actual value of the associated response variables, $y^{(1)}, \ldots, y^{(N)}$, then our *prediction errors* or *residuals* are

$$r^{(i)} = y^{(i)} - \hat{y}^{(i)}, \quad i = 1, \ldots, N.$$

(Some authors define the prediction errors as $\hat{y}^{(i)} - y^{(i)}$.)

We can express this using compact matrix-vector notation. We form the $n \times N$ feature matrix X with columns $x^{(1)}, \ldots, x^{(N)}$. We let y^{d} denote the N-vector whose entries are the actual values of the response for the N examples. (The superscript 'd' stands for 'data'.) We let $\hat{y}^{\mathrm{d}}$ denote the N-vector of regression model predictions for the N examples, and we let r^{d} denote the N-vector of residuals or prediction errors. We can then express the regression model predictions for this data set in matrix-vector form as

$$\hat{y}^{\mathrm{d}} = X^T\beta + v\mathbf{1}.$$

The vector of N prediction errors for the examples is given by

$$r^{\mathrm{d}} = y^{\mathrm{d}} - \hat{y}^{\mathrm{d}} = y^{\mathrm{d}} - X^T\beta - v\mathbf{1}.$$

We can include the offset v in the regression model by including an additional feature equal to one as the first entry of each feature vector:

$$\hat{y}^{\mathrm{d}} = \begin{bmatrix} \mathbf{1}^T \\ X \end{bmatrix}^T \begin{bmatrix} v \\ \beta \end{bmatrix} = \tilde{X}^T\tilde{\beta},$$

where $\tilde{X}$ is the new feature matrix, with a new first row of ones, and $\tilde{\beta} = (v, \beta)$ is the vector of regression model parameters. This is often written without the tildes, as $\hat{y}^{\mathrm{d}} = X^T\beta$, by simply including the feature one as the first feature.

The equation above shows that the N-vector of predictions for the N examples is a linear function of the model parameters (v, β). The N-vector of prediction errors is an affine function of the model parameters.

8.3 Systems of linear equations

Consider a set (also called a system) of m linear equations in n variables or unknowns $x_1, \ldots, x_n$:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= b_m. \end{aligned}$$

The numbers A_{ij} are called the *coefficients* in the linear equations, and the numbers b_i are called the *right-hand sides* (since by tradition, they appear on the right-hand

side of the equation). These equations can be written succinctly in matrix notation as

$$Ax = b. \tag{8.5}$$

In this context, the $m \times n$ matrix A is called the *coefficient matrix*, and the m-vector b is called the *right-hand side*. An n-vector x is called a *solution* of the linear equations if $Ax = b$ holds. A set of linear equations can have no solutions, one solution, or multiple solutions.

Examples.

- The set of linear equations

$$x_1 + x_2 = 1, \quad x_1 = -1, \quad x_1 - x_2 = 0$$

  is written as $Ax = b$ with

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

  It has no solutions.

- The set of linear equations

$$x_1 + x_2 = 1, \quad x_2 + x_3 = 2$$

  is written as $Ax = b$ with

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

  It has multiple solutions, including $x = (1, 0, 2)$ and $x = (0, 1, 1)$.

Over-determined and under-determined systems of linear equations. The set of linear equations is called *over-determined* if $m > n$, *under-determined* if $m < n$, and *square* if $m = n$; these correspond to the coefficient matrix being tall, wide, and square, respectively. When the system of linear equations is over-determined, there are more equations than variables or unknowns. When the system of linear equations is under-determined, there are more unknowns than equations. When the system of linear equations is square, the numbers of unknowns and equations is the same. A set of equations with zero right-hand side, $Ax = 0$, is called a *homogeneous* set of equations. Any homogeneous set of equations has $x = 0$ as a solution.

In chapter 11 we will address the question of how to determine if a system of linear equations has a solution, and how to find one when it does. For now, we give a few interesting examples.

8.3.1 Examples

Coefficients of linear combinations. Let $a_1, \ldots, a_n$ denote the columns of A. The system of linear equations $Ax = b$ can be expressed as

$$x_1 a_1 + \cdots + x_n a_n = b,$$

i.e., b is a linear combination of $a_1, \ldots, a_n$ with coefficients $x_1, \ldots, x_n$. So solving $Ax = b$ is the same as finding coefficients that express b as a linear combination of the vectors $a_1, \ldots, a_n$.

Polynomial interpolation. We seek a polynomial p of degree at most $n-1$ that interpolates a set of m given points (t_i, y_i), $i = 1, \ldots, m$. (This means that $p(t_i) = y_i$.) We can express this as a set of m linear equations in the n unknowns c, where c is the n-vector of coefficients: $Ac = y$. Here the matrix A is the Vandermonde matrix (6.7), and the vector c is the vector of polynomial coefficients, as described in the example on page 120.

Balancing chemical reactions. A chemical reaction involves p reactants (molecules) and q products, and can be written as

$$a_1 R_1 + \cdots + a_p R_p \longrightarrow b_1 P_1 + \cdots + b_q P_q.$$

Here $R_1, \ldots, R_p$ are the reactants, $P_1, \ldots, P_q$ are the products, and the numbers $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ are positive numbers that tell us how many of each of these molecules is involved in the reaction. They are typically integers, but can be scaled arbitrarily; we could double all of these numbers, for example, and we still have the same reaction. As a simple example, we have the electrolysis of water,

$$2\mathrm{H_2O} \longrightarrow 2\mathrm{H_2} + \mathrm{O_2},$$

which has one reactant, water ($\mathrm{H_2O}$), and two products, molecular hydrogen ($\mathrm{H_2}$) and molecular oxygen ($\mathrm{O_2}$). The coefficients tell us that 2 water molecules create 2 hydrogen molecules and 1 oxygen molecule. The coefficients in a reaction can be multiplied by any nonzero numbers; for example, we could write the reaction above as $3\mathrm{H_2O} \longrightarrow 3\mathrm{H_2} + (3/2)\mathrm{O_2}$. By convention reactions are written with all coefficients integers, with least common divisor one.

In a chemical reaction the numbers of constituent atoms must balance. This means that for each atom appearing in any of the reactants or products, the total amount on the left-hand side must equal the total amount on the right-hand side. (If any of the reactants or products is charged, *i.e.*, an ion, then the total charge must also balance.) In the simple water electrolysis reaction above, for example, we have 4 hydrogen atoms on the left (2 water molecules, each with 2 hydrogen atoms), and 4 on the right (2 hydrogen molecules, each with 2 hydrogen atoms). The oxygen atoms also balance, so this reaction is balanced.

Balancing a chemical reaction with specified reactants and products, *i.e.*, finding the numbers $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$, can be expressed as a system of linear equations. We can express the requirement that the reaction balances as a set of

m equations, where m is the number of different atoms appearing in the chemical reaction. We define the $m \times p$ matrix R by

$$R_{ij} = \text{number of atoms of type } i \text{ in } R_j, \quad i = 1, \ldots, m, \quad j = 1, \ldots, p.$$

(The entries of R are nonnegative integers.) The matrix R is interesting; for example, its jth column gives the chemical formula for reactant R_j. We let a denote the p-vector with entries $a_1, \ldots, a_p$. Then, the m-vector Ra gives the total number of atoms of each type appearing in the reactants. We define an $m \times q$ matrix P in a similar way, so the m-vector Pb gives the total number of atoms of each type that appears in the products.

We write the balance condition using vectors and matrices as $Ra = Pb$. We can express this as

$$\begin{bmatrix} R & -P \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0,$$

which is a set of m homogeneous linear equations.

A simple solution of these equations is $a = 0$, $b = 0$. But we seek a nonzero solution. We can set one of the coefficients, say a_1, to be one. (This might cause the other quantities to be fractional-valued.) We can add the condition that $a_1 = 1$ to our system of linear equations as

$$\begin{bmatrix} R & -P \\ e_1^T & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = e_{m+1}.$$

Finally, we have a set of $m + 1$ equations in $p + q$ variables that expresses the requirement that the chemical reaction balances. Finding a solution of this set of equations is called *balancing* the chemical reaction.

For the example of electrolysis of water described above, we have $p = 1$ reactant (water) and $q = 2$ products (molecular hydrogen and oxygen). The reaction involves $m = 2$ atoms, hydrogen and oxygen. The reactant and product matrices are

$$R = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad P = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

The balancing equations are then

$$\begin{bmatrix} 2 & -2 & 0 \\ 1 & 0 & -2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These equations are easily solved, and have the solution $(1, 1, 1/2)$. (Multiplying these coefficients by 2 gives the reaction given above.)

Diffusion systems. A *diffusion system* is a common model that arises in many areas of physics to describe *flows* and *potentials*. We start with a directed graph with n nodes and m edges. (See §6.1.) Some quantity (like electricity, heat, energy, or mass) can flow across the edges, from one node to another.

With edge j we associate a flow (rate) f_j, which is a scalar; the vector of all m flows is the flow m-vector f. The flows f_j can be positive or negative: Positive

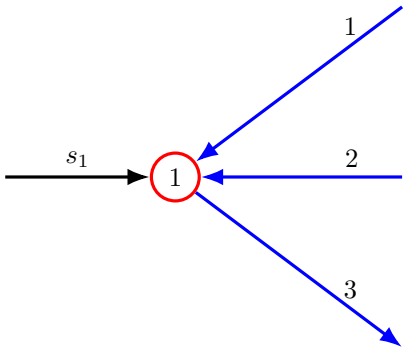

Figure 8.1 A node in a diffusion system with label 1, exogenous flow s_1 and three incident edges.

f_j means the quantity flows in the direction of edge j, and negative f_j means the quantity flows in the opposite direction of edge j. The flows can represent, for example, heat flow (in units of Watts) in a thermal model, electrical current (Amps) in an electrical circuit, or movement (diffusion) of mass (such as, for example, a pollutant). We also have a source (or exogenous) flow s_i at each node, with $s_i > 0$ meaning that an exogenous flow is injected into node i, and $s_i < 0$ means that an exogenous flow is removed from node i. (In some contexts, a node where flow is removed is called a *sink*.) In a thermal system, the sources represent thermal (heat) sources; in an electrical circuit, they represent electrical current sources; in a system with diffusion, they represent external injection or removal of the mass.

In a diffusion system, the flows must satisfy (flow) *conservation*, which means that at each node, the total flow entering each node from adjacent edges and the exogenous source, must be zero. This is illustrated in figure 8.1, which shows three edges adjacent to node 1, two entering node 1 (flows 1 and 2), and one (flow 3) leaving node 1, and an exogenous flow. Flow conservation at this node is expressed as

$$f_1 + f_2 - f_3 + s_1 = 0.$$

Flow conservation at every node can be expressed by the simple matrix-vector equation

$$Af + s = 0, \tag{8.6}$$

where A is the incidence matrix described in §7.3. (This is called *Kirchhoff's current law* in an electrical circuit, after the physicist Gustav Kirchhoff; when the flows represent movement of mass, it is called *conservation of mass*.)

With node i we associate a potential e_i; the n-vector e gives the potential at all nodes. (Note that here, e represents the n-vector of potentials; e_i is the scalar potential at node i, and not the standard ith unit vector.) The potential might represent the node temperature in a thermal model, the electrical potential (voltage) in an electrical circuit, and the concentration in a system that involves mass diffusion.

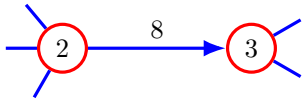

Figure 8.2 The flow through edge 8 is equal to $f_8 = (e_2 - e_3)/r_8$.

In a diffusion system *the flow on an edge is proportional to the potential difference across its adjacent nodes*. This is typically written as $r_j f_j = e_k - e_l$, where edge j goes from node k to node l, and r_j (which is typically positive) is called the *resistance* of edge j. In a thermal model, r_j is called the thermal resistance of the edge; in an electrical circuit, it is called the electrical resistance. This is illustrated in figure 8.2, which shows edge 8, connecting node 2 and node 3, corresponding to an edge flow equation

$$r_8 f_8 = e_2 - e_3.$$

We can write the edge flow equations in a compact way as

$$Rf = -A^T e, \tag{8.7}$$

where $R = \mathbf{diag}(r)$ is called the *resistance matrix*.

The diffusion model can be expressed as one set of block linear equations in the variables f, s, and e:

$$\begin{bmatrix} A & I & 0 \\ R & 0 & A^T \end{bmatrix} \begin{bmatrix} f \\ s \\ e \end{bmatrix} = 0.$$

This is a set of $n+m$ homogeneous equations in $m+2n$ variables. To these underdetermined equations we can add others, for example, by specifying some of the entries of f, s, and e.

Leontief input-ouput model. We consider an economy with n different industrial sectors. We let x_i be the economic activity level, or total production output, of sector i, for $i = 1, \ldots, n$, measured in a common unit, such as (billions of) dollars. The output of each sector flows to other sectors, to support their production, and also to consumers. We denote the total consumer demand for sector i as d_i, for $i = 1, \ldots, n$.

Supporting the output level x_j for sector j requires $A_{ij}x_j$ output for sector i. We refer to $A_{ij}x_j$ as the sector i *input* that flows to sector j. (We can have $A_{ii} \neq 0$; for example, it requires some energy to support the production of energy.) Thus, $A_{i1}x_1 + \cdots + A_{in}x_n$ is the total sector i output required by, or flowing into, the n industrial sectors. The matrix A is called the *input-output matrix* of the economy, since it describes the flows of sector outputs to the inputs of itself and other sectors. The vector Ax gives the sector outputs required to support the production levels given by x. (This sounds circular, but isn't.)

Finally, we require that for each sector, the total production level matches the demand plus the total amount required to support production. This leads to the balance equations,

$$x = Ax + d.$$

Suppose the demand vector d is given, and we wish to find the sector output levels that will support it. We can write this as a set of n equations in n unknowns,

$$(I - A)x = d.$$

This model of the sector inputs and outputs of an economy was developed by Wassily Leontief in the late 1940s, and is now known as Leontief input-output analysis. He was awarded the Nobel Prize in economics for this work in 1973.