7

APPLICATIONS OF LINEAR ALGEBRA TO NON-LINEAR FUNCTIONS

A

Introduction

From many points of view this chapter should be the last few sections of Chapter 6. You may recall we motivated that chapter by looking at the system of equations:

$$\begin{aligned} y_1 &= f_1(x_1,\dots,x_n) \\ &\;\;\vdots \\ y_m &= f_m(x_1,\dots,x_n) \end{aligned} \tag{1}$$

and studying the special case in which $f_1,\dots,$ and f_m were <u>linear</u> functions of the n variables $x_1,\dots,$ and x_n.

The reason for emphasizing linearity was based on the fundamental result that if $f_1,\dots,$ and f_m were not linear but were at least <u>continuously differentiable functions</u> of $x_1,\dots,$ and x_n, then increments $\Delta x_1,\dots,$ and Δx_n would produce increments $\Delta y_1,\dots,$ and Δy_m which could be approximated by the differentials $dy_1,\dots,$ and dy_m, where

$$\begin{aligned} dy_1 &= \frac{\partial y_1}{\partial x_1}\,dx_1 + \dots + \frac{\partial y_1}{\partial x_n}\,dx_n \\ &\;\;\vdots \\ dy_m &= \frac{\partial y_m}{\partial x_1}\,dx_1 + \dots + \frac{\partial y_m}{\partial x_n}\,dx_n \end{aligned} \tag{2}$$

i.e.,

$$\begin{aligned} \Delta y_1 &\approx \frac{\partial y_1}{\partial x_1}\,\Delta x_1 + \dots + \frac{\partial y_1}{\partial x_n}\,\Delta x_n \\ &\;\;\vdots \\ \Delta y_m &\approx \frac{\partial y_m}{\partial x_1}\,\Delta x_1 + \dots + \frac{\partial y_m}{\partial x_n}\,\Delta x_n \end{aligned}$$

Recall that $x_1,\dots,$ and x_n refer to increments measured from a given point; say, $x_1 = a_1,\dots,$ $x_n = a_n$, so that each of the coefficients, $\partial y_i/\partial x_j$, in (2) is a <u>constant</u>, namely

$$\left.\frac{\partial y_i}{\partial x_j}\right|^{*} \quad , \ \underline{x} = \underline{a}\ . \tag{2'}$$

* Actually, had we been consistent with our notation in (1) we would have written $f_{i_{x_j}}(a_1,\dots,a_n)$ rather than (2'). The confusion arising from the use of multiple subscripts encouraged us to adopt the shorter notation used in (2) and (2').

[where $\underline{x} = (x_1,\ldots,x_n)$ and $\underline{a} = (a_1,\ldots,a_n)$]

In any event, the main point is that while system (1) is not necessarily linear, system (2) is. That is, system (2) expresses $dy_1,\ldots,$ and dy_m as linear combinations of $dx_1,\ldots,$ and dx_n.

Thus, the study of linear algebra, introduced in Chapter 6 to help us study system (1) in the case that the functions were linear, can now be applied to system (2) provided only that the functions are continuously differentiable (a far weaker [i.e., more general] condition than being linear). In summary, while $y_1,\ldots,$ and y_m need not be linear combinations of $x_1,\ldots,$ and x_n; $dy_1,\ldots,$ and dy_m are linear combinations of $dx_1,\ldots,$ and dx_n; so that we may view $\Delta y_1,\ldots,$ and Δy_m as linear combinations of $\Delta x_1,\ldots, \Delta x_n$ in sufficiently small neighborhoods of a point $(a_1 \ldots, a_n)$.

Our main reason for beginning a new chapter at this time is so that we may better emphasize the implications of system (2).

B

The Jacobian Matrix

In terms of <u>structure</u>, system (1) may be viewed as an example of $\underline{f}(\underline{x})$. That is, system (1) may be identified with the <u>vector</u> function

$$\underline{f} : E^n \rightarrow E^m \tag{3}$$

defined by

$$\underline{f}(x_1,\ldots,x_n) = (y_1,\ldots,y_m) \tag{4}$$

where

$y_1,\ldots,$ and y_m are as in (1).

[Equation (4) is abbreviated as usual by $\underline{y} = \underline{f}(\underline{x})$ where $\underline{y} = (y_1,\ldots,y_m)$ and $\underline{x} = (x_1,\ldots,x_n)$]

In Chapter 6 we introduced such questions as whether $\underline{f}$ was onto and/or 1-1. We could extend our inquiry still further by asking what it would mean to talk about, say,

$$\lim_{\underline{x} \to \underline{a}} \underline{f}(\underline{x})$$

For example, in line with our previous strategies, it would make sense to say that

$$\lim_{\underline{x} \to \underline{a}} \underline{f}(\underline{x}) = \underline{L} \quad (\text{where } \underline{x} \in E^r \text{ but } \underline{L} \in E^m \text{ since } \underline{f}(\underline{x}) \in E^m)$$

means

Given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < || \underline{x} - \underline{a} || < \delta \rightarrow || \underline{f}(\underline{x}) - \underline{L} || < \varepsilon .$$

We also might have noticed that $\underline{f}$ may itself be viewed as an m-triple. Namely, if we substitute the values of $y_1, \ldots,$ and y_m as given by (1) into (4) we obtain

$$\underline{f}(x_1, \ldots, x_n) = [f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)]$$

or

$$\underline{f}(\underline{x}) = [f_1(\underline{x}), \ldots, f_m(\underline{x})] \tag{5}$$

Since each component of the m-triple in equation (5) is a <u>scalar</u> function of $\underline{x}$ we already know how to compute limits involving these components. [By way of review $\lim_{\underline{x} \to \underline{a}} f(\underline{x}) = L$ means given $\varepsilon > 0$ we can find $\delta > 0$ such that $0 < || \underline{x} - \underline{a} || < \delta \rightarrow | f(\underline{x}) - L| < \varepsilon$]

It would then seem natural to use (5) to define $\lim_{\underline{x} \to \underline{a}} \underline{f}(\underline{x})$; i.e.,

$$\lim_{\underline{x} \to \underline{a}} \underline{f}(\underline{x}) = [\lim_{\underline{x} \to \underline{a}} f_1(\underline{x}), \ldots, \lim_{\underline{x} \to \underline{a}} f_m(\underline{x})] \tag{6}$$

(Definition (6) can be shown to be equivalent to the ε, δ - definition, but this is not too important in the present context.)

Returning to (5), notice that we may now write

$$\underline{f} = (f_1, \ldots, f_m) \tag{7}$$

with the understanding that $(f_1, \ldots, f_m)$ means $[f_1(\underline{x}), \ldots, f_m(\underline{x})]$.

If we were now to become interested in the calculus of $\underline{f}$ it would seem natural to extend the notation in (7) by defining

$$\left.\begin{aligned} \underline{dy} &= (dy_1,\ldots,dy_m) \\ &\text{and} \\ \underline{dx} &= (dx_1,\ldots,dx_n) \end{aligned}\right\} \tag{8}$$

To define $\underline{f}'(\underline{x})$ we might then try to copy the structure

$$dy = f'(x)dx \tag{9}$$

used in the calculus of a single variable.

That is, we shall try to relate $\underline{dy}$ and $\underline{dx}$, as defined by (8), in the form

$$\underline{dy} = M\ \underline{dx} \tag{10}$$

and then to capture the structure in (9), define M to be $\underline{f}'(\underline{x})$.

Observe that equation (10) is a "strange animal", Neither $\underline{dy}$ nor $\underline{dx}$ is a number. In fact, each is a vector and they need not have the same dimension. That is, $\underline{dy}$ is an m-triple, $\underline{dx}$ is an n-triple and m need not equal n. With this in mind, equation (10) should suggest matrix arithmetic. (For example, viewing $\underline{dy}$ and $\underline{dx}$ as column matrices, we see that $\underline{dy}$ is m by 1 while $\underline{dx}$ is n by 1. Hence, by the usual properties of matrix multiplication M [in equation (10)] must be an m by n matrix.)

If we re-examine system (2) in an equivalent matrix form, we see that

$$\underbrace{\begin{bmatrix} dy_1 \\ \cdot \\ \cdot \\ \cdot \\ dy_m \end{bmatrix}}_{\underline{dy}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \underbrace{\begin{bmatrix} dx_1 \\ \cdot \\ \cdot \\ \cdot \\ dx_n \end{bmatrix}}_{\underline{dx}} \tag{11}$$

Equation (11) now tells us how to define $\underline{f}'(\underline{x})$. Namely we define $\underline{f}'(\underline{x})$ to be the m by n matrix $[\partial y_i/\partial k_j]$ whereupon (11) becomes

$\underline{dg} = \underline{f}'(\underline{x})\underline{dx}$.

The matrix $\left[\dfrac{\partial y_i}{\partial x_j}\right]$ is so important that it is given a special name, independently of whether we elect to interpret it as $\underline{f}'(\underline{x})$. Namely if $y_1, \ldots,$ and y_m are <u>continuously differentiable functions</u> of n independent variables $x_1, \ldots,$ and x_n, we define the <u>Jacobian Matrix</u> of $y_1, \ldots, y_m$ with respect to $x_1, \ldots, x_n$ to be the m by n matrix

$$\left[\frac{\partial y_i}{\partial x_j}\right] = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial y_m}{\partial x_1} & & \dfrac{\partial y_m}{\partial x_n} \end{bmatrix} .$$

This matrix is often denoted by

$$J\left(\frac{y_1, \ldots y_m}{x_1, \ldots x_n}\right)$$

or

$$\frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_n)} \; *$$

The remaining sections of this chapter describe applications of the Jacobian.

C

The Inverse Function Theorem

In the previous unit, we discussed <u>informally</u>** how to invert system (1) in the case m = n. Namely, given the system

---

* In many texts $\partial(y_1, \ldots y_n)/\partial(x_1, \ldots, x_n)$ is reserved to name the <u>determinant</u> of the Jacobian matrix. Notice that determinants apply only to square matrices, hence, we prefer our notation and we will write $|\partial(y_1, \ldots y_n)/\partial(x_1, \ldots, x_n)|$ when we mean the determinant.

**That is, we used *loose* expressions such as "$dy \approx \Delta y$" and "sufficiently small" without attempting to show what these meant in rigorous, computational terms.

$$\begin{cases} y_1 = f_1(x_1,\ldots,x_n) \\ \cdot \\ \cdot \\ y_n = f_n(x_1,\ldots,x_n) \end{cases}$$

we formed

$$\begin{cases} dy_1 = \dfrac{\partial y_1}{\partial x_1}\,dx_1 + \ldots + \dfrac{\partial y_1}{\partial x_n}\,dx_n \\ \cdot \\ \cdot \\ dy_n = \dfrac{\partial y_n}{\partial x_1}\,dx_1 + \ldots + \dfrac{\partial y_n}{\partial x_n}\,dx_n \end{cases}$$

from which we concluded (from our studies of linear algebra) that $dx_1,\ldots,$ and dx_n could be expressed uniquely as linear combinations of $dy_1,\ldots$ and dy_n provided that $[\partial y_i/\partial k_j]$ was non-singular (i.e., $[\partial y_i/\partial x_j]^{-1}$ existed, or in the language of determinants, det $[\partial y_i/\partial x_j] \neq 0$).

Restated in the language of vector calculus, from system (1) we obtained system (2) which is equivalent to

$$\underline{dy} = \underline{f}'(\underline{x})\underline{dx}, \qquad \text{where } \underline{f}'(\underline{x}) = \left[\frac{\partial y_i}{\partial x_j}\right],$$

and we concluded that we could solve for $\underline{dx}$ in terms of $\underline{dy}$ provided that $|\underline{f}'(\underline{x})|$ * $\neq 0$. In fact, if $|\underline{f}'(\underline{x})| \neq 0$ then

$$\underline{dx} = [\underline{f}'(\underline{x})]^{-1}\,\underline{dy}\ .$$

This result can be proven rigorously and once proven it goes under the name of the Inversion Theorem or the Inverse Function Theorem.

The theorem corroborates what we already suspected from our informed treatment. More specifically;

The Inverse Function Theorem
Suppose $\underline{f}: E^n \to E^n$ is continuously differentiable** at $\underline{x} = \underline{a} \in E^n$.

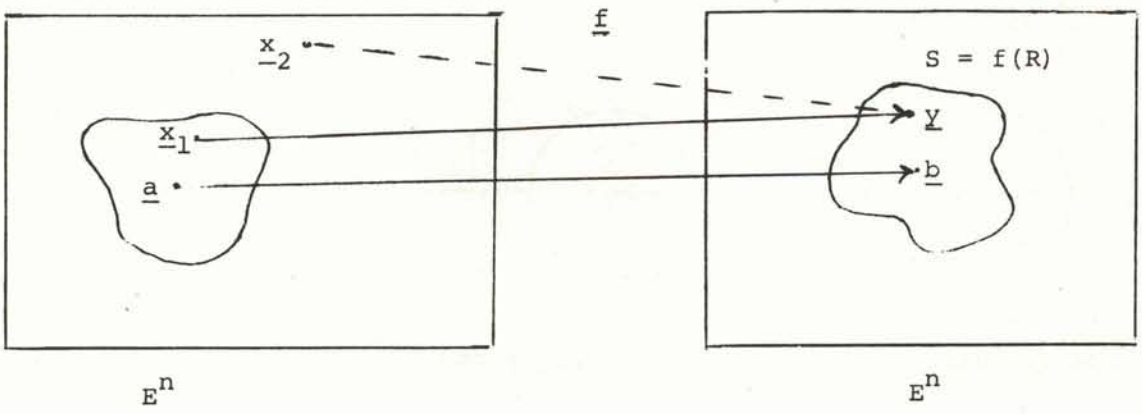
* Keep in mind that $\underline{f}'(\underline{x})$ is, in the present context, an n by n matrix so it makes sense to talk about the determinant of $\underline{f}'(\underline{x})$.

**In terms of $\underline{f} = (f_1,\ldots,f_n)$, we define $\underline{f}$ to be continuously differentiable $\leftrightarrow$ each of its scaler-function components $f_1,\ldots,$ and f_n is continuously differentiable.

Then if $\underline{f}'(\underline{g}) \neq 0$ there is a (sufficiently small) neighborhood R of $\underline{x} = \underline{a}$ such that, when restricted to R, $\underline{f}$ is 1-1 and onto (i.e., $\underline{f}^{-1}$ exists). In other words if $S = \underline{f}(R)$ then $\underline{f}: R \rightarrow S$ is invertible, i.e., $\underline{f}^{-1}: S \rightarrow R$ exists. Moreover $\underline{g} = \underline{f}^{-1}$ is then a continuously differentiable function of $y_1, \ldots, y_n$ in S and

$$|\underline{g}'(\underline{b})| = \frac{1}{|\underline{f}'(a)|} \quad \text{where } \underline{b} = \underline{f}(\underline{a}).$$

The theorem is given a computational application in Exercise 4.6.1. Pictorially, the theorem says:



Summary

1. $\underline{f}$ is 1-1 on R, but
2. If $\underline{f}(\underline{x}_1) = \underline{y} \in S$ where $\underline{x}_1 \in R$ it is possible that there is another vector $\underline{x}_2 \in E^n$ such that $\underline{f}(\underline{x}_1) = \underline{f}(\underline{x}_2)$. However $\underline{x}_2$ cannot be in R (since $\underline{f}$ is 1-1 on R).
3. This is why the inverse function theorem is said to involve a local property. That is, all "bets" about $\underline{f}$ being 1-1 are "off" once R gets too large.

D

A New Look at the Chain Rule

Suppose

$$\left.\begin{array}{l} z_1 = g_1(y_1, \ldots, y_k) \\ \vdots \\ z_m = g_m(y_1, \ldots, y_k) \end{array}\right\} \qquad (12)$$

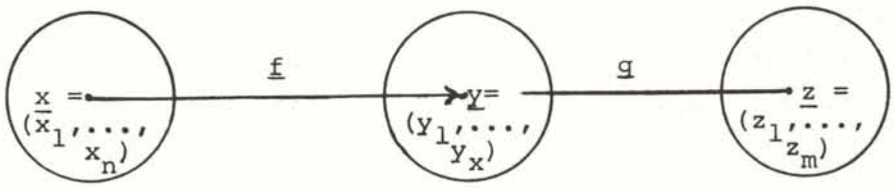
and

$$\left.\begin{array}{l} y_1 = f_1(x_1,\ldots,x_n) \\ \vdots \qquad\quad \vdots \\ y_k = f_k(x_1,\ldots,x_n) \end{array}\right\} \tag{13}$$

Then clearly, systems (12) and (13) allow us to conclude that $z_1,\ldots,z_m$ are functions of $x_1,\ldots,x_n$.

Notice that if we elect to write (12) and (13) in such a way as to emphasize the idea of vector functions, we have

$\underline{f} : E^n \to E^k$, $\underline{g} : E^k \to E^m$



(14)

where $\underline{f} = (f_1,\ldots,f_k)$, $\underline{g} = (g_1,\ldots,g_m)$; and $f_1,\ldots,f_k$, $g_1,\ldots,g_m$ are as in (12) and (13).

In terms of (14), we see that if $\underline{h} = \underline{g} \cdot \underline{f}$ then $\underline{h} : E^n \to E^m$; i.e., $\underline{z} = \underline{h}(\underline{x})$, where $\underline{h}(\underline{x}) = \underline{g}(\underline{f}(\underline{x}))$, and this in turn seems to suggest the chain role. In fact if we assume that $\underline{f}$ and $\underline{g}$ are continuously differentiable, systems (12) and (13) lead to

$$\left.\begin{array}{l} dz_1 = \dfrac{\partial z_1}{\partial y_1}\, dy_1 + \ldots + \dfrac{\partial z_1}{\partial y_k}\, dy_k \\ \vdots \\ dz_m = \dfrac{\partial z_m}{\partial y_1}\, dy_1 + \ldots + \dfrac{\partial z_m}{\partial y_k}\, dy_k \end{array}\right\} \tag{15}$$

and

$$\left.\begin{array}{l} dy_1 = \dfrac{\partial y_1}{\partial x_1}\, dx_1 + \ldots + \dfrac{\partial y_1}{\partial k_n}\, dx_n \\ \vdots \\ dy_k = \dfrac{\partial y_k}{\partial k_1}\, dx_1 + \ldots + \dfrac{\partial y_k}{\partial x_n}\, dx_n \end{array}\right\} \tag{16}$$

Since (15) and (16) are linear systems, we may apply our knowledge of matrix algebra to conclude that

$$\begin{bmatrix} dz_1 \\ \vdots \\ dz_m \end{bmatrix} = \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_1}{\partial y_k} \\ \vdots & & \vdots \\ \frac{\partial z_m}{\partial y_1} & \cdots & \frac{\partial z_m}{\partial y_k} \end{bmatrix} \begin{bmatrix} dy_1 \\ \vdots \\ dy_k \end{bmatrix} = \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_1}{\partial y_k} \\ \vdots & & \vdots \\ \frac{\partial z_m}{\partial y_1} & \cdots & \frac{\partial z_m}{\partial y_k} \end{bmatrix} \left(\begin{bmatrix} \frac{\partial y_1}{\partial k_1} & \cdots & \frac{\partial y_1}{\partial k_n} \\ \vdots & & \vdots \\ \frac{\partial y_k}{\partial x_1} & \cdots & \frac{\partial y_k}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} \right)$$

and since matrix multiplication is associative this says that

$$\begin{bmatrix} dz_1 \\ \vdots \\ dz_m \end{bmatrix} = \left(\begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_1}{\partial y_k} \\ \vdots & & \vdots \\ \frac{\partial z_m}{\partial y_1} & \cdots & \frac{\partial z_m}{\partial y_k} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_k}{\partial x_1} & & \frac{\partial y_k}{\partial x_n} \end{bmatrix} \right) \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} . \tag{17}$$

If we now introduce the Jacobian notation as well as the notations $\underline{dx}$ and $\underline{dz}$ into (17) into (17) we obtain

$$\underline{dz} = \left[\frac{\partial(z_1,\ldots,z_m)}{\partial(y_1,\ldots,y_k)} \right] \left[\frac{\partial(y_1,\ldots,y_k)}{\partial(x_1,\ldots,x_n)} \right] \underline{dx} . \tag{18}$$

With this notation, notice also that (15) and (16) take the forms

$$\left. \begin{aligned} \underline{dz} &= \frac{\partial(z_1,\ldots,z_m)}{\partial(y_1,\ldots,y_k)} \underline{dy} \\ \underline{dy} &= \frac{\partial(y_1,\ldots,y_k)}{\partial(x_1,\ldots,x_n)} \underline{dx} \end{aligned} \right\} \tag{19}$$

To identify this discussion with the derivatives of vector functions notice that equation (19) says

$$\left. \begin{aligned} \underline{dz} &= \underline{g}'(\underline{y})\underline{dy} \\ \underline{dy} &= \underline{f}'(\underline{x})\underline{dx} \end{aligned} \right\} \tag{20}$$

while equation (19) says

$$\underline{dz} = [\underline{g}'(\underline{y})][\underline{f}'(\underline{x})] \ \underline{dx} . \tag{21}$$

Since $\underline{z} = \underline{h}(\underline{x})$, it also follows that

$$\underline{dz} = \underline{h}'(\underline{x})\underline{dx} \tag{22}$$

Comparing (21) and (22) we see that the chain rule as it exists in the study of the calculus of a single real variable extends to the calculus of vector functions.

By way of review, recall that the "old chain rule said:

> If $f : E^1 \to E^1$, $g : E^1 \to E^1$ are differentiable and $h = g \circ f$, then h is also differentiable. Moreover if $x_0 \in \text{dom } f$, $h'(x_0) = g'(y_0) \cdot f'(x_0)$ where $y_0 = f(x_0)$.

Now what we have is:

> If $\underline{f} : E^n \to E^k$ and $\underline{g} : E^k \to E^m$ are continuously differentiable and $\underline{h} = \underline{g} \circ \underline{f}$ then $\underline{h} : E^n \to E^m$ is also continuously differentiable. Moreover, if $\underline{x}_0 \in \text{dom } \underline{f}$ and $\underline{y}_0 = \underline{f}(\underline{x}_0)$, then $\underline{h}'(\underline{x}_0) = [\underline{g}'(\underline{y}_0)][\underline{f}'(\underline{x}_0)]$

In fact if we identify 1 by 1 matrices with numbers, it is easily seen that the "old" chain rule is a special case of the new chain rule. In addition notice that the Jacobian matrix notation $\partial(y_1,\ldots,y_k)/\partial(x_1,\ldots,x_n)$ is an extension of the idea of writing $f'(x)$ as dy/dx. Namely, if we look at (18) and (22) we see that

$$\underline{dz} = \frac{\partial(z_1,\ldots,z_m)}{\partial(x_1,\ldots,x_n)}\,\underline{dx} \text{ and } \underline{dz} = \left[\frac{\partial(z_1,\ldots,z_m)}{\partial(y_1,\ldots,y_k)}\right]\left[\frac{\partial(y_1,\ldots,y_k)}{\partial(x_1,\ldots,x_n)}\right] \text{so that}$$

$$\frac{\partial(z_1,\ldots,z_m)}{\partial(x_1,\ldots,x_n)} = \left[\frac{\partial(z_1,\ldots,z_m)}{\partial(y_1,\ldots,y_k)}\right]\left[\frac{\partial(y_1,\ldots,y_k)}{\partial(x_1,\ldots,x_n)}\right]. \tag{23}$$

Equation (23) indicates that we may treat $(y_1,\ldots,y_k)$ as a number and cancel it from both numerator and denominator on the right hand side of (23), so that the fraction-like notation for the Jacobian is justified.

Again, as reinforcement, notice in the special case that $r = k = m = 1$, equation (23) becomes the familiar

$$\frac{\partial z}{\partial x} = [\frac{dz}{dy}][\frac{dy}{dx}] .$$

As another special case if $r = k = m$ and $\underline{g} = \underline{f}^{-1}$ (assuming $\underline{f}^{-1}$ exists), equation (23) becomes

$$\frac{\partial(x_1,\ldots,x_m}{\partial(x_1,\ldots,x_n)} = \left[\frac{\partial(x_1,\ldots,x_n)}{\partial(y_1,\ldots,y_n)}\right]\left[\frac{\partial(y_1,\ldots,y_n)}{\partial(x_1,\ldots,x_n)}\right] \tag{24}$$

and since

$$\frac{\partial(x_1,\ldots,x_n)}{\partial(x_1,\ldots,x_n)} = I \quad \text{(see Exercise 4.6.8)}$$

equation (24) reaffirms the fact that

$$\left|\frac{\partial(x_1,\ldots,x_n)}{\partial(y_1,\ldots,y_n)}\right| = \left|\frac{\partial(y_1,\ldots,y_n)}{\partial(x_1,\ldots,x_n)}\right|^{-1} .$$

Other computations are left to Exercise 4.6.4, and this concludes our introductory remarks concerning the Jacobian as an extension of the chain rule.

E

Functional Dependence

In our study of linear systems of equations, we saw that a crucial point was whether any of the equations was a linear combination of the others. For example, we saw that the system

$$\begin{aligned} y_1 &= x_1 + 2x_2 + 3x_3 \\ y_2 &= 4x_1 + 5x_2 + 6x_3 \\ y_3 &= 7x_1 + 8x_2 + 9x_3 \end{aligned} \tag{25}$$

was not invertible, even though it was three equations in three unknowns, because the third equation was twice the second minus the first. That is, $y_3 + 2y_2 - y_1$; or equivalently, $y_1 = 2y_2 - y_3$, etc., but the important point is that at least one of the equations can be written as a linear combination of the others. For this reason we say that the system of equations (25) is linearly dependent. The concept of linear dependence will occur again in our course in Block 7 when we discuss differential equations and in Block 8 when we talk more about linear algebra. For now, however, all we want to point out is that the question of invertibility of a system of n equations in n unknowns was resolved in terms of whether the system of equations was linearly dependent.

This idea could be extended to the case in which the number of equations and the number of unknowns were unequal. For example, if we were given four linear equations in seven unknowns, we would, in general, expect to be able to choose three of the unknowns at random and this would have the effect of reducing our system to four equations in four unknowns, from which we could then determine the value of the other four unknowns. The success of this procedure, of course, required that the resulting system of four equations in four unknowns was not linearly dependent.

Now, if we leave out any reference to the equations being linear, the same type of questions is still suggested. For example, suppose we are given the non-linear system of three equations in three unknowns,

$$\left.\begin{aligned} u &= x^2 + y^2 + z^2 \\ v &= 2x^2y^2 \\ w &= x^4 + 2x^2z^2 + z^4 + 2y^2z^2 + y^4 \end{aligned}\right\} \qquad (26)$$

then we might be tempted to ask whether the system can be inverted. That is, does (26) define x,y, and z (at least, implicitly) as functions of u,v, and w? Without trying to establish any analysis of how we might obtain the result, the fact is that from (26) we can show that

$$w = u^2 - v\ . \qquad (27)$$

That is, w is dependent on u and v, even though now the dependence is no longer linear. What this means, by use of (27), is that any function f(u,v,w) is actually a function only of the two independent variables u and v (and notice that we have not proven that u and v are independent, but a glance at (26) should convince you that they are*).

In particular

$$f(u,v,w) = f(u,v,u^2-v) = g(u,v)$$

*But in the event you are not convinced a more indepth statement will be made later.

In any event, when a condition such as (27) holds, we say that the functions (variables) u,v, and w are functionally dependent. This is a generalization of linear dependence. Namely, every case in which we have linear dependence we also have functional dependence, but we may have functional dependence without having linear dependence. Indeed, (27) shows us that we have functional dependence, but the fact that u^2 appears on the right side of (27) tells us that the dependence is nonlinear.

There is a very concise mathematical way of stating what we mean by functions being functionally dependent. For the sake of concreteness we will illustrate the definition in terms of three equations in three unknowns, and then state the more general definition.

Suppose u, v, and w are functions of x, y, and z. Then u, v, and w are said to be functionally dependent if and only if there exists a function $f : E^3 \to E$, such that $f(u,v,w) = 0$ but $f \neq 0$, otherwise u, v, and w are called functionally independent.

A complete understanding of this definition again requires that we know the difference between an identity and an equation.

For example, if $f = 0$, then $f(y_1,y_2,y_3) = 0$ for every 3-triple (y_1,y_2,y_3). What our definition says that if $f \neq 0$ but $f(u,v,w) = 0$ then u, v, and w are functionally dependent.

To illustrate this in terms of the specific system (26), we see from (27) that

$$u^2 - v - w = 0 . \tag{27'}$$

Using (27') as motivation, consider the function f, defined by

$$f(y_1,y_2,y_3) = y_1^{\,2} - y_2 - y_3 . \tag{28}$$

Clearly f is not the zero function, since among other things, if we let $y_1 = y_2 = y_3 = 1$ in (28) we obtain $f(1,1,1) = 1^2 - 1 - 1 = -1 \neq 0$. On the other hand, (28) says that $f(u,v,w) = u^2 - v - w$, and this is identically zero from (27').

More intuitively, the mathematical definition of functional dependence simply states that there is a non-trivial* relationship between u, v, and w that makes these variables dependent.

Our aim in this section is to show how functional dependence is related to a Jacobian matrix (or determinant). Again, omitting proofs, the major result is:

> If $u = u(x,y,z)$, $v = v(x,y,z)$, and $w = w(x,y,z)$ where x, y, and z are independent variables and u,v,w are continuously differentiable. Then u, v, and w are functional dependent in some region R if and only if $\left|\partial(u,v,w)/\partial(x,y,z)\right| \equiv 0$ in the region R (i.e., if and only if $\partial(u,v,w)/\partial(x,y,z)$ is singular for all $(x,y,z) \in R$)

Again, from an intuitive point of view, this condition is the only one that prevents us from solving for dx, dy, and dz in terms of du, dv, and dw once we have that

$$\begin{cases} du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz \\ dv = \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy + \frac{\partial v}{\partial z}\,dz \\ dw = \frac{\partial w}{\partial x}\,dx + \frac{\partial w}{\partial y}\,dy + \frac{\partial w}{\partial z}\,dz \end{cases}$$

With respect to (26), observe that

$$\frac{\partial(u,v,w)}{\partial(x,y,z)} = \begin{bmatrix} 2x & 2y & 2z \\ 4xy^2 & 4x^2y & 0 \\ 4x^3 + 4xz^2 & 4yz^2 + 4y^3 & 4x^2z + 4z^3 + 4y^2z \end{bmatrix}$$

therefore,

$$\begin{aligned} \left|\frac{\partial(u,v,w)}{\partial(x,y,z)}\right| &= 2x[4x^2y(4x^2z + 4z^3 + 4y^2z)] \\ &\quad -2y[4xy^2(4x^2z + 4z^3 + 4y^2z)] \\ &\quad +2z[4xy^2(4yz^2 + 4y^3) - 4x^2y(4x^3 + 4xz^2)] \\ &= 32x^5yz + 32x^3yz^3 + 32x^3y^3z \\ &\quad -32x^3y^3z - 32xy^3z^3 - 32xy^5z \\ &\quad +32xy^3z^3 + 32xy^5z - 32x^5yz - 32x^3yz^3 \end{aligned}$$

*By "trivial" we mean that for any three variables u,v, and w; $0u + 0v + 0w \equiv 0$. Non-trivial, therefore, is reflected in the definition by the requirement that $f \neq 0$.

$\equiv 0,$

which tells us that u, v, and w are functionally dependent, even though it may not reveal the specific relationship that exists between u, v, and w.

This result carries over into the case in which we have more unknowns than equations. For example suppose that $f_1, \ldots,$ and f_n are continuously differentiable functions of the $n + k$ variables $x_1, \ldots, x_{n+k}$. Then the system of n-equations

$$\left.\begin{array}{l} f_1(x_1, \ldots, x_{n+k}) = 0 \\ \cdot \\ \cdot \\ \cdot \\ f_n(x_1, \ldots, x_{n+k}) = 0 \end{array}\right\} \tag{29}$$

implicitly defines any n of the unknowns as differentiable functions of the other k unknowns in a neighborhood of a point satisfying (29) if the Jacobian determinant of $f_1, \ldots, f_n$ with respect to the n dependent variables is not zero at that point.

Again, further details are left for the exercises.

Summary

The Jacobian is to the study of the calculus of vector functions as the derivative is to the study of calculus of a single real variable.

In order to appreciate the Jacobian, we first introduced linear algebra and applied this study to systems of equations involving total differentials.

In this chapter we have tried to give an inkling as to what role the Jacobian, in particular, and matrix algebra, in general, play in the development of the calculus of several real variables.