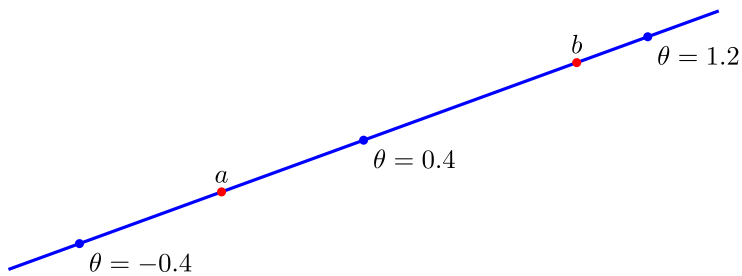

Figure 1.12 The affine combination $(1-\theta)a + \theta b$ for different values of θ. These points are on the line passing through a and b; for θ between 0 and 1, the points are on the line segment between a and b.

1.4 Inner product

The (standard) *inner product* (also called *dot product*) of two n-vectors is defined as the scalar

$$a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n,$$

the sum of the products of corresponding entries. (The origin of the superscript 'T' in the inner product notation $a^T b$ will be explained in chapter 6.) Some other notations for the inner product (that we will not use in this book) are $\langle a, b\rangle$, $\langle a|b\rangle$, (a, b), and $a \cdot b$. (In the notation used in this book, (a, b) denotes a stacked vector of length $2n$.) As you might guess, there is also a vector *outer product*, which we will encounter later, in §10.1. As a specific example of the inner product, we have

$$\left[\begin{array}{r} -1 \\ 2 \\ 2 \end{array}\right]^T \left[\begin{array}{r} 1 \\ 0 \\ -3 \end{array}\right] = (-1)(1) + (2)(0) + (2)(-3) = -7.$$

When $n = 1$, the inner product reduces to the usual product of two numbers.

Properties. The inner product satisfies some simple properties that are easily verified from the definition. If a, b, and c are vectors of the same size, and γ is a scalar, we have the following.

- *Commutativity.* $a^T b = b^T a$. The order of the two vector arguments in the inner product does not matter.
- *Associativity with scalar multiplication.* $(\gamma a)^T b = \gamma(a^T b)$, so we can write both as $\gamma a^T b$.
- *Distributivity with vector addition.* $(a+b)^T c = a^T c + b^T c$. The inner product can be distributed across vector addition.

These can be combined to obtain other identities, such as $a^T(\gamma b) = \gamma(a^T b)$, or $a^T(b + \gamma c) = a^T b + \gamma a^T c$. As another useful example, we have, for any vectors a, b, c, d of the same size,

$$(a+b)^T(c+d) = a^T c + a^T d + b^T c + b^T d.$$

This formula expresses an inner product on the left-hand side as a sum of four inner products on the right-hand side, and is analogous to expanding a product of sums in algebra. Note that on the left-hand side, the two addition symbols refer to vector addition, whereas on the right-hand side, the three addition symbols refer to scalar (number) addition.

General examples.

- *Unit vector.* $e_i^T a = a_i$. The inner product of a vector with the ith standard unit vector gives (or 'picks out') the ith element a.
- *Sum.* $\mathbf{1}^T a = a_1 + \cdots + a_n$. The inner product of a vector with the vector of ones gives the sum of the elements of the vector.
- *Average.* $(\mathbf{1}/n)^T a = (a_1 + \cdots + a_n)/n$. The inner product of an n-vector with the vector $\mathbf{1}/n$ gives the average or mean of the elements of the vector. The average of the entries of a vector is denoted by $\mathbf{avg}(x)$. The Greek letter μ is a traditional symbol used to denote the average or mean.
- *Sum of squares.* $a^T a = a_1^2 + \cdots + a_n^2$. The inner product of a vector with itself gives the sum of the squares of the elements of the vector.
- *Selective sum.* Let b be a vector all of whose entries are either 0 or 1. Then $b^T a$ is the sum of the elements in a for which $b_i = 1$.

Block vectors. If the vectors a and b are block vectors, and the corresponding blocks have the same sizes (in which case we say they *conform*), then we have

$$a^T b = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}^T \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} = a_1^T b_1 + \cdots + a_k^T b_k.$$

The inner product of block vectors is the sum of the inner products of the blocks.

Applications. The inner product is useful in many applications, a few of which we list here.

- *Co-occurrence.* If a and b are n-vectors that describe occurrence, *i.e.*, each of their elements is either 0 or 1, then $a^T b$ gives the total number of indices for which a_i and b_i are both one, that is, the total number of co-occurrences. If we interpret the vectors a and b as describing subsets of n objects, then $a^T b$ gives the number of objects in the intersection of the two subsets. This is illustrated in figure 1.13, for two subsets A and B of 7 objects, labeled $1, \ldots, 7$, with corresponding occurrence vectors

  $$a = (0, 1, 1, 1, 1, 1, 1), \qquad b = (1, 0, 1, 0, 1, 0, 0).$$

  Here we have $a^T b = 2$, which is the number of objects in both A and B (*i.e.*, objects 3 and 5).

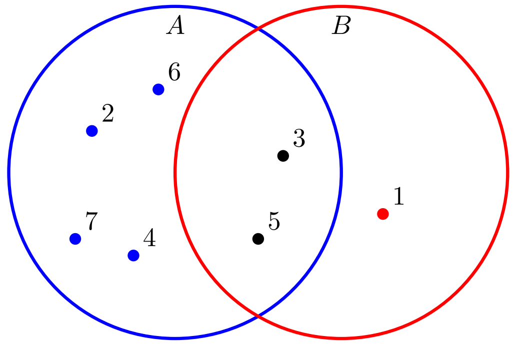

Figure 1.13 Two sets A and B, containing seven objects.

- *Weights, features, and score.* When the vector f represents a set of features of an object, and w is a vector of the same size (often called a *weight vector*), the inner product $w^T f$ is the sum of the feature values, scaled (or weighted) by the weights, and is sometimes called a *score*. For example, if the features are associated with a loan applicant (*e.g.*, age, income, ...), we might interpret $s = w^T f$ as a credit score. In this example we can interpret w_i as the weight given to feature i in forming the score.

- *Price-quantity.* If p represents a vector of prices of n goods, and q is a vector of quantities of the n goods (say, the bill of materials for a product), then their inner product $p^T q$ is the total cost of the goods given by the vector q.

- *Speed-time.* A vehicle travels over n segments with constant speed in each segment. Suppose the n-vector s gives the speed in the segments, and the n-vector t gives the times taken to traverse the segments. Then $s^T t$ is the total distance traveled.

- *Probability and expected values.* Suppose the n-vector p has nonnegative entries that sum to one, so it describes a set of proportions among n items, or a set of probabilities of n outcomes, one of which must occur. Suppose f is another n-vector, where we interpret f_i as the value of some quantity if outcome i occurs. Then $f^T p$ gives the expected value or mean of the quantity, under the probabilities (or fractions) given by p.

- *Polynomial evaluation.* Suppose the n-vector c represents the coefficients of a polynomial p of degree $n-1$ or less:

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

  Let t be a number, and let $z = (1, t, t^2, \ldots, t^{n-1})$ be the n-vector of powers of t. Then $c^T z = p(t)$, the value of the polynomial p at the point t. So the inner product of a polynomial coefficient vector and vector of powers of a number evaluates the polynomial at the number.

- *Discounted total.* Let c be an n-vector representing a cash flow, with c_i the cash received (when $c_i > 0$) in period i. Let d be the n-vector defined as

$$d = (1, 1/(1+r), \ldots, 1/(1+r)^{n-1}),$$

where $r \geq 0$ is an interest rate. Then

$$d^T c = c_1 + c_2/(1+r) + \cdots + c_n/(1+r)^{n-1}$$

is the discounted total of the cash flow, *i.e.*, its *net present value* (NPV), with interest rate r.

- *Portfolio value.* Suppose s is an n-vector representing the holdings in shares of a portfolio of n different assets, with negative values meaning short positions. If p is an n-vector giving the prices of the assets, then $p^T s$ is the total (or net) value of the portfolio.

- *Portfolio return.* Suppose r is the vector of (fractional) returns of n assets over some time period, *i.e.*, the asset relative price changes

$$r_i = \frac{p_i^{\mathrm{final}} - p_i^{\mathrm{initial}}}{p_i^{\mathrm{initial}}}, \quad i = 1, \ldots, n,$$

where p_i^{initial} and p_i^{final} are the (positive) prices of asset i at the beginning and end of the investment period. If h is an n-vector giving our portfolio, with h_i denoting the dollar value of asset i held, then the inner product $r^T h$ is the total return of the portfolio, in dollars, over the period. If w represents the fractional (dollar) holdings of our portfolio, then $r^T w$ gives the total return of the portfolio. For example, if $r^T w = 0.09$, then our portfolio return is 9%. If we had invested \$10000 initially, we would have earned \$900.

- *Document sentiment analysis.* Suppose the n-vector x represents the histogram of word occurrences in a document, from a dictionary of n words. Each word in the dictionary is assigned to one of three sentiment categories: *Positive*, *Negative*, and *Neutral*. The list of positive words might include 'nice' and 'superb'; the list of negative words might include 'bad' and 'terrible'. Neutral words are those that are neither positive nor negative. We encode the word categories as an n-vector w, with $w_i = 1$ if word i is positive, with $w_i = -1$ if word i is negative, and $w_i = 0$ if word i is neutral. The number $w^T x$ gives a (crude) measure of the sentiment of the document.

1.5 Complexity of vector computations

Computer representation of numbers and vectors. Real numbers are stored in computers using *floating point format*, which represents a real number using a block of 64 *bits* (0s and 1s), or 8 *bytes* (groups of 8 bits). Each of the 2^{64} possible sequences of bits corresponds to a specific real number. The floating point numbers