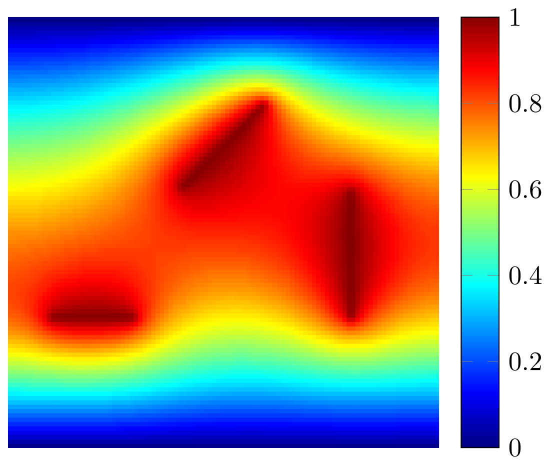

Figure 11.3 Temperature distribution on a 100×100 grid of nodes. Nodes in the top and bottom rows are held at zero temperature. The three sets of nodes with rectilinear shapes are held at temperature one.

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, the *pseudo-inverse* of A. It is denoted $A^\dagger$ (or A^+):

$$A^\dagger = (A^TA)^{-1}A^T. \tag{11.5}$$

The pseudo-inverse is also called the *Moore–Penrose inverse*, after the mathematicians Eliakim Moore and Roger Penrose.

When A is square, the pseudo-inverse $A^\dagger$ reduces to the ordinary inverse:

$$A^\dagger = (A^TA)^{-1}A^T = A^{-1}A^{-T}A^T = A^{-1}I = A^{-1}.$$

Note that this equation does not make sense (and certainly is not correct) when A is not square.

Pseudo-inverse of a square or wide matrix. Transposing all the equations, we can show that a (square or wide) matrix A has a right inverse if and only if its rows are linearly independent. Indeed, one right inverse is given by

$$A^T(AA^T)^{-1}. \tag{11.6}$$

(The matrix AA^T is invertible if and only if the rows of A are linearly independent.)

The matrix in (11.6) is also referred to as the pseudo-inverse of A, and denoted $A^\dagger$. The only possible confusion in defining the pseudo-inverse using the two different formulas (11.5) and (11.6) occurs when the matrix A is square. In this case, however, they both reduce to the ordinary inverse:

$$A^T(AA^T)^{-1} = A^TA^{-T}A^{-1} = A^{-1}.$$

Pseudo-inverse in other cases. The pseudo-inverse $A^\dagger$ is defined for any matrix, including the case when A is tall but its columns are linearly dependent, the case when A is wide but its rows are linearly dependent, and the case when A is square but not invertible. In these cases, however, it is not a left inverse, right inverse, or inverse, respectively. We mention it here since the reader may encounter it. (We will see what $A^\dagger$ means in these cases in exercise 15.11.)

Pseudo-inverse via QR factorization. The QR factorization gives a simple formula for the pseudo-inverse. If A is left-invertible, its columns are linearly independent and the QR factorization $A = QR$ exists. We have

$$A^TA = (QR)^T(QR) = R^TQ^TQR = R^TR,$$

so

$$A^\dagger = (A^TA)^{-1}A^T = (R^TR)^{-1}(QR)^T = R^{-1}R^{-T}R^TQ^T = R^{-1}Q^T.$$

We can compute the pseudo-inverse using the QR factorization, followed by back substitution on the columns of Q^T. (This is exactly the same as algorithm 11.3 when A is square and invertible.) The complexity of this method is $2n^2m$ flops (for the QR factorization), and mn^2 flops for the m back substitutions. So the total is $3mn^2$ flops.

Similarly, if A is right-invertible, the QR factorization $A^T = QR$ of its transpose exists. We have $AA^T = (QR)^T(QR) = R^TQ^TQR = R^TR$ and

$$A^\dagger = A^T(AA^T)^{-1} = QR(R^TR)^{-1} = QRR^{-1}R^{-T} = QR^{-T}.$$

We can compute it using the method described above, using the formula

$$(A^T)^\dagger = (A^\dagger)^T.$$

Solving over- and under-determined systems of linear equations. The pseudo-inverse gives us a method for solving over-determined and under-determined systems of linear equations, provided the columns of the coefficient matrix are linearly independent (in the over-determined case), or the rows are linearly independent (in the under-determined case). If the columns of A are linearly independent, and the over-determined equations $Ax = b$ have a solution, then $x = A^\dagger b$ is it. If the rows of A are linearly independent, the under-determined equations $Ax = b$ have a solution for any vector b, and $x = A^\dagger b$ is a solution.

Numerical example. We illustrate these ideas with a simple numerical example, using the 3×2 matrix A used in earlier examples on pages 199 and 201,

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix}.$$

This matrix has linearly independent columns, and QR factorization with (to 4 digits)

$$Q = \begin{bmatrix} -0.5883 & 0.4576 \\ 0.7845 & 0.5230 \\ 0.1961 & -0.7191 \end{bmatrix}, \qquad R = \begin{bmatrix} 5.0990 & 7.2563 \\ 0 & 0.5883 \end{bmatrix}.$$

It has pseudo-inverse (to 4 digits)

$$A^\dagger = R^{-1}Q^T = \begin{bmatrix} -1.2222 & -1.1111 & 1.7778 \\ 0.7778 & 0.8889 & -1.2222 \end{bmatrix}.$$

We can use the pseudo-inverse to check if the over-determined systems of equations $Ax = b$, with $b = (1, -2, 0)$, has a solution, and to find a solution if it does. We compute $x = A^\dagger(1, -2, 0) = (1, -1)$ and check whether $Ax = b$ holds. It does, so we have found the unique solution of $Ax = b$.