Chapter 4: Smooth Functions

4.1 Along the demand curve, price and quantity are related according to the equation

$$p = 10 - x$$

This is called the *inverse demand function*. Total revenue $R(x)$ (price times quantity) is given by

$$\begin{aligned} R(x) &= px \\ &= (10 - x)x \\ &= 10x - x^2 \\ &= f(x) \end{aligned}$$

$g(x)$ can be rewritten as

$$g(x) = 21 + 4(x - 3)$$

At $x = 3$, the price is 7 but the marginal revenue of an additional unit is only 4. The function g decomposes (approximately) the total revenue into two components — the revenue from the sale of 3 units ($21 = 3 \times 7$) plus the marginal revenue from the sale of additional units ($4(x - 3)$).

4.2 If your answer is 5 per cent, obtained by subtracting the inflation rate from the growth rate of nominal GDP, you are implicitly using a linear approximation. To see this, let

$$\begin{aligned} p &= \text{price level at the beginning of the year} \\ q &= \text{real GDP at the beginning of the year} \\ dp &= \text{change in prices during year} \\ dq &= \text{change in output during year} \end{aligned}$$

We are told that nominal GDP at the end of the year, $(p + dp)(q + dq)$, equals 1.10 times nominal GDP at the beginning of the year, pq. That is

$$(p + dp)(q + dq) = 1.10pq \tag{4.1}$$

Furthermore, the price level at the end of the year, $p + dp$ equals 1.05 times the price level of the start of year, p:

$$p + dp = 1.05p$$

Substituting this in equation (4.1) yields

$$1.05p(q + dq) = 1.10pq$$

which can be solved to give

$$dq = (\frac{1.10}{1.05} - 1)q = 0.0476$$

The growth rate of real GDP (dq/q) is equal to 4.76 per cent.

To show how the estimate of 5 per cent involves a linear approximation, we expand the expression for real GDP at the end of the year.

$$(p + dp)(q + dq) = pq + pdq + qdp + dpdq$$

Dividing by pq

$$\frac{(p + dp)(q + dq)}{pq} = 1 + \frac{dq}{q} + \frac{dp}{p} + \frac{dpdq}{pq}$$

The growth rate of nominal GDP is

$$\begin{aligned}\frac{(p + dp)(q + dq) - pq}{pq} &= \frac{(p + dp)(q + dq)}{pq} - 1 \\ &= \frac{dq}{q} + \frac{dp}{dp} + \frac{dpdq}{pq} \\ &= \text{Growth rate of output} \\ &+ \text{Inflation rate} \\ &+ \text{Error term}\end{aligned}$$

For small changes, the error term $dpdq/pq$ is insignificant, and we can approximate the growth rate of output according to the sum

$$\text{Growth rate of nominal GDP} = \text{Growth rate of output} + \text{Inflation rate}$$

This is a *linear* approximation since it approximates the function $(p + dp)(q + dq)$ by the linear function $pq + pdq + qdp$. In effect, we are evaluating the change output at the old prices, and the change in prices at the old output, and ignoring in interaction between changes in prices and changes in quantities. The use of linear approximation in growth rates is extremely common in practice.

4.3 From (4.2)

$$\|\mathbf{x}\| \eta(\mathbf{x}) = f(\mathbf{x}_0 + \mathbf{x}) - f(\mathbf{x}_0) - g(\mathbf{x})$$

and therefore

$$\eta(\mathbf{x}) = \frac{f(\mathbf{x}_0 + \mathbf{x}) - f(\mathbf{x}_0) - g(\mathbf{x})}{\|\mathbf{x}\|}$$

$$\eta(\mathbf{x}) \to \mathbf{0}_Y \text{ as } \mathbf{x} \to \mathbf{0}_X$$

can be expressed as

$$\lim_{\mathbf{x} \to \mathbf{0}_X} \eta(\mathbf{x}) = \mathbf{0}_Y$$

4.4 Suppose not. That is, there exist two linear maps such that

$$f(\mathbf{x}_0 + \mathbf{x}) = f(\mathbf{x}_0) + g_1(\mathbf{x}) + \|\mathbf{x}\| \eta_1(\mathbf{x})$$
$$f(\mathbf{x}_0 + \mathbf{x}) = f(\mathbf{x}_0) + g_2(\mathbf{x}) + \|\mathbf{x}\| \eta_2(\mathbf{x})$$

with

$$\lim_{\mathbf{x} \to 0} \eta_i(\mathbf{x}) = 0, \quad i = 1, 2$$

Subtracting we have

$$L_1(\mathbf{x}) - L_2(\mathbf{x}) = \|\mathbf{x}\| \left(\eta_1(\mathbf{x}) - \eta_2(\mathbf{x})\right)$$

and

$$\lim_{\mathbf{x} \to 0} \frac{g_1(\mathbf{x}) - g_2(\mathbf{x})}{\|\mathbf{x}\|} = 0$$

Since $g_1 - g_2$ is linear, (4) implies that $g_1(\mathbf{x}) = g_2(\mathbf{x})$ for all $\mathbf{x} \in X$.

To see this, we proceed by contradiction. Again, suppose not. That is, suppose there exists some $\mathbf{x} \in X$ such that

$$g_1(\mathbf{x}) \neq g_2(\mathbf{x})$$

For this $\mathbf{x}$, let

$$\eta = \frac{g_1(\mathbf{x}) - g_2(\mathbf{x})}{\|\mathbf{x}\|}$$

By linearity,

$$\frac{g_1(t\mathbf{x}) - g_2(t\mathbf{x})}{\|t\mathbf{x}\|} = \eta \text{ for every } \forall t > 0$$

and therefore

$$\lim_{t \to 0} \frac{g_1(t\mathbf{x}) - g_2(t\mathbf{x})}{\|t\mathbf{x}\|} = \eta \neq 0$$

which contradicts (4). Therefore $g_1(\mathbf{x}) = g_2(\mathbf{x})$ for all $\mathbf{x} \in X$.

4.5 If $f \colon X \to Y$ is differentiable at $\mathbf{x}_0$, then

$$f(\mathbf{x}_0 + \mathbf{x}) = f(\mathbf{x}_0) + g(\mathbf{x}) + \eta(\mathbf{x}) \|\mathbf{x}\|$$

where $\eta(\mathbf{x}) \to \mathbf{0}_Y$ as $\mathbf{x} \to \mathbf{0}_X$. Since g is a continuous linear function, $g(\mathbf{x}) \to \mathbf{0}_Y$ as $\mathbf{x} \to \mathbf{0}_X$. Therefore

$$\begin{aligned} \lim_{\mathbf{x} \to 0} f(\mathbf{x}_0 + \mathbf{x}) &= \lim_{\mathbf{x} \to 0} f(\mathbf{x}_0) + \lim_{\mathbf{x} \to 0} g(\mathbf{x}) + \lim_{\mathbf{x} \to 0} \eta(\mathbf{x}) \|\mathbf{x}\| \\ &= f(\mathbf{x}_0) \end{aligned}$$

f is continuous.

4.6

4.7

4.8 The approximation error at the point $(2, 16)$ is

$f(2,16)$	=8.0000
$g(2,16)$	=11.3333
Absolute error	=-3.3333
Percentage error	=-41.6667
Relative error	=-4.1667

By contrast, $h(2,16) = 8 = f(2,16)$. Table 4.1 shows that h gives a good approximation to f in the neighborhood of $(2, 16)$.

Table 4.1: Approximating the Cobb-Douglas function at $(2, 16)$

$\mathbf{x}$	$\mathbf{x}_0 + \mathbf{x}$	$f(\mathbf{x}_0 + \mathbf{x})$	$h(\mathbf{x}_0 + \mathbf{x})$	Approximation Error Percentage	Approximation Error Relative
At their intersection:					
(0.0, 0.0)	(2.0, 16.0)	8.0000	8.0000	0.0000	NIL
Around the unit circle:					
(1.0, 0.0)	(3.0, 16.0)	9.1577	9.3333	-1.9177	-0.1756
(0.7, 0.7)	(2.7, 16.7)	9.1083	9.1785	-0.7712	-0.0702
(0.0, 1.0)	(2.0, 17.0)	8.3300	8.3333	-0.0406	-0.0034
(-0.7, 0.7)	(1.3, 16.7)	7.1196	7.2929	-2.4342	-0.1733
(-1.0, 0.0)	(1.0, 16.0)	6.3496	6.6667	-4.9934	-0.3171
(-0.7, -0.7)	(1.3, 15.3)	6.7119	6.8215	-1.6323	-0.1096
(0.0, -1.0)	(2.0, 15.0)	7.6631	7.6667	-0.0466	-0.0036
(0.7, -0.7)	(2.7, 15.3)	8.5867	8.7071	-1.4018	-0.1204
Around a smaller circle:					
(0.10, 0.00)	(2.1, 16.0)	8.1312	8.1333	-0.0266	-0.0216
(0.07, 0.07)	(2.1, 16.1)	8.1170	8.1179	-0.0103	-0.0083
(0.00, 0.10)	(2.0, 16.1)	8.0333	8.0333	-0.0004	-0.0003
(-0.07, 0.07)	(1.9, 16.1)	7.9279	7.9293	-0.0181	-0.0143
(-0.10, 0.00)	(1.9, 16.0)	7.8644	7.8667	-0.0291	-0.0229
(-0.07, -0.07)	(1.9, 15.9)	7.8813	7.8821	-0.0110	-0.0087
(0.00, -0.10)	(2.0, 15.9)	7.9666	7.9667	-0.0004	-0.0003
(0.07, -0.07)	(2.1, 15.9)	8.0693	8.0707	-0.0171	-0.0138
Parallel to the $\mathbf{x}_1$ axis:					
(-2.0, 0.0)	(0.0, 16.0)	0.0000	5.3333	NIL	-2.6667
(-1.0, 0.0)	(1.0, 16.0)	6.3496	6.6667	-4.9934	-0.3171
(-0.5, 0.0)	(1.5, 16.0)	7.2685	7.3333	-0.8922	-0.1297
(-0.1, 0.0)	(1.9, 16.0)	7.8644	7.8667	-0.0291	-0.0229
(0.0, 0.0)	(2.0, 16.0)	8.0000	8.0000	0.0000	NIL
(0.1, 0.0)	(2.1, 16.0)	8.1312	8.1333	-0.0266	-0.0216
(0.5, 0.0)	(2.5, 16.0)	8.6177	8.6667	-0.5678	-0.0979
(1.0, 0.0)	(3.0, 16.0)	9.1577	9.3333	-1.9177	-0.1756
(2.0, 0.0)	(4.0, 16.0)	10.0794	10.6667	-5.8267	-0.2936
(4.0, 0.0)	(6.0, 16.0)	11.5380	13.3333	-15.5602	-0.4488
Parallel to the $\mathbf{x}_2$ axis:					
(0.0, -4.0)	(2.0, 12.0)	6.6039	6.6667	-0.9511	-0.0157
(0.0, -2.0)	(2.0, 14.0)	7.3186	7.3333	-0.2012	-0.0074
(0.0, -1.0)	(2.0, 15.0)	7.6631	7.6667	-0.0466	-0.0036
(0.0, -0.5)	(2.0, 15.5)	7.8325	7.8333	-0.0112	-0.0018
(0.0, -0.1)	(2.0, 15.9)	7.9666	7.9667	-0.0004	-0.0003
(0.0, 0.0)	(2.0, 16.0)	8.0000	8.0000	0.0000	NIL
(0.0, 0.1)	(2.0, 16.1)	8.0333	8.0333	-0.0004	-0.0003
(0.0, 0.5)	(2.0, 16.5)	8.1658	8.1667	-0.0105	-0.0017
(0.0, 1.0)	(2.0, 17.0)	8.3300	8.3333	-0.0406	-0.0034
(0.0, 2.0)	(2.0, 18.0)	8.6535	8.6667	-0.1522	-0.0066
(0.0, 4.0)	(2.0, 20.0)	9.2832	9.3333	-0.5403	-0.0125

4.9 To show that r is nonlinear, consider

$$\begin{aligned} r((1,2,3,4,5) + (66,55,75,81,63)) &= r(67,57,78,85,68) \\ &= (85,78,68,67,58) \\ &\neq (5,4,3,2,1) + (81,75,67,63,55) \end{aligned}$$

To show that r is differentiable, consider a particular point, say $(66, 55, 75, 81, 63)$. Consider the permutation $g\colon \Re^n \to \Re^n$ defined by

$$g(x_1, x_2, \dots, x_5) = (x_4, x_3, x_1, x_5, x_2)$$

g is linear and

$$g(66, 55, 75, 81, 63) = (81, 75, 67, 63, 55) = r(66, 55, 75, 81, 63)$$

Furthermore, $g(\mathbf{x}) = r(\mathbf{x})$ for all $\mathbf{x}$ close to $(66, 55, 75, 81, 63)$. Hence, $g(\mathbf{x})$ approximates $r(\mathbf{x})$ in a neighborhood of $(66, 55, 75, 81, 63)$ and so r is differentiable at $(66, 55, 75, 81, 63)$. The choice of $(66, 55, 75, 81, 63)$ was arbitrary, and the argument applies at every $\mathbf{x}$ such that $\mathbf{x}_i \neq \mathbf{x}_j$.

In summary, each application of r involves a permutation, although the particular permutation depends upon the argument, $\mathbf{x}$. However, for any given $\mathbf{x}^0$ with $\mathbf{x}_i^0 \neq \mathbf{x}_j^0$, the same permutation applies to all $\mathbf{x}$ in the neighborhood of $\mathbf{x}^0$, so that the permutation (which is a linear function) is the derivative of r at $\mathbf{x}^0$.

4.10 Using (4.3), we have for any $\mathbf{x}$

$$\lim_{t\mathbf{x}\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{x}) - f(\mathbf{x}^0) - Df[\mathbf{x}^0](t\mathbf{x})}{\|t\mathbf{x}\|} = 0$$

or

$$\lim_{t\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{x}) - f(\mathbf{x}^0) - tDf[\mathbf{x}^0](\mathbf{x})}{t\,\|\mathbf{x}\|} = 0$$

For $\|\mathbf{x}\| = 1$, this implies

$$\lim_{t\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{x}) - f(\mathbf{x}^0)}{t} = \frac{tDf[\mathbf{x}^0](\mathbf{x})}{t}$$

that is

$$\vec{D}_{\mathbf{x}} f[\mathbf{x}^0] = \lim_{t\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{x}) - f(\mathbf{x}^0)}{t} = Df[\mathbf{x}^0](\mathbf{x})$$

4.11 By direct calculation

$$\begin{aligned} D_{x_i} f[\mathbf{x}^0] &= \lim_{t\to 0} \frac{h(\mathbf{x}_i^0 + t) - h(\mathbf{x}_i^0)}{t} \\ &= \lim_{t\to 0} \frac{f(\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_i^0 + t, \dots, \mathbf{x}_n^0) - f(\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_i^0, \dots, \mathbf{x}_n^0)}{t} \\ &= \lim_{t\to 0} \frac{f(\mathbf{x}^0 + t\mathbf{e}_i) - f(\mathbf{x}^0)}{t} \\ &= \vec{D}_{\mathbf{e}_i} f[\mathbf{x}^0] \end{aligned}$$

4.12 Define the function

$$\begin{aligned} h(t) &= f\big((8,8) + t(1,1)\big) \\ &= (8+t)^{1/3}(8+t)^{2/3} \\ &= 8+t \end{aligned}$$

The directional derivative of f in the direction $(1,1)$ is

$$\begin{aligned} \vec{D}_{(1,1)} f(8,8) &= \lim_{t \to 0} \frac{h(t) - h(0)}{t} \\ &= 1 \end{aligned}$$

Generalization of this example reveals that the directional derivative of f along any ray through the origin equals 1, that is $\vec{D}_{\mathbf{x}^0} f[\mathbf{x}^0] = 1$ for every $\mathbf{x}^0$. Economically, this means that increasing inputs in the same proportions leads to a proportionate increase in output, which is the property of constant returns to scale. We will study this property of *homogeneity* is some depth in Section 4.6.

4.13 Let $\mathbf{p} = \nabla f(\mathbf{x}^0)$. Each component of $\mathbf{p}$ represents the action of the derivative on an element of the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$(see proof of Theorem 3.4)

$$p_i = Df[\mathbf{x}^0](\mathbf{e}_i)$$

Since $\|\mathbf{e}_i\| = 1$, $Df[\mathbf{x}^0](\mathbf{e}_i)$ is the directional derivative at $\mathbf{x}^0$ in the direction $\mathbf{e}_i$ (Exercise 4.10)

$$p_i = Df[\mathbf{x}^0](\mathbf{e}_i) = \vec{D}_{\mathbf{e}_i}(\mathbf{x}^0)$$

But this is simply the i partial derivative of f (Exercise 4.11)

$$p_i = Df[\mathbf{x}^0](\mathbf{e}_i) = \vec{D}_{\mathbf{e}_i}(\mathbf{x}^0) = D_{x_i} f(\mathbf{x}_0)$$

4.14 Using the standard inner product on $\Re^n$ (Example 3.26) and Exercise 4.13

$$< \nabla f(\mathbf{x}^0), \mathbf{x} >= \sum_{i=1}^{n} D_{x_i} f[\mathbf{x}^0] \mathbf{x}_i = Df[\mathbf{x}^0](\mathbf{x})$$

4.15 Since f is differentiable

$$f(\mathbf{x}^1 + t\mathbf{x}) = f(\mathbf{x}^1) + \nabla f(\mathbf{x}^0)^T t\mathbf{x} + \eta(t\mathbf{x}) \|t\mathbf{x}\|$$

with $\eta(t\mathbf{x}) \to 0$ as $t\mathbf{x} \to \mathbf{0}$. If f is increasing, $f(\mathbf{x}^1 + t\mathbf{x}) \geq f(\mathbf{x}^1)$ for every $\mathbf{x} \geq \mathbf{0}$ and $t > 0$. Therefore

$$\nabla f(\mathbf{x}^0)^T t\mathbf{x} + \eta(t\mathbf{x}) \|t\mathbf{x}\| = t\nabla f(\mathbf{x}^0)^T \mathbf{x} + t\eta(t\mathbf{x}) \|\mathbf{x}\| \geq 0$$

Dividing by t and letting $t \to 0$

$$\nabla f(\mathbf{x}^0)^T \mathbf{x} \geq 0 \text{ for every } \ \mathbf{x} \geq \mathbf{0}$$

In particular, this applies for unit vectors $\mathbf{e}_i$. Therefore

$$D_{x_i} f(\mathbf{x}^1) \geq 0, \qquad i = 1, 2, \ldots, n$$

4.16 The directional derivative $\vec{D}_{\mathbf{x}} f(\mathbf{x}^0)$ measures the rate of increase of f in the direction $\mathbf{x}$. Using Exercises 4.10, 4.14 and 3.61, assuming $\mathbf{x}$ has unit norm,

$$\vec{D}_{\mathbf{x}} f(\mathbf{x}^0) = Df[\mathbf{x}^0](\mathbf{x}) = < \nabla f(\mathbf{x}^0), \mathbf{x} > \leq \left\| \nabla f(\mathbf{x}^0) \right\|$$

This bound is attained when $\mathbf{x} = \nabla f(\mathbf{x}^0) / \left\| \nabla f(\mathbf{x}^0) \right\|$ since

$$\vec{D}_{\mathbf{x}} f(\mathbf{x}^0) = < \nabla f(\mathbf{x}^0), \frac{\nabla f(\mathbf{x}^0)}{\|\nabla f(\mathbf{x}^0)\|} > = \frac{\left\| \nabla f(\mathbf{x}^0) \right\|^2}{\|\nabla f(\mathbf{x}^0)\|} = \left\| \nabla f(\mathbf{x}^0) \right\|$$

The directional derivative is maximized when $\nabla f(\mathbf{x}^0)$ and $\mathbf{x}$ are aligned.

4.17 Using Exercise 4.14

$$H = \{ \, \mathbf{x} \in X : < \nabla f[\mathbf{x}^0], \mathbf{x} > = 0 \, \}$$

4.18 Assume each f_j is differentiable at $\mathbf{x}_0$ and let

$$Df[\mathbf{x}_0] = (Df_1[\mathbf{x}_0], Df_2[\mathbf{x}_0], \dots, Df_m[\mathbf{x}_0])$$

Then

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{x}) - \mathbf{f}[\mathbf{x}_0] - D\mathbf{f}[\mathbf{x}_0]\mathbf{x} = \begin{pmatrix} f_1(\mathbf{x}_0 + \mathbf{x}) - f_1[\mathbf{x}_0] - Df_1[\mathbf{x}_0]\mathbf{x} \\ f_2(\mathbf{x}_0 + \mathbf{x}) - f_2[\mathbf{x}_0] - Df_2[\mathbf{x}_0]\mathbf{x} \\ \vdots \\ f_m(\mathbf{x}_0 + \mathbf{x}) - f_m(\mathbf{x}_0) - Df_m[\mathbf{x}_0]\mathbf{x} \end{pmatrix}$$

and

$$\frac{f_j(\mathbf{x}_0 + \mathbf{x}) - f_j(\mathbf{x}_0) - Df_j[\mathbf{x}_0]\mathbf{x}}{\|\mathbf{x}\|} \to 0 \text{ as } \|\mathbf{x}\| \to 0$$

for every j implies

$$\frac{\mathbf{f}(\mathbf{x}_0 + \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - D\mathbf{f}[\mathbf{x}_0](\mathbf{x})}{\|\mathbf{x}\|} \to 0 \text{ as } \|\mathbf{x}\| \to 0 \tag{4.2}$$

Therefore $\mathbf{f}$ is differentiable with derivative

$$D\mathbf{f}[\mathbf{x}_0] = L = (Df_1(\mathbf{x}_0), Df_2[\mathbf{x}_0], \dots, Df_m[\mathbf{x}_0])$$

Each $Df_j[\mathbf{x}_0]$ is represented by the gradient $\nabla f_j[\mathbf{x}_0]$ (Exercise 4.13) and therefore $Df[\mathbf{x}_0]$ is represented by the matrix

$$J = \begin{pmatrix} \nabla f_1[\mathbf{x}_0] \\ \nabla f_2[\mathbf{x}_0] \\ \vdots \\ \nabla f_m[\mathbf{x}_0] \end{pmatrix} = \begin{pmatrix} D_{x_1} f_1[\mathbf{x}_0] & D_{x_2} f_1[\mathbf{x}_0] & \dots & D_{x_n} f_1[\mathbf{x}_0] \\ D_{x_1} f_2[\mathbf{x}_0] & D_{x_2} f_2[\mathbf{x}_0] & \dots & D_{x_n} f_2[\mathbf{x}_0] \\ \vdots & \vdots & \ddots & \vdots \\ D_{x_1} f_m[\mathbf{x}_0] & D_{x_2} f_m[\mathbf{x}_0] & \dots & D_{x_n} f_m[\mathbf{x}_0] \end{pmatrix}$$

Conversely, if $\mathbf{f}$ is differentiable, its derivative $D\mathbf{f}[\mathbf{x}_0] \colon \Re^n \to \Re^m$ be decomposed into m component $Df_1[\mathbf{x}_0], Df_2[\mathbf{x}_0], \dots, Df_m[\mathbf{x}_0]$ functionals such that

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - D\mathbf{f}[\mathbf{x}_0]\mathbf{x} = \begin{pmatrix} f_1(\mathbf{x}_0 + \mathbf{x}) - f_1(\mathbf{x}_0) - Df_1[\mathbf{x}_0]\mathbf{x} \\ f_2(\mathbf{x}_0 + \mathbf{x}) - f_2(\mathbf{x}_0) - Df_2[\mathbf{x}_0]\mathbf{x} \\ \vdots \\ f_m(\mathbf{x}_0 + \mathbf{x}) - f_m(\mathbf{x}_0) - Df_m[\mathbf{x}_0]\mathbf{x} \end{pmatrix}$$

(4.2) implies that

$$\frac{f_j(\mathbf{x}_0 + \mathbf{x}) - f_j(\mathbf{x}_0) - Df_j[\mathbf{x}_0]\mathbf{x}}{\|\mathbf{x}\|} \to 0 \text{ as } \|\mathbf{x}\| \to 0$$

for every j.

4.19 If $Df[\mathbf{x}_0]$ has full rank, then it is one-to-one (Exercise 3.25) and onto (Exercise 3.16). Therefore $Df[\mathbf{x}_0]$ is nonsingular. The Jacobian $J_f(\mathbf{x}_0)$ represents $Df[\mathbf{x}_0]$, which is therefore nonsingular if and only if $\det J_f(\mathbf{x}_0) \neq 0$.

4.20 When f is a functional, rank $X \geq rankY = 1$. If $Df[\mathbf{x}_0]$ has full rank (1), then $Df[\mathbf{x}_0]$ maps X onto $\Re$ (Exercise 3.16), which requires that $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$.

4.21

4.23 If $f\colon X \times Y \to Z$ is bilinear

$$f(\mathbf{x}_0 + \mathbf{x}, \mathbf{y}_0 + \mathbf{y}) = f(\mathbf{x}_0, \mathbf{y}_0) + f(\mathbf{x}_0, \mathbf{y}) + f(\mathbf{x}, \mathbf{y}_0) + f(\mathbf{x}, \mathbf{y})$$

Defining

$$Df[\mathbf{x}_0, \mathbf{y}_0](\mathbf{x}, \mathbf{y}) = f(\mathbf{x}_0, \mathbf{y}) + f(\mathbf{x}, \mathbf{y}_0)$$

$$f(\mathbf{x}_0 + \mathbf{x}, \mathbf{y}_0 + \mathbf{y}) = f(\mathbf{x}_0, \mathbf{y}_0) + Df[\mathbf{x}_0, \mathbf{y}_0](\mathbf{x}, \mathbf{y}) + f(\mathbf{x}, \mathbf{y})$$

Since f is continuous, there exists M such that

$$f(\mathbf{x}, \mathbf{y}) \leq M \|\mathbf{x}\| \|\mathbf{y}\| \qquad \text{for every } \mathbf{x} \in X \text{ and } \mathbf{y} \in Y$$

and therefore

NOTE *This is not quite right. See Spivak p. 23. Avez (Tilburg) has*

$$\|f(\mathbf{x}, \mathbf{y})\| \leq M \|\mathbf{x}\| \|\mathbf{y}\| \leq M\big(\|\mathbf{x}\| + \|\mathbf{y}\|\big)^2 \leq M \|(\mathbf{x}, \mathbf{y})\|^2$$

which implies that

$$\frac{\|f(\mathbf{x}, \mathbf{y})\|}{\|(\mathbf{x}, \mathbf{y})\|} \to 0 \textit{ as } (\mathbf{x}, \mathbf{y}) \to \mathbf{0}$$

$$\lim_{\mathbf{x}_1, \mathbf{x}_2 \to 0} \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{\|\mathbf{x}_1\| \|\mathbf{x}_2\|} = 0$$

Therefore f is differentiable with derivative

$$Df[\mathbf{x}_0, \mathbf{y}_0] = f(\mathbf{x}_0, \mathbf{y}) + f(\mathbf{x}, \mathbf{y}_0)$$

4.24 Define $m\colon \Re^2 \to \Re$ by

$$m(z_1, z_2) = z_1 z_2$$

Then m is bilinear (Example 3.23) and continuous (Exercise 2.79) and therefore differentiable (Exercise 4.23) with derivative

$$Dm[z_1, \mathbf{z}_2] = m(\mathbf{z}_1, \cdot) + m(\cdot, \mathbf{z}_2)$$

The function fg is the composition of m with f and g,

$$fg(\mathbf{x}, \mathbf{y}) = m(f(\mathbf{x}), g(\mathbf{y}))$$

By the chain rule, the derivative of fg is

$$\begin{aligned} Dfg[\mathbf{x}, \mathbf{y}] &= Dm[z_1, \mathbf{z}_2]\big(Df[\mathbf{x}], Dg[\mathbf{x}]\big) \\ &= m(\mathbf{z}_1, Dg[\mathbf{y}]) + m(Df[\mathbf{x}], \mathbf{z}_2) \\ &= f[\mathbf{x}]Dg[\mathbf{y}]) + g(\mathbf{y})Df[\mathbf{x}] \end{aligned}$$

where $\mathbf{z}_1 = f(\mathbf{x})$ and $\mathbf{z}_2 = g(\mathbf{y})$.

4.25 For $n = 1$, $f(x) = x$ is linear and therefore (Exercise 4.6) $Df[x] = 1$ ($Df[x](x) = x$). For $n = 2$, let $g(x) = x$ so that $f(x) = x^2 = g(x)g(x)$. Using the product rule

$$Df[\mathbf{x}] = g(x)Dg(x) + g(x)Dg(x) = 2x$$

Now assume it is true for $n-1$ and let $g(x) = x^{n-1}$, so that $f(\mathbf{x}) = xg(x)$. By the product rule

$$Df[\mathbf{x}] = xDg[x] + g(x)1$$

By assumption $Dg[x] = (n-1)x^{n-2}$ and therefore

$$Df[\mathbf{x}] = xDg[x] + g(x)1 = x(n-1)x^{n-2} + x^{n-1} = nx^{n-1}$$

4.26 Using the product rule (Exercise 4.24)

$$\begin{aligned} D_x R(x_0) &= f(x_0)D_x x + x_0 D_x f(x_0) \\ &= p_0 + x_0 D_x f(x_0) \end{aligned}$$

where $p_0 = f(x_0)$. Marginal revenue equals one unit at the current price *minus* the reduction in revenue caused by reducing the price on existing sales.

4.27 Fix some $\mathbf{x}^0$ and let $g = \big(Df[\mathbf{x}^0]\big)^{-1}$. Let $\mathbf{y}^0 = f(\mathbf{x}^0)$. For any $\mathbf{y}$, let $\mathbf{x} = f^{-1}(\mathbf{y}^0 + \mathbf{y}) - f^{-1}(\mathbf{y}^0)$ so that $g(\mathbf{y}) = f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^)$ and

$$\big\|f^{-1}(\mathbf{y}^0 + \mathbf{y}) - f^{-1}(\mathbf{y}^0) - g(\mathbf{y})\big\| = \big\|(\mathbf{x} - g\big(f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0)\big))\big\|$$

Since f is differentiable at $\mathbf{x}^0$ with $Df[\mathbf{x}^0] = g^{-1}$

$$f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) = g^{-1}(\mathbf{x}) + \eta(\mathbf{x}) \|\mathbf{x}\|$$

Substituting

$$\begin{aligned} \big\|f^{-1}(\mathbf{y}^0 + \mathbf{y}) - f^{-1}(\mathbf{y}^0) - g(\mathbf{y})\big\| &= \Big\|\mathbf{x} - g\Big(g^{-1}(\mathbf{x}) + \eta(\mathbf{x}) \|\mathbf{x}\|\Big)\Big\| \\ &= \Big\|g\Big(\eta(\mathbf{x}) \|\mathbf{x}\|\Big)\Big\| \\ &= \|\mathbf{x}\| \Big\|g\Big(\eta(\mathbf{x})\Big)\Big\| \end{aligned}$$

with $\eta(\mathbf{x}) \to 0_Y$ as $\mathbf{x} \to \mathbf{0}_X$. Since f^{-1} and g are continuous, $g\Big(\eta(\mathbf{x})\Big) \to 0_X$ as $\mathbf{y} \to \mathbf{0}$. We conclude that f^{-1} is differentiable with derivative $g = \Big(Df[\mathbf{x}^0]\Big)^{-1}$.

4.28

$$\log f(x) = x \log a$$

and therefore

$$f(x) = \exp\Big(\log f(x)\Big) = e^{x \log a}$$

By the Chain Rule, f is differentiable with derivative

$$D_x f(x) = e^{x \log a} \log a = a^x \log a$$

4.29 By Exercise 4.15, the function $g\colon \Re \to \Re$ defined by $g(y) = \frac{1}{y} = y^{-1}$ is differentiable with derivative

$$D_y g[y] = -y^{-2} = -\frac{1}{y^2}$$

Applying the Chain Rule, $1/f = g \circ f$ is differentiable with derivative

$$D\frac{1}{f}[\mathbf{x}] = Dg[f(\mathbf{x})]Df[\mathbf{x}] = -\frac{Df[\mathbf{x}]}{\big(f(\mathbf{x})\big)^2}$$

4.30 Applying the Product Rule to $f \times (1/g)$

$$\begin{aligned} D\frac{f}{g}[\mathbf{x},\mathbf{y}] &= f(\mathbf{x})D\frac{1}{g}[\mathbf{y}] + \frac{1}{g(\mathbf{y})}Df[\mathbf{x}] \\ &= -f(\mathbf{x})\frac{Dg[\mathbf{y}]}{\big(g(\mathbf{y})\big)^2} + \frac{1}{g(\mathbf{y})}Df[\mathbf{x}] \\ &= \frac{g(\mathbf{y})Df[\mathbf{x}] - f(\mathbf{x})Dg[\mathbf{y}]}{\big(g(\mathbf{y})\big)^2} \end{aligned}$$

4.31 In the particular case where

$$f(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^{1/3}\mathbf{x}_2^{2/3}$$

the partial derivatives at the point $(8, 8)$ are

$$D_{x_1} f[(8,8)] = \frac{2}{3} \text{ and } D_{x_2} F[(8,8)] = \frac{1}{3}$$

4.32 The partial derivatives of $f(\mathbf{x})$ are from Table 4.4

$$\begin{aligned} D_{x_i} f[\mathbf{x}] &= x_1^{a_1} x_2^{a_2} \dots a_i x_i^{a_i - 1} \dots x_n^{a_n} \\ &= a_i \frac{f(\mathbf{x})}{x_i} \end{aligned}$$

so that the gradient is

$$\nabla f(\mathbf{x}) = \left(\frac{a_1}{x_1}, \frac{a_2}{x_2}, \dots, \frac{a_n}{x_n}\right) f(\mathbf{x})$$

4.33 Applying the chain rule (Exercise 4.22) to general power function (Example 4.15), the partial derivatives of the CES function are

$$\begin{aligned} D_{x_i} f[\mathbf{x}] &= \frac{1}{\rho}\left(a_1 x_1^\rho + a_2 x_2^\rho + \dots + a_n x_n^\rho\right)^{\frac{1}{\rho}-1} a_i \rho x_i^{\rho-1} \\ &= a_i x_i^{\rho-1}\left(a_1 x_1^\rho + a_2 x_2^\rho + \dots + a_n x_n^\rho\right)^{\frac{1-\rho}{\rho}} \\ &= a_i \left(\frac{f(\mathbf{x})}{x_i}\right)^{1-\rho} \end{aligned}$$

4.34 Define

$$h(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Then h is continuous on $[a, b]$ and differentiable on (a, b) with

$$h(b) = f(b) - \frac{f(b) - f(a)}{b - a}(b - a)f(a) = h(a)$$

By Rolle's theorem (Exercise 5.8), there exists $x \in (a, b)$ such that

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a} = 0$$

4.35 Assume $\nabla f(\mathbf{x}) \geq \mathbf{0}$ for every $\mathbf{x} \in X$. By the mean value theorem, for any $\mathbf{x}^2 \geq \mathbf{x}^1$ in X, there exists $\bar{\mathbf{x}} \in (\mathbf{x}^1, \mathbf{x}^2)$ such that

$$f(\mathbf{x}^2) = f(\mathbf{x}^1) + Df[\bar{\mathbf{x}}](\mathbf{x}^2 - \mathbf{x}^1)$$

Using (4.6)

$$f(\mathbf{x}^2) = f(\mathbf{x}^1) + \sum_{i=1}^{n} D_{x_i} f(\bar{\mathbf{x}})(x_i^2 - x_i^1) \tag{4.3}$$

$\nabla f(\bar{\mathbf{x}}) \geq \mathbf{0}$ and $\mathbf{x}^2 \geq \mathbf{x}^1$ implies that

$$\sum_{i=1}^{n} D_{x_i} f(\bar{\mathbf{x}})(x_i^2 - x_i^1) \geq 0$$

and therefore $f(\mathbf{x}^2) \geq f(\mathbf{x}^1)$. f is increasing. The converse was established in Exercise 4.15

4.36 $\nabla f(\bar{\mathbf{x}}) > \mathbf{0}$ and $\mathbf{x}^2 \geq \mathbf{x}^1$ implies that

$$\sum_{i=1}^{n} D_{x_i} f(\bar{\mathbf{x}})(x_i^2 - x_i^1) > 0$$

Substituting in (4.3)

$$f(\mathbf{x}^2) = f(\mathbf{x}^1) + \sum_{i=1}^{n} D_{x_i} f(\bar{\mathbf{x}})(x_i^2 - x_i^1) > f(\mathbf{x}^1)$$

f is strictly increasing.

4.37 Differentiability implies the existence of the gradient and hence the partial derivatives of f (Exercise 4.13). Continuity of $Df[\mathbf{x}]$ implies the continuity of the partial derivatives.

To prove the converse, choose some $\mathbf{x}^0 \in S$ and define for the partial functions

$$h_i(t) = f(x_1^0, x_2^0, \dots, x_{i-1}^0, t, x_{i+1}^0 + x_{i+1}, \dots, x_n^0 + x_n) \qquad i = 1, 2, \dots, n$$

so that $h_i'(t) = D_{x_i} f(\mathbf{x}^i)$ where $\mathbf{x}^i = (x_1^0, x_2^0, \dots, x_i^0, t, x_{i+1}^0 + x_{i+1}, \dots, x_n^0 + x_n)$. Further, $h_1(x_1^0 + x_1) = f(\mathbf{x}^0 + \mathbf{x})$, $h_n(x_n^0) = f(\mathbf{x}^0)$, and $h_i(x_i^0 + x_i) = h_{i-1}(x_i^0)$ so that

$$f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) = \sum_{i=1}^{n} \left(h_i(x_i^0 + x_i) - h_i(x_i^0) \right)$$

By the mean value theorem, there exists, for each i, $\bar{t}_i$ between $x_i^0 + x_i$ and x_i such that

$$h_i(x_i^0 + x_i) - h_i(x_i) = D_{x_i} f(\bar{\mathbf{x}}^i) x_i$$

where $\bar{\mathbf{x}}^i = (x_1^0, x_2^0, \dots, x_i^0, \bar{t}, x_{i+1}^0 + x_{i+1}, \dots, x_n^0 + x_n)$. Therefore

$$f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) = \sum_{i=1}^n D_{x_i} f(\bar{\mathbf{x}}^i) x_i$$

Define the linear functional

$$g(\mathbf{x}) = \sum_{i=1}^n D_{x_i} f(\mathbf{x}^0) x_i$$

Then

$$f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) - g(\mathbf{x}) = \sum_{i=1}^n \Big(D_{x_i} f(\bar{\mathbf{x}}^i) - D_{x_i} f(\mathbf{x}^0) \Big) x_i$$

and

$$\big\| f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) - g(\mathbf{x}) \big\| \leq \sum_{i=1}^n \big\| (D_{x_i} f(\bar{\mathbf{x}}^i) - D_{x_i} f(\mathbf{x}^0) \big\| \, |x_i|$$

so that

$$\begin{aligned} \lim_{\mathbf{x} \to 0} \frac{\big\| f(\mathbf{x}^0 + \mathbf{x}) - f(\mathbf{x}^0) - g(\mathbf{x}) \big\|}{\|\mathbf{x}\|} &\leq \sum_{i=1}^n \big\| (D_{x_i} f(\bar{\mathbf{x}}^i) - D_{x_i} f(\mathbf{x}^0) \big\| \frac{|x_i|}{\|\mathbf{x}\|} \\ &\leq \sum_{i=1}^n \big\| (D_{x_i} f(\bar{\mathbf{x}}^i) - D_{x_i} f(\mathbf{x}^0) \big\| \\ &= 0 \end{aligned}$$

since the partial derivatives $D_{x_i} f(\mathbf{x})$ are continuous. Therefore f is differentiable with derivative

$$g(\mathbf{x}) = \sum_{i=1}^n D_{x_i} f[\mathbf{x}^0] x_i$$

4.38 For every $\mathbf{x}_1, \mathbf{x}_2 \in S$

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \leq \sup_{\mathbf{x} \in [\mathbf{x}_1, \mathbf{x}_2]} \|Df(\mathbf{x})\| \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

by Corollary 4.1.1. If $Df[\mathbf{x}] = \mathbf{0}$ for every $\mathbf{x} \in X$, then

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| = 0$$

which implies that $f(\mathbf{x}_1) = f(\mathbf{x}_2)$. We conclude that f is constant on S. The converse was established in Exercise 4.7.

4.39 For any $\mathbf{x}_0 \in S$, let $B \subseteq S$ be an open ball of radius of radius r centered on $\mathbf{x}_0$. Applying the mean value inequality (Corollary 4.1.1) to $f_n - f_m$ we have

$$\begin{aligned} \big\| f_n(\mathbf{x}) - f_m(\mathbf{x}) - \big(f_n(\mathbf{x}_0) - f_m(\mathbf{x}_0)\big) \big\| &\leq \sup_{\bar{\mathbf{x}} \in B} \|Df_n[\bar{\mathbf{x}}] - Df_m[\bar{\mathbf{x}}]\| \, \|\mathbf{x} - \mathbf{x}_0\| \\ &\leq r \sup_{\bar{\mathbf{x}} \in B} \|Df_n[\bar{\mathbf{x}}] - Df_m[\bar{\mathbf{x}}]\| \end{aligned}$$

for every $\mathbf{x} \in B$. Given $\epsilon > 0$, there exists N such that for every $m, n > N$

$$\|Df_n - Df_m\| < \epsilon/r \text{ and } \|Df_n - g\| < \epsilon$$

Letting $m \to \infty$

$$\left\| f_n(\mathbf{x}) - f(\mathbf{x}) - \big(f_n(\mathbf{x}_0) - f(\mathbf{x}_0)\big) \right\| \leq \epsilon \|\mathbf{x} - \mathbf{x}_0\| \tag{4.4}$$

for $n \geq N$ and $\mathbf{x} \in B$. Applying the mean value inequality to f_n, there exists δ such that

$$\|f_n(\mathbf{x}) - f_n(\mathbf{x}_0)\| \leq \epsilon \|\mathbf{x} - \mathbf{x}_0\| \tag{4.5}$$

Using (4.4) and (4.5) and the fact that $\|Df_n - g\| < \epsilon$ we deduce that

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - g(\mathbf{x}_0)\| \leq 3\epsilon \|\mathbf{x} - \mathbf{x}_0\|$$

f is differentiable with derivative g.

4.40 Define

$$f(x) = \frac{e^{x+y}}{e^y}$$

By the chain rule (Exercise 4.22)

$$f'(x) = \frac{e^{x+y}}{e^y} = f(x)$$

which implies (Example 4.21) that

$$f(x) = \frac{e^{x+y}}{e^y} = Ae^x \text{ for some } A \in \Re$$

Evaluating at $x = 0$ using $e^0 = 1$ gives

$$f(0) = \frac{e^y}{e^y} = A \text{ for some } A \in \Re$$

so that

$$f(x) = \frac{e^{x+y}}{e^y} = \frac{e^y}{e^y}e^x$$

which implies that

$$e^{x+y} = e^x e^y$$

4.41 If $f = Ax^a$, $f'(x) = aAx^{a-1}$ and

$$E(x) = x\frac{aAx^{a-1}}{Ax^a} = a$$

To show that this is the only function with constant elasticity, define

$$g(x) = \frac{f(x)}{x^a}$$

g is differentiable (Exercise 4.30) with derivative

$$g'(x) = \frac{x^a f'(x) - f(x)ax^{a-1}}{x^{2a}} = \frac{xf'(x) - af(x)}{x^{a+1}} \tag{4.6}$$

If

$$E(x) = x\frac{f'(x)}{f(x)} = a$$

then

$$xf'(x) = af(x)$$

Substituting in (4.6)

$$g'(x) = \frac{xf'(x) - af(x)}{x^{a+1}} = 0 \text{ for every } x \in \Re$$

Therefore, g is a constant function (Exercise 4.38). That is, there exists $A \in \Re$ such that

$$g(x) = \frac{f(x)}{x^a} = A \text{ or } f(x) = Ax^a$$

4.42 Define $g\colon S \to Y$ by

$$g(\mathbf{x}) = f(\mathbf{x}) - Df[\mathbf{x}_0](\mathbf{x})$$

g is differentiable with

$$Dg[\mathbf{x}] = Df[\mathbf{x}] - Df[\mathbf{x}_0]$$

Applying Corollary 4.1.1 to g,

$$\|g(\mathbf{x}_1) - g(\mathbf{x}_2)\| \leq \sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} \|Dg[\mathbf{x}]\| \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in S$. Substituting for g and Dg

$$\begin{aligned}\|f(\mathbf{x}_1) - Df[\mathbf{x}_0](\mathbf{x}_1) - f(\mathbf{x}_2) + Df[\mathbf{x}_0](\mathbf{x}_2)\| &= \|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \\ &\leq \sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} \|Df[\mathbf{x}] - Df[\mathbf{x}_0]\| \, \|\mathbf{x}_1 - \mathbf{x}_2\|\end{aligned}$$

4.43 Since Df is continuous, there exists a neighborhood S of $\mathbf{x}_0$ such that

$$\|Df[\mathbf{x}] - Df[\mathbf{x}_0]\| < \epsilon \text{ for every } \mathbf{x} \in S$$

and therefore for every $\mathbf{x}_1, \mathbf{x}_2 \in S$

$$\sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} \|Df[\mathbf{x}] - Df[\mathbf{x}_0]\| < \epsilon$$

By the previous exercise (Exercise 4.42)

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \leq \epsilon \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

4.44 By the previous exercise (Exercise 4.43), there exists a neighborhood such that

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \leq \epsilon \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

The Triangle Inequality (Exercise 1.200) implies

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| - \|Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \leq \epsilon \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

and therefore

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \leq \|Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| + \epsilon \, \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|Df[\mathbf{x}_0] + \epsilon\| \, \|\mathbf{x}_1 - \mathbf{x}_2\|$$

4.45 Assume not. That is, assume that

$$\mathbf{y} = f(\mathbf{x}_1) - f(\mathbf{x}_2) \notin \overline{\text{conv}}\, A$$

Then by the (strong) separating hyperplane theorem (Proposition 3.14) there exists a linear functional φ on Y such that

$$\varphi(\mathbf{y}) > \varphi(\mathbf{a}) \qquad \text{for every} \quad \mathbf{a} \in A \tag{4.7}$$

where

$$\varphi(y) = \varphi(f(\mathbf{x}_1) - f(\mathbf{x}_2)) = \varphi(f(\mathbf{x}_1)) - \varphi(f(\mathbf{x}_2))$$

φf is a functional on S. By the mean value theorem (Theorem 4.1), there exists some $\bar{\mathbf{x}} \in [\mathbf{x}_1, \mathbf{x}_2]$ such that

$$\varphi \circ f(\mathbf{x}_1) - \varphi \circ f(\mathbf{x}_2)) = D(\varphi \circ f)[\bar{\mathbf{x}}](\mathbf{x}_1 - \mathbf{x}_2) = \varphi \circ Df[\bar{\mathbf{x}}](\mathbf{x} - \mathbf{x}_2) = \varphi(a)$$

for some $a \in A$ contradicting (4.7).

4.46 Define $h\colon [a,b] \to \Re$ by

$$h(x) = \big(f(b) - f(a)\big)g(x) - \big(g(b) - g(a)\big)f(x)$$

$h \in C[a,b]$ and is differentiable on $a,b)$ with

$$h(a) = \big(f(b) - f(a)\big)g(a) - \big(g(b) - g(a)\big)f(a) = f(b)g(a) - f(a)g(b) = h(b)$$

By Rolle's theorem (Exercise 5.8), there exists $x \in (a,b)$ such that

$$h'(x) = \big(f(b) - f(a)\big)g'(x) - \big(g(b) - g(a)\big)f'(x) = 0$$

4.47 The hypothesis that $\lim_{x\to a} Df(x)/Dg(x)$ exists contains two implicit assumptions, namely

- f and g are differentiable on a neighborhood S of a (except perhaps at a)
- $g'(x) \neq 0$ in this neighborhood (except perhaps at a).

Applying the Cauchy mean value theorem, for every $x \in S$, there exists some $y_x \in (a,x)$ such that

$$\frac{f'(y_x)}{g'(y_x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)}$$

and therefore

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(y_x)}{g'(y_x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

4.48 Let $A = a_1 + a_2 + \cdots + a_n \neq 1$. Then from (4.12)

$$\begin{aligned}\lim_{\rho\to 0} g(\rho) &= \frac{a_1 \log x_1 + a_2 \log x_2 + \dots a_n \log x_n}{A} \\ &= \frac{a_1}{A}\log x_1 + \frac{a_2}{A}\log x_2 + \dots \frac{a_n}{A}\log x_n\end{aligned}$$

and therefore

$$\lim_{\rho\to 0} \log f(\rho,\mathbf{x}) = \frac{a_1}{A}\log x_1 + \frac{a_2}{A}\log x_2 + \dots \frac{a_n}{A}\log x_n$$

so that

$$\lim_{\rho\to 0} f(\rho,\mathbf{x}) = x_1^{\frac{a_1}{A}} x_2^{\frac{a_1}{A}} \dots x_n^{\frac{a_1}{A}}$$

which is homogeneous of degree one.

4.49 Average cost is given by $c(y)/y$ which is undefined at $y = 0$. We seek $\lim_{y\to 0} c(y)/y$. By L'Hôpital's rule

$$\begin{aligned}\lim_{y\to 0}\frac{c(y)}{y} &= \lim_{y\to 0}\frac{c'(y)}{1}\\ &= c'(0)\end{aligned}$$

which is marginal cost at zero output.

4.50 1. Since $\lim_{x\to\infty} f'(x)/g'(x) = \mathrm{k}$, for every $\epsilon > 0$ there exists a such that

$$\left|\frac{f'(\bar{x})}{g'(\bar{x})} - k\right| < \epsilon/2 \text{ for every } \ \bar{x} > a \tag{4.8}$$

For every $x > a$, there exists (Exercise 4.46) $\bar{x} \in (a, x)$ such that

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(\bar{x})}{g'(\bar{x})}$$

and therefore by (4.8)

$$\left|\frac{f(x) - f(a)}{g(x) - g(a)} - k\right| < \epsilon/2 \text{ for every } \ x > a$$

2.

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{f(x) - f(a)}{g(x) - g(a)} \times \frac{f(x)}{f(x) - f(a)} \times \frac{g(x) - g(a)}{g(x)}\\ &= \frac{f(x) - f(a)}{g(x) - g(a)} \times \frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}}\end{aligned}$$

For fixed a

$$\lim_{x\to\infty}\frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} = 1$$

and therefore there exists a_2 such that

$$\frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} < 2 \text{ for every } \ x > a_2$$

which implies that

$$\left|\frac{f(x)}{g(x)} - k\right| < \frac{\epsilon}{2} \times 2 \text{ for every } \ x > a = \max\{a_1, a_2\}$$

4.51 We know that the result holds for $n = 1$ (Exercise 4.22). Assume that the result holds for $n - 1$. By the chain rule

$$D(g \circ f)[\mathbf{x}] = Dg[f(\mathbf{x})] \circ Df[\mathbf{x}]$$

If $f, g \in C^n$, the $Df, Dg \in C^{n-1}$ and therefore (by assumption) $D(g \circ f) \in C^{n-1}$, which implies that $g \circ f \in C^n$.

4.52 The partial derivatives of the quadratic function are

$$D_1 f = 2ax_1 + 2bx_2$$
$$D_2 f = 2bx_1 + 2cx_2$$

The second-order partial derivatives are

$$D_{11} f = 2a \qquad D_{21} f = 2b$$
$$D_{12} f = 2b \qquad D_{22} f = 2c$$

4.53 Apply Exercise 4.37 to each partial derivative $D_i f[\mathbf{x}]$.

4.54

$$H(\mathbf{x}^0) = \begin{pmatrix} D_{11} ff[\mathbf{x}^0] & D_{12} ff[\mathbf{x}^0] \\ D_{21} ff[\mathbf{x}^0] & D_{22} ff[\mathbf{x}^0] \end{pmatrix} = 2 \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

4.55

4.56 For any $x_1 \in S$, define $g : S \to \Re$ by

$$g(t) = f(t) + f'[t](x_1 - t) + a_2(x_1 - t)^2$$

g is differentiable on S with

$$p'(t) = f'[t] - f'[t] + f''[t](x_1 - t) - 2a_2(x_1 - t) = f''[t](x_1 - t) - 2a_2(x_1 - t)$$

Note that $g(x_1) = f(x_1)$ and

$$g(x_0) = f(x_0) + f'(x_0)(x_1 - x_0) + a_2(x_1 - x_0)^2 \tag{4.9}$$

is a quadratic approximation for f near x_0. If we require that this be exact at $x_1 \neq x_0$, then $g(x_0) = f(x_1) = g(x_1)$. By the mean value theorem (Theorem 4.1), there exists some $\bar{x}$ between x_0 and x_1 such that

$$g(x_1) - g(x_0) = p'(\bar{x})(x_1 - x_0) = f''(\bar{x})(x_1 - x_0) - 2a_2(x_1 - t) = 0$$

which implies that

$$a_2 = \frac{1}{2} f''(\bar{x})$$

Setting $x = x_1 - x_0$ in (4.9) gives the required result.

4.57 For any $x_1 \in S$, define $g : S \to \Re$ by

$$\begin{aligned} g(t) = f(t) + f'[t](x_1 - t) + \frac{1}{2} f''[t](x_1 - t)^2 + \frac{1}{3!} f^{(3)}[t](x_1 - t)^3 + \dots \\ + \frac{1}{n!} f^{(n)}[t](x_1 - t)^n + a_{n+1}(x_1 - t)^{n+1} \end{aligned}$$

g is differentiable on S with

$$\begin{aligned} g'(t) = f'[t] - f'[t] + f''[t](x_1 - t) - f''[t](x_1 - t) + \frac{1}{2} f^{(3)}[t](x_1 - t)^2 - \frac{1}{2} f^{(3)}[t](x_1 - x_0)^2 + \dots \\ + \frac{1}{(n-1)!} f^{(n)}[t](x_1 - t)^{n-1} + \frac{1}{n!} f^{(n+1)}[t](x_1 - t)^n - (n+1)a_{n+1}(x_1 - t)^n \end{aligned}$$

All but the last two terms cancel, so that

$$g'(t) = \frac{1}{n!} f^{(n+1)}[t](x_1 - t)^n - (n+1)a_{n+1}(x_1 - t)^n = \left(\frac{1}{n!} f^{(n+1)}[t] - (n+1)a_{n+1} \right) (x_1 - t)^n$$

Note that $g(x_1) = f(x_1)$ and

$$g(x_0) = f(x_0) + f'[x_0](x_1 - x_0) + \frac{1}{2}f''[x_0](x_1 - x_0)^2 + \frac{1}{3!}f^{(3)}[x_0](x_1 - x_0)^3 + \dots \\ + \frac{1}{n!}f^{(n+1)}[x_0](x_1 - x_0)^n + a_{n+1}(x_1 - x_0)^{n+1} \quad (4.10)$$

is a polynomial approximation for f near x_0. If we require that a_{n+1} be such that $g(x_0) = f(x_1) = g(x_1)$, there exists (Theorem 4.1) some $\bar{x}$ between x_0 and x_1 such that

$$g(x_1) - g(x_0) = g'(\bar{x})(x_1 - x_0) = 0$$

which for $x_1 \neq x_0$ implies that

$$g'(\bar{x}) = \frac{1}{n!}f^{n+1}[\bar{x}] - (n+1)a_{n+1} = 0$$

or

$$a_{n+1} = \frac{1}{(n+1)!}f^{n+1}[\bar{x}]$$

Setting $x = x_1 - x_0$ in (4.10) gives the required result.

4.58 By Taylor's theorem (Exercise 4.57), for every $x \in S - x_0$, there exists $\bar{x}$ between 0 and x such that

$$f(x_0 + x) = f(x_0) + f'[x_0]x + \frac{1}{2}f''[x_0]x^2 + \epsilon(x)$$

where

$$\epsilon(x) = \frac{1}{3!}f^{(3)}[\bar{x}]x^3$$

and

$$\frac{\epsilon(x)}{x^2} = \frac{1}{3!}f^{(3)}[\bar{x}](x)$$

Since $f \in C^3$, $f^{(3)}[\bar{x}]$ is bounded on $[0, x]$ and therefore

$$\lim_{x \to 0} |\frac{e(x)}{x^2}| = \lim_{x \to 0} \frac{1}{3!}|f^{(3)}[\bar{x}](x)| = 0$$

4.59 The function $g\colon \Re \to S$ defined by

$$g(t) = t\mathbf{x}_0 + (1 - t)\mathbf{x}$$

g is C^∞ with $Dg[t] = \mathbf{x}$ and $D^k g(t) = 0$ for $k = 2, 3, \dots$. By Exercise 4.51, the composite function $h = f \circ g$ is C^{n+1}. By the Chain rule

$$h'(t) = Df[g(t)] \circ Dg[t] = Df[g(t)](\mathbf{x})$$

Similarly

$$\begin{aligned} h''(t) &= D\Big(Df[g(t)](\mathbf{x})\Big) \\ &= D^2 f[g(t)] \circ Dg[t](\mathbf{x} - \mathbf{x}_0) \\ &= D^2 f[g(t)](\mathbf{x})^{(2)} \end{aligned}$$

and for all $1 \leq k \leq n+1$

$$\begin{aligned} h^{(k)}(t) &= D\Big(D^{(k-1)} f[g(t)](\mathbf{x})^{(k-1)}\Big) \\ &= D^k f[g(t)] \circ Dg[t](\mathbf{x} - \mathbf{x}_0)^{(k-1)} \\ &= D^k f[g(t)](\mathbf{x})^{(k)} \end{aligned}$$

4.60 From Exercise 4.54, the Hessian of f is

$$H(\mathbf{x}) = 2 \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and the gradient of f is

$$\nabla f(\mathbf{x}) = (2ax_1, 2cx_2) \text{ with } \nabla f\big((0,0)\big) = 0$$

so that the second order Taylor series at $(0,0)$ is

$$\begin{aligned} f(\mathbf{x}) &= f(0,0) + \nabla f(0,0)\mathbf{x} + \frac{1}{2} 2\mathbf{x}^T \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{x} \\ &= ax_1^2 + 2bx_1x_2 + cx_2^2 \end{aligned}$$

Not surprisingly, we conclude that the best quadratic approximation of a quadratic function is the function itself.

4.61 1. Since $Df[\mathbf{x}_0]$ is continuous and one-to-one (Exercise 3.36), there exists a constant m such that

$$m \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \tag{4.11}$$

Let $\epsilon = m/2$. By Exercise 4.43, there exists a neighborhood S such that

$$\|Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2) - (f(\mathbf{x}_1) - f(\mathbf{x}_2))\| = \|f(\mathbf{x}_1) - f(\mathbf{x}_2) - Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| \leq \epsilon \|\mathbf{x}_1 - \mathbf{x}_2\|$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in S$. The Triangle Inequality (Exercise 1.200) implies

$$\|Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_2)\| - \|(f(\mathbf{x}_1) - f(\mathbf{x}_2))\| \leq \epsilon \|\mathbf{x}_1 - \mathbf{x}_2\|$$

Substituting (4.11)

$$2\epsilon \|\mathbf{x}_1 - \mathbf{x}_2\| - \|(f(\mathbf{x}_1) - f(\mathbf{x}_2))\| \leq \epsilon \|\mathbf{x}_1 - \mathbf{x}_2\|$$

That is

$$\epsilon \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|(f(\mathbf{x}_1) - f(\mathbf{x}_2))\| \tag{4.12}$$

and therefore

$$f(\mathbf{x}_1) = f(\mathbf{x}_2) \implies \mathbf{x}_1 = \mathbf{x}_2$$

2. Let $T = f(S)$. Since the restriction of f to S is one-to-one and onto, and therefore there exists an inverse $f^{-1}\colon T \to S$. For any $\mathbf{y}_1, \mathbf{y}_2 \in T$, let $\mathbf{x}_1 = f^{-1}(\mathbf{y}_1)$ and $\mathbf{x}_2 = f^{-1}(\mathbf{y}_2)$. Substituting in (4.12)

$$\epsilon \left\|f^{-1}(\mathbf{y}_1) - f^{-1}(\mathbf{y}_2)\right\| \leq \|\mathbf{y}_1 - \mathbf{y}_2\|$$

so that

$$\left\|f^{-1}(\mathbf{y}_1) - f^{-1}(\mathbf{y}_2)\right\| \leq \frac{1}{\epsilon} \|\mathbf{y}_1 - \mathbf{y}_2\|$$

f^{-1} is continuous.

3. Since S is open, $T = f^{-1}(S)$ is open. Therefore, $T = f(S)$ is a neighborhood of $f(\mathbf{x}_0)$. Therefore, f is locally onto.

4.62 Assume to the contrary that there exists $\mathbf{x}_0 \neq \mathbf{x}_1 \in S$ with $f(\mathbf{x}_0) = f(\mathbf{x}_1)$. Let $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_0$. Define $g\colon [0,1] \to S$ by $g(t) = (1-t)\mathbf{x}_0 + t\mathbf{x}_1 = \mathbf{x}_0 + t\mathbf{x}$. Then

$$g(0) = \mathbf{x}_0 \qquad g(1) = \mathbf{x}_1 \qquad g'(t) = \mathbf{x}$$

Define

$$h(t) = \mathbf{x}^T \Big(f\big(g(t)\big) - f(\mathbf{x}_0) \Big)$$

Then

$$h(0) = 0 = h(1)$$

By the mean value theorem (Mean value theorem), there exists $0 < \alpha < 1$ such that $g(\alpha) \in S$ and

$$h'(\alpha) = \mathbf{x}^T Df[g(\alpha)]\mathbf{x} = \mathbf{x}^T J_f(g(\alpha))\mathbf{x} = 0$$

which contradicts the definiteness of J_f.

4.63 Substituting the linear functions in (4.35) and (4.35), the IS-LM model can be expressed as

$$\begin{aligned}(1-C_y)y - I_r r &= C_0 + I_0 + G - C_y T \\ L_y y + L_r r &= M/P\end{aligned}$$

which can be rewritten in matrix form as

$$\begin{pmatrix} 1-C_y & I_r \\ L_y & L_r \end{pmatrix} \begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} Z - C_y T \\ M/P \end{pmatrix}$$

where $Z = C_0 + I_0 + G$. Provided the system is nonsingular, that is

$$D = \begin{vmatrix} 1-C_y & I_r \\ L_y & L_r \end{vmatrix} \neq 0$$

the system can be solved using Cramer's rule (Exercise 3.103) to yield

$$\begin{aligned} r &= \frac{(1-C_y)M/P - L_y(Z - C_y T)}{D} \\ y &= \frac{L_r(Z - C_y)T - I_r M/P}{D}\end{aligned}$$

4.64 The kernel

$$\text{kernel } DF[(\mathbf{x}_0, \boldsymbol{\theta}_0)] = \{\ (\mathbf{x}, \boldsymbol{\theta}) : DF[(\mathbf{x}_0, \boldsymbol{\theta}_0)](\mathbf{x}, \boldsymbol{\theta}) = \mathbf{0}\ \}$$

is the set of solutions to the equation

$$DF[\mathbf{x}_0, \boldsymbol{\theta}_0] \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} D_{\mathbf{x}} f(\mathbf{x}_0, \boldsymbol{\theta}_0)\mathbf{x} + D_{\boldsymbol{\theta}} f(\mathbf{x}_0, \boldsymbol{\theta}_0)\boldsymbol{\theta} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

Only $\boldsymbol{\theta} = \mathbf{0}$ satisfies this equation. Substituting $\boldsymbol{\theta} = \mathbf{0}$, the equation reduces to

$$D_{\mathbf{x}} f(\mathbf{x}_0, \boldsymbol{\theta}_0)\mathbf{x} = \mathbf{0}$$

which has a unique solution $\mathbf{x} = \mathbf{0}$ since $D_{\mathbf{x}} f[\mathbf{x}_0, \boldsymbol{\theta}_0]$ is nonsingular. Therefore the kernel of $DF[\mathbf{x}_0, \boldsymbol{\theta}_0]$ consists of the single point $(\mathbf{0}, \mathbf{0})$ which implies that $DF[\mathbf{x}_0, \boldsymbol{\theta}_0]$ is nonsingular (Exercise 3.19).

4.65 The IS curve is horizontal if its slope is zero, that is

$$D_y g = -\frac{1 - D_y C}{-D_r I}$$

This requires either

1. unit marginal propensity to consume ($D_y C = 1$)
2. infinite interest elasticity of investment ($D_r I = \infty$)

4.66 The LM curve $r = h(y)$ is implicitly defined by the equation

$$f(r, y; G, T, M) = L(y, r) - M/P = 0$$

the slope of which is given by

$$\begin{aligned} D_y h &= -\frac{D_y f}{D_r f} \\ &= -\frac{D_y L}{D_r L} \end{aligned}$$

Economic considerations dictate that the numerator ($D_y f$) is positive while the denominator ($D_r L$) is negative. Preceded by a negative sign, the slope of the LM curve is positive. The LM curve would be vertical (infinite slope) if the interest elasticity of the demand for money was zero ($D_r L = 0$).

4.67 Suppose f is convex. For any $\mathbf{x}, \mathbf{x}_0 \in S$ let

$$h(t) = f\Big(t\mathbf{x} + (1-t)\mathbf{x}_0\Big) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{x}_0)$$

for $0 < t < 1$. Subtracting $h(0) = f(\mathbf{x}_0)$

$$h(t) - h(0) \leq tf(\mathbf{x}) - tf(\mathbf{x}_0)$$

and therefore

$$f(\mathbf{x}) - f(\mathbf{x}_0) \geq \frac{h(t) - h(0)}{t}$$

Using Exercise 4.10

$$f(\mathbf{x}) - f(\mathbf{x}_0) \geq \lim_{t \to 0} \frac{h(t) - h(0)}{t} = \vec{D}_{\mathbf{x}} f[\mathbf{x}_0] = Df[\mathbf{x}_0](\mathbf{x} - \mathbf{x}_0)$$

Conversely, let $\mathbf{x}_0 = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ for any $\mathbf{x}_1, \mathbf{x}_2 \in S$. If f satisfies (4.29) on S, then

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\mathbf{x}_0) + Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_0) \\ f(\mathbf{x}_2) &\geq f(\mathbf{x}_0) + Df[\mathbf{x}_0](\mathbf{x}_2 - \mathbf{x}_0) \end{aligned}$$

and therefore for any $0 \leq \alpha \leq 1$

$$\begin{aligned} \alpha f(\mathbf{x}_1) &\geq \alpha f(\mathbf{x}_0 + \alpha Df[\mathbf{x}_0](\mathbf{x}_1 - \mathbf{x}_0) \\ (1-\alpha) f(\mathbf{x}_2) &\geq (1-\alpha) f(\mathbf{x}_0 + (1-\alpha) Df[\mathbf{x}_0](\mathbf{x}_2 - \mathbf{x}_0) \end{aligned}$$

Adding and using the linearity of Df (Exercise 4.21)

$$\begin{aligned} \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) &\geq f(\mathbf{x}_0) + Df[\mathbf{x}_0](\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 - \mathbf{x}_0) \qquad (4.13) \\ &= f(\mathbf{x}_0) = f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \end{aligned}$$

That is, f is convex. If (4.29) is strict, so is (4.13).

4.68 Since h is convex, it has a subgradient $g \in X^*$ (Exercise 3.181) such that

$$h(\mathbf{x}) \geq h(\mathbf{x}_0) + g(\mathbf{x} - \mathbf{x}_0) \text{ for every } \mathbf{x} \in X$$

(4.31) implies that g is also a subgradient of f on S

$$f(\mathbf{x}) \geq f(x_0) + g(\mathbf{x} - \mathbf{x}_0) \text{ for every } \mathbf{x} \in S$$

Since f is differentiable, this implies that g is unique (Remark 4.14) and equal to the derivative of f. Hence h is differentiable at $\mathbf{x}_0$ with $Dh[\mathbf{x}_0] = Df[\mathbf{x}_0]$.

4.69 Assume f is convex. For every $\mathbf{x}, \mathbf{x}_0 \in S$, Exercise 4.67 implies

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) \\ f(\mathbf{x}_0) &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{x}_0 - \mathbf{x}) \end{aligned}$$

Adding

$$f(\mathbf{x}) + f(\mathbf{x}_0) \geq f(\mathbf{x}) + f(\mathbf{x}_0) + \nabla f(\mathbf{x})^T(\mathbf{x}_0 - \mathbf{x}) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0)$$

or

$$\nabla f(\mathbf{x})^T(\mathbf{x} - \mathbf{x}_0) \geq \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0)$$

and therefore

$$\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}_0)^T \mathbf{x} - \mathbf{x}_0 \geq 0$$

When f is strictly convex, the inequalities are strict.

Conversely, assume (4.32). By the mean value theorem (Theorem 4.1), there exists $\bar{\mathbf{x}} \in (\mathbf{x}, \mathbf{x}_0)$ such that

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \nabla f(\bar{\mathbf{x}})^T \mathbf{x} - \mathbf{x}_0$$

By assumption

$$\nabla f(\bar{\mathbf{x}}) - \nabla f(\mathbf{x}_0)^T \bar{\mathbf{x}} - \mathbf{x}_0 \geq 0$$

But

$$\bar{\mathbf{x}} - \mathbf{x}_0 = \alpha \mathbf{x}_0 + (1-\alpha)\mathbf{x} - \mathbf{x}_0 = (1-\alpha)(\mathbf{x} - \mathbf{x}_0)$$

and therefore

$$(1-\alpha)\nabla f(\bar{\mathbf{x}}) - \nabla f(\mathbf{x}_0)^T \mathbf{x} - \mathbf{x}_0 \geq 0$$

so that

$$\nabla f(\bar{\mathbf{x}})^T \mathbf{x} - \mathbf{x}_0 \geq \nabla f(\mathbf{x}_0)^T \mathbf{x} - \mathbf{x}_0 \geq 0$$

and therefore

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \nabla f(\bar{\mathbf{x}})^T \mathbf{x} - \mathbf{x}_0 \geq \nabla f(\mathbf{x}_0)^T \mathbf{x} - \mathbf{x}_0$$

Therefore f is convex by Exercise 4.67.

4.70 For $S \subseteq \Re$, $\nabla f(x) = f'(x)$ and (4.32) becomes

$$(f'(x_2) - f'(x_1)(x_2 - x_1) \geq 0$$

for every $x_1, x_2 \in S$. This is equivalent to

$$f'(x_2)(x_2 - x_1) \geq f'(x_1)(x_2 - x_1)0$$

or

$$x_2 > x_1 \implies f'(x_2) \geq f'(x_1)$$

f is strictly convex if and only if the inequalities are strict.

4.71 f' is increasing if and only if $f'' = Df' \geq 0$ (Exercise 4.35). f' is strictly increasing if $f'' = Df' > 0$ (Exercise 4.36).

4.72 Adapting the previous example

$$f''(x) = n(n-1)x^n - 2 = \begin{cases} = 0 & \text{if } n = 1 \\ \geq 0 & \text{if } n = 2, 4, 6, dots \\ \text{indeterminate} & \text{otherwise} \end{cases}$$

Therefore, the power function is convex if n is even, and neither convex if $n \geq 3$ is odd. It is both convex and concave when $n = 1$.

4.73 Assume f is quasiconcave, and $f(\mathbf{x}) \geq f(\mathbf{x}_0)$. Differentiability at $\mathbf{x}_0$ implies for all $0 < t < 1$

$$f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)t(\mathbf{x} - \mathbf{x}_0) + \eta(t) \left\| t(\mathbf{x} - \mathbf{x}_0) \right\|$$

where $\eta(t) \to \mathbf{0}$ and $t \to 0$. Quasiconcavity implies

$$f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0) \geq f(\mathbf{x}_0)$$

and therefore

$$\nabla f(\mathbf{x}_0)t(\mathbf{x} - \mathbf{x}_0) + \eta(t) \left\| t(\mathbf{x} - \mathbf{x}_0) \right\| \geq 0$$

Dividing by t and letting $t \to 0$, we get

$$\nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geq 0$$

Conversely, assume f is a differentiable functional satisfying (4.36). For any $\mathbf{x}_1, \mathbf{x}_2 \in S$ with $f(\mathbf{x}_1) \geq f(\mathbf{x}_2$ for every $\mathbf{x}, \mathbf{x}_0 \in S)$, define $h\colon [0,1] \to \Re$ by

$$h(t) = f\big((1-t)\mathbf{x}_1 + t\mathbf{x}_2\big) = f\big(\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)\big)$$

We need to show that $h(t) \geq h(1)$ for every $t \in (0,1)$. Suppose to the contrary that $h(t_1) < h(1)$. Then (see below) there exists t_0 with $h(t_0) < h(1)$ and $h'(t_0) < 0$. By the Chain Rule, this implies

$$h'(t_0) = \nabla f(\mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_1) < 0$$

critical where $\mathbf{x}_0 = \mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)$. Since $\mathbf{x}_2 - \mathbf{x}_0 = (1-t)(\mathbf{x}_2 - \mathbf{x}_1)$ this implies that

$$h'(t_0) = \frac{1}{1-t} \nabla f(\mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_0) \tag{4.14}$$

On the other hand, since $f(\mathbf{x}_0) \geq f(\mathbf{x}_2)$, (4.36) implies

$$\nabla f(\mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_0) \geq 0$$

contradicting (4.14).

To show that there exists t_0 with $h(t_0) < h(1)$ and $h'(t_0) < 0$: Since f is continuous, there exists an open interval (a, b) with $a < t_1 < b$ with $h(a) = h(b) = h(1)$ and $h(t) < h(1)$ for every $t \in (a, b)$. By the Mean Value Theorem, there exist $t_0 \in (a, t_1)$ such that

$$0 < h(t_1) - h(a) = h'(t_0)(t_1 - a)$$

which implies that $h'(t_0) > 0$.

4.74 Suppose to the contrary that

$$f(\mathbf{x}) > f(\mathbf{x}_0) \text{ and } \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \leq 0$$

critical Let $\mathbf{x}_1 = -\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. For every $t \in \Re_+$

$$\begin{aligned}\nabla f(\mathbf{x}_0)(\mathbf{x} + t\mathbf{x}_1 - \mathbf{x}_0) &= \nabla f(\mathbf{x}_0)t\mathbf{x}_1 + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\\ &\leq t\nabla f(\mathbf{x}_0)\mathbf{x}_1\\ &= -t\left\|\nabla f(\mathbf{x}_0)\right\|^2 < 0\end{aligned}$$

Since f is continuous, there exists $t > 0$ such that

$$f(\mathbf{x} + t\mathbf{x}_1) > f(\mathbf{x}_0) \text{ and } \nabla f(\mathbf{x}_0)(\mathbf{x} + t\mathbf{x}_1 - \mathbf{x}_0) < 0$$

contradicting the quasiconcavity of f (4.36).

4.75 Suppose

$$f(\mathbf{x}) < f(\mathbf{x}_0) \implies \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) < 0$$

This implies that

$$-f(\mathbf{x}) > -f(\mathbf{x}_0) \implies \nabla - f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) > 0$$

and $-f$ is pseudoconcave.

4.76

1. If $f \in F[S]$ is concave (and differentiable)

$$f(\mathbf{x}) \leq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0)$$

for every $\mathbf{x}, \mathbf{x}_0 \in S$(equation 4.30). Therefore

$$f(\mathbf{x}) > f(\mathbf{x}_0) \implies \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) > 0$$

f is pseudoconcave.

2. Assume to the contrary that f is pseudoconcave but not quasiconcave. Then, there exists $\bar{\mathbf{x}} = \alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$, $\mathbf{x}_1, \mathbf{x}_2 \in S$ such that

$$f(\bar{\mathbf{x}}) < \min\{f(\mathbf{x}_1), f(\mathbf{x}_2)\} \tag{4.15}$$

Assume without loss of generality that

$$f(\bar{\mathbf{x}}) < f(\mathbf{x}_1) \leq f(\mathbf{x}_2)$$

Pseudoconcavity (4.38) implies

$$\nabla f(\bar{\mathbf{x}})(\mathbf{x}_2 - \bar{\mathbf{x}}) > 0 \tag{4.16}$$

Since $\mathbf{x}_1 = (\bar{\mathbf{x}} - (1-\alpha)\mathbf{x}_2)/\alpha$

$$\mathbf{x}_1 - \bar{\mathbf{x}} = \frac{1}{\alpha}\big(\bar{\mathbf{x}} - (1-\alpha)\mathbf{x}_2 - \alpha\bar{\mathbf{x}}\big) = -\frac{1-\alpha}{\alpha}(\mathbf{x}_2 - \bar{\mathbf{x}})$$

Substituting in (4.16) gives

$$\nabla f(\bar{\mathbf{x}})(\mathbf{x}_1 - \bar{\mathbf{x}}) < 0$$

which by pseudoconcavity implies $f(\mathbf{x}_1) \leq f(\bar{\mathbf{x}})$ contradicting our assumption (4.15) .

3. Exercise 4.74.

4.77 The CES function is quasiconcave provided $\rho \leq 1$ (Exercise 3.58). Since $D_{x_i} f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \Re^n_+ +$, the CES function with $\rho \leq 1$ is pseudoconcave on $\Re^n_{++}$.

4.78 Assume that $f\colon S \to \Re$ is homogeneous of degree k, so that for every $\mathbf{x} \in S$

$$f(t\mathbf{x}) = t^n f(\mathbf{x}) \text{ for every } \ t > 0$$

Differentiating both sides of this identity with respect to x_i

$$D_{x_i} f(t\mathbf{x}) t = t^n D_{x_i} f(\mathbf{x})$$

and dividing by $t > 0$

$$D_{x_i} f(t\mathbf{x}) = t^{k-1} D_{x_i} f(\mathbf{x})$$

4.79 If f is homogeneous of degree k

$$\begin{aligned}
\vec{D}_{\mathbf{x}} f(\mathbf{x}) &= \lim_{t\to 0} \frac{f(\mathbf{x} + t\mathbf{x}) - f(\mathbf{x})}{t} \\
&= \lim_{t\to 0} \frac{f((1+t)\mathbf{x}) - f(\mathbf{x})}{t} \\
&= \lim_{t\to 0} \frac{(1+t)^n f(\mathbf{x}) - f(\mathbf{x})}{t} \\
&= \lim_{t\to 0} \frac{(1+t)^n - 1}{t} f(\mathbf{x})
\end{aligned}$$

Applying L'Hôpital's Rule (Exercise 4.47)

$$\lim_{t\to 0} \frac{(1+t)^{k-1}}{t} f(\mathbf{x}) = \lim_{t\to 0} \frac{k(1+t)^{k-1}}{1} = k$$

and therefore

$$\vec{D}_{\mathbf{x}} f(\mathbf{x}) = k f(\mathbf{x}) \tag{4.17}$$

4.80 For fixed $\mathbf{x}$, define

$$h(t) = f(t\mathbf{x})$$

By the Chain Rule

$$h'(t) = tDf[t\mathbf{x}](\mathbf{x}) = tkf(t\mathbf{x}) = tkh(t) \tag{4.18}$$

using(4.40). Differentiating the product $h(t)t^{-k}$

$$D_t\left(h(t)t^{-k}\right) = -kh(t)t^{-k-1} + t^{-k}h'(t) = t^{-k}\left(h'(t) - kth(t)\right) = 0$$

from (4.18). Since this holds for every t, $h(t)t^{-k}$ must be constant (Exercise 4.38), that is there exists $c \in \Re$ such that

$$h(t)t^{-k} = c \implies h(t) = ct^k$$

Evaluating at $t = 1$, $h(1) = c$ and therefore

$$h(t) = t^k h(1)$$

Since $h(t) = f(t\mathbf{x})$ and $h(1) = f(\mathbf{x})$, this implies

$$f(t\mathbf{x}) = t^k f(\mathbf{x}) \text{ for every } \mathbf{x} \text{ and } t > 0$$

f is homogeneous of degree k.

4.81 If f is linearly homogeneous and quasiconcave, then f is concave (Proposition 3.12). Therefore, its Hessian is nonpositive definite (Proposition 4.1). and its diagonal elements $D^2_{x_ix_i}f(\mathbf{x})$ are nonpositive (Exercise 3.95). By Wicksell's law, $D^2_{x_ix_j}f(\mathbf{x})$ is nonnegative.

4.82 Assume f is homogeneous of degree k, that is

$$f(t\mathbf{x}) = t^k f(\mathbf{x}) \text{ for every } \mathbf{x} \in S \text{ and } t > 0$$

By Euler's theorem

$$D_t f[t\mathbf{x}](t\mathbf{x}) = kf(t\mathbf{x})$$

and therefore the elasticity of scale is

$$E(\mathbf{x}) = \frac{t}{f(t\mathbf{x})} D_t f(t\mathbf{x})\bigg|_{t=1} = \frac{t}{f(t\mathbf{x})} kf(t\mathbf{x}) = k$$

Conversely, assume that

$$E(\mathbf{x}) = \frac{t}{f(t\mathbf{x})} D_t f(t\mathbf{x})\bigg|_{t=1} = k$$

that is

$$D_t f(t\mathbf{x}) = kf(t\mathbf{x})$$

By Euler's theorem, f is homogeneous of degree k.

4.83 Assume $f \in F(S)$ is differentiable and homogeneous of degree $k \neq 0$. By Euler's theorem

$$Df[\mathbf{x}](\mathbf{x}) = kf(\mathbf{x}) \neq 0$$

for every $\mathbf{x} \in S$ such that $f(\mathbf{x}) \neq 0$.

4.84 f satisfies Euler's theorem

$$kf(\mathbf{x}) = \sum_{i=1}^n D_i f(\mathbf{x}) x_i$$

Differentiating with respect to x_j

$$kD_j f(\mathbf{x}) = \sum_{i=1}^{n} D_{ij} f(\mathbf{x}) x_i + D_j f(\mathbf{x})$$

or

$$(k-1) D_j f(\mathbf{x}) = \sum_{i=1}^{n} D_{ij} f(\mathbf{x}) x_i \qquad j = 1, 2, \dots, n$$

Multiplying each equation by x_j and summing

$$(k-1) \sum_{j=1}^{n} D_j f(\mathbf{x}) x_j = \sum_{j=1}^{n} \sum_{i=1}^{n} D_{ij} f(\mathbf{x}) x_i x_j = \mathbf{x}' H \mathbf{x}$$

By Euler's theorem, the left hand side is

$$(k-1) k f(\mathbf{x}) = \mathbf{x}' H \mathbf{x}$$

4.85 If f is homothetic, there exists strictly increasing g and linearly homogeneous h such that $f = g \circ h$ (Exercise 3.175). Using the Chain Rule and Exercise 4.78

$$D_{x_i} f(t\mathbf{x}) = g'(f(t\mathbf{x})) D_{x_i} h(t\mathbf{x}) = t g'(f(t\mathbf{x})) D_{x_i} h(\mathbf{x})$$

and therefore

$$\begin{aligned}
\frac{D_{x_i} f(t\mathbf{x})}{D_{x_j} f(t\mathbf{x})} &= \frac{t g'(f(t\mathbf{x})) D_{x_i} h(\mathbf{x})}{D_{x_j} t g'(f(t\mathbf{x})) D_{x_j} h(\mathbf{x})} \\
&= \frac{D_{x_i} h(\mathbf{x})}{D_{x_j}} D_{x_j} h(\mathbf{x}) \\
&= \frac{g'(f(\mathbf{x}) D_{x_i} h(\mathbf{x})}{D_{x_j} g'(f(\mathbf{x})) D_{x_j} h(\mathbf{x})} \\
&= \frac{D_{x_i} f(\mathbf{x})}{D_{x_j}} D_{x_j} f(\mathbf{x})
\end{aligned}$$