Chapter 8

Linear equations

In this chapter we consider vector-valued linear and affine functions, and systems of linear equations.

8.1 Linear and affine functions

Vector-valued functions of vectors. The notation $f : \mathbf{R}^n \to \mathbf{R}^m$ means that f is a function that maps real n-vectors to real m-vectors. The value of the function f, evaluated at an n-vector x, is an m-vector $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$. Each of the components f_i of f is itself a scalar-valued function of x. As with scalar-valued functions, we sometimes write $f(x) = f(x_1, x_2, \ldots, x_n)$ to emphasize that f is a function of n scalar arguments. We use the same notation for each of the components of f, writing $f_i(x) = f_i(x_1, x_2, \ldots, x_n)$ to emphasize that f_i is a function mapping the scalar arguments $x_1, \ldots, x_n$ into a scalar.

The matrix-vector product function. Suppose A is an $m \times n$ matrix. We can define a function $f : \mathbf{R}^n \to \mathbf{R}^m$ by $f(x) = Ax$. The inner product function $f : \mathbf{R}^n \to \mathbf{R}$, defined as $f(x) = a^T x$, discussed in §2.1, is the special case with $m = 1$.

Superposition and linearity. The function $f : \mathbf{R}^n \to \mathbf{R}^m$, defined by $f(x) = Ax$, is *linear*, *i.e.*, it satisfies the superposition property:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{8.1}$$

holds for all n-vectors x and y and all scalars α and β. It is a good exercise to parse this simple looking equation, since it involves overloading of notation. On the left-hand side, the scalar-vector multiplications αx and βy involve n-vectors, and the sum $\alpha x + \beta y$ is the sum of two n-vectors. The function f maps n-vectors to m-vectors, so $f(\alpha x + \beta y)$ is an m-vector. On the right-hand side, the scalar-vector multiplications and the sum are those for m-vectors. Finally, the equality sign is equality between two m-vectors.

We can verify that superposition holds for f using properties of matrix-vector and scalar-vector multiplication:

$$\begin{array}{rcl} f(\alpha x+\beta y) & = & A(\alpha x+\beta y) \\ & = & A(\alpha x)+A(\beta y) \\ & = & \alpha(Ax)+\beta(Ay) \\ & = & \alpha f(x)+\beta f(y) \end{array}$$

Thus we can associate with every matrix A a linear function $f(x)=Ax$.

The converse is also true. Suppose f is a function that maps n-vectors to m-vectors, and is linear, *i.e.*, (8.1) holds for all n-vectors x and y and all scalars α and β. Then there exists an $m\times n$ matrix A such that $f(x)=Ax$ for all x. This can be shown in the same way as for scalar-valued functions in §2.1, by showing that if f is linear, then

$$f(x)=x_1f(e_1)+x_2f(e_2)+\cdots+x_nf(e_n), \tag{8.2}$$

where e_k is the kth unit vector of size n. The right-hand side can also be written as a matrix-vector product Ax, with

$$A=\left[\begin{array}{cccc} f(e_1) & f(e_2) & \cdots & f(e_n) \end{array}\right].$$

The expression (8.2) is the same as (2.3), but here $f(x)$ and $f(e_k)$ are vectors. The implications are exactly the same: A linear vector-valued function f is completely characterized by evaluating f at the n unit vectors $e_1,\ldots,e_n$.

As in §2.1 it is easily shown that the matrix-vector representation of a linear function is unique. If $f:\mathbf{R}^n\rightarrow\mathbf{R}^m$ is a linear function, then there exists exactly one matrix A such that $f(x)=Ax$ for all x.

Examples of linear functions. In the examples below we define functions f that map n-vectors x to n-vectors $f(x)$. Each function is described in words, in terms of its effect on an arbitrary x. In each case we give the associated matrix multiplication representation.

- *Negation.* f changes the sign of x: $f(x)=-x$.

  Negation can be expressed as $f(x)=Ax$ with $A=-I$.

- *Reversal.* f reverses the order of the elements of x: $f(x)=(x_n,x_{n-1},\ldots,x_1)$.

  The reversal function can be expressed as $f(x)=Ax$ with

  $$A=\left[\begin{array}{cccc} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \iddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{array}\right].$$

  (This is the $n\times n$ identity matrix with the order of its columns reversed. It is the *reverser matrix* introduced in §7.2.)

- *Running sum.* f forms the running sum of the elements in x:

$$f(x) = (x_1,\, x_1 + x_2,\, x_1 + x_2 + x_3,\, \ldots,\, x_1 + x_2 + \cdots + x_n).$$

The running sum function can be expressed as $f(x) = Ax$ with

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix},$$

i.e., $A_{ij} = 1$ if $i \geq j$ and $A_{ij} = 0$ otherwise. This is the running sum matrix defined in (6.6).

- *De-meaning.* f subtracts the mean from each entry of a vector x: $f(x) = x - \mathbf{avg}(x)\mathbf{1}$.

The de-meaning function can be expressed as $f(x) = Ax$ with

$$A = \begin{bmatrix} 1 - 1/n & -1/n & \cdots & -1/n \\ -1/n & 1 - 1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1 - 1/n \end{bmatrix}.$$

Examples of functions that are not linear. Here we list some examples of functions f that map n-vectors x to n-vectors $f(x)$ that are *not* linear. In each case we show a superposition counterexample.

- *Absolute value.* f replaces each element of x with its absolute value: $f(x) = (|x_1|, |x_2|, \ldots, |x_n|)$.

The absolute value function is not linear. For example, with $n = 1$, $x = 1$, $y = 0$, $\alpha = -1$, $\beta = 0$, we have

$$f(\alpha x + \beta y) = 1 \neq \alpha f(x) + \beta f(y) = -1,$$

so superposition does not hold.

- *Sort.* f sorts the elements of x in decreasing order.

The sort function is not linear (except when $n = 1$, in which case $f(x) = x$). For example, if $n = 2$, $x = (1, 0)$, $y = (0, 1)$, $\alpha = \beta = 1$, then

$$f(\alpha x + \beta y) = (1, 1) \neq \alpha f(x) + \beta f(y) = (2, 0).$$

Affine functions. A vector-valued function $f : \mathbf{R}^n \to \mathbf{R}^m$ is called affine if it can be expressed as $f(x) = Ax + b$, where A is an $m \times n$ matrix and b is an m-vector. It can be shown that a function $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine if and only if

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

holds for all n-vectors x, y, and all scalars α, β that satisfy $\alpha + \beta = 1$. In other words, superposition holds for affine combinations of vectors. (For linear functions, superposition holds for any linear combinations of vectors.)

The matrix A and the vector b in the representation of an affine function as $f(x) = Ax + b$ are unique. These parameters can be obtained by evaluating f at the vectors $0, e_1, \ldots, e_n$, where e_k is the kth unit vector in $\mathbf{R}^n$. We have

$$A = \left[\begin{array}{cccc} f(e_1) - f(0) & f(e_2) - f(0) & \cdots & f(e_n) - f(0) \end{array} \right], \qquad b = f(0).$$

Just like affine scalar-valued functions, affine vector-valued functions are often called linear, even though they are linear only when the vector b is zero.

8.2 Linear function models

Many functions or relations between variables that arise in natural science, engineering, and social sciences can be *approximated* as linear or affine functions. In these cases we refer to the linear function relating the two sets of variables as a *model* or an *approximation*, to remind us that the relation is only an approximation, and not exact. We give a few examples here.

- *Price elasticity of demand.* Consider n goods or services with prices given by the n-vector p, and demands for the goods given by the n-vector d. A change in prices will induce a change in demands. We let δ^{price} be the n-vector that gives the fractional change in the prices, *i.e.*, $\delta_i^{\mathrm{price}} = (p_i^{\mathrm{new}} - p_i)/p_i$, where p^{new} is the n-vector of new (changed) prices. We let δ^{dem} be the n-vector that gives the fractional change in the product demands, *i.e.*, $\delta_i^{\mathrm{dem}} = (d_i^{\mathrm{new}} - d_i)/d_i$, where d^{new} is the n-vector of new demands. A linear demand elasticity model relates these vectors as $\delta^{\mathrm{dem}} = E^{\mathrm{d}}\delta^{\mathrm{price}}$, where E^{d} is the $n \times n$ *demand elasticity matrix.* For example, suppose $E_{11}^{\mathrm{d}} = -0.4$ and $E_{21}^{\mathrm{d}} = 0.2$. This means that a 1% increase in the price of the first good, with other prices kept the same, will cause demand for the first good to drop by 0.4%, and demand for the second good to increase by 0.2%. (In this example, the second good is acting as a *partial substitute* for the first good.)

- *Elastic deformation.* Consider a steel structure like a bridge or the structural frame of a building. Let f be an n-vector that gives the forces applied to the structure at n specific places (and in n specific directions), sometimes called a *loading.* The structure will deform slightly due to the loading. Let d be an m-vector that gives the displacements (in specific directions) of m points in the structure, due to the load, *e.g.*, the amount of sag at a specific point on a bridge. For small displacements, the relation between displacement and loading is well approximated as linear: $d = Cf$, where C is the $m \times n$ *compliance matrix.* The units of the entries of C are m/N.