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

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

3-columns length 3 $n=3$

Matrix must have same number of columns as vector has length

Rows

Matrix $A \doteq m \times n$ $x \doteq n$ $n = n$

columns length

Row and column interpretations. We can express the matrix-vector product in terms of the rows or columns of the matrix. From (6.4) we see that y_i is the inner product of x with the ith row of A:

$$y_i = b_i^T x, \quad i = 1, \ldots, m,$$

Inner product of row w/ column

where b_i^T is the row i of A. The matrix-vector product can also be interpreted in terms of the columns of A. If a_k is the kth column of A, then $y = Ax$ can be written

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n.$$

This shows that $y = Ax$ is a linear combination of the columns of A; the coefficients in the linear combination are the elements of x.

General examples. In the examples below, A is an $m \times n$ matrix and x is an n-vector.

- *Zero matrix.* When $A = 0$, we have $Ax = 0$. In other words, $0x = 0$. (The left-hand 0 is an $m \times n$ matrix, and the right-hand zero is an m-vector.)
- *Identity.* We have $Ix = x$ for any vector x. (The identity matrix here has dimension $n \times n$.) In other words, multiplying a vector by the identity matrix gives the same vector.
- *Picking out columns and rows.* An important identity is $Ae_j = a_j$, the jth column of A. Multiplying a unit vector by a matrix 'picks out' one of the columns of the matrix. $A^T e_i$, which is an n-vector, is the ith row of A, transposed. (In other words, $(A^T e_i)^T$ is the ith row of A.)
- *Summing or averaging columns or rows.* The m-vector $A\mathbf{1}$ is the sum of the columns of A; its ith entry is the sum of the entries in the ith row of A. The m-vector $A(\mathbf{1}/n)$ is the average of the columns of A; its ith entry is the average of the entries in the ith row of A. In a similar way, $A^T\mathbf{1}$ is an n-vector, whose jth entry is the sum of the entries in the jth column of A.
- *Difference matrix.* The $(n-1) \times n$ matrix

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \qquad (6.5)$$

(where entries not shown are zero, and entries with diagonal dots are 1 or -1, continuing the pattern) is called the *difference matrix*. The vector Dx is the $(n-1)$-vector of differences of consecutive entries of x:

$$Dx = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix}.$$

Identity matrix has to be a Square matrix (n×n)

Zero matrix can be any dimension (m×n)

Difference matrix is a matrix of differences between consecutive entries

- *Running sum matrix.* The $n \times n$ matrix

Similar to difference matrix but is the sum instead

$$S = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \tag{6.6}$$

  is called the *running sum matrix*. The ith entry of the n-vector Sx is the sum of the first i entries of x:

Constantly summing values

$$Sx = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \\ \vdots \\ x_1 + \cdots + x_n \end{bmatrix}.$$

Application examples.

See the weight of values

- *Feature matrix and weight vector.* Suppose X is a feature matrix, where its N columns $x_1, \ldots, x_N$ are feature n-vectors for N objects or examples. Let the n-vector w be a *weight vector*, and let $s_i = x_i^T w$ be the score associated with object i using the weight vector w. Then we can write $s = X^T w$, where s is the N-vector of scores of the objects.

Similar to feature matrix

- *Portfolio return time series.* Suppose that R is a $T \times n$ asset return matrix, that gives the returns of n assets over T periods. A common trading strategy maintains constant investment weights given by the n-vector w over the T periods. For example, $w_4 = 0.15$ means that 15% of the total portfolio value is held in asset 4. (Short positions are denoted by negative entries in w.) Then Rw, which is a T-vector, is the time series of the portfolio returns over the periods $1, \ldots, T$.

  As an example, consider a portfolio of the 4 assets in table 6.1, with weights $w = (0.4, 0.3, -0.2, 0.5)$. The product $Rw = (0.00213, -0.00201, 0.00241)$ gives the portfolio returns over the three periods in the example.

- *Polynomial evaluation at multiple points.* Suppose the entries of the n-vector c are the coefficients of a polynomial p of degree $n-1$ or less:

$$p(t) = c_1 + c_2 t + \cdots + c_{n-1} t^{n-2} + c_n t^{n-1}.$$

  Let $t_1, \ldots, t_m$ be m numbers, and define the m-vector y as $y_i = p(t_i)$. Then we have $y = Ac$, where A is the $m \times n$ matrix

$$A = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-2} & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-2} & t_2^{n-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & t_m & \cdots & t_m^{n-2} & t_m^{n-1} \end{bmatrix}. \tag{6.7}$$

Each row of matrix is multiplied with coefficient vector to get polynomial

So multiplying a vector c by the matrix A is the same as evaluating a polynomial with coefficients c at m points. The matrix A in (6.7) comes up often, and is called a *Vandermonde matrix* (of degree $n-1$, at the points $t_1, \ldots, t_m$), named for the mathematician Alexandre-Théophile Vandermonde.

All generate vectors

- *Total price from multiple suppliers.* Suppose the $m \times n$ matrix P gives the prices of n goods from m suppliers (or in m different locations). If q is an n-vector of quantities of the n goods (sometimes called a *basket* of goods), then $c = Pq$ is an N-vector that gives the total cost of the goods, from each of the N suppliers.
- *Document scoring.* Suppose A in an $N \times n$ document-term matrix, which gives the word counts of a corpus of N documents using a dictionary of n words, so the rows of A are the word count vectors for the documents. Suppose that w in an n-vector that gives a set of weights for the words in the dictionary. Then $s = Aw$ is an N-vector that gives the scores of the documents, using the weights and the word counts. A search engine, for example, might choose w (based on the search query) so that the scores are predictions of relevance of the documents (to the search).
- *Audio mixing.* Suppose the k columns of A are vectors representing audio signals or tracks of length T, and w is a k-vector. Then $b = Aw$ is a T-vector representing the mix of the audio signals, with track weights given by the vector w.

Inner product. When a and b are n-vectors, a^Tb is exactly the inner product of a and b, obtained from the rules for transposing matrices and forming a matrix-vector product. We start with the (column) n-vector a, consider it as an $n \times 1$ matrix, and transpose it to obtain the n-row-vector a^T. Now we multiply this $1 \times n$ matrix by the n-vector b, to obtain the 1-vector a^Tb, which we also consider a scalar. So the notation a^Tb for the inner product is just a special case of matrix-vector multiplication.

Similar to vectors →

Linear dependence of columns. We can express the concepts of linear dependence and independence in a compact form using matrix-vector multiplication. The columns of a matrix A are linearly dependent if $Ax = 0$ for some $x \neq 0$. The columns of a matrix A are linearly independent if $Ax = 0$ implies $x = 0$.

Expansion in a basis. If the columns of A are a basis, which means A is square with linearly independent columns $a_1, \ldots, a_n$, then for any n-vector b there is a unique n-vector x that satisfies $Ax = b$. In this case the vector x gives the coefficients in the expansion of b in the basis $a_1, \ldots, a_n$.

Properties of matrix-vector multiplication. Matrix-vector multiplication satisfies several properties that are readily verified. First, it distributes across the vector argument: For any $m \times n$ matrix A and any n-vectors u and v, we have

$$A(u + v) = Au + Av.$$

↑

Distributes across vector addition

Matrix-vector multiplication, like ordinary multiplication of numbers, has higher precedence than addition, which means that when there are no parentheses to force the order of evaluation, multiplications are to be carried out before additions. This means that the right-hand side above is to be interpreted as $(Au) + (Av)$. The equation above looks innocent and natural, but must be read carefully. On the left-hand side, we first add the vectors u and v, which is the addition of n-vectors. We then multiply the resulting n-vector by the matrix A. On the right-hand side, we first multiply each of n-vectors by the matrix A (this is two matrix-vector multiplies); and then add the two resulting m-vectors together. The left- and right-hand sides of the equation above involve very different steps and operations, but the final result of each is the same m-vector.

Matrix-vector multiplication also distributes across the matrix argument: For any $m \times n$ matrices A and B, and any n-vector u, we have

Also distributes across matrix addition

$$(A + B)u = Au + Bu.$$

On the left-hand side the plus symbol is matrix addition; on the right-hand side it is vector addition.

Another basic property is, for any $m \times n$ matrix A, any n-vector u, and any scalar α, we have

Commutative with multiplication $\longrightarrow$

$$(\alpha A)u = \alpha(Au)$$

(and so we can write this as αAu). On the left-hand side, we have scalar-matrix multiplication, followed by matrix-vector multiplication; on the right-hand side, we start with matrix-vector multiplication, and then perform scalar-vector multiplication. (Note that we also have $\alpha Au = A(\alpha u)$.)

Input-output interpretation. We can interpret the relation $y = Ax$, with A an $m \times n$ matrix, as a mapping from the n-vector x to the m-vector y. In this context we might think of x as an input, and y as the corresponding output. From equation (6.4), we can interpret A_{ij} as the factor by which y_i depends on x_j. Some examples of conclusions we can draw are given below.

- If A_{23} is positive and large, then y_2 depends strongly on x_3, and increases as x_3 increases.
- If A_{32} is much larger than the other entries in the third row of A, then y_3 depends much more on x_2 than the other inputs.
- If A is square and lower triangular, then y_i only depends on $x_1, \ldots, x_i$.

6.5 Complexity

Computer representation of matrices. An $m \times n$ matrix is usually represented on a computer as an $m \times n$ array of floating point numbers, which requires $8mn$ bytes. In some software systems symmetric matrices are represented in a more efficient way, by only storing the upper triangular elements in the matrix, in some