APPENDIX A

Symbolic Differentiation and Optimization of Multivariable Functions

A.1 INTRODUCTION

In our earlier discussions of multiple regression, principal components analysis, and multiple discriminant analysis, matrix equations were employed to solve for the various parameter values of interest. However, relatively little has been said so far about the characteristics of the functions being optimized and the process by which the matrix equations are derived.

In each of the three preceding cases, it is the calculus that provides the rationale and specific techniques for optimization. Accordingly, this appendix provides a selective review of those topics from the calculus that bear on problems of optimizing functions of multivariable arguments. No exhaustive treatment is attempted; rather, we confine our discussion to specific aspects of optimization involving only the case of differentiable variables where all appropriate partial derivatives can be assumed to exist.

We first provide a rapid review of formulas from the calculus that involve functions of one argument. This is followed by a similar discussion that covers functions of two variables. At this point, optimization subject to side conditions is introduced, and the topic of Lagrange multipliers is described and illustrated numerically.

The next main section of the appendix deals with symbolic differentiation of multivariable functions, with respect to vectors and matrices. Constrained optimization in this most general of cases is also discussed.

We then turn to each of the three major techniques described in the book:

1. multiple regression,
2. principal components analysis,
3. multiple discriminant analysis,

and show how their respective matrix equations are obtained from application of optimization procedures drawn from the calculus.

A.2 DIFFERENTIATION OF FUNCTIONS OF ONE ARGUMENT

In Section 6.2.1, a matrix equation (involving the so-called normal equations of multiple regression) was described:

$$\boxed{\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}$$

where $\mathbf{b}$ is the to-be-solved-for vector of regression coefficients, $\mathbf{X}$ is the data matrix of predictor variables (augmented by a column of unities), and $\mathbf{y}$ is the data vector representing the criterion variable. The parameter vector $\mathbf{b}$ is chosen so as to minimize the sum of squared deviations between the original criterion vector and the fitted values obtained from the regression equation.

In Section 6.4, the matrix equation

$$\boxed{(\mathbf{C}-\lambda_1\mathbf{I})\mathbf{t}_1 = \mathbf{0}}$$

was set up to find the eigenvalue λ_1 and its associated eigenvector $\mathbf{t}_1$ that maximized the variance of point projections of the deviation-from-mean data. The axis itself was obtained by considering the entries of $\mathbf{t}_1$ as direction cosines defining the first principal component. $\mathbf{C}$ denotes the covariance matrix.

In Section 6.5 we sought a vector $\mathbf{v}_1$ that maximized the ratio

$$\boxed{\lambda_1 = \frac{\mathbf{v}_1'\mathbf{A}\mathbf{v}_1}{\mathbf{v}_1'\mathbf{W}\mathbf{v}_1}}$$

where λ_1 is a scalar (actually an eigenvalue), $\mathbf{A}$ denotes the among-group SSCP matrix, and $\mathbf{W}$ denotes the pooled within-group SSCP matrix. One solves for λ_1 via the matrix equation

$$\boxed{(\mathbf{A}-\lambda_1\mathbf{W})\mathbf{v}_1 = \mathbf{0}}$$

The discriminant analysis problem involves finding the eigenstructure of $\mathbf{W}^{-1}\mathbf{A}$, a matrix that is nonsymmetric.[1]

Note that in all three cases we are trying to optimize some function that involves multiple arguments. Also recall that in the case of principal components and multiple discriminant analysis, certain side conditions, such as $\mathbf{t}_1'\mathbf{t}_1$ or $\mathbf{v}_1'\mathbf{v}_1 = 1$, are imposed. Appendix A is motivated by the desire to provide a rationale for the preceding matrix equations. As such, we shall need to draw upon various tools from the calculus, starting with the simplest case of functions involving one argument and then working up to more complex problems involving several variables.

[1] In the cases of principal components and (multiple) discriminant analysis, we shall generally find successive λ_i's, subject to meeting stated side conditions with regard to their associated eigenvectors.

A.2.1 Derivatives of Functions of One Argument

By way of introduction, assume that we have some function of one argument, such as the quadratic

$$y = f(x) = 2x^2$$

We can find the value of $y = f(x)$ for each value x of interest. For a given value of x, let us next imagine taking a somewhat larger value, such as $x_1 = x_0 + \Delta x$. If so, the function y will change, as well, from y to $y + \Delta y$. That is,

$$y + \Delta y = f(x_1) = f(x_0 + \Delta x)$$

If we plot y versus x, the ratio $\Delta y/\Delta x$ can be viewed as the tangent of the angle between the x axis and the chord joining the point (x_0, y) to the point $(x_1, y + \Delta y)$. Furthermore, if Δx is made smaller and smaller, the angle that the chord makes with the x axis will approximate the angle between the x axis and the tangent line of the point (x_0, y). This appears as a dotted line in Fig. A.1.

If we let dy/dx denote the limit of the ratio $\Delta y/\Delta x$ as x_1 approaches x_0, then we can call this limit the (first) derivative of $f(x)$, denoted variously as dy/dx, y', or $f'(x)$:

$$\boxed{\frac{dy}{dx} = y' = f'(x) = \lim_{x_1 \Rightarrow x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}}$$

Note that as $x_1 \Rightarrow x_0$, $\Delta x \Rightarrow 0$.

To illustrate this notion numerically, let us go through the computations for the preceding example:

$$\begin{aligned} y &= f(x) = 2x^2 \\ y + \Delta y &= 2(x + \Delta x)^2 \\ &= 2x^2 + 4x\,\Delta x + 2(\Delta x)^2 \end{aligned}$$

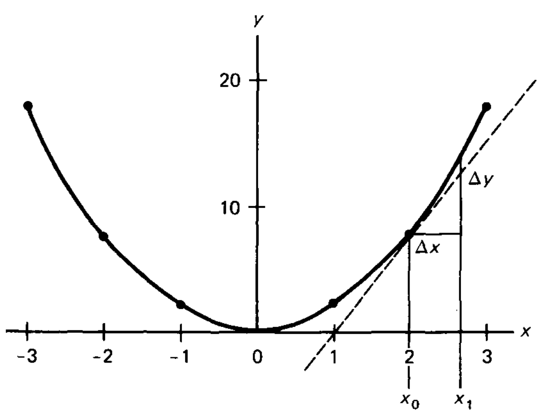

Fig. A.1 Graph of function $y = 2x^2$ $(-3 \leqslant x \leqslant 3)$. Dashed line is tangent line.

TABLE A.1

Derivatives of Elementary Functions

Constant

The derivative of a constant is 0. For example, if

$$f(x) = 12$$

then

$$f'(x) = 0$$

Algebraic

The derivative of ax^n is nax^{n-1} (for $n \neq 0$). For example, if

$$f(x) = 3x^4$$

then

$$f'(x) = 12x^3$$

Exponential

The derivative of the exponential e^x is e^x. The derivative of $e^{f(x)}$ is $f'(x)e^{f(x)}$. The derivative of a^x is $(\ln a)(a^x)$. As one example, if

$$f(x) = e^{-2x}$$

then

$$f'(x) = -2e^{-2x}$$

Logarithmic

The derivative of the natural logarithm $\ln x$ is $1/x$. The derivative of $\ln v(x)$ is $1/v \cdot dv/dx$. For example, if

$$v = x^2$$

then

$$\frac{d}{dx}(\ln v) = \left[\frac{1}{v}\right]\left[\frac{dv}{dx}\right] = \left[\frac{1}{x^2}\right](2x) = \frac{2}{x}$$

Since $y = 2x^2$, we can simplify the above expression to

$$\Delta y = 4x\,\Delta x + 2(\Delta x)^2$$

Dividing both sides by Δx gives us

$$\Delta y/\Delta x = 4x + 2\,\Delta x$$

However, as Δx becomes smaller and smaller, the second term on the right can be neglected, and we get the (first) derivative

$$\frac{dy}{dx} = f'(x) = 4x$$

TABLE A.2

Basic Rules of Differentiation Involving Simple Functions

Sum or Difference of Two Functions

The derivative of a sum (difference) of two functions $u(x)$ and $v(x)$ equals the sum (difference) of their derivatives. If u and v are functions of x, then

$$\frac{d}{dx}(u \pm v) = \frac{du}{dx} \pm \frac{dv}{dx}$$

I

$$u = 2x + 2; \qquad v = 3x - 4$$

$$\frac{du}{dx} = 2; \qquad \frac{dv}{dx} = 3$$

$$\frac{d}{dx}(u + v) = \frac{d}{dx}(5x - 2) = 5 = 2 + 3$$

Product of Two Functions

The derivative of the product of two functions $u(x)$ and $v(x)$ is given by

$$\frac{d}{dx}(uv) = v\frac{du}{dx} + u\frac{dv}{dx}$$

Let

$$u = 2x; \qquad v = 3x - 4$$

$$\frac{du}{dx} = 2; \qquad \frac{dv}{dx} = 3$$

$$\frac{d}{dx}(uv) = \frac{d}{dx}(6x^2 - 8x) = 12x - 8 = 2(3x - 4) + 3(2x)$$

Quotient of Two Functions

The derivative of the quotient of two functions $u(x)$ and $v(x)$ is given by

$$\frac{d}{dx}\left[\frac{u}{v}\right] = \frac{v(du/dx) - u(dv/dx)}{v^2}$$

$$u = 2x^4; \qquad v = 3x^3$$

$$\frac{du}{dx} = 8x^3; \qquad \frac{dv}{dx} = 9x^2$$

$$\frac{d}{dx}\left[\frac{u}{v}\right] = \frac{d}{dx}\left[\frac{2x^4}{3x^3}\right] = 2/3 = \frac{3x^3(8x^3) - 2x^4(9x^2)}{(3x^3)^2}$$

We could, in turn, take the derivative of the function $f'(x) = 4x$ and obtain the second derivative $f''(x) = 4$. (Higher-order derivatives are all zero since 4 is a constant.) That is, second-order derivatives are simply derivatives of first-order derivatives; third-order derivatives are derivatives of second-order derivatives, and so on.

The above procedure can be generalized to the case of finding the derivative of any algebraic expression of the form $y = ax^n$. In general, if $y = ax^n$, we have[2]

$$\boxed{\frac{dy}{dx} = nax^{n-1}}$$

By way of review, Table A.1 lists this formula and others that involve various elementary functions of interest to the applied researcher.[3]

Not only are derivatives of the more common elementary functions of use, but we may also be interested in the derivatives of simple functions of these. Accordingly, Table A.2 lists the basic rules that are applicable to simple functions of two elementary functions, such as their sum or product.

A.2.2 The Chain Rule

Another concept of the elementary calculus that should be reviewed is the chain rule. The chain rule applies to functions of functions. Suppose we have a function $f(g)$ where g, in turn, is $g(x)$, a function of x. If such is the case, the chain rule states that

$$\boxed{\frac{df}{dx} = \left[\frac{df}{dg}\right]\left[\frac{dg}{dx}\right]}$$

To illustrate application of the chain rule, consider the function $f(x) = 2(x^2 + 3x)^2 - 1$, which in turn, can be written as

$$f(g) = 2g^2 - 1$$

$$g(x) = x^2 + 3x$$

The chain rule states that

$$\frac{df}{dx} = \left[\frac{df}{dg}\right]\left[\frac{dg}{dx}\right] = 4g(2x + 3)$$

Next, we substitute the expression for $g(x)$ to get

$$\frac{df}{dx} = 4(x^2 + 3x)(2x + 3) = 8x^3 + 36x^2 + 36x$$

We can verify this result directly by making the substitution of $g(x)$ in $f(g)$ to get

$$f(x) = 2(x^2 + 3x)^2 - 1 = 2x^4 + 12x^3 + 18x^2 - 1$$

and

$$\frac{df}{dx} = 8x^3 + 36x^2 + 36x$$

as desired.

[2] We assume that n is a real number not equal to 0 and that $f(x)$ is defined and differentiable.

[3] Although not shown in Table A.1, the derivative of $\sin x$ is $\cos x$, the derivative of $\cos x$ is $-\sin x$, and the derivative of $\tan x$ is $1/\cos^2 x$.

As a second example, consider the function $f(x) = \ln(x^2 + 3)$. This, in turn, can be expressed as

$$f(g) = \ln g; \qquad g(x) = x^2 + 3$$

Then, by application of the chain rule, we have

$$\frac{df}{dx} = \left[\frac{df}{dg}\right]\left[\frac{dg}{dx}\right] = \frac{1}{g}(2x)$$

and, substituting $g(x) = x^2 + 3$ for g, we have

$$\frac{df}{dx} = \frac{2x}{x^2 + 3}$$

The chain rule can be easily extended to three or more functions in terms of the following:

$$\frac{df}{dx} = \left[\frac{df}{dg}\right]\left[\frac{dg}{dh}\right]\left[\frac{dh}{dx}\right]$$

and so on, for additional functions.

As an example of the case involving three functions, consider the expression

$$f(x) = [\ln(x + 1)]^2$$

This, in turn, can be expressed as

$$f(g) = g^2; \qquad g(h) = \ln h; \qquad h(x) = x + 1$$

Applying the chain rule leads to

$$\frac{df}{dx} = \left[\frac{df}{dg}\right]\left[\frac{dg}{dh}\right]\left[\frac{dh}{dx}\right] = 2g\left(\frac{1}{h}\right)(1) = \frac{2\ln(x + 1)}{x + 1}$$

The chain rule, augmented by the formulas shown in Tables A.1 and A.2, provides a flexible procedure for differentiating the more common functions encountered in applied research.

A.2.3 Optimization of Functions of One Argument

As recalled from the elementary calculus, a function of one argument has a local maximum at some point x_0 if the values of the function on either side of x_0 are less than $f(x_0)$. On the other hand, if the values of the function on either side of x_0 are greater than $f(x_0)$, the function has a local minimum. Maxima and minima are called extreme values, and the values of x for which $f(x)$ takes on an extreme value are called *extreme points.*

Suppose that $f(x)$ has a continuously varying first derivative in an interval that includes x_0. If $f(x_0)$ is a maximum, the first derivative must then change from positive to negative. Conversely, if $f(x_0)$ is a minimum, the first derivative must change from negative to positive. These facts relate, of course, to the basic definition of $f'(x)$ as the slope of the curve $y = f(x)$ at the point x_0. Under the preceding conditions then, $f'(x)$ is zero at the point x_0 where the curvature of $f(x)$ changes direction.

More generally, $f'(x) = 0$ is the necessary condition for a *stationary point* (that includes the instances of maxima and minima as special cases). At a stationary point, the function may have either a maximum, a minimum, or neither. For example, each of the following functions displays a stationary point at $x = 0$, but we note that for

$f(x) = x^2$, the stationary point is a minimum

$f(x) = -(x^2)$, the stationary point is a maximum

$f(x) = x^3$, the stationary point is neither a maximum nor minimum[4]

Finding local extreme points for differentiable functions of one argument involves the following steps:

1. Since local extreme points can only occur at stationary points, where $f'(x) = 0$, first find all solutions to the equation

$$f'(x) = 0$$

2. For each of the stationary points obtained from the above solutions, compute higher-order derivatives $f''(x)$, $f'''(x)$, etc., as needed, so as to find the value of the *lowest*-order derivative that is not zero at the stationary point in question.
3. Examine the lowest-order derivative that is nonzero to determine if
   a. its order is even.
      (i) If the value of the derivative of this order is positive, the function exhibits a local minimum at the stationary point under evaluation.
      (ii) If the value of the derivative of this order is negative, the function exhibits a local maximum at the stationary point under evaluation.
   b. its order is odd; if so, the stationary point is an inflection point.[5]

As a simple numerical illustration of the above procedure, consider the function

$$f(x) = x^3 + 2x^2 + x$$

Its first derivative is

$$f'(x) = 3x^2 + 4x + 1$$

Next, we solve for the stationary points by setting $f'(x)$ equal to zero:

$$f'(x) = 3x^2 + 4x + 1 = 0$$

and find, as solutions,

$$x_1 = -1; \qquad x_2 = -\tfrac{1}{3}$$

Next, let us find the second derivative of $f(x)$. This is, of course, the derivative of $f'(x)$. That is,

$$f''(x) = 6x + 4$$

[4] In this case the stationary point is an inflection point where the first derivative $f'(x_0)$ is zero and, furthermore, the second derivative $f''(x)$ changes sign as the function goes through x_0.

[5] As noted earlier, the necessary condition for x_0 to be a stationary point is that $f'(x_0) = 0$. The sufficient conditions appearing above can be obtained by examining successive terms of the Taylor series expansion (Lang, 1964).

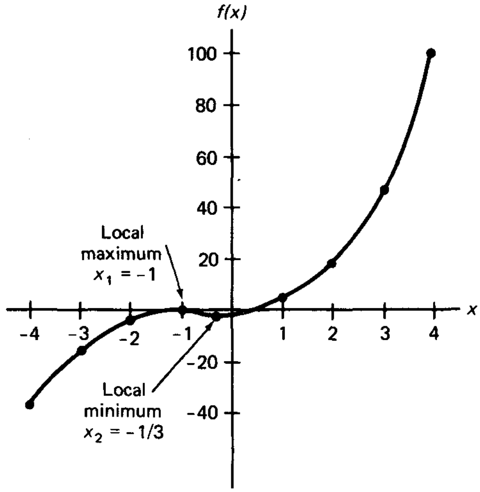

Fig. A.2 Graph of function $f(x) = x^3 + 2x^2 + x$ $(-4 \leqslant x \leqslant 4)$.

We then evaluate $f''(x)$ at the first stationary point x_1 to see if the value of $f''(x)$ is nonzero:

$$f''(-1) = -6 + 4 = -2$$

Since $f''(-1) = -2$ is nonzero, the order of the first nonzero derivative is even; moreover, we have a local maximum at this point since $f''(-1)$ is negative.

Next, the same thing is done for the second stationary point x_2:

$$f''(-\tfrac{1}{3}) = -2 + 4 = 2$$

Since $f''(-1/3) = 2$ is also nonzero, the stationary point is an extreme point; since $f''(-1/3)$ is positive, we have a local minimum at this point.

Figure A.2 shows a plot of the function

$$f(x) = x^3 + 2x^2 + x$$

over the (illustrative) domain $-4 \leqslant x \leqslant 4$. At the point $x_1 = -1, f(x) = 0$, which is a local maximum. At the point $x_2 = -\frac{1}{3}, f(x) = -4/27$, which is a local minimum.

It should be kept in mind that what is being found are *local* stationary points in which interest centers on the behavior of the function in a relatively small interval. As indicated in Fig. A.2, neither x_1 nor x_2 represents a global extremum over the domain $(-4 \leqslant x \leqslant 4)$ of interest.

Figure A.3 further illustrates the distinction between global and local extrema. If we consider local extrema *within* the interval of $x_0 < x < x_6$, a local maximum is found at x_2 and a local minimum at x_4. Also, other stationary points (viz., inflection points) are found at x_1, x_3, and x_5. However, when the end points x_0 and x_6 are also considered, the global maximum turns out to be at x_6, while the global minimum is still at x_4. The

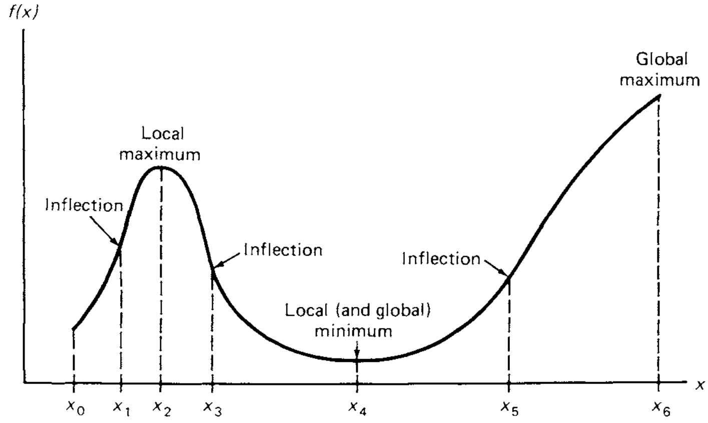

Fig. A.3 Global versus local extrema.

search for global extrema can be quite tedious, particularly if various discontinuities appear in the function.[6] However, further discussion of this specialized process exceeds our intended coverage.

A.3 DIFFERENTIATION OF FUNCTIONS OF TWO ARGUMENTS

The next step in this review discussion of the calculus involves functions of two arguments:

$$\boxed{z = f(x, y)}$$

In the case of functions of one argument we were able to represent $f(x)$ versus x by a curve in two dimensions. Analogously, in the present case $f(x, y)$ versus x and y is represented by a surface embedded in three dimensions. Graphical devices, such as contour lines and projective drawings, are useful in portraying certain three-dimensional relationships in two-dimensional space.

In this part of the appendix we discuss the concepts of level curve, partial differentiation, unconstrained optimization, and optimization subject to equality constraints.

A.3.1 Level Curves and Partial Differentiation

The notion of level curves is employed in a variety of applications. For example, in map making one may use a series of contour lines to represent altitudes, as schematized in Panel I of Fig. A.4. We note that the level of 300 feet consists of all of those points on the hill that involve

$$f(x, y) = 300$$

[6] If discontinuities appear, the function must be examined at each of these points (as well as at end points and stationary points).

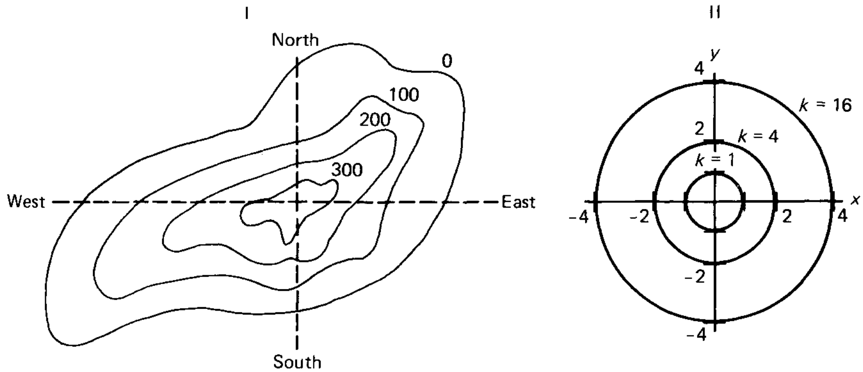

Fig. A.4 Illustrations of level curves. Key: I, altitude of land; II, $z = x^2 + y^2$.

If one were to hike around the hill on that level curve, one would remain at a constant height of 300 feet. Similarly, other contour lines show other altitudes, such as 200 feet, 100 feet, and sea level.

Level curves appear in many fields, such as pictorial displays of air pressure (as isobars), temperature (as isotherms), and consumer utility (as indifference curves). As inferred from Panel I, we could "build up" the hill if we imagined that each level curve were made of cardboard and we stacked one piece on top of another in building up the surface.

Panel II shows level curves for the function

$$z = f(x, y) = x^2 + y^2$$

In this case we can find any level curve of interest by choosing a fixed number k and then finding the set of points x_i and y_j for which

$$k = f(x, y) = x^2 + y^2$$

For example, if we set $k = 9$, we have

$$x^2 + y^2 = 9; \qquad x = \pm\sqrt{9 - y^2}; \quad y = \pm\sqrt{9 - x^2}$$

If we let $y = 0$, then $x = \pm 3$; conversely if $x = 0$, then $y = \pm 3$. As can be inferred from examining the various level curves of Panel II in Fig. A.4, the surface of the function looks like that of a bowl with its center point at the origin of the x, y plane. The level curves are, of course, circles of varying radius.

As a third example of level curves, consider the surface, depicted in Panel I of Fig. A.5, representing the function

$$f(x, y) = 3xy$$

where we assume that $x, y \geqslant 0$. Panel I shows the surface itself as a conelike figure that is cut in half.[7] Panel II shows selected level curves for

$$k_1 = 0; \qquad k_2 = 1; \qquad k_3 = 3; \qquad k_4 = 6; \qquad k_5 = 9; \qquad k_6 = 12$$

We return to this function after a brief review of partial differentiation.

[7] The plane ABC, depicted as a slice through the surface in Panel I, and dotted line AC in Panel II are discussed later (in Section A.3.3) in the context of Lagrange multipliers.

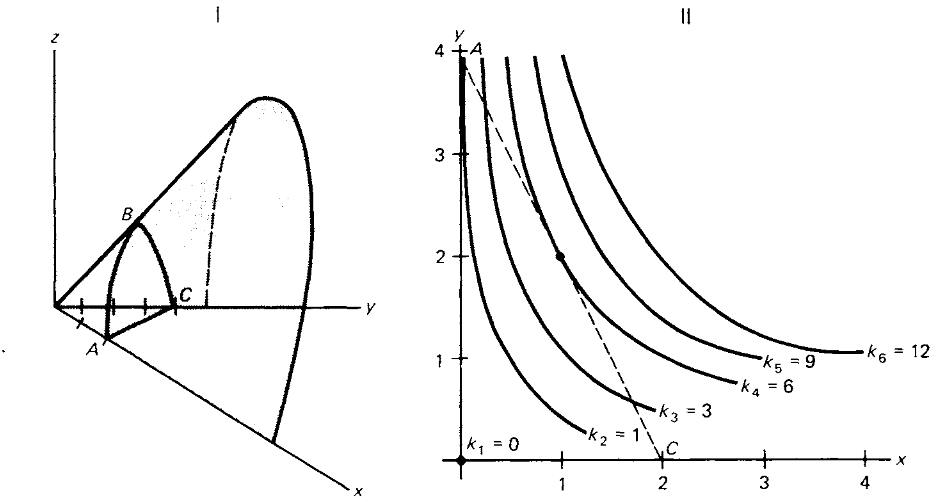

Fig. A.5 Surface and level curve plots of $f(x, y) = 3xy$ $(x, y \geqslant 0)$. Key: I, response surface, $f(x, y) = 3xy$ $(x, y \geqslant 0)$; II, level curves, $k = 3xy$.

Partial differentiation is a straightforward generalization of simple differentiation. Partial derivatives are found by differentiating the function $f(x, y)$ with respect to each variable separately. The variable not involved in the differentiation is treated as a constant. For example, for the function portrayed in Fig. A.5, we have

$$f(x, y) = 3xy; \qquad \frac{\partial f}{\partial x} = 3y; \quad \frac{\partial f}{\partial y} = 3x$$

Partial derivatives are usually denoted by $\partial f/\partial x$ or f_x.

Second-order partial derivatives involve differentiation of first-order partial derivatives, since partial derivatives are, themselves, functions. That is,

$$\frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}(x, y)\right] \qquad \text{and} \qquad \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}(x, y)\right]$$

are second-order derivatives, denoted by the symbols

$$\frac{\partial^2 f}{\partial x^2}(x, y) \quad \text{or} \quad f_{xx}(x, y); \quad \frac{\partial^2 f}{\partial y^2}(x, y) \quad \text{or} \quad f_{yy}(x, y)$$

Moreover, the function $(\partial f/\partial x)(x, y)$ may also be differentiated with respect to y, and $(\partial f/\partial y)(x, y)$ may be differentiated with respect to x. These are called *mixed* partial derivatives and are usually denoted by

$$\frac{\partial^2 f}{\partial x\, \partial y} \quad (\text{or } f_{xy}) \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y\, \partial x} \quad (\text{or } f_{yx})$$

respectively. If the mixed partial derivatives of $f(x, y)$ are continuous, then

$$\frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial^2 f}{\partial y\, \partial x}$$

and the order in which differentiation proceeds is irrelevant. Continuing with the example

$$f(x, y) = 3xy$$

we have

$$\frac{\partial f}{\partial x} = 3y; \qquad \frac{\partial^2 f}{\partial x^2} = 0; \qquad \frac{\partial^2 f}{\partial x\, \partial y} = 3$$

$$\frac{\partial f}{\partial y} = 3x; \qquad \frac{\partial^2 f}{\partial y^2} = 0; \qquad \frac{\partial^2 f}{\partial y\, \partial x} = 3$$

As an additional example, consider the more elaborate polynomial

$$f(x, y) = 2x^2 + y^2 + 3xy + x - y + 3$$

The first- and second-order partial derivatives are

$$\frac{\partial f}{\partial x} = 4x + 3y + 1; \qquad \frac{\partial^2 f}{\partial x^2} = 4; \qquad \frac{\partial^2 f}{\partial x\, \partial y} = 3$$

$$\frac{\partial f}{\partial y} = 3x + 2y - 1; \qquad \frac{\partial^2 f}{\partial y^2} = 2; \qquad \frac{\partial^2 f}{\partial y\, \partial x} = 3$$

Since all second-order partial derivatives are constants, all third- (and higher) order partial derivatives are zero.

A.3.2 Unconstrained Optimization of Functions of Two Arguments

Analogous to the case involving functions of a single argument, conditions for local extrema can be listed for the two-argument case. To be specific, let us continue to consider the case of the function

$$f(x, y) = 2x^2 + y^2 + 3xy + x - y + 3$$

A necessary condition that (x_0, y_0) be a local stationary point is that the following equations are satisfied:

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0; \qquad \frac{\partial f}{\partial y}(x_0, y_0) = 0$$

In the above example we have

$$\frac{\partial f}{\partial x}(x, y) = 4x + 3y + 1 = 0; \qquad \frac{\partial f}{\partial y}(x, y) = 3x + 2y - 1 = 0$$

On solving these equations simultaneously we get

$$x = 5; \qquad y = -7$$

Therefore, (5, −7) is a stationary point.

Sufficiency conditions for local extrema involving functions of two arguments are somewhat more complex than the counterpart case for one variable. To summarize what these are, one first sets up the determinant of second-order partial derivatives as follows:

$$\delta_2 = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x\, \partial y} \\[2ex] \dfrac{\partial^2 f}{\partial y\, \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{vmatrix}$$

which, in the illustrative problem, is

$$|\mathbf{A}| = \begin{vmatrix} 4 & 3 \\ 3 & 2 \end{vmatrix} = (4 \times 2) - (3 \times 3) = -1$$

Next, we examine $\delta_1 = \partial^2 f/\partial x^2 = 4$ and note that its sign is positive. Sufficiency rules, for the general case, can now be stated in terms of δ_1 and δ_2:

1. A local extreme point exists if $\delta_2 > 0$.
   a. The local extreme point is a minimum if $\delta_1 > 0$.
   b. The local extreme point is a maximum if $\delta_1 < 0$.
2. A local extreme point does not exist if $\delta_2 < 0$.
3. A local extreme point may or may not exist if $\delta_2 = 0$. In general, additional examination of the function at the stationary point values is needed to see if an extreme point exists and, if so, whether it is a minimum or a maximum.[8]

In the illustrative problem we note that $\delta_2 = -1$ (<0), and, hence, no local extreme point exists for this function.

Since quadratic functions of the general form

$$f(x, y) = ax^2 + bxy + cy^2 + dx + ey + f$$

appear so frequently in multivariate statistical work, it is useful to examine their properties more generally. First, we shall find, as noted earlier, that all second-order partial derivatives are constants. Stationary points are found by solving the equations $\partial f/\partial x = 0$, $\partial f/\partial y = 0$, simultaneously:

$$\frac{\partial f}{\partial x} = 2ax + by + d = 0; \qquad \frac{\partial^2 f}{\partial x^2} = 2a$$

$$\frac{\partial f}{\partial y} = 2cy + bx + e = 0; \qquad \frac{\partial^2 f}{\partial y^2} = 2c$$

which can be expressed as

$$2ax + by = -d; \qquad bx + 2cy = -e$$

[8] In particular, if both $\partial^2 f/\partial x^2$ and $\partial^2 f/\partial y^2$ are equal to zero, the stationary point is not a local extreme point. If $\partial^2 f/\partial x\, \partial y = \partial^2 f/\partial y\, \partial x = 0$ and if $\partial^2 f/\partial x^2$ and $\partial^2 f/\partial y^2$ have the same sign, the stationary point is a local extreme point. Still, specific examination of the function is generally needed to see whether the point is a minimum or a maximum.

to obtain the solutions

$$x_0 = \frac{2cd-eb}{b^2-4ac}; \qquad y_0 = \frac{2ae-bd}{b^2-4ac}$$

Moreover, since the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial x^2} = 2a; \qquad \frac{\partial^2 f}{\partial y^2} = 2c; \qquad \frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial^2 f}{\partial y\, \partial x} = b$$

we have, as the expression for δ_2,

$$\delta_2 = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x\, \partial y} \\ \dfrac{\partial^2 f}{\partial y\, \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{vmatrix} = 4ac - b^2$$

If $4ac - b^2 > 0$, an extreme point is indicated. That is,

1. if $\delta_2 = 4ac - b^2 > 0$, the stationary point is an extreme point.
   a. If $\delta_1 = \partial^2 f/\partial x^2 = 2a > 0$, then the extreme point is a local minimum.
   b. If $\delta_1 = 2a < 0$, then the extreme point is a local maximum.
2. if $\delta_2 < 0$, no extreme point exists.
3. if $\delta_2 = 0$, additional examination of the function is needed to see if an extreme point exists and, if so, whether it is a minimum or a maximum.

In the preceding numerical case, $x_0 = 5$ and $y_0 = -7$. Substituting these values leads to

$$4ac - b^2 = 4(2)(1) - (3)^2 = -1$$

and the stationary point $(5, -7)$ is neither a minimum nor a maximum, as was indicated earlier.

A.3.3 Constrained Optimization and Lagrange Multipliers

In the previous section our concern was with finding local extrema without constraining the domain over which x and y, the two arguments of $f(x, y)$, might vary. However, cases frequently arise where we are interested in setting up certain side conditions that must be satisfied in the course of optimizing some function $f(x, y)$.

Optimization of functions subject to constraints is a vast topic which goes well beyond our coverage. Here we concern ourselves with only one technique, the method of Lagrange multipliers for optimizing functions subject to equality constraints.

The basic idea behind Lagrange multipliers involves setting up a more general function *that includes the constraint* and optimizing this more general function. Suppose we have a function $f(x, y)$ and a side condition, expressed as $g(x, y) = 0$. If so, we can define a new function, composed of *three* variables:

$$u(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

where λ is the Lagrange multiplier. (The variable λ is an artificial variable that is employed to provide as many unknowns as there are equations.) Having set up the general function $u(x, y, \lambda)$, we can find its partial derivatives with respect to x, y, and λ and set each of them equal to zero:

$$\frac{\partial u}{\partial x} = 0; \qquad \frac{\partial u}{\partial y} = 0; \qquad \frac{\partial u}{\partial \lambda} = 0$$

Any point that satisfies the above necessary conditions is a stationary point. We continue to assume, of course, that both $f(x, y)$ and $g(x, y)$, the latter being the constraint equation, are differentiable in the neighborhood of the stationary point.

Let us illustrate the Lagrange multiplier technique by returning to the function

$$f(x, y) = 3xy$$

as first depicted in Fig. A.5. However, now let us impose the constraint equation that

$$g(x, y) = 2x + y = 4$$

or, equivalently,

$$g(x, y) = 2x + y - 4 = 0$$

As shown in Fig. A.5, application of this constraint results in a plane (labelled as *ABC*) intersecting the surface in Panel I. Furthermore, *AC* in Panel II represents the same linear constraint as a dotted straight line; this line is the locus of all points on the xy plane that satisfy the constraint equation $2x + y = 4$.

We next set up the function $u(x, y, \lambda)$ that incorporates the Lagrange multiplier

$$u(x, y, \lambda) = 3xy - \lambda(2x + y - 4)$$

Then $u(x, y, \lambda)$ is differentiated with respect to each argument, in turn, and each derivative is set equal to zero:

$$\frac{\partial u}{\partial x} = 3y - 2\lambda = 0; \qquad \frac{\partial u}{\partial y} = 3x - \lambda = 0; \qquad -\frac{\partial u}{\partial \lambda} = 2x + y - 4 = 0$$

Notice that the partial derivative with respect to λ is just the constraint equation itself. The next step is to find the stationary point by solving the three preceding equations for x, y, and λ. This is easily done by first expressing both y and x in terms of λ:

$$y = 2\lambda/3; \qquad x = \lambda/3$$

and solving for λ in the third equation:

$$2(\lambda/3) + 2\lambda/3 - 4 = 0; \qquad \lambda = 3$$

We then find x and y to be

$$x = 1; \quad y = 2$$

Also we note that in terms of the original function $f(x, y) = 3xy$, the value of the function at the stationary point $(1, 2)$ is

$$f(1, 2) = 3(1)(2) = 6$$

Panel II of Fig. A.5 provides a graphical representation of what is going on. First, we examine the constraint equation, represented by the dotted line $AC = g(x, y) = 2x + y = 4$. The one level curve for which AC represents the tangent line is

$$k_4 = f(x, y) = 6$$

Thus, when the stationary point $(x_0, y_0) = (1, 2)$ is found with $f(1, 2) = 6$, we see that its tangent line is represented by the constraint equation $g(x, y) = 2x + y = 4$. The value of the function $f(1, 2) = 6$ coincides with the highest level curve that can be reached via AC. Furthermore, in looking at the plane ABC, slicing through the surface in Panel I, we also see that the local extreme value is a maximum point B on the arch traced out by ABC.

One additional point of interest concerns the value of λ itself; in this case $\lambda = 3$. If we examine the original function

$$f(x, y) = 3xy$$

and assume that the (negative) constraint equation $2x + y = 4$ were "relaxed" to $2x + y = 5$, we would have an extra unit of (say) y to work with. If so, we could compute the value of $f(x, y)$ before and after allowing for one unit increase in y. Letting $x_0 = 1$, the x coordinate at the original stationary point, we have

$$f(x, y) = 3(1)y = 3y$$

$$f(x, y + 1) = 3(1)(y + 1) = 3y + 3$$

Hence, an increase in $f(x, y)$ of 3 units could be effected if one more unit of y were available. This is equal to λ, the value of the Lagrange multiplier, as found in the earlier computations.

In summary, Lagrange multipliers are useful in handling one (or possibly more) equality constraints in cases where it would be difficult to solve the problem via direct substitution of the constraint equation(s) into $f(x, y)$.[9] Since a necessary condition for a stationary point is that each first-order partial derivative equals zero, introduction of the Lagrange multiplier λ adds a needed artificial variable to balance out the number of equations with the number of unknowns. In terms of the numerical example

$$u(x, y, \lambda) = 3xy - \lambda(2x + y - 4)$$

we see that partial differentiation with respect to x, y, and λ separately leads to three equations in three unknowns.

Furthermore, if the constraint equation is always met, then the term $(2x + y - 4)$ in the general function $u(x, y, \lambda)$ will always equal zero so that $u(x, y, \lambda)$ will behave in the same way as $f(x, y)$, the original function.[10]

[9] With only two original variables, one would not generally deal with more than a single constraint equation since two constraint equations in two variables would normally have only a single point in common. With a large number of original variables, however, two (or more) equality constraints might be employed.

[10] Sufficiency conditions for local extrema in the context of Lagrange multipliers are found in Hancock (1960).

A.4 SYMBOLIC DIFFERENTIATION

The most complex cases in partial differentiation and function optimization involve functions of several arguments. Often, this situation is portrayed by vector and matrix notation. The term "symbolic differentiation" has been coined to refer to partial differentiation of vector or matrix functions whose results are also described in the same format. For example, symbolic differentiation of a function with respect to a vector involves finding the partial derivative of the function with respect to each entry of the vector; the partial derivatives themselves are then arranged in vector form.

To illustrate, if $y = f(\mathbf{x})$, where $\mathbf{x}$ is a column vector with elements $x_1, x_2, \ldots, x_n$, then one can express its symbolic derivative by means of the column vector

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}$$

Notice that each entry of $\partial f/\partial \mathbf{x}$ is a partial derivative of $f(\mathbf{x})$ with respect to a specific variable. By the same token, one can find a row vector of partial derivatives:[11]

$$\frac{\partial f}{\partial \mathbf{x}'} = (\partial f/\partial x_1, \partial f/\partial x_2, \ldots, \partial f/\partial x_n)$$

In applied multivariate analysis, we frequently have occasion to find the symbolic derivatives of functions that are bilinear or quadratic forms, such as

$u = \mathbf{x}'\mathbf{A}\mathbf{y}$ (where $\mathbf{A}$ may be rectangular)

$v = \mathbf{x}'\mathbf{A}\mathbf{x}$ (where $\mathbf{A}$ is square, nonsymmetric)

$w = \mathbf{x}'\mathbf{A}\mathbf{x}$ (where $\mathbf{A}$ is symmetric)

$t = \mathbf{x}'\mathbf{I}\mathbf{x}$ (where $\mathbf{I}$ is the identity matrix)

We first consider symbolic differentiation with respect to vectors. We can then turn to problems of optimization of functions involving multivariable arguments.

A.4.1 Symbolic Differentiation with Respect to Vectors

If we take one of the simplest cases first, namely, the linear combination $y = \mathbf{a}'\mathbf{x}$, where $\mathbf{a}'$ is a row vector of coefficients, the symbolic derivative of y with respect to $\mathbf{x}$ is simply the row vector

$$\frac{\partial y}{\partial \mathbf{x}'} = \mathbf{a}'$$

[11] Note that $\partial \mathbf{x}'$ appears as a row vector in the denominator since $\partial f/\partial \mathbf{x}'$ is being expressed explicitly in this form. That is, we shall adopt the notation of $\partial \mathbf{x}$ or $\partial \mathbf{x}'$ on the basis of how the final vector of derivatives is displayed—in column or row form, respectively.

This follows from the fact that we can write y explicitly as

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

Then, each partial derivative of y is found, in turn: $\partial y/\partial x_1 = a_1$, $\partial y/\partial x_2 = a_2, \ldots, \partial y/\partial x_n = a_n$. These elements can then be arranged in the row vector $\mathbf{a}'$.

Next, suppose we have the bilinear form

$$u = \mathbf{x}'\mathbf{A}\mathbf{y}$$

which can be written out explicitly (for $\mathbf{A}$ of order $m \times n$) as

$$\begin{aligned} u = {} & x_1 a_{11} y_1 + x_1 a_{12} y_2 + \cdots + x_1 a_{1n} y_n \\ & + x_2 a_{21} y_1 + x_2 a_{22} y_2 + \ldots + x_2 a_{2n} y_n \\ & + \cdots \\ & + x_m a_{m1} y_1 + x_m a_{m2} y_2 + \cdots + x_m a_{mn} y_n \end{aligned}$$

The partial derivative $\partial u/\partial x_1$ of $u = \mathbf{x}'\mathbf{A}\mathbf{y}$ with respect to the first element x_1 is

$$\frac{\partial u}{\partial x_1} = a_{11} y_1 + a_{12} y_2 + \cdots + a_{1n} y_n$$

which can be written as the scalar product

$$\frac{\partial u}{\partial x_1} = \mathbf{a}_1{}'\mathbf{y}$$

By the same procedure the other partial derivatives are obtained:

$$\frac{\partial u}{\partial x_2} = \mathbf{a}_2{}'\mathbf{y}$$

$$\vdots$$

$$\frac{\partial u}{\partial x_m} = \mathbf{a}_m{}'\mathbf{y}$$

which can all be arranged in the $m \times 1$ column vector

$$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y}$$

By a similar rationale we can obtain $\partial u/\partial \mathbf{y}$ as the $1 \times n$ row vector

$$\frac{\partial u}{\partial \mathbf{y}'} = \mathbf{x}'\mathbf{A}$$

By taking appropriate transposes of the two preceding equations, we could also write

$$\frac{\partial u}{\partial \mathbf{x}'} = \mathbf{y}'\mathbf{A}'; \qquad \frac{\partial u}{\partial \mathbf{y}} = \mathbf{A}'\mathbf{x}$$

as a row and column vector, respectively.

By similar reasoning we can find that the symbolic derivative of

$$v = \mathbf{x}'\mathbf{A}\mathbf{x} \qquad \text{(for square, nonsymmetric } \mathbf{A}\text{)}$$

with respect to x is the column vector

$$\frac{\partial v}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$$

Furthermore, the symbolic derivative of

$$w = \mathbf{x}'\mathbf{A}\mathbf{x} \qquad \text{(for symmetric } \mathbf{A}\text{, so that } \mathbf{A} = \mathbf{A}'\text{)}$$

is the column vector

$$\frac{\partial w}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

In particular, if $\mathbf{A} = \mathbf{I}$, we have the special case of the sum of squares

$$t = \mathbf{x}'\mathbf{I}\mathbf{x} = \mathbf{x}'\mathbf{x}; \qquad \frac{\partial t}{\partial \mathbf{x}} = 2\mathbf{x}$$

As a numerical illustration of the case involving a bilinear form, consider the function

$$u = (x_1, x_2)\begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= x_1y_1 + 2x_2y_1 + 3x_1y_2 + 4x_2y_2 + x_1y_3 + 3x_2y_3$$

If we differentiate u with respect to x, we have the column vector

$$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y} = \begin{bmatrix} y_1 + 3y_2 + y_3 \\ 2y_1 + 4y_2 + 3y_3 \end{bmatrix}$$

In a similar way, we could find symbolic derivatives of other functions with respect to x or y. As would be surmised, however, no new principles are involved.[12]

A.4.2 Some Aspects of Optimization in Matrix Notation

Extreme values can be found for functions of vectors in much the same way as described earlier for functions of scalars. In particular, suppose we had the function

$$y = 2x_1^2 + 3x_2^2$$

subject to the constraint equation

$$g(x) = x_1 - x_2 - 1 = 0$$

[12] Bilinear and quadratic forms can also be differentiated with respect to A, the matrix of the form. For example, for nonsymmetric A, $(\partial/\partial\mathbf{A})(\mathbf{x}'\mathbf{A}\mathbf{x})$ is $\mathbf{x}\mathbf{x}'$. However, this more advanced topic exceeds our scope. The interested reader is referred to Tatsuoka (1971).

We could, of course, apply the same Lagrange multiplier procedure described in Section A.3.3 to find the extreme values (if any) of this function. However, let us express the equations in vector or matrix form and work through their solution in this format:

$$y = \mathbf{x}'\mathbf{A}\mathbf{x} = (x_1, x_2)\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

subject to

$$g = \mathbf{c}'\mathbf{x} - 1 = 0 = \left\{(1, -1)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right\} - 1 = 0$$

In matrix equation form, the general function is, analogously,

$$u = \mathbf{x}'\mathbf{A}\mathbf{x} - \lambda(\mathbf{c}'\mathbf{x} - 1)$$

We set its partial derivatives equal to the zero column vector:

$$\frac{\partial u}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - \lambda\mathbf{c} = \mathbf{0}$$

and find the vector solution

$$\mathbf{x} = \lambda\mathbf{A}^{-1}\mathbf{c}/2$$

Moreover, since $\mathbf{c}'\mathbf{x} - 1 = 0$, then $\mathbf{c}'\mathbf{x} = 1$, so that after multiplying both sides of the preceding equation by $\mathbf{c}'$, we have

$$1 = \mathbf{c}'\mathbf{x} = \lambda\mathbf{c}'\mathbf{A}^{-1}\mathbf{c}/2$$

and λ can then be found from

$$\lambda = 2(\mathbf{c}'\mathbf{A}^{-1}\mathbf{c})^{-1}$$

Substituting this expression for λ in $\mathbf{x} = \lambda\mathbf{A}^{-1}\mathbf{c}/2$ gives us

$$\mathbf{x} = (\mathbf{c}'\mathbf{A}^{-1}\mathbf{c})^{-1}\mathbf{A}^{-1}\mathbf{c}$$

In the simple numerical illustration shown above, we have

$$\lambda = 2\left\{\underset{\mathbf{c}'}{(1, -1)}\underset{\mathbf{A}^{-1}}{\begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}}\underset{\mathbf{c}}{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}\right\}^{-1} = 12/5$$

$$\mathbf{x} = \frac{12}{5} \cdot \frac{1}{2}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3/5 \\ -2/5 \end{bmatrix}$$

At the stationary point $\mathbf{x}' = (3/5, -2/5)$, the value of the function is

$$y = 2(3/5)^2 + 3(-2/5)^2 = 0$$

and it is noted that the constraint equation

$$g(\mathrm{x}) = 3/5-(-2/5)-1 = 0$$

is also satisfied.[13]

In summary, use of matrix notation provides a compact way to set down procedures for function optimization, in this case optimization under a constraint equation.

A.4.3 Conditions for the Optimization of Functions Involving Multivariable Arguments

In preceding sections of the appendix, necessary and sufficient conditions for identifying local extreme points have been listed for the single-argument and (unconstrained) two-argument cases. Things become considerably more complicated when we consider functions of multivariable arguments. Accordingly, we do not delve into the topic in much detail; in particular, all proofs are omitted. The reader interested in a more detailed discussion is referred to books by Beveridge and Schechter (1970) and Wilde and Beightler (1967).

In the case of a multivariable function, the necessary condition for a stationary point continues to be the vanishing of all first-order derivatives. That is, at a staionary point, we have the condition

$$\frac{\partial f}{\partial x_1} = 0; \quad \frac{\partial f}{\partial x_2} = 0; \ldots ; \quad \frac{\partial f}{\partial x_n} = 0$$

This condition holds for either unconstrained or constrained functions (in the context of Lagrange multipliers).

However, as was observed in the cases of one or two arguments, a multivariable function does not necessarily have a local extremum at the stationary point of interest. To examine sufficiency conditions for a local extreme point, use is again made of the determinant of second-order partial derivatives:

$$\delta_n = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1{}^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2{}^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n{}^2} \end{vmatrix}$$

[13] Although we do not delve into details, the stationary point (3/5, −2/5) is a minimum.

As was the case with two arguments, we set up the principal minors of δ_n as follows:[14]

$$\delta_1 = \frac{\partial^2 f}{\partial x_1^2}; \quad \delta_2 = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{vmatrix}; \quad \delta_3 = \begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_3} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \dfrac{\partial^2 f}{\partial x_2\,\partial x_3} \\ \dfrac{\partial^2 f}{\partial x_3\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_3\,\partial x_2} & \dfrac{\partial^2 f}{\partial x_3^2} \end{vmatrix}$$

..., up to, and including, δ_n.

Having done this, we evaluate each of the n determinants. *In order for a stationary point to be an extreme point, all of the n determinants must be nonzero.* A local minimum is distinguished from a local maximum in terms of the pattern of signs of the (evaluated) determinants:

1. If $\delta_j > 0$ for *all* $j = 1, 2, \ldots, n$, then the stationary point is a local minimum.
2. If $\delta_1 < 0$, $\delta_2 > 0$, $\delta_3 < 0$, $\delta_4 > 0$, ..., then the stationary point is a local maximum.
3. If neither situation occurs, one must examine the specific nature of the stationary point by computing values of the function in the neighborhood of the point.

Notice, then, that these conditions generalize what was discussed earlier for functions of one and two arguments.[15]

In case the function is subject to an equality constraint of the type illustrated in the context of Lagrange multipliers, the necessary condition for an extreme point involves finding a vector $\mathbf{x}_0$ that satisfies the $n + 1$ equations

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} - \frac{\lambda\,\partial g(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}; \qquad g(\mathbf{x}) = 0$$

that are obtained by setting the derivatives of

$$u(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$$

with respect to $\mathbf{x}$ and λ each equal to zero.

Sufficiency conditions for a local extremum in the case of Lagrange constraints become rather complex, particularly if more than one constraint equation is involved. Accordingly, the interested reader is referred to more specialized books on the subject, such as the book by Hancock (1960).

[14] By the term principal minors is meant successive determinants computed for submatrices of order 1, order 2, etc., formed along the main (principal) diagonal of the original $n \times n$ matrix.

[15] It should also be mentioned that a rather elegant approach to examining sufficiency conditions utilizes the matrix of partial derivatives as the matrix of a quadratic form. One then checks on whether the form is positive definite, negative definite, etc., and the type of definiteness is related to the type of extremum represented by the stationary point. This approach is fully compatible with the principal minor procedure, described above.

A.5 APPLICATION OF THE CALCULUS TO MULTIVARIATE ANALYSIS

At this point we have discussed (albeit selectively and rapidly) a number of concepts from the calculus that relate to the development of various matrix equations that arise in solving problems in multiple regression, principal components, and multiple discriminant analysis. It is now time to examine the specific nature of these central equations in multivariate analysis.

A.5.1 Multiple Regression Equations

The so-called normal equations of multiple regression theory represent a straightforward application of function minimization that utilizes the least-squares criterion. In multiple regression we have the case in which the matrix equation

$$\boxed{\mathbf{y} \cong \mathbf{X}\mathbf{b}}$$

has more equations (one for each case) than unknowns. As recalled from Chapters 1 and 6, $\mathbf{X}$ is the data matrix of predictors (augmented by a column vector of unities); $\mathbf{y}$ is the data vector representing the criterion variable; $\mathbf{b}$ is the to-be-solved-for vector of regression coefficients (including the intercept term); and $\cong$ denotes least-squares approximation.

The vector of prediction errors can be written as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

where $\hat{\mathbf{y}}$ denotes the set of predicted values for $\mathbf{y}$. As we know, the least-squares criterion seeks a vector $\mathbf{b}$ that minimizes

$$f = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$$

Since $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$, we have

$$\begin{aligned} f &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{y}'\mathbf{y} \end{aligned}$$

where $\mathbf{y}'\mathbf{X}\mathbf{b} = \mathbf{b}'\mathbf{X}'\mathbf{y}$ since each term denotes the same scalar. Our objective is to find a vector of parameters $\mathbf{b}$ that minimizes f. This suggests finding the symbolic derivative and setting it equal to the $\mathbf{0}$ column vector:

$$\frac{\partial f}{\partial \mathbf{b}} = 2\mathbf{X}'\mathbf{X}\mathbf{b} - 2\mathbf{X}'\mathbf{y} = \mathbf{0}$$

We note that $\mathbf{X}'\mathbf{X}$ in $\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$ is symmetric, with derivative $2\mathbf{X}'\mathbf{X}\mathbf{b}$. Furthermore, we observe that the partial derivative with respect to the row vector $\mathbf{b}'$ is being found; hence, we take the transpose of $2\mathbf{y}'\mathbf{X}$ to obtain $2\mathbf{X}'\mathbf{y}$, the second term in the preceding equation. Dividing both sides by 2 and transposing leads to

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

and solving for $\mathbf{b}$, we get

$$\boxed{\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}$$

We now recognize the matrix equation as that appearing in the discussion of multiple regression in Chapter 6. Although no check of sufficiency conditions has been made here, it turns out that $\mathbf{b}$ is the vector of parameters that does, indeed, minimize the function f.

A.5.2 Principal Components Analysis

In principal components analysis, we recall that interest centers on rotation of a deviation-from-mean data matrix $\mathbf{X}_d$ so as to maximize the quadratic form

$$f = \mathbf{t}'(\mathbf{X}_d{}'\mathbf{X}_d)\mathbf{t}$$

where we denote the SSCP matrix by $\mathbf{X}_d{}'\mathbf{X}_d$, the minor product moment of $\mathbf{X}_d$. (Alternatively, we could use the raw cross products, covariance, or correlation matrix.) Furthermore, we want to restrict the vector $\mathbf{t}$ to be a set of direction cosines that define the vector of linear composites:

$$\mathbf{y} = \mathbf{X}_d\mathbf{t}, \qquad \text{where} \quad \mathbf{t}'\mathbf{t} = 1$$

If we let $\mathbf{A} = \mathbf{X}_d{}'\mathbf{X}_d$, the principal components problem is to maximize

$$f = \mathbf{t}'\mathbf{A}\mathbf{t}$$

subject to the constraint that $\mathbf{t}'\mathbf{t} = 1$.

Based on our discussion of Lagrange multipliers, we can formalize the task by writing

$$u = \mathbf{t}'\mathbf{A}\mathbf{t} - \lambda(\mathbf{t}'\mathbf{t} - 1)$$

where $\mathbf{t}'\mathbf{t} - 1 = 0$ represents the constraint equation. As we know, the problem is to find the symbolic partial derivative of u with respect to $\mathbf{t}$ and set this equal to the $\mathbf{0}$ vector. Remembering that $\mathbf{A}$ is symmetric, we obtain

$$\frac{\partial u}{\partial \mathbf{t}} = 2\mathbf{A}\mathbf{t} - 2\lambda\mathbf{t} = \mathbf{0}$$

Next, dividing through by 2 and factoring out $\mathbf{t}$, we get

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{t} = \mathbf{0}$$

This represents the necessary condition to be satisfied by a stationary point $\mathbf{t}$ in which the constraint equation $\mathbf{t}'\mathbf{t} = 1$ is also satisfied. Again, we do not delve into the more complex topic of checking on sufficiency conditions, other than to say that the eigenvector $\mathbf{t}_1$ associated with the largest eigenvalue λ_1 of $\mathbf{A}$ is the vector of direction cosines that maximizes the function u.

A.5.3 Multiple Discriminant Analysis

As recalled from Chapter 6, in multiple discriminant analysis we seek a vector v with the property of maximizing the ratio

$$\lambda = \frac{\mathbf{v}'\mathbf{A}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}}$$

where **A** is the among-group SSCP matrix and **W** is the pooled within-group SSCP matrix. (Again, we could place some restriction on the vector v, such as $\mathbf{v}'\mathbf{v} = 1$.) Note, however, that λ, the discriminant ratio in the present context, is simply the quotient of two functions (as illustrated in Table A.2). We can then find the symbolic derivative of λ with respect to v, by means of the quotient rule, and set it equal to the **0** vector:

$$\frac{\partial \lambda}{\partial \mathbf{v}} = \frac{2[(\mathbf{A}\mathbf{v})(\mathbf{v}'\mathbf{W}\mathbf{v}) - (\mathbf{v}'\mathbf{A}\mathbf{v})(\mathbf{W}\mathbf{v})]}{(\mathbf{v}'\mathbf{W}\mathbf{v})^2} = \mathbf{0}$$

This can be simplified by dividing numerator and denominator by ($\mathbf{v}'\mathbf{W}\mathbf{v}$) and making the substitution

$$\lambda = \frac{\mathbf{v}'\mathbf{A}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}}$$

to obtain

$$\frac{2[\mathbf{A}\mathbf{v} - \lambda\mathbf{W}\mathbf{v}]}{\mathbf{v}'\mathbf{W}\mathbf{v}} = \mathbf{0}$$

Next, we divide both sides by the scalar 2 and further simplify to

$$(\mathbf{A} - \lambda\mathbf{W})\mathbf{v} = \mathbf{0}$$

Next, assuming that **W** is nonsingular, we have the familiar expression of Chapter 6:

$$(\mathbf{W}^{-1}\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$$

where, as we know, $\mathbf{W}^{-1}\mathbf{A}$ is nonsymmetric. Again, we omit discussion of the sufficiency conditions, indicating that a maximum has been found. Suffice it to say that all three procedures:

1. multiple regression,
2. principal components analysis, and
3. multiple discriminant analysis

involve aspects of the calculus that deal with the optimization of functions of multivariable arguments. The concept of symbolic differentiation is central to the topic as well as the techniques of function optimization, either unconstrained or constrained optimization, as the case may be.

A.6 SUMMARY

This appendix has dealt with those aspects of the calculus—particularly symbolic differentiation and optimization theory—related to the matrix equations that appear in various multivariate methods, such as multiple regression, principal components, and multiple discriminant analysis.

The review was brief and selective. We first discussed the differentiation of functions of one argument, including the statement of necessary and sufficient conditions for local extrema. This was followed by a similar discussion of the case involving functions of two arguments. Also, the technique of Lagrange multipliers was introduced at this point.

We next described the most general case of functions of multivariable arguments and the concept of symbolic differentiation. Symbolic derivatives of common matrix functions were illustrated, and necessary and sufficient conditions for local extrema of multivariable functions were also listed. We concluded the appendix with applications of the calculus to the derivation of matrix equations in multiple regression, principal components, and multiple discriminant analysis.

REVIEW QUESTIONS

1. By means of the chain rule, find the derivative of the following functions:

$$\text{a.}\quad \ln(2x-x^2) \qquad \text{b.}\quad \frac{1-2x}{x^2+4} \qquad \text{c.}\quad e^{x^{1/2}} \qquad \text{d.}\quad \left[\frac{x^3-1}{2x^3+1}\right]^4$$

2. Find (and identify) extreme points for the function

$$f(x) = \frac{6x}{x^2+1}$$

over the domain $-2 \leqslant x \leqslant 2$.

3. Find the partial derivative of $f(x, y)$ at $f(1, 3)$ where

$$f(x, y) = x^2 + 2x + 4y + \ln(x^2 + y^2)$$

4. Find the minimum of $f(x, y) = 2x^2 + 4x + 8y + y^2$.
5. Find (and identify) a stationary point of the function

$$f(x, y) = y^2 + 4y + 2x - x^2$$

subject to the constraint

$$x + 2y = 2$$

6. Find the derivative with respect to x of the quadratic form

$$g(\mathbf{x}) = (x_1, x_2, x_3)\begin{bmatrix} 2 & 3 & 4 \\ 3 & 1 & 2 \\ 4 & 2 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Evaluate the derivative at $\mathbf{x}' = (3, 1, 2)$.

7. If $g(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + c$ where $\mathbf{A}$ is symmetric, then it can be shown that the derivative of this general quadratic function with respect to $\mathbf{x}$ is

$$\frac{\partial g}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} + \mathbf{b}$$

Furthermore, a stationary point is given by

$$2\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}; \qquad \mathbf{x} = -\tfrac{1}{2}\mathbf{A}^{-1}\mathbf{b}$$

If

$$\underset{\mathbf{x}'}{(x_1, x_2)} \underset{\mathbf{A}}{\begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} + \underset{\mathbf{b}'}{(2, 2)} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} \underset{c}{-4}$$

is the function $g(\mathbf{x})$, find (and identify) a stationary point of $g(\mathbf{x})$.