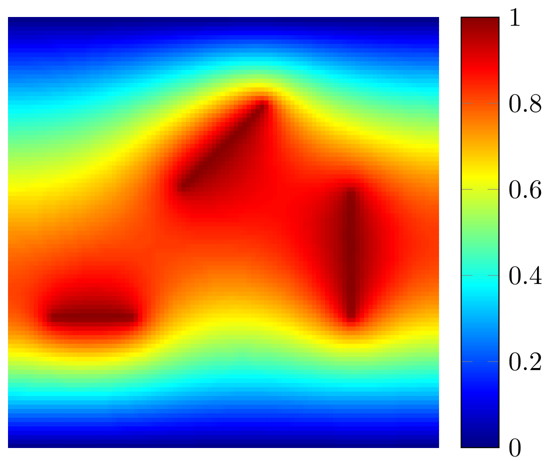

Figure 11.3 Temperature distribution on a 100×100 grid of nodes. Nodes in the top and bottom rows are held at zero temperature. The three sets of nodes with rectilinear shapes are held at temperature one.

11.5 Pseudo-inverse

Linearly independent columns and Gram invertibility. We first show that an $m \times n$ matrix A has linearly independent columns if and only if its $n \times n$ Gram matrix $A^T A$ is invertible.

First suppose that the columns of A are linearly independent. Let x be an n-vector which satisfies $(A^T A)x = 0$. Multiplying on the left by x^T we get

$$0 = x^T 0 = x^T (A^T A x) = x^T A^T A x = \|Ax\|^2,$$

which implies that $Ax = 0$. Since the columns of A are linearly independent, we conclude that $x = 0$. Since the only solution of $(A^T A)x = 0$ is $x = 0$, we conclude that $A^T A$ is invertible.

Now let's show the converse. Suppose the columns of A are linearly dependent, which means there is a nonzero n-vector x which satisfies $Ax = 0$. Multiply on the left by A^T to get $(A^T A)x = 0$. This shows that the Gram matrix $A^T A$ is singular.

Pseudo-inverse of square or tall matrix. We show here that if A has linearly independent columns (and therefore, is square or tall) then it has a left inverse. (We already have observed the converse, that a matrix with a left inverse has linearly independent columns.) Assuming A has linearly independent columns, we know that $A^T A$ is invertible. We now observe that the matrix $(A^T A)^{-1} A^T$ is a left inverse of A:

$$\left((A^T A)^{-1} A^T\right) A = (A^T A)^{-1} (A^T A) = I.$$

This particular left-inverse of A will come up in the sequel, and has a name,