A triangular $n \times n$ matrix A has up to $n(n+1)/2$ nonzero entries, *i.e.*, around half its entries are zero. Triangular matrices are generally not considered sparse matrices, since their density is around 50%, but their special sparsity pattern will be important in the sequel.

6.3 Transpose, addition, and norm

6.3.1 Matrix transpose

If A is an $m \times n$ matrix, its *transpose*, denoted A^T (or sometimes A' or A^*), is the $n \times m$ matrix given by $(A^T)_{ij} = A_{ji}$. In words, the rows and columns of A are transposed in A^T. For example,

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}.$$

If we transpose a matrix twice, we get back the original matrix: $(A^T)^T = A$. (The superscript T in the transpose is the same one used to denote the inner product of two n-vectors; we will soon see how they are related.)

Row and column vectors. Transposition converts row vectors into column vectors and vice versa. It is sometimes convenient to express a row vector as a^T, where a is a column vector. For example, we might refer to the m rows of an $m \times n$ matrix A as $\tilde{a}_i^T, \ldots, \tilde{a}_m^T$, where $\tilde{a}_1, \ldots, \tilde{a}_m$ are (column) n-vectors. As an example, the second row of the matrix

$$\begin{bmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}$$

can be written as (the row vector) $(4, 0, 1)^T$.

It is common to extend concepts from (column) vectors to row vectors, by applying the concept to the transposed row vectors. We say that a collection of row vectors is linearly dependent (or independent) if their transposes (which are column vectors) are linearly dependent (or independent). For example, 'the rows of a matrix A are linearly independent' means that the columns of A^T are linearly independent. As another example, 'the rows of a matrix A are orthonormal' means that their transposes, the columns of A^T, are orthonormal. 'Clustering the rows of a matrix X' means clustering the columns of X^T.

Transpose of block matrix. The transpose of a block matrix has the simple form (shown here for a 2×2 block matrix)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^T = \begin{bmatrix} A^T & C^T \\ B^T & D^T \end{bmatrix},$$

where A, B, C, and D are matrices with compatible sizes. The transpose of a block matrix is the transposed block matrix, with each element transposed.

Document-term matrix. Consider a corpus (collection) of N documents, with word count vectors for a dictionary with n words. The *document-term* matrix associated with the corpus is the $N \times n$ matrix A, with A_{ij} the number of times word j appears in document i. The rows of the document-term matrix are $a_1^T, \ldots, a_N^T$, where the n-vectors $a_1, \ldots, a_N$ are the word count vectors for documents $1, \ldots, N$, respectively. The columns of the document-term matrix are also interesting. The jth column of A, which is an N-vector, gives the number of times word j appears in the corpus of N documents.

Data matrix. A collection of N n-vectors, for example feature n-vectors associated with N objects, can be given as an $n \times N$ matrix whose N columns are the vectors, as described on page 111. It is also common to describe this collection of vectors using the transpose of that matrix. In this case, we give the vectors as an $N \times n$ matrix X. Its ith row is x_i^T, the transpose of the ith vector. Its jth column gives the value of the jth entry (or feature) across the collection of N vectors. When an author refers to a *data matrix* or *feature matrix*, it can usually be determined from context (for example, its dimensions) whether they mean the data matrix organized by rows or columns.

Symmetric matrix. A square matrix A is *symmetric* if $A = A^T$, *i.e.*, $A_{ij} = A_{ji}$ for all i, j. Symmetric matrices arise in several applications. For example, suppose that A is the adjacency matrix of a graph or relation (see page 112). The matrix A is symmetric when the relation is symmetric, *i.e.*, whenever $(i, j) \in \mathcal{R}$, we also have $(j, i) \in \mathcal{R}$. An example is the *friend relation* on a set of n people, where $(i, j) \in \mathcal{R}$ means that person i and person j are friends. (In this case the associated graph is called the 'social network graph'.)

6.3.2 Matrix addition

Two matrices of the same size can be added together. The result is another matrix of the same size, obtained by adding the corresponding elements of the two matrices. For example,

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}.$$

Matrix subtraction is similar. As an example,

$$\begin{bmatrix} 1 & 6 \\ 9 & 3 \end{bmatrix} - I = \begin{bmatrix} 0 & 6 \\ 9 & 2 \end{bmatrix}.$$

(This gives another example where we have to figure out the size of the identity matrix. Since we can only add or subtract matrices of the same size, I refers to a 2×2 identity matrix.)

Properties of matrix addition. The following important properties of matrix addition can be verified directly from the definition. We assume here that A, B, and C are matrices of the same size.

- *Commutativity.* $A + B = B + A$.
- *Associativity.* $(A + B) + C = A + (B + C)$. We therefore write both as $A + B + C$.
- *Addition with zero matrix.* $A + 0 = 0 + A = A$. Adding the zero matrix to a matrix has no effect.
- *Transpose of sum.* $(A + B)^T = A^T + B^T$. The transpose of a sum of two matrices is the sum of their transposes.

6.3.3 Scalar-matrix multiplication

Scalar multiplication of matrices is defined in a similar way as for vectors, and is done by multiplying every element of the matrix by the scalar. For example

$$(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}.$$

As with scalar-vector multiplication, the scalar can also appear on the right. Note that $0\,A = 0$ (where the left-hand zero is the scalar zero, and the right-hand 0 is the zero matrix).

Several useful properties of scalar multiplication follow directly from the definition. For example, $(\beta A)^T = \beta(A^T)$ for a scalar β and a matrix A. If A is a matrix and β, γ are scalars, then

$$(\beta + \gamma)A = \beta A + \gamma A, \qquad (\beta\gamma)A = \beta(\gamma A).$$

It is useful to identify the symbols appearing in these two equations. The + symbol on the left of the left-hand equation is addition of scalars, while the + symbol on the right of the left-hand equation denotes matrix addition. On the left side of the right-hand equation we see scalar-scalar multiplication ($\alpha\beta$) and scalar-matrix multiplication; on the right we see two cases of scalar-matrix multiplication.

Finally, we mention that scalar-matrix multiplication has higher precedence than matrix addition, which means that we should carry out multiplication before addition (when there are no parentheses to fix the order). So the right-hand side of the left equation above is to be interpreted as $(\beta A) + (\gamma A)$.

6.3.4 Matrix norm

The norm of an $m \times n$ matrix A, denoted $\|A\|$, is the squareroot of the sum of the squares of its entries,

$$\|A\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}. \qquad (6.3)$$

This agrees with our definition for vectors when A is a vector, *i.e.*, $n = 1$. The norm of an $m \times n$ matrix is the norm of an mn-vector formed from the entries of the matrix (in any order). Like the vector norm, the matrix norm is a quantitative measure of the magnitude of a matrix. In some applications it is more natural to use the RMS values of the matrix entries, $\|A\|/\sqrt{mn}$, as a measure of matrix size. The RMS value of the matrix entries tells us the typical size of the entries, independent of the matrix dimensions.

The matrix norm (6.3) satisfies the properties of any norm, given on page 46. For any $m \times n$ matrix A, we have $\|A\| \geq 0$ (*i.e.*, the norm is nonnegative), and $\|A\| = 0$ only if $A = 0$ (definiteness). The matrix norm is nonnegative homogeneous: For any scalar γ and $m \times n$ matrix A, we have $\|\gamma A\| = |\gamma| \|A\|$. Finally, for any two $m \times n$ matrices A and B, we have the triangle inequality,

$$\|A + B\| \leq \|A\| + \|B\|.$$

(The plus symbol on the left-hand side is matrix addition, and the plus symbol on the right-hand side is addition of numbers.)

The matrix norm allows us to define the distance between two matrices as $\|A - B\|$. As with vectors, we can say that one matrix is close to, or near, another one if their distance is small. (What qualifies as small depends on the application.)

In this book we will only use the matrix norm (6.3). Several other norms of a matrix are commonly used, but are beyond the scope of this book. In contexts where other norms of a matrix are used, the norm (6.3) is called the *Frobenius norm*, after the mathematician Ferdinand Georg Frobenius, and is usually denoted with a subscript, as $\|A\|_F$.

One simple property of the matrix norm is $\|A\| = \|A^T\|$, *i.e.*, the norm of a matrix is the same as the norm of its transpose. Another one is

$$\|A\|^2 = \|a_1\|^2 + \cdots + \|a_n\|^2,$$

where $a_1, \ldots, a_n$ are the columns of A. In other words: The squared norm of a matrix is the sum of the squared norms of its columns.

6.4 Matrix-vector multiplication

If A is an $m \times n$ matrix and x is an n-vector, then the *matrix-vector product* $y = Ax$ is the m-vector y with elements

$$y_i = \sum_{k=1}^{n} A_{ik} x_k = A_{i1} x_1 + \cdots + A_{in} x_n, \quad i = 1, \ldots, m. \tag{6.4}$$

As a simple example, we have

$$\begin{bmatrix} 0 & 2 & -1 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} (0)(2) + (2)(1) + (-1)(-1) \\ (-2)(2) + (1)(1) + (1)(-1) \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}.$$