

Chapter 2

Matrix Calculus and Notation



2.1 Introduction: Can It Possibly Be That Simple?

In October of 2005, I scribbled in a notebook, “can it possibly be *that* simple?” I was referring to the sensitivity of transient dynamics (the eventual results appear in Chap. 7), and had just begun to use matrix calculus as a tool. The answer to my question was yes. It can be that simple.

This book relies on this set of mathematical techniques. This chapter introduces the basics, which will be used throughout the text. For more information, I recommend four sources in particular. The most complete treatment, but not the easiest starting point, is the book by Magnus and Neudecker (1988). More accessible introductions can be found in the paper by Magnus and Neudecker (1985) and especially the text by Abadir and Magnus (2005). A review paper by Nel (1980) is helpful in placing the Magnus-Neudecker formulation in the context of other attempts at a calculus of matrices.

Sensitivity analysis asks how much change in an outcome variable y is caused by a change in some parameter x . At its most basic level, and with some reasonable assumptions about the continuity and differentiability of the functional relationships involved, the solution is given by differential calculus. If y is a function of x , then the derivative

$$\frac{dy}{dx}$$

tells how y responds to a change in x , i.e., the sensitivity of y to a change in x .

However, the outcomes of a demographic calculation may be scalar-valued (e.g., the population growth rate λ), vector-valued (e.g., the stable stage distribution), or matrix-valued (e.g., the fundamental matrix). Any of these outcomes may be

functions of scalar-valued parameters (e.g., the Gompertz aging rate), vector-valued parameters (e.g., the mortality schedule), or matrix-valued parameters (e.g., the transition matrix) parameters. Thus, sensitivity analysis in demography requires more than the simple derivative in (2.1). We want a consistent and flexible approach to differentiating

$$\left\{ \begin{array}{l} \text{scalar-valued} \\ \text{vector-valued} \\ \text{matrix-valued} \end{array} \right\} \text{ functions of } \left\{ \begin{array}{l} \text{scalar} \\ \text{vector} \\ \text{matrix} \end{array} \right\} \text{ arguments}$$

2.2 Notation and Matrix Operations

2.2.1 Notation

Matrices are denoted by upper case bold symbols (e.g., \mathbf{A}), vectors (usually) by lower case bold symbols (\mathbf{n}). The (i, j) entry of the matrix \mathbf{A} is a_{ij} , and the i th entry of the vector \mathbf{n} is n_i . Sometimes we will use MATLAB notation, and write

$$\mathbf{X}(i, :) = \text{row } i \text{ of } \mathbf{X} \quad (2.1)$$

$$\mathbf{X}(:, j) = \text{column } j \text{ of } \mathbf{X} \quad (2.2)$$

The notation

$$(x_{ij})$$

denotes a matrix whose (i, j) entry is x . For example,

$$\left(\frac{dy_i}{dx_j} \right)$$

is the matrix whose (i, j) entry is the derivative of y_i with respect to x_j .

The transpose of \mathbf{X} is \mathbf{X}^T . Logarithms are natural logarithms. The vector norm $\|\mathbf{x}\|$ is, unless noted otherwise, the 1-norm. The symbol $\mathcal{D}(\mathbf{x})$ denotes the square matrix with \mathbf{x} on the diagonal and zeros elsewhere. The symbol $\mathbf{1}$ denotes a vector of ones. The vector \mathbf{e}_i is a unit vector with 1 in the i th entry and zeros elsewhere. The identity matrix is \mathbf{I} . Where necessary for clarity, the dimension of matrices or vectors will be indicated by a subscript. Thus \mathbf{I}_s is a $s \times s$ identity matrix, $\mathbf{1}_s$ is an $s \times 1$ vector of ones, and $\mathbf{X}_{m \times n}$ is a $m \times n$ matrix.

In some places (Chaps. 6 and 10) block-structured matrices appear; these are denoted by either \mathbb{A} or $\hat{\mathbf{A}}$, depending on the context and the role of the matrix.

2.2.2 Operations

In addition to the familiar matrix product \mathbf{AB} , we will also use the Hadamard, or elementwise product

$$\mathbf{A} \circ \mathbf{B} = (a_{ij}b_{ij}) \quad (2.3)$$

and the Kronecker product

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B}) \quad (2.4)$$

The Hadamard product requires that \mathbf{A} and \mathbf{B} be the same size. The Kronecker product is defined for any sizes of \mathbf{A} and \mathbf{B} . Some useful properties of the Kronecker product include

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) \quad (2.5)$$

$$(\mathbf{A} \otimes \mathbf{B})^T = (\mathbf{A}^T \otimes \mathbf{B}^T) \quad (2.6)$$

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) \quad (2.7)$$

and, provided that the matrices are of the right size for the products to be defined,

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = (\mathbf{A}_1\mathbf{A}_2 \otimes \mathbf{B}_1\mathbf{B}_2). \quad (2.8)$$

2.2.3 The Vec Operator and Vec-Permutation Matrix

The vec operator transforms a $m \times n$ matrix \mathbf{A} into a $mn \times 1$ vector by stacking the columns one above the next,

$$\text{vec } \mathbf{A} = \begin{pmatrix} \mathbf{A}(:, 1) \\ \vdots \\ \mathbf{A}(:, n) \end{pmatrix} \quad (2.9)$$

For example,

$$\text{vec} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}. \quad (2.10)$$

The vec of \mathbf{A} and the vec of \mathbf{A}^\top are rearrangements of the same entries; they are related by

$$\text{vec } \mathbf{A}^\top = \mathbf{K}_{m,n} \text{vec } \mathbf{A} \quad (2.11)$$

where \mathbf{A} is $m \times n$ and $\mathbf{K}_{m,n}$ is the *vec-permutation matrix* (Henderson and Searle 1981) or *commutation matrix* (Magnus and Neudecker 1979). The vec -permutation matrix can be calculated as

$$\mathbf{K}_{m,n} = \sum_{i=1}^m \sum_{j=1}^n \left(\mathbf{E}_{ij} \otimes \mathbf{E}_{ij}^\top \right) \quad (2.12)$$

where \mathbf{E}_{ij} is a matrix, of dimension $m \times n$, with a 1 in the (i, j) entry and zeros elsewhere. Like any permutation matrix, $\mathbf{K}^{-1} = \mathbf{K}^\top$.

The vec operator and the vec -permutation matrix are particularly important in multistate models (e.g., age \times stage-classified models), where they are used in both the formulation and analysis of the models (e.g., Caswell 2012, 2014; Caswell and Salguero-Gómez 2013; Caswell et al. 2018); see also Chap. 6. Extensions to an arbitrary number of dimensions, so-called hyperstate models, have been presented by Roth and Caswell (2016).

2.2.4 Roth's Theorem

The vec operator and the Kronecker product are connected by a theorem due to Roth (1934):

$$\text{vec } (\mathbf{ABC}) = \left(\mathbf{C}^\top \otimes \mathbf{A} \right) \text{vec } \mathbf{B}. \quad (2.13)$$

We will often want to obtain the vec of a matrix that appears in the middle of a product; we will use Roth's theorem repeatedly.

2.3 Defining Matrix Derivatives

The derivative of a scalar y with respect to a scalar x is familiar. What, however, does it mean to speak of the derivative of a scalar with respect to a vector, or of a vector with respect to another vector, or any other combination? These can be defined in more than one way and the choice is critical (Nel 1980; Magnus and Neudecker 1985). This book relies on the notation due to Magnus and Neudecker, because it makes certain operations possible and consistent.

- If x and y are scalars, the derivative of y with respect to x is the familiar derivative dy/dx .

- If \mathbf{y} is a $n \times 1$ vector and x a scalar, the derivative of \mathbf{y} with respect to x is the $n \times 1$ column vector

$$\frac{d\mathbf{y}}{dx} = \begin{pmatrix} \frac{dy_1}{dx} \\ \vdots \\ \frac{dy_n}{dx} \end{pmatrix}. \tag{2.14}$$

- If y is a scalar and \mathbf{x} a $m \times 1$ vector, the derivative of y with respect to \mathbf{x} is the $1 \times m$ row vector (called the gradient vector)

$$\frac{dy}{d\mathbf{x}^\top} = \left(\frac{\partial y}{\partial x_1} \cdots \frac{\partial y}{\partial x_m} \right). \tag{2.15}$$

Note the orientation of dy/dx as a column vector and $dy/d\mathbf{x}^\top$ as a row vector.

- If \mathbf{y} is a $n \times 1$ vector and \mathbf{x} a $m \times 1$ vector, the derivative of \mathbf{y} with respect to \mathbf{x} is defined to be the $n \times m$ matrix whose (i, j) entry is the derivative of y_i with respect to x_j , i.e.,

$$\frac{d\mathbf{y}}{d\mathbf{x}^\top} = \left(\frac{dy_i}{dx_j} \right) \tag{2.16}$$

(this matrix is called the Jacobian matrix).

- Derivatives involving matrices are written by first transforming the matrices into vectors using the `vec` operator, and then applying the rules for vector differentiation to the resulting vectors. Thus, the derivative of the $m \times n$ matrix \mathbf{Y} with respect to the $p \times q$ matrix \mathbf{X} is the $mn \times pq$ matrix

$$\frac{d\text{vec } \mathbf{Y}}{d(\text{vec } \mathbf{X})^\top}. \tag{2.17}$$

From now on, I will write $\text{vec }^\top \mathbf{X}$ for $(\text{vec } \mathbf{X})^\top$.

2.4 The Chain Rule

The chain rule for differentiation is your friend. The Magnus-Neudecker notation, unlike some alternatives, extends the familiar scalar chain rule to derivatives of vectors and matrices (Nel 1980; Magnus and Neudecker 1985). If \mathbf{u} (size $m \times 1$) is a function of \mathbf{v} (size $n \times 1$) and \mathbf{v} is a function of \mathbf{x} (size $p \times 1$), then

$$\underbrace{\frac{d\mathbf{u}}{d\mathbf{x}^\top}}_{m \times p} = \underbrace{\left(\frac{d\mathbf{u}}{d\mathbf{v}^\top} \right)}_{m \times n} \underbrace{\left(\frac{d\mathbf{v}}{d\mathbf{x}^\top} \right)}_{n \times p} \tag{2.18}$$

Notice that the dimensions are correct, and that the order of the multiplication matters. Checking dimensional consistency in this way is a useful way to find errors.

2.5 Derivatives from Differentials

The key to the matrix calculus of Magnus and Neudecker (1988) is the relationship between the differential and the derivative of a function. Experience suggests that, for many readers of this book, this relationship is shrouded in the mists of long-ago calculus classes.

2.5.1 Differentials of Scalar Function

Start with scalars. Suppose that $y = f(x)$ is a differentiable function at $x = x_0$. Then the derivative of y with respect to x at the point x_0 is defined as

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}. \quad (2.19)$$

Now define the differential of y . This is customarily denoted dy , but for the moment, I will denote it by cy . The differential of y at x_0 is a function of h , defined by

$$cy(x_0, h) = f'(x_0)h. \quad (2.20)$$

There is no requirement that h be “small.” Since x is a function of itself, $x = g(x)$, with $g'(x) = 1$, we also have $cx(x_0, h) = g'(x_0)h = h$. Thus the ratio of the differential of y and the differential of x is

$$\frac{cy(x_0, h)}{cx(x_0, h)} = \frac{f'(x_0)h}{h} = f'(x_0). \quad (2.21)$$

That is, *the derivative is equal to the ratio of the differentials*.

Now, return to the standard notation of dy for the differential of y . This gives two meanings to the familiar notation for derivatives,

$$\left. \frac{dy}{dx} \right|_{x_0} = f'(x_0). \quad (2.22)$$

The left hand side can be regarded either as equivalent to the limit (2.19) or the ratio of the differentials given by (2.21). Mathematicians are strangely unconcerned with this ambiguity (e.g., Hardy 1952).

All this leads to a set of familiar rules for calculating differentials that guarantee that they can be used to create derivatives. A few of these, for scalars, are

$$d(u + v) = du + dv \quad (2.23)$$

$$d(cu) = c du \quad (2.24)$$

$$d(uv) = u(dv) + (du)v \quad (2.25)$$

$$d(e^u) = e^u du \quad (2.26)$$

$$d(\log u) = \frac{1}{u} du \quad (2.27)$$

If $y = f(x_1, x_2)$, then the total differential is

$$dy = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2. \quad (2.28)$$

Derivatives can be constructed from these expressions at will by dividing by differentials. For example, dividing (2.23) by dx gives $d(u + v)/dx = du/dx + dv/dx$. From (2.28), we have

$$\frac{dy}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dx_1} \quad (2.29)$$

$$\frac{dy}{dx_2} = \frac{\partial f}{\partial x_1} \frac{dx_1}{dx_2} + \frac{\partial f}{\partial x_2}. \quad (2.30)$$

2.5.2 Differentials of Vectors and Matrices

To extend these concepts to matrices, we define the differential of a matrix (or vector) as the matrix (or vector) of differentials of the elements; i.e.,

$$d\mathbf{X} = (dx_{ij}). \quad (2.31)$$

This definition leads to some basic rules for differentials of matrices:

$$d(c\mathbf{U}) = c(d\mathbf{U}) \quad (2.32)$$

$$d(\mathbf{U} + \mathbf{V}) = d\mathbf{U} + d\mathbf{V} \quad (2.33)$$

$$d(\mathbf{UV}) = (d\mathbf{U})\mathbf{V} + \mathbf{U}(d\mathbf{V}) \quad (2.34)$$

$$d(\mathbf{U} \otimes \mathbf{V}) = (d\mathbf{U}) \otimes \mathbf{V} + \mathbf{U} \otimes (d\mathbf{V}) \quad (2.35)$$

$$d(\mathbf{U} \circ \mathbf{V}) = (d\mathbf{U}) \circ \mathbf{V} + \mathbf{U} \circ (d\mathbf{V}) \quad (2.36)$$

$$d\text{vec } \mathbf{U} = \text{vec } d\mathbf{U} \quad (2.37)$$

where c is a constant, and, of course, the dimensions of \mathbf{U} and \mathbf{V} must be conformable. The differentials of an operators applied elementwise to a vector can be obtained from the differentials of the elements. For example, suppose \mathbf{u} is a $s \times 1$ vector, and the exponential is applied elementwise. Then

$$d(\exp(\mathbf{u})) = \begin{pmatrix} e^{u_1} du_1 \\ \vdots \\ e^{u_s} du_s \end{pmatrix} \quad (2.38)$$

$$= \mathcal{D} [\exp(\mathbf{u})] d\mathbf{u}. \quad (2.39)$$

If y is a function of \mathbf{x}_1 and \mathbf{x}_2 , the total differential is given just as in (2.28), by

$$d\mathbf{y} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1^T} d\mathbf{x}_1 + \frac{\partial \mathbf{y}}{\partial \mathbf{x}_2^T} d\mathbf{x}_2 \quad (2.40)$$

2.6 The First Identification Theorem

For scalar y and x ,

$$dy = qdx \implies \frac{dy}{dx} = q. \quad (2.41)$$

That much is easy. But, suppose that \mathbf{y} is a $n \times 1$ vector function of the $m \times 1$ vector \mathbf{x} . The differential $d\mathbf{y}$ is the $n \times 1$ vector

$$d\mathbf{y} = \begin{pmatrix} dy_1 \\ \vdots \\ dy_n \end{pmatrix} \quad (2.42)$$

which, by the total derivative rule, is

$$d\mathbf{y} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} dx_1 + \cdots + \frac{\partial y_1}{\partial x_m} dx_m \\ \vdots \\ \frac{\partial y_n}{\partial x_1} dx_1 + \cdots + \frac{\partial y_n}{\partial x_m} dx_m \end{pmatrix} \quad (2.43)$$

$$= \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \dots & \frac{\partial y_n}{\partial x_m} \end{pmatrix} \begin{pmatrix} dx_1 \\ \vdots \\ dx_m \end{pmatrix} \quad (2.44)$$

$$= \mathbf{Q} \, d\mathbf{x}. \quad (2.45)$$

If these were scalars, dividing both sides by $d\mathbf{x}$ would give \mathbf{Q} as the derivative of \mathbf{y} with respect to \mathbf{x} . But, one cannot divide by a vector. Instead, Magnus and Neudecker proved that if it can be shown that

$$d\mathbf{y} = \mathbf{Q} \, d\mathbf{x} \quad (2.46)$$

then the derivative is

$$\frac{d\mathbf{y}}{d\mathbf{x}^\top} = \mathbf{Q}. \quad (2.47)$$

This is the First Identification Theorem of Magnus and Neudecker (1988).¹

2.6.1 The Chain Rule and the First Identification Theorem

Suppose that $d\mathbf{y}$ is given by (2.46), and that \mathbf{x} is in turn a function of some vector $\boldsymbol{\theta}$. Then

$$d\mathbf{x} = \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\top} d\boldsymbol{\theta} \quad (2.48)$$

and

$$\frac{d\mathbf{y}}{d\boldsymbol{\theta}^\top} = \mathbf{Q} \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\top}. \quad (2.49)$$

In other words, the differential expression (2.46) can be transformed into a derivative with respect to any vector by careful use of the chain rule. This applies equally to more complicated expressions for the differential. Suppose that

$$d\mathbf{y} = \mathbf{Q}d\mathbf{x} + \mathbf{R}d\mathbf{z}. \quad (2.50)$$

¹There is also a second identification theorem that provides the second derivatives of matrix functions. See Shyu and Caswell (2014) for applications of this theory to the second derivatives of measures of population growth rate.

Applying the chain rule to the differentials on the right hand side gives

$$dy = \mathbf{Q} \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\top} d\boldsymbol{\theta} + \mathbf{R} \frac{dz}{d\boldsymbol{\theta}^\top} d\boldsymbol{\theta} \quad (2.51)$$

for any vector $\boldsymbol{\theta}$. Thus

$$dy = \left(\mathbf{Q} \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\top} + \mathbf{R} \frac{dz}{d\boldsymbol{\theta}^\top} \right) d\boldsymbol{\theta}, \quad (2.52)$$

and the First Identification Theorem gives

$$\frac{dy}{d\boldsymbol{\theta}^\top} = \left(\mathbf{Q} \frac{d\mathbf{x}}{d\boldsymbol{\theta}^\top} + \mathbf{R} \frac{dz}{d\boldsymbol{\theta}^\top} \right). \quad (2.53)$$

2.7 Elasticity

When parameters are measured on different scales, it is sometimes helpful to calculate proportional effects of proportional perturbations, also called elasticities. The elasticity of y_i to θ_j is

$$\frac{\epsilon y_i}{\epsilon \theta_j} = \frac{\theta_j}{y_i} \frac{dy_i}{d\theta_j}. \quad (2.54)$$

For vectors \mathbf{y} and $\boldsymbol{\theta}$, this becomes

$$\frac{\epsilon \mathbf{y}}{\epsilon \boldsymbol{\theta}^\top} = \mathcal{D}(\mathbf{y})^{-1} \frac{d\mathbf{y}}{d\boldsymbol{\theta}^\top} \mathcal{D}(\boldsymbol{\theta}). \quad (2.55)$$

There seems to be no accepted notation for elasticities; the notation used here is adapted from that in Samuelson (1947).

2.8 Some Useful Matrix Calculus Results

Several matrix calculus results will be used repeatedly. Many more can be found in Magnus and Neudecker (1988) and Abadir and Magnus (2005).

1. The matrix product $\mathbf{Y} = \mathbf{A}\mathbf{B}$. Differentiate,

$$d\mathbf{Y} = (d\mathbf{A})\mathbf{B} + \mathbf{A}(d\mathbf{B}). \quad (2.56)$$

Then write (or imagine writing; with practice one does not actually need this step explicitly)

$$(d\mathbf{A})\mathbf{B} = \mathbf{I}(d\mathbf{A})\mathbf{B} \quad (2.57)$$

$$\mathbf{A}(d\mathbf{B}) = \mathbf{A}(d\mathbf{B})\mathbf{I} \quad (2.58)$$

and apply the vec operator and Roth's theorem, to obtain

$$d\text{vec } \mathbf{Y} = (\mathbf{B}^\top \otimes \mathbf{I}) d\text{vec } \mathbf{A} + (\mathbf{I} \otimes \mathbf{A}) d\text{vec } \mathbf{B}. \quad (2.59)$$

The chain rule gives, for any vector variable $\boldsymbol{\theta}$

$$\frac{d\text{vec } \mathbf{Y}}{d\boldsymbol{\theta}^\top} = (\mathbf{B}^\top \otimes \mathbf{I}) \frac{d\text{vec } \mathbf{A}}{d\boldsymbol{\theta}^\top} + (\mathbf{I} \otimes \mathbf{A}) \frac{d\text{vec } \mathbf{B}}{d\boldsymbol{\theta}^\top}. \quad (2.60)$$

2. The Hadamard product $\mathbf{Y} = \mathbf{A} \circ \mathbf{B}$. Differentiate the product,

$$d\mathbf{Y} = d\mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ d\mathbf{B}, \quad (2.61)$$

then vec

$$d\text{vec } \mathbf{Y} = d\text{vec } \mathbf{A} \circ \text{vec } \mathbf{B} + \text{vec } \mathbf{A} \circ d\text{vec } \mathbf{B}. \quad (2.62)$$

It will be useful to replace the Hadamard products, which we do using the fact that $\mathbf{x} \circ \mathbf{y} = \mathcal{D}(\mathbf{x})\mathbf{y}$, to get

$$d\text{vec } \mathbf{Y} = \mathcal{D}(\text{vec } \mathbf{B})d\text{vec } \mathbf{A} + \mathcal{D}(\text{vec } \mathbf{A})d\text{vec } \mathbf{B}. \quad (2.63)$$

The chain rule gives the derivative from the differential,

$$\frac{d\text{vec } \mathbf{Y}}{d\boldsymbol{\theta}^\top} = \mathcal{D}(\text{vec } \mathbf{B}) \frac{d\text{vec } \mathbf{A}}{d\boldsymbol{\theta}^\top} + \mathcal{D}(\text{vec } \mathbf{A}) \frac{d\text{vec } \mathbf{B}}{d\boldsymbol{\theta}^\top}. \quad (2.64)$$

3. Diagonal matrices. The diagonal matrix $\mathcal{D}(\mathbf{x})$, with the vector \mathbf{x} on the diagonal and zeros elsewhere, can be written

$$\mathcal{D}(\mathbf{x}) = \mathbf{I} \circ (\mathbf{1} \mathbf{x}^\top) \quad (2.65)$$

Differentiate both sides,

$$d\mathcal{D}(\mathbf{x}) = \mathbf{I} \circ (\mathbf{1} d\mathbf{x}^\top) \quad (2.66)$$

and vec the result

$$d\text{vec } \mathcal{D}(\mathbf{x}) = \mathcal{D}(\text{vec } \mathbf{I}) \text{vec} \left(\mathbf{1} d\mathbf{x}^T \right) \quad (2.67)$$

$$= \mathcal{D}(\text{vec } \mathbf{I}) (\mathbf{I} \otimes \mathbf{1}) d\mathbf{x} \quad (2.68)$$

The First Identification Theorem gives

$$\frac{d\text{vec } \mathcal{D}(\mathbf{x})}{d\boldsymbol{\theta}^T} = \mathcal{D}(\text{vec } \mathbf{I}) (\mathbf{I} \otimes \mathbf{1}) \frac{d\mathbf{x}}{d\boldsymbol{\theta}^T}. \quad (2.69)$$

The identity matrix in (2.65) *masks* the matrix $(\mathbf{1} \mathbf{x}^T)$, setting to zero all but the diagonal elements. Matrices other than \mathbf{I} can be used in this way to mask entries of a matrix. For example, the transition matrix for a Leslie matrix, with a vector of survival probabilities \mathbf{p} on the subdiagonal, is obtained by setting $\mathbf{x} = \mathbf{p}$ and replacing \mathbf{I} with a matrix \mathbf{Z} that contains ones on the subdiagonal and zeros elsewhere (see, e.g., Chap. 4).

Some Markov chain calculations (Chaps. 5 and 11) involve a matrix \mathbf{N}_{dg} , which contains the diagonal elements of \mathbf{N} on the diagonal and zeros elsewhere. This can be written

$$\mathbf{N}_{\text{dg}} = \mathbf{I} \circ \mathbf{N}. \quad (2.70)$$

Differentiating and applying the vec operator yields

$$d\text{vec } \mathbf{N}_{\text{dg}} = \mathcal{D}(\text{vec } \mathbf{I}) d\text{vec } \mathbf{N}. \quad (2.71)$$

4. The Kronecker product. Differentiating the Kronecker product is a bit more complicated (Magnus and Neudecker 1985, Theorem 11). We want an expression for the differential of the product in terms of the differentials of the components, something of the form

$$d\text{vec} (\mathbf{A} \otimes \mathbf{B}) = \mathbf{Z}_1 d\text{vec } \mathbf{A} + \mathbf{Z}_2 d\text{vec } \mathbf{B} \quad (2.72)$$

for some matrices \mathbf{Z}_1 and \mathbf{Z}_2 .

This requires a result for the vec of the Kronecker product. Let \mathbf{A} be of dimension $m \times p$ and \mathbf{B} be $r \times s$. Then

$$\text{vec} (\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r) (\text{vec } \mathbf{A} \otimes \text{vec } \mathbf{B}). \quad (2.73)$$

Let $\mathbf{Y} = \mathbf{A} \otimes \mathbf{B}$. Differentiate,

$$d\mathbf{Y} = (d\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes d\mathbf{B}) \quad (2.74)$$

and vec

$$d\text{vec } \mathbf{Y} = (\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r) \left[(d\text{vec } \mathbf{A} \otimes \text{vec } \mathbf{B}) + (\text{vec } \mathbf{A} \otimes d\text{vec } \mathbf{B}) \right]. \tag{2.75}$$

With some ingenious simplifications (Magnus and Neudecker 1985), this reduces to (2.72) with

$$\mathbf{Z}_1 = (\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r) (\mathbf{I}_m \otimes \text{vec } \mathbf{B}) \tag{2.76}$$

$$\mathbf{Z}_2 = (\mathbf{I}_p \otimes \mathbf{K}_{s,m} \otimes \mathbf{I}_r) (\text{vec } \mathbf{A} \otimes \mathbf{I}_{rs}). \tag{2.77}$$

Substituting \mathbf{Z}_1 and \mathbf{Z}_2 into (2.72) gives the differential of the Kronecker product in terms of the differentials of its component matrices.

5. The matrix inverse. The inverse of \mathbf{X} satisfies

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}. \tag{2.78}$$

Differentiate both sides

$$(d\mathbf{X})\mathbf{X}^{-1} + \mathbf{X}(d\mathbf{X}^{-1}) = \mathbf{0}, \tag{2.79}$$

then vec

$$\left[(\mathbf{X}^{-1})^T \otimes \mathbf{I} \right] d\text{vec } \mathbf{X} + [\mathbf{I} \otimes \mathbf{X}] d\text{vec } \mathbf{X}^{-1} = \mathbf{0} \tag{2.80}$$

and finally solve for $d\text{vec } \mathbf{X}^{-1}$

$$d\text{vec } \mathbf{X}^{-1} = -[\mathbf{I} \otimes \mathbf{X}]^{-1} \left[(\mathbf{X}^{-1})^T \otimes \mathbf{I} \right] d\text{vec } \mathbf{X} \tag{2.81}$$

The properties (2.5) and (2.8) of the Kronecker product let this be simplified to

$$d\text{vec } \mathbf{X}^{-1} = - \left[(\mathbf{X}^{-1})^T \otimes \mathbf{X}^{-1} \right] d\text{vec } \mathbf{X} \tag{2.82}$$

6. The square root and ratios. In calculating standard deviations and coefficients of variation it is useful to calculate the elementwise square root and the elementwise ratio of two vectors. If \mathbf{x} is a non-negative vector, and the square root $\sqrt{\mathbf{x}}$ is taken elementwise, then

$$d\sqrt{\mathbf{x}} = \frac{1}{2} \mathcal{D}(\sqrt{\mathbf{x}})^{-1} d\mathbf{x}. \tag{2.83}$$

For the elementwise ratio, let \mathbf{x} and \mathbf{y} be $m \times 1$ vectors, with \mathbf{y} nonzero. Let \mathbf{w} be a vector whose i th element is x_i/y_i ; i.e., $\mathbf{w} = \mathcal{D}(\mathbf{y})^{-1}\mathbf{x}$. Then

$$d\mathbf{w} = \mathcal{D}(\mathbf{y})^{-1}d\mathbf{x} - \left[\mathbf{x}^T \mathcal{D}(\mathbf{y})^{-1} \otimes \mathcal{D}(\mathbf{y})^{-1} \right] \mathcal{D}(\text{vec } \mathbf{I}_m) (\mathbf{I}_m \otimes \mathbf{1}_m) d\mathbf{y}. \quad (2.84)$$

This list could go on. The books by Magnus and Neudecker (1988) and Abadir and Magnus (2005) contain many other results, and demographically relevant derivations appear throughout this book, especially in Chap. 5.

2.9 LTRE Decomposition of Demographic Differences

The LTRE decomposition in Sect. 1.3.1 extends readily to matrix calculus. Suppose that a demographic outcome ξ , dimension $(s \times 1)$, is a function of a vector θ of parameters, dimension $(p \times 1)$. Suppose that results are obtained under two “conditions,” with parameters $\theta^{(1)}$ and $\theta^{(2)}$. Define the parameter difference as $\Delta\theta = \theta^{(2)} - \theta^{(1)}$ and the effect as $\Delta\xi = \xi^{(2)} - \xi^{(1)}$. Then, to first order,

$$\Delta\xi \approx \sum_{i=1}^p \frac{d\xi}{d\theta_i} \Delta\theta_i \quad (2.85)$$

$$= \frac{d\xi}{d\theta^T} \Delta\theta. \quad (2.86)$$

Writing

$$\Delta\theta = \mathcal{D}(\Delta\theta)\mathbf{1}_p, \quad (2.87)$$

we create a *contribution matrix* \mathbf{C} , of dimension $s \times p$,

$$\mathbf{C} = \frac{d\xi}{d\theta^T} \mathcal{D}(\Delta\theta). \quad (2.88)$$

The (i, j) entry of \mathbf{C} is the contribution of $\Delta\theta_j$ to the difference ξ_i , for $i = 1, \dots, s$ and $j = 1, \dots, p$. The rows and columns of \mathbf{C} give

$$\mathbf{C}(i, :) = \text{contributions of } \Delta\theta \text{ to } \Delta\xi_i \quad (2.89)$$

$$\mathbf{C}(:, j) = \text{contributions of } \theta_j \text{ to } \Delta\xi \quad (2.90)$$

When calculating C , the derivative of ξ must be evaluated somewhere. Experience suggests that the evaluating it at the midpoint between $\theta^{(1)}$ and $\theta^{(2)}$ gives good results (Logofet and Lesnaya 1997; Caswell 2001).

2.10 A Protocol for Sensitivity Analysis

The calculations may grow to be complex, but the protocol is simple:

1. write a matrix expression for the outcome,
2. differentiate,
3. vec,
4. simplify,
5. calculate derivatives from the differentials, and
6. extend using the chain rule

The rest of this book shows what can be done with this simple procedure.

Bibliography

- Abadir, K. M., and J. R. Magnus. 2005. *Matrix algebra. Econometric exercises 1*. Cambridge University Press, Cambridge, United Kingdom.
- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd edition. Sinauer Associates, Sunderland, MA.
- Caswell, H. 2012. Matrix models and sensitivity analysis of populations classified by age and stage: a vec-permutation matrix approach. *Theoretical Ecology* **5**:403–417.
- Caswell, H. 2014. A matrix approach to the statistics of longevity in heterogeneous frailty models. *Demographic Research* **31**:553–592.
- Caswell, H., C. de Vries, N. Hartemink, G. Roth, and S. F. van Daalen. 2018. Age \times stage-classified demographic analysis: a comprehensive approach. *Ecological Monographs* **88**:560–584.
- Caswell, H., and R. Salguero-Gómez. 2013. Age, stage and senescence in plants. *Journal of Ecology* **101**:585–595.
- Hardy, G. H. 1952. *A course of pure mathematics*. 10th edition. Cambridge University Press, Cambridge, United Kingdom.
- Henderson, H. V., and S. R. Searle. 1981. The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear and Multilinear Algebra* **9**:271–288.
- Logofet, D. O., and E. V. Lesnaya. 1997. Why are the middle points the most sensitive in the sensitivity experiments? *Ecological Modelling* **104**:303–306.
- Magnus, J. R., and H. Neudecker. 1979. The commutation matrix: some properties and applications. *Annals of Statistics* **7**:381–394.
- Magnus, J. R., and H. Neudecker. 1985. Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. *Journal of Mathematical Psychology* **29**:474–492.
- Magnus, J. R., and H. Neudecker. 1988. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley and Sons, New York, New York.
- Nel, D. G. 1980. On matrix differentiation in statistics. *South African Statistical Journal* **14**:137–193.

- Roth, G., and H. Caswell. 2016. Hyperstate matrix models: extending demographic state spaces to higher dimensions. *Methods in Ecology and Evolution* 7:1438–1450.
- Roth, W. E. 1934. On direct product matrices. *Bulletin of the American Mathematical Society* 40:461–468.
- Samuelson, P. A. 1947. *Foundations of economic analysis*. Harvard University Press, Cambridge, Massachusetts, USA.
- Shyu, E., and H. Caswell. 2014. Calculating second derivatives of population growth rates for ecology and evolution. *Methods in Ecology and Evolution* 5:473–482.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

