Chapter 4
Multivariate Calculus

4.1 Partial Derivatives

Definition Let $f\colon \boldsymbol{S} \to \boldsymbol{R}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a function. The ***partial derivative*** of *f* at **x** with respect to x_j is given by,

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = f_j(\mathbf{x})$$
$$= \lim_{h \to 0} \frac{f(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x_1, \ldots, x_j, \ldots, x_n)}{h}.$$

If $y = f(\mathbf{x})$, we can also write $\dfrac{\partial f(\mathbf{x})}{\partial x_j} = \dfrac{\partial y}{\partial x_j}$. This means the partial derivative is just derivative with all other variables being fixed. All the formulae of derivatives in Chapter 2 will thus apply in partial derivatives with all other variables treated as constants.

4.2 Second-Order Partial Derivatives and Cross Partial Derivatives

Definition Let $f\colon \boldsymbol{S} \to \boldsymbol{R}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a function. The ***second-order partial derivative of f with respect to*** x_j, $j = 1, 2, \ldots, n$, is given by

$$f_{jj}(\mathbf{x}) = \frac{\partial}{\partial x_j}\left(\frac{\partial f(\mathbf{x})}{\partial x_j}\right)$$
$$= \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j^2},$$

and the ***cross partial derivatives with respect to*** x_j ***and*** x_i is given by

$$f_{ij}(\mathbf{x}) = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\mathbf{x})}{\partial x_j}\right) = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}.$$

Note that when f is a function of just two variables, it might be easier to write $z = f(x,y)$, and thus the first -order, second-order and cross partial derivatives can be written respectively as $f_x(x,y)$, $f_{xx}(x,y)$ and $f_{xy}(x,y)$.

HW Baldani, p. 136, #5.1.

Young's Theorem

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right) = \frac{\partial}{\partial x_i}\left(\frac{\partial f(\mathbf{x})}{\partial x_j}\right).$$

4.3 Total Differentials

The change in y, $y = f(\mathbf{x})$, as caused by a change in x_j , Δx_j , while all other variables are held constant, can be approximated by,

$$\begin{aligned}\Delta y &= f\left(x_1,\ldots,x_{j-1},x_j+\Delta x_j,x_{j+1},\ldots,x_n\right) - f\left(x_1,\ldots,x_j,\ldots,x_n\right) \\ &\approx \frac{\partial f(\mathbf{x})}{\partial x_j}\Delta x_j.\end{aligned}$$

When all variables change at the same time by the quantities Δx_j , $j = 1,2,\ldots,n$, the total effect in the change of the variable y will be approximately just the sum of all changes caused by each $\Delta x_j, j = 1,2,\ldots,n$. That is,

$$\begin{aligned}\Delta y &\approx \frac{\partial f(\mathbf{x})}{\partial x_1}\Delta x_1 + \frac{\partial f(\mathbf{x})}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial f(\mathbf{x})}{\partial x_n}\Delta x_n \\ &= \sum_{j=1}^{n}\frac{\partial f(\mathbf{x})}{\partial x_j}\Delta x_j = \sum_{j=1}^{n} f_j(\mathbf{x})\Delta x_j.\end{aligned}$$

The total differential is thus given by dy as obtained by letting Δx_j approach zero, for $j = 1,2,\ldots,n$.

Definition Let $f : \boldsymbol{S} \to \boldsymbol{R}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a function, and $y = f(\mathbf{x})$. The ***total differential*** of y at **x**, denoted by dy, is given by,

$$dy = \sum_{j=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_j} dx_j = \sum_{j=1}^{n} f_j(\mathbf{x}) dx_j ,$$

where dx_i is just Δx_i as it is approaching zero, $i = 1,2,\ldots,n$.

Example The total differential of $y = f(x_1, x_2) = x_1^2 x_2$ is

$$dy = 2x_1x_2dx_1 + x_1^2dx_2 .$$

When there are changes $\Delta x_1 = 0.2$ and $\Delta x_2 = -0.3$, the change Δy, where $x_1 = 10$ and $x_2 = 20$, is approximately

$$\begin{aligned} \Delta y &\approx 2x_1x_2\Delta x_1 + x_1^2\Delta x_2 \\ &= 400(0.2) + 100(-0.3) = 50. \end{aligned}$$

Example For the production function,

$$Q = 0.07L^2 + LK + 54K ,$$

the total differential of Q is

$$\begin{aligned} dQ &= \frac{\partial Q}{\partial L} dL + \frac{\partial Q}{\partial K} dK \\ &= (0.14L + K)dL + (L + 54)dK. \end{aligned}$$

With the current use of $L = 100$ and $K = 50$, the output Q is changed, when $\Delta L = -4$ and $\Delta K = 2$, by approximately

$$\Delta Q \approx (0.14L + K)\Delta L + (L + 54)\Delta K$$
$$= (64)(-4) + (154)(2) = 52.$$

HW Baldani, p. 136, #5.2, 5.3.

HW Compute the total differential of the following functions and estimate the its changes when there are changes in *x* and *y* as specified.

a) $z = f(x,y) = x\sqrt{x+y}$, $\Delta x = 0.01$, $\Delta y = 0.02$
where $x = 3, y = 1$.

b) $z = f(x,y) = Ax^{0.3}y^{0.5}$, $\Delta x = -0.2$, $\Delta y = -0.4$
where $x = 1, y = 4$.

c) $z = f(x,y) = \dfrac{x^2 + 3xy + y^2}{\sqrt{x + y^2}}$, $\Delta x = 1$, $\Delta y = 0$
where $x = 9, y = 4$.

d) $z = f(x,y) = 5y\,\boldsymbol{ln}(x^2 + y)$, $\Delta x = 0.1$, $\Delta y = 0.1$
where $x = 0, y = e^4$

Example Let $f : \boldsymbol{R}_+^2 \to \boldsymbol{R}$ be a production function, given by

$$y = f(K,L).$$

The total change in the output when there are small changes in the labor and capital is given by,

$$dy = f_L(K,L)dL + f_K(K,L)dK ,$$

which is just the change in labor multiplied by its marginal product of labor, plus the change in capital multiplied by the marginal product of capital. On a given isoquant, the change in labor and capital is such that the final effect on the output is no change, i.e.,

$$dy = 0$$
$$= f_L(K,L)dL + f_K(K,L)dK.$$

Thus, we have the differential of *K*

$$dK = -\frac{f_L(K,L)}{f_K(K,L)}dL\ ,$$

and the derivative of K with respect to L, when the output is held constant, is given by

$$\frac{dK}{dL} = -\frac{f_L(K,L)}{f_K(K,L)}.$$

The slope of isoquant is thus the negative of the ratio of the marginal products.

4.4 Conventions of Matrix Notations for Derivatives of Functions of Serveral Variables

In dealing with partial derivatives of functions of several variables, it is convenient to adopt the matrix language. As we will later be computing the derivatives of functions that map points in $\boldsymbol{R}^n$ to points in $\boldsymbol{R}^m$, we will use the following conventions that will be useful when we apply chain rules to deal with the derivatives of composite functions.

a) If $f:\boldsymbol{R}^n \to \boldsymbol{R}$, then

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f(\mathbf{x})}{\partial x_1} \\ \dfrac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}.$$

Example The gradient of $f(\mathbf{x}) = \mathbf{c}^{\mathrm{T}}\mathbf{x}$ is $\nabla f(\mathbf{x}) = \mathbf{c}$.

Example Let $f : \boldsymbol{S} \to \boldsymbol{R}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a function, and $y = f(\mathbf{x})$. The total differential of y at $\mathbf{x}$ can be written as,

$$dy = \nabla f(\mathbf{x})^{\mathrm{T}} \mathbf{dx},$$

where $\mathbf{dx} = \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$.

• f is differentiable at $\mathbf{x}$ if $\nabla f(\mathbf{x})$ exists.

• f is differentiable if $\nabla f(\mathbf{x})$ exists at each $\mathbf{x} \in \boldsymbol{S}$.

HW Compute the gradient of $f(\mathbf{x}) = a\mathbf{x}^{\mathrm{T}}\mathbf{x}$, and $g(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix.

HW Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$ and $g : \boldsymbol{R}^n \to \boldsymbol{R}$. Show that

a) $\nabla(f(\mathbf{x}) + g(\mathbf{x})) = \nabla f(\mathbf{x}) + \nabla g(\mathbf{x})$,

b) $\nabla(f(\mathbf{x})g(\mathbf{x})) = f(\mathbf{x})\nabla g(\mathbf{x}) + g(\mathbf{x})\nabla f(\mathbf{x})$, and

c) $\nabla\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) = \frac{1}{g(\mathbf{x})^2}\left(g(\mathbf{x})\nabla f(\mathbf{x}) - f(\mathbf{x})\nabla g(\mathbf{x})\right)$, for $g(\mathbf{x}) \neq 0$.

b) If $\mathbf{x} : \boldsymbol{R} \to \boldsymbol{R}^n$, where $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$, and $x_i : \boldsymbol{R} \to \boldsymbol{R}$, then

$$\mathbf{x}'(t) = \frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}.$$

Example For a given point $\mathbf{x}_0$ and some direction **d**, define $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{d}$, $t \in \boldsymbol{R}$. The derivative of $\mathbf{x}(t)$ is just $\mathbf{x}'(t) = \mathbf{d}$.

c) If $\mathbf{f} : \boldsymbol{R}^n \to \boldsymbol{R}^m$, where $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f^1(\mathbf{x}) \\ f^2(\mathbf{x}) \\ \vdots \\ f^m(\mathbf{x}) \end{bmatrix}$ and

$f^i : \boldsymbol{R}^n \to \boldsymbol{R}$, $i = 1,2,\ldots,m$, then

$$\nabla \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \nabla f^1(\mathbf{x})^{\mathrm{T}} \\ \nabla f^2(\mathbf{x})^{\mathrm{T}} \\ \vdots \\ \nabla f^m(\mathbf{x})^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1(\mathbf{x})}{\partial x_1} & \frac{\partial f^1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f^1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f^2(\mathbf{x})}{\partial x_1} & \frac{\partial f^2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f^2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f^m(\mathbf{x})}{\partial x_1} & \frac{\partial f^m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f^m(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \boldsymbol{R}^{m\times n}$$

is called the ***Jacobian*** of **f**, though some may just say gradient of **f**.

Example The gradient of $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is $\nabla \mathbf{f}(\mathbf{x}) = \mathbf{A}$.

HW Baldani, p. 136, #5.5.

By this last convention, as the partial derivatives are also functions of **x,** we can take the gradient of the gradient of *f*. The resulting matrix is called the Hessian of *f*.

Definition If each of the partial derivatives $f_i(\mathbf{x})$, $i = 1,2,\ldots,n$, are also differentiable at **x**, the ***Hessian*** of *f* at **x** is given by,

$$\begin{aligned}
\mathbf{H}(\mathbf{x}) &= \nabla^2 f(\mathbf{x}) = \nabla\left(\nabla f(\mathbf{x})\right) \\
&= \nabla \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \nabla f_1(\mathbf{x})^{\mathrm{T}} \\ \nabla f_2(\mathbf{x})^{\mathrm{T}} \\ \vdots \\ \nabla f_n(\mathbf{x})^{\mathrm{T}} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \frac{\partial f_n(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) & \cdots & f_{1n}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) & \cdots & f_{2n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1}(\mathbf{x}) & f_{n2}(\mathbf{x}) & \cdots & f_{nn}(\mathbf{x}) \end{bmatrix} = \left[f_{ij}(\mathbf{x})\right]_{n\times n} \in \boldsymbol{R}^{n\times n},
\end{aligned}$$

- By Young's Theorem, Hessian is a symmetric matrix.
- *f* is twice differentiable at **x** if $\mathbf{H}(\mathbf{x})$ exists.
- *f* is twice differentiable if $\mathbf{H}(\mathbf{x})$ exists at each $\mathbf{x} \in \boldsymbol{S}$.

4.5 Chain Rules of Composite Functions of Several Variables

Chain Rule I Let $g : \boldsymbol{R} \to \boldsymbol{R}$ be the composite function where $g(t) = f(\mathbf{x}(t)) = f(x_1(t), x_2(t), \ldots, x_n(t))$. Since *g* is a function of single variable, we can write

$$\begin{aligned}
g'(t) &= \frac{\partial f(\mathbf{x}(t))}{\partial x_1}\frac{dx_1(t)}{dt} + \frac{\partial f(\mathbf{x}(t))}{\partial x_2}\frac{dx_2(t)}{dt} + \cdots + \frac{\partial f(\mathbf{x}(t))}{\partial x_n}\frac{dx_n(t)}{dt} \\
&= \nabla_{\mathbf{x}} f(\mathbf{x}(t))^{\mathrm{T}} \mathbf{x}'(t).
\end{aligned}$$

This is also called the ***total derivative*** of f with respect to t.

Example Let $f(x, y, z) = 3x + y^2 + z$, and

$$x(t) = t^2$$
$$y(t) = t$$
$$z(t) = \sqrt{t}.$$

We have

$$g'(t) = \frac{\partial f(x,y,z)}{\partial x}\frac{dx(t)}{dt} + \frac{\partial f(x,y,z)}{\partial y}\frac{dy(t)}{dt} + \frac{\partial f(x,y,z)}{\partial z}\frac{dz(t)}{dt}$$
$$= 3(2t) + 2y(1) + 1\left(\frac{1}{2}t^{-0.5}\right).$$

HW Baldani, p. 136, compute $g'(t)$ for the functions f as given in problem 5.1 with w, x, y, and z as functions of t as given in problem 5.2.

Chain Rule II Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$ and $\mathbf{x} : \boldsymbol{R}^s \to \boldsymbol{R}^n$ be functions in $\boldsymbol{C}^1$. Define the composite function $g : \boldsymbol{R}^s \to \boldsymbol{R}$, where $g(\mathbf{t}) = f(\mathbf{x}(\mathbf{t}))$.

$$\nabla g(\mathbf{t})^{\mathrm{T}} = \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{t}))^{\mathrm{T}} \nabla \mathbf{x}(\mathbf{t}),$$

where

$$\nabla \mathbf{x}(\mathbf{t}) = \begin{bmatrix} \nabla x_1(\mathbf{t})^{\mathrm{T}} \\ \nabla x_2(\mathbf{t})^{\mathrm{T}} \\ \vdots \\ \nabla x_n(\mathbf{t})^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \frac{\partial x_1(\mathbf{t})}{\partial t_1} & \frac{\partial x_1(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_1(\mathbf{t})}{\partial t_s} \\ \frac{\partial x_2(\mathbf{t})}{\partial t_1} & \frac{\partial x_2(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_2(\mathbf{t})}{\partial t_s} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial x_n(\mathbf{t})}{\partial t_1} & \frac{\partial x_n(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_n(\mathbf{t})}{\partial t_s} \end{bmatrix} \in \boldsymbol{R}^{n\times s},$$

and $\nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{t})) = \begin{bmatrix} \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_1} \\ \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_n} \end{bmatrix} \in \boldsymbol{R}^n$.

Proof Each element of the gradient $\nabla g(\mathbf{t})$ is the partial derivative of g with respect to t_i, $i = 1,2,\ldots,s$, which is just the derivative with $t_1, t_2, \ldots, t_{i-1}, t_{i+1}, \ldots t_s$ being constant. Thus, we can apply Chain Rule I and have

$$\frac{\partial g(\mathbf{t})}{\partial t_i} = \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_1}\frac{\partial x_1(\mathbf{t})}{\partial t_i} + \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_2}\frac{\partial x_2(\mathbf{t})}{\partial t_i} + \cdots + \frac{\partial f(\mathbf{x}(\mathbf{t}))}{\partial x_n}\frac{\partial x_n(\mathbf{t})}{\partial t_i} = \nabla_{\mathbf{x}} f(\mathbf{x}(\mathbf{t}))^{\mathrm{T}} \begin{bmatrix} \frac{\partial x_1(\mathbf{x})}{\partial t_i} \\ \frac{\partial x_2(\mathbf{t})}{\partial t_i} \\ \vdots \\ \frac{\partial x_n(\mathbf{t})}{\partial t_i} \end{bmatrix},$$

where the column vector in the last term is just the i^{th} column of $\nabla \mathbf{x}(\mathbf{t})$.❑

Chain Rule III Let $\mathbf{f}: \boldsymbol{R}^n \to \boldsymbol{R}^m$ and $\mathbf{x}: \boldsymbol{R} \to \boldsymbol{R}^n$ be functions in $\boldsymbol{C}^1$. Define the composite function $\mathbf{g}: \boldsymbol{R} \to \boldsymbol{R}^m$, where $\mathbf{g}(t) = \mathbf{f}(\mathbf{x}(t))$. Then

$$\mathbf{g}'(t) = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}(t))\mathbf{x}'(t),$$

where,

$$\nabla_{\mathbf{x}}\mathbf{f}\big(\mathbf{x}(t)\big)=\begin{bmatrix}\nabla f^{1}\big(\mathbf{x}(t)\big)^{\mathrm{T}}\\ \nabla f^{2}\big(\mathbf{x}(t)\big)^{\mathrm{T}}\\ \vdots\\ \nabla f^{m}\big(\mathbf{x}(t)\big)^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}\frac{\partial f^{1}(\mathbf{x}(t))}{\partial x_1} & \frac{\partial f^{1}(\mathbf{x}(t))}{\partial x_2} & \cdots & \frac{\partial f^{1}(\mathbf{x}(t))}{\partial x_n}\\ \frac{\partial f^{2}(\mathbf{x}(t))}{\partial x_1} & \frac{\partial f^{2}(\mathbf{x}(t))}{\partial x_2} & \cdots & \frac{\partial f^{2}(\mathbf{x}(t))}{\partial x_n}\\ \vdots & \vdots & \cdots & \vdots\\ \frac{\partial f^{m}(\mathbf{x}(t))}{\partial x_1} & \frac{\partial f^{m}(\mathbf{x}(t))}{\partial x_2} & \cdots & \frac{\partial f^{m}(\mathbf{x}(t))}{\partial x_n}\end{bmatrix}\in \boldsymbol{R}^{m\times n},$$

and

$$\mathbf{x}'(t)=\begin{bmatrix}x_1'(t)\\ x_2'(t)\\ \vdots\\ x_n'(t)\end{bmatrix}\in \boldsymbol{R}^{n}.$$

Proof Since $\mathbf{g}'(t)=\begin{bmatrix}\frac{dg^{1}(t)}{dt}\\ \frac{dg^{2}(t)}{dt}\\ \vdots\\ \frac{dg^{m}(t)}{dt}\end{bmatrix}$, and $g^{i}:\boldsymbol{R}\to\boldsymbol{R}$, $i=1,2,\ldots,m$, where $g^{i}(t)=f^{i}\big(\mathbf{x}(t)\big)$. By Chain Rule I, $\frac{dg^{i}(t)}{dt}=\nabla_{\mathbf{x}}f^{i}\big(\mathbf{x}(t)\big)^{\mathrm{T}}\mathbf{x}'(t)$, and the theorem follows.❑

Chain Rule IV Let $\mathbf{f}:\boldsymbol{R}^{n}\to\boldsymbol{R}^{m}$ and $\mathbf{x}:\boldsymbol{R}^{s}\to\boldsymbol{R}^{n}$ be functions in $\boldsymbol{C}^{1}$. Define the composite function $\mathbf{g}:\boldsymbol{R}^{s}\to\boldsymbol{R}^{m}$, where $\mathbf{g}(\mathbf{t})=\mathbf{f}\big(\mathbf{x}(\mathbf{t})\big)$. Then

$$\nabla\mathbf{g}(\mathbf{t})=\nabla_{\mathbf{x}}\mathbf{f}\big(\mathbf{x}(\mathbf{t})\big)\nabla\mathbf{x}(\mathbf{t}),$$

where

$$\nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x}(\mathbf{t}))=\begin{bmatrix}\nabla f^1(\mathbf{x}(\mathbf{t}))^{\mathrm{T}}\\ \nabla f^2(\mathbf{x}(\mathbf{t}))^{\mathrm{T}}\\ \vdots\\ \nabla f^m(\mathbf{x}(\mathbf{t}))^{\mathrm{T}}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{\partial f^1(\mathbf{x}(\mathbf{t}))}{\partial x_1} & \frac{\partial f^1(\mathbf{x}(\mathbf{t}))}{\partial x_2} & \cdots & \frac{\partial f^1(\mathbf{x}(\mathbf{t}))}{\partial x_n}\\ \frac{\partial f^2(\mathbf{x}(\mathbf{t}))}{\partial x_1} & \frac{\partial f^2(\mathbf{x}(\mathbf{t}))}{\partial x_2} & \cdots & \frac{\partial f^2(\mathbf{x}(\mathbf{t}))}{\partial x_n}\\ \vdots & \vdots & \cdots & \vdots\\ \frac{\partial f^m(\mathbf{x}(\mathbf{t}))}{\partial x_1} & \frac{\partial f^m(\mathbf{x}(\mathbf{t}))}{\partial x_2} & \cdots & \frac{\partial f^m(\mathbf{x}(\mathbf{t}))}{\partial x_n}\end{bmatrix}\in \boldsymbol{R}^{m\times n},$$

and

$$\nabla\mathbf{x}(\mathbf{t})=\begin{bmatrix}\nabla x_1(\mathbf{t})^{\mathrm{T}}\\ \nabla x_2(\mathbf{t})^{\mathrm{T}}\\ \vdots\\ \nabla x_n(\mathbf{t})^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}\frac{\partial x_1(\mathbf{t})}{\partial t_1} & \frac{\partial x_1(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_1(\mathbf{t})}{\partial t_s}\\ \frac{\partial x_2(\mathbf{t})}{\partial t_1} & \frac{\partial x_2(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_2(\mathbf{t})}{\partial t_s}\\ \vdots & \vdots & \cdots & \vdots\\ \frac{\partial x_n(\mathbf{t})}{\partial t_1} & \frac{\partial x_n(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial x_n(\mathbf{t})}{\partial t_s}\end{bmatrix}\in \boldsymbol{R}^{n\times s}.$$

Proof Since

$$\nabla\mathbf{g}(\mathbf{t})=\begin{bmatrix}\nabla g^1(\mathbf{t})^{\mathrm{T}}\\ \nabla g^2(\mathbf{t})^{\mathrm{T}}\\ \vdots\\ \nabla g^m(\mathbf{t})^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}\frac{\partial g^1(\mathbf{t})}{\partial t_1} & \frac{\partial g^1(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial g^1(\mathbf{t})}{\partial t_s}\\ \frac{\partial g^2(\mathbf{t})}{\partial t_1} & \frac{\partial g^2(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial g^2(\mathbf{t})}{\partial t_s}\\ \vdots & \vdots & \cdots & \vdots\\ \frac{\partial g^m(\mathbf{t})}{\partial t_1} & \frac{\partial g^m(\mathbf{t})}{\partial t_2} & \cdots & \frac{\partial g^m(\mathbf{t})}{\partial t_s}\end{bmatrix},$$

and $g^i:\boldsymbol{R}^s\to\boldsymbol{R}$, $i=1,2,\ldots,s$, where $g^i(\mathbf{t})=f^i(\mathbf{x}(\mathbf{t}))$.

By Chain Rule II, $\nabla g^i(\mathbf{t})^{\mathrm{T}}=\nabla_{\mathbf{x}}f^i(\mathbf{x}(\mathbf{t}))^{\mathrm{T}}\nabla\mathbf{x}(\mathbf{t})$, and the theorem follows.❑

HW (Chain Rule V) Let $\mathbf{f}:\boldsymbol{R}^n \to \boldsymbol{R}^m$, $\mathbf{x}:\boldsymbol{R}^s \to \boldsymbol{R}^n$ and $\mathbf{t}:\boldsymbol{R}^p \to \boldsymbol{R}^s$ be functions in $\boldsymbol{C}^1$. Define the composite function $\mathbf{g}:\boldsymbol{R}^p \to \boldsymbol{R}^m$, where $\mathbf{g}(\mathbf{r}) = \mathbf{f}(\mathbf{x}(\mathbf{t}(\mathbf{r})))$. Then $\mathbf{g}$ is also a function in $\boldsymbol{C}^1$. Show that $\nabla\mathbf{g}(\mathbf{r}) = \nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x}(\mathbf{t}(\mathbf{r})))\nabla_{\mathbf{t}}\mathbf{x}(\mathbf{t}(\mathbf{r}))\nabla\mathbf{t}(\mathbf{r})$.

Solution Let $\mathbf{y}(\mathbf{r}) = \mathbf{x}(\mathbf{t}(\mathbf{r}))$ and by Chain Rule IV, $\nabla\mathbf{y}(\mathbf{r}) = \nabla_{\mathbf{t}}\mathbf{x}(\mathbf{t}(\mathbf{r}))\nabla\mathbf{t}(\mathbf{r})$. Then write $\mathbf{g}(\mathbf{r}) = \mathbf{f}(\mathbf{y}(\mathbf{r}))$ and reapply Chain Rule IV, we have

$$\nabla\mathbf{g}(\mathbf{r}) = \nabla_{\mathbf{y}}\mathbf{f}(\mathbf{y}(\mathbf{r}))\nabla\mathbf{y}(\mathbf{r}) = \nabla_{\mathbf{y}}\mathbf{f}(\mathbf{y}(\mathbf{r}))\nabla_{\mathbf{t}}\mathbf{x}(\mathbf{t}(\mathbf{r}))\nabla\mathbf{t}(\mathbf{r}) = \nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x}(\mathbf{t}(\mathbf{r})))\nabla\mathbf{x}(\mathbf{t}(\mathbf{r}))\nabla\mathbf{t}(\mathbf{r}).$$

HW Find the gradient of the product of functions $g(\mathbf{x})f(\mathbf{x})$ when both are functions from $\boldsymbol{R}^n$ to $\boldsymbol{R}$.

HW Find the second order derivatives of $g(\mathbf{x})f(\mathbf{x})$ by chain rules.

HW Let $g:\boldsymbol{R} \to \boldsymbol{R}$ and $f:\boldsymbol{R}^n \to \boldsymbol{R}$ be functions in $\boldsymbol{C}^1$. Show that the composite function $h(\mathbf{x}) = g(f(\mathbf{x}))$ has the gradient and the Hessian given by

$$\nabla h(\mathbf{x}) = g'(f(\mathbf{x}))\nabla f(\mathbf{x})$$
$$\nabla^2 h(\mathbf{x}) = g'(f(\mathbf{x}))\nabla^2 f(\mathbf{x}) + \nabla f(\mathbf{x})g''(f(\mathbf{x}))\nabla f(\mathbf{x})^{\mathrm{T}}.$$

4.6 Directional Derivatives

4.6.1 First-order Directional Derivatives

We can apply Chain Rule I to find directional derivative of *f* at some point $\mathbf{x}_0$ in the direction **d**. This directional derivative of *f* is just the rate of change of *f* when we move away from $\mathbf{x}_0$ in the direction **d**. Some examples will help visualize this derivative.

Suppose $f : \boldsymbol{R} \to \boldsymbol{R}$ is a function. At any point x_0 in the domain, there are only two possible direction to move away from x_0 ; to increase or decrease x_0. If we chose to increase x_0, the directional derivative of *f* is just $f'(x_0)$, which is just the rate *f* changes at x_0. And if we decrease x_0, the directional derivative is $-f'(x_0)$.

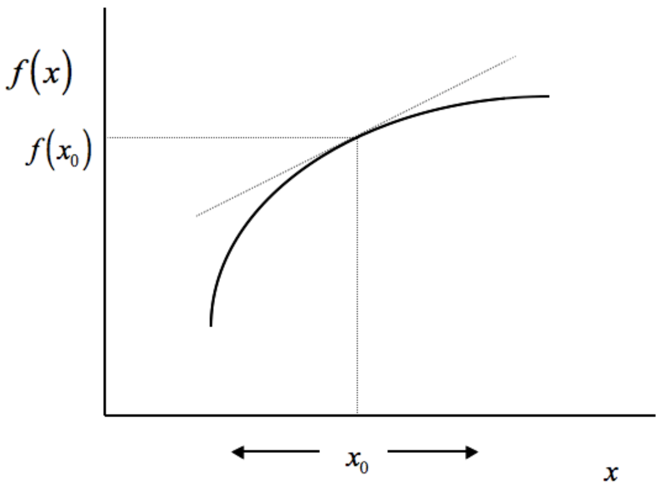

Figure 4.1 Directional derivative of *f* when the domain is $\boldsymbol{R}$.

Now let $f : \boldsymbol{R}^2 \to \boldsymbol{R}$, and given a point $\mathbf{x}_0$ and direction $\mathbf{d}$. If we move away from $\mathbf{x}_0$ in the direction $\mathbf{d}$, it means we move according to the function $\mathbf{x} : \boldsymbol{R} \to \boldsymbol{R}^2$, where

$$D^k(\mathbf{x}(t),\mathbf{d}) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} d_{i_1} d_{i_2} \cdots d_{i_k} f_{i_1 i_2 \cdots i_k}(\mathbf{x}(t)).$$

So $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{x}(1) = \mathbf{x}_0 + \mathbf{d}$ and $\mathbf{x}(-1) = \mathbf{x}_0 - \mathbf{d}$. The value of the function *f* for a given value of *t* is given according to a function of $g(0) = f(\mathbf{a})$, where

$$g(t) = f(\mathbf{x}(t)) = f(\mathbf{x}_0 + t\mathbf{d}),$$

which is a function of single variable *t*.

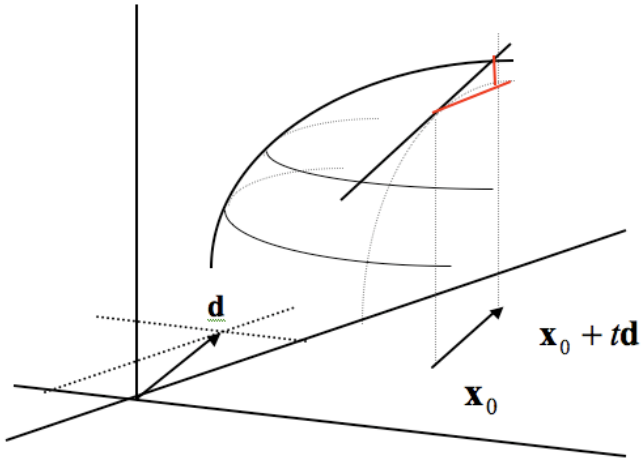

Figure 4.2 Directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{d}$ and the graph of $g(t)$.

We can take the derivative of g with respect to t, and this derivative at $t = 0$ will be the directional derivative of f at $\mathbf{x}_0$ in the direction **d**.

Theorem Let $f : \boldsymbol{S} \to \boldsymbol{R}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a function that is differentiable at $\mathbf{x}_0 \in \operatorname{int} \boldsymbol{S}$. Let $\mathbf{d} \in \boldsymbol{R}^n$ and define $g : \mathbf{R} \to \mathbf{R}$, where $g(t) = f(\mathbf{x}(t)) = f(\mathbf{x}_0 + t\mathbf{d})$. The ***first-order directional derivative*** of f at $\mathbf{x}_0$ in the direction **d** is given by $\nabla f(\mathbf{x}_0)^{\mathrm{T}} \mathbf{d}$.

Proof By Chain Rule I, the derivative of the composite function g with respect to t is given by,

$$
\begin{aligned}
g'(t) &= \frac{\partial}{\partial x_1} f(\mathbf{x}(t)) \frac{d}{dt}(x_1^0 + td_1) + \cdots + \frac{\partial}{\partial x_n} f(\mathbf{x}(t)) \frac{d}{dt}(x_n^0 + td_n) \\
&= \frac{\partial}{\partial x_1} f(\mathbf{x}(t)) d_1 + \cdots + \frac{\partial}{\partial x_n} f(\mathbf{x}(t)) d_n \\
&= \nabla_{\mathbf{x}} f(\mathbf{x}(t))^{\mathrm{T}} \mathbf{d}.
\end{aligned}
$$

Set $t = 0$, $g'(0)$ is the rate of change of the graph $(\mathbf{x}(t), f(\mathbf{x}(t)))$ at $\mathbf{x}_0$, and is given by,

$$
g'(0) = \nabla f(\mathbf{x}_0)^{\mathrm{T}} \mathbf{d}. \square
$$

Problem The vectors **d** and 2**d** are vectors of the same direction. Thus, are the directional derivatives with respect to direction **d** and 2**d** equal?

4.6.2 Directional Derivatives and Optimization

Directional derivative is essential in optimization. For example, if $\mathbf{x}_0$ is a minimum point of *f*, and if *f* is differentiable at $\mathbf{x}_0$, then the directional derivative at $\mathbf{x}_0$ in any direction **d** necessarily vanishes to zero. That is,

$$\nabla f\left(\mathbf{x}_0\right)^{\mathrm{T}}\mathbf{d} \geq 0\text{, for any } \mathbf{d} \in \boldsymbol{R}^n \iff \nabla f\left(\mathbf{x}_0\right) = \mathbf{0}\ .$$

Problem Show that if $\mathbf{x}_0$ is a minimum point of *f*, and if *f* is differentiable at $\mathbf{x}_0$, then $\nabla f\left(\mathbf{x}_0\right)^{\mathrm{T}}\mathbf{d} \geq 0$, for any $\mathbf{d} \in \boldsymbol{R}^n$, and this is equivalent to saying that $\nabla f\left(\mathbf{x}_0\right) = \mathbf{0}$.

The directional derivative also indicates the direction that *f* can be increase at the fastest rate. (See also the discussion of gradient and level set below.) To state the problem explicitly, if $\nabla f\left(\mathbf{x}_0\right)$ exists at some point $\mathbf{x}_0$ and $\nabla f\left(\mathbf{x}_0\right) \neq \mathbf{0}$, what is the direction $\mathbf{d} \in \boldsymbol{R}^n$ that maximizes the directional derivative $\nabla f\left(\mathbf{x}_0\right)^{\mathrm{T}}\mathbf{d}$, where $\|\mathbf{d}\| = 1$? The condition that $\|\mathbf{d}\| = 1$ is needed to ensure that an optimal solution exists. Why? We can write the constrained optimization problem as

$$\begin{aligned} &\max \quad \nabla f\left(\mathbf{x}_0\right)^{\mathrm{T}}\mathbf{d} \\ &\text{st} \quad \sum_{i=1}^{n} d_i^2 = 1. \end{aligned}$$

It can be shown that optimal solution of this problem is given by $\mathbf{d}* = \dfrac{\nabla f\left(\mathbf{x}_0\right)}{\left[\nabla f\left(\mathbf{x}_0\right)^{\mathrm{T}}\nabla f\left(\mathbf{x}_0\right)\right]^{\frac{1}{2}}}$, which is just the normalization of the gradient $\nabla f(\mathbf{x}_0)$ and has the same direction as the gradient $\nabla f(\mathbf{x}_0)$.

Problem Show that the result obtained in the previous paragraph is true for the two-variable function *f*.

Problem **Simon & Blume** [1994], page 350, #15.12, 15.13.

15.12 Consider the function $f(x,y) = x^2e^y$. In what direction should one more from the point $(2,0)$ in order to increase f most quickly? Express your answer as a vector of length 1.

15.13 A firm uses *x* hours of unskilled labor and *y* hours of skilled labor each day to produce $Q(x,y) = 60x^{\frac{2}{3}}y^{\frac{1}{3}}$ units of output per day. It currently employs 64 hours of unskilled labor and 27 hours of skilled labor.

a) What is the current output?
b) In what direction (expressed as a unit vector) should it change (x,y) if it wants to increase output most rapidly?
c) The firm is planning to hire an additional hour and a half of skilled labor. Use calculus to estimate the corresponding change in unskilled labor that would keep its output at its current level.

Problem Show that if the gradient vector $\nabla f(\mathbf{x}_0)$ and direction vector $\mathbf{d}$ make an angle less than 90 degrees, the value of *f* increases along the direction $\mathbf{d}$. (This is used in showing that the Lagrange Multipliers associated with the binding constraints are strictly positive)

4.6.3 Second-Order Directional Derivatives

Theorem Let $f:\boldsymbol{S}\to\boldsymbol{R}$, $\boldsymbol{S}\subseteq\boldsymbol{R}^n$, be a function that is twice differentiable at $\mathbf{x}_0\in\operatorname{int}\boldsymbol{S}$. Let $\mathbf{d}\in\boldsymbol{R}^n$ and define $g(0)=f(\mathbf{a})$, where $g(t)=f(\mathbf{x}(t))=f(\mathbf{x}_0+t\mathbf{d})$. The ***second-order directional derivative*** of f at $\mathbf{x}_0$ in the direction $\mathbf{d}$ is given by $\mathbf{d}^{\mathrm{T}}\nabla^2 f(\mathbf{x}_0)\mathbf{d}$.

Proof From the proof the the first-order derivative, $g'(t)=\nabla f(\mathbf{x}(t))^{\mathrm{T}}\mathbf{d}$, the second order derivative of g is given by

$$\begin{aligned}
g''(t) &= \frac{d}{dt}\nabla f(\mathbf{x}(t))^{\mathrm{T}}\mathbf{d} \\
&= \frac{d}{dt}f_1(\mathbf{x}(t))d_1+\cdots+\frac{d}{dt}f_n(\mathbf{x}(t))d_n \\
&= d_1\nabla f_1(\mathbf{x}(t))^{\mathrm{T}}\mathbf{d}+\cdots+d_n\nabla f_n(\mathbf{x}(t))^{\mathrm{T}}\mathbf{d} \\
&= d_1\sum_{j=1}^{n}f_{1j}(\mathbf{x}(t))d_j+\cdots+d_n\sum_{j=1}^{n}f_{nj}(\mathbf{x}(t))d_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}f_{ij}(\mathbf{x}(t))d_i d_j \\
&= \mathbf{d}^{\mathrm{T}}\nabla^2 f(\mathbf{x}(t))\mathbf{d}.
\end{aligned}$$

The second-order directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{d}$ is given by $g''(0)=\mathbf{d}^{\mathrm{T}}\nabla^2 f(\mathbf{x}_0)\mathbf{d}$.❑

Definition Define the k^{th}-order directional derivative of f at $\mathbf{x}(t)$ in the direction of $\mathbf{d}$ is given by $D^k(\mathbf{x}(t),\mathbf{d})$. That is,

$$\begin{aligned}
D^1(\mathbf{x}(t),\mathbf{d}) &= \nabla f(\mathbf{x}(t))^{\mathrm{T}}\mathbf{d} \\
D^2(\mathbf{x}(t),\mathbf{d}) &= \mathbf{d}^{\mathrm{T}}\nabla^2 f(\mathbf{x}(t))\mathbf{d}.
\end{aligned}$$

Problem Show that the k^{th}-order directional derivative of f at $\mathbf{x}(t)$ in the direction of $\mathbf{d}$ is given by

$$D^k(\mathbf{x}(t),\mathbf{d}) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} d_{i_1} d_{i_2}\cdots d_{i_k} f_{i_1 i_2\cdots i_k}(\mathbf{x}(t)).$$

4.7 Mean Value Theorem in $\boldsymbol{R}^n$

Theorem Let $f:\boldsymbol{S}\to\boldsymbol{T}$, $\boldsymbol{S}\subseteq\boldsymbol{R^n}$ be a function in $\boldsymbol{C}^1$. Let $\mathbf{a},\mathbf{b}\in\operatorname{int}\boldsymbol{S}$ such that any convex combination $(1-\lambda)\mathbf{a}+\lambda\mathbf{b}=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})\in\operatorname{int}\boldsymbol{S}$, for all $0\le\lambda\le 1$. Then, there exists some $\mathbf{c}=\mathbf{a}+\lambda_0(\mathbf{b}-\mathbf{a})$, for some , such that

$$f(\mathbf{b})-f(\mathbf{a})=\nabla f(\mathbf{c})^{\mathrm{T}}(\mathbf{b}-\mathbf{a}).$$

Proof Define $\mathbf{x}(\lambda)=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})$, and $g(\lambda)=f(\mathbf{x}(\lambda))$. Thus, $g:[0,1]\to\boldsymbol{R}$ is also a function in $\boldsymbol{C}^1$, where $g(0)=f(\mathbf{a})$ and $g(1)=f(\mathbf{b})$. By the Mean Value Theorem of functions of single variable, there exists $\lambda_0\in(0,1)$ such that,

$$g(1)-g(0)=g'(\lambda_0)(1-0)=g'(\lambda_0).$$

The derivative of g with respect to is given by,

$$\begin{aligned} g'(\lambda) &= \frac{d}{d\lambda}f(\mathbf{x}(\lambda))=\nabla f(\mathbf{x}(\lambda))^{\mathrm{T}}\mathbf{x}'(\lambda) \\ &= \nabla f(\mathbf{x}(\lambda))^{\mathrm{T}}(\mathbf{b}-\mathbf{a}). \end{aligned}$$

At λ_0, $g'(\lambda_0)=\nabla f(\mathbf{c})^{\mathrm{T}}(\mathbf{b}-\mathbf{a})$, where $\mathbf{c}=\mathbf{a}+\lambda_0(\mathbf{b}-\mathbf{a})$, and the theorem follows.❑

4.8 Taylor's Polynomials and Taylor's Approximation in $\boldsymbol{R}^n$

Theorem Let $f : \boldsymbol{S} \to \boldsymbol{T}$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$ be a function in $\boldsymbol{C}^k$. Then, for any $\mathbf{a} \in \operatorname{int}\boldsymbol{S}$ and $\mathbf{h} \in \boldsymbol{R}^n$ such that all the points $\mathbf{a} + t\mathbf{h} \in \operatorname{int}\boldsymbol{S}$, for any $0 \le t \le 1$,

$$f(\mathbf{a}+\mathbf{h}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^{\mathrm{T}}\mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\nabla^2 f(\mathbf{a})\mathbf{h} + \cdots + \frac{1}{k!}D^k(\mathbf{a},\mathbf{h})$$

.

Proof For a given point , define the function $g(0) = f(\mathbf{a})$, where

$$g(t) = f(\mathbf{a} + t\mathbf{h}) .$$

Since this function g is also a function in $\boldsymbol{C}^k$, by Taylor's Polynomials of functions of single variables,

$$g(1) \approx g(0) + g'(0) + \frac{1}{2}g''(0) + \cdots + \frac{1}{k!}g^k(0),$$

where,

$$\begin{aligned} g(0) &= f(\mathbf{a}) \\ g'(0) &= \nabla f(\mathbf{a})^{\mathrm{T}}\mathbf{h} \\ g''(0) &= \mathbf{h}^{\mathrm{T}}\nabla^2 f(\mathbf{a})\mathbf{h} = \mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{a})\mathbf{h}, \text{ and} \\ g^k(0) &= D^k(\mathbf{a},\mathbf{h}). \end{aligned}$$

❑

Theorem Let $f\!:\!\boldsymbol{S} \to \boldsymbol{R}$, $\mathrm{S} \subseteq \boldsymbol{R}^n$ be a function in $\boldsymbol{C}^2$. Then, for any $\mathbf{a} \in \operatorname{int}\boldsymbol{S}$ and $\mathbf{h} \in \boldsymbol{R}^n$ such that all the points $\mathbf{a} + t\mathbf{h} \in \operatorname{int}\boldsymbol{S}$, for any $0 \le t \le 1$,

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a})^{\mathrm{T}}\mathbf{h} + R_2(\mathbf{h};\mathbf{a}) ,$$

where $R_2(\mathbf{h};\mathbf{a}) = \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{c})\mathbf{h}$, $\mathbf{c} = \mathbf{a} + t_0\mathbf{h}$ for some $t_0 \in (0,1)$.

Proof For a given point $\mathbf{a} \in \operatorname{int} \boldsymbol{S}$, define the function $g: \boldsymbol{R} \to \boldsymbol{R}$, where

$$g(t) = f(\mathbf{a} + t\mathbf{h}).$$

Since this function g is also a function in $\boldsymbol{C}^2$, by Taylor's Approximation of functions of single variables, there exists $t_0 \in (0,1)$ such that

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(t_0).$$

By Chain Rule I, we have from the problem above,

$$g'(t) = \nabla f(\mathbf{a} + t\mathbf{h})^{\mathrm{T}} \mathbf{h},$$
$$g''(t_0) = \mathbf{h}^{\mathrm{T}} \mathbf{H}(\mathbf{a} + t_0\mathbf{h})\mathbf{h}.$$

Thus,

$$g(1) = f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a})^{\mathrm{T}} \mathbf{h} + \frac{1}{2} \mathbf{h}^{\mathrm{T}} \mathbf{H}(\mathbf{c})\mathbf{h},$$

where $\mathbf{c} = \mathbf{a} + t_0\mathbf{h}$.❑

The following Corollary shows that the remainder term converges to zero faster than **h** does.

Corollary Let $R_2(\mathbf{h};\mathbf{a})$ be defined as in the previous theorem. Then,

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{R_2(\mathbf{h};\mathbf{a})}{\|\mathbf{h}\|} = 0.$$

Proof Let $\mathbf{h} = t\mathbf{d}$, where $\|\mathbf{d}\| = 1$. Thus $\mathbf{h} \to \mathbf{0}$ if, and only if, $t \to 0$. We have,

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{R_2(\mathbf{h};\mathbf{a})}{\|\mathbf{h}\|}=\lim_{\mathbf{h}\to\mathbf{0}}\frac{\frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{c})\mathbf{h}}{\|\mathbf{h}\|}$$

$$=\lim_{t\to 0}\frac{\frac{1}{2}t\mathbf{d}^{\mathrm{T}}\mathbf{H}(\mathbf{c})t\mathbf{d}}{\|t\mathbf{d}\|}$$

$$=\lim_{t\to 0}\frac{\frac{1}{2}t^2\mathbf{d}^{\mathrm{T}}\mathbf{H}(\mathbf{c})\mathbf{d}}{t\|\mathbf{d}\|}$$

$$=\lim_{t\to 0}\frac{\frac{1}{2}t\mathbf{d}^{\mathrm{T}}\mathbf{H}(\mathbf{c})\mathbf{d}}{\|\mathbf{d}\|}=0.$$

❑

Theorem Let $f:\boldsymbol{S}\to\boldsymbol{R}$, $\boldsymbol{S}\subseteq\boldsymbol{R}^n$ be a function in $\boldsymbol{C}^3$. Then, for any $\mathbf{a}\in\operatorname{int}\boldsymbol{S}$ and $\mathbf{h}\in\boldsymbol{R}^n$ such that all the points $\mathbf{a}+t\mathbf{h}\in\operatorname{int}\boldsymbol{S}$, for any $0\le t\le 1$,

$$f(\mathbf{a}+\mathbf{h})=f(\mathbf{a})+\nabla f(\mathbf{a})^{\mathrm{T}}\mathbf{h}+\frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{a})\mathbf{h}+R_3(\mathbf{h};\mathbf{a}),$$

where $R_3(\mathbf{h};\mathbf{a})=\frac{1}{3!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}h_ih_jh_kf_{ijk}(\mathbf{c})$, $\mathbf{c}=\mathbf{a}+t_0\mathbf{h}$ for some $t_0\in(0,1)$.

Proof Similar to the proof of first-order Taylor's Approximation, for a given point $\mathbf{a}\in\operatorname{int}\boldsymbol{S}$, define the function $g:\boldsymbol{R}\to\boldsymbol{R}$, where

$$g(t)=f(\mathbf{a}+t\mathbf{h}).$$

Since this function g is also a function in $\boldsymbol{C}^3$, by Taylor's Approximation of functions of single variables, there exists $t_0\in(0,1)$ such that

$$g(1)=g(0)+g'(0)+\frac{1}{2}g''(0)+\frac{1}{3!}g'''(t_0).$$

By chain rule I, $g'(0)$ and $g''(0)$ have been found above. We will compute the third-order derivative of g with respect to t as follows,

$$\begin{aligned}
g'''(t) = \frac{dg''(t)}{dt} &= \frac{d}{dt}\mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{a}+t\mathbf{h})\mathbf{h} \\
&= \frac{d}{dt}\sum_{i=1}^{n}\sum_{j=1}^{n}h_ih_jf_{ij}(\mathbf{a}+t\mathbf{h}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}h_ih_j\frac{d}{dt}f_{ij}(\mathbf{a}+t\mathbf{h}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}h_ih_j\nabla f_{ij}(\mathbf{a}+t\mathbf{h})^{\mathrm{T}}\mathbf{h} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}h_ih_jh_kf_{ijk}(\mathbf{a}+t\mathbf{h}).
\end{aligned}$$

Thus,

$$g(1) = f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a})^{\mathrm{T}}\mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathbf{H}(\mathbf{a})\mathbf{h} + \frac{1}{3!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}h_ih_jh_kf_{ijk}(\mathbf{c})$$

,

where $\mathbf{c} = \mathbf{a} + t_0\mathbf{h}$.❑

Corollary Let $R_3(\mathbf{h};\mathbf{a})$ be defined as in the previous theorem. Then,

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{R_3(\mathbf{h};\mathbf{a})}{\|\mathbf{h}\|^2} = 0.$$

Proof Let $\mathbf{h} = t\mathbf{d}$, where $\|\mathbf{d}\| = 1$. Thus $\mathbf{h}\to\mathbf{0}$ if, and only if, $t\to 0$. We have,

$$\begin{aligned}
\lim_{\mathbf{h}\to\mathbf{0}}\frac{R_3(\mathbf{h};\mathbf{a})}{\|\mathbf{h}\|^2} &= \lim_{\mathbf{h}\to\mathbf{0}}\frac{\frac{1}{3!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}h_ih_jh_kf_{ijk}(\mathbf{c})}{\|\mathbf{h}\|^2} \\
&= \lim_{t\to 0}\frac{\frac{1}{3!}t^3\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}d_id_jd_kf_{ijk}(\mathbf{c})}{t^2\|\mathbf{d}\|} \\
&= \lim_{t\to 0}\frac{1}{3!}t\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}d_id_jd_kf_{ijk}(\mathbf{c})}{\|\mathbf{d}\|^2} = 0.
\end{aligned}$$

❑

4.9 Implicit Functions and Implicit Function Theorem

Functions that we have discussed so far are in the ***explicit*** form $y = f\left(x_1, x_2, \ldots, x_n\right)$, where y is called ***endogenous*** variable, and $x_1, x_2, \ldots, x_n$ ***exogenous*** variables. The exogenous variables determine the value of the endogenous variable according to the function *f*.

There are functions in implicit form $f\left(x_1, x_2, \ldots, x_n, y\right) = c$, where the endogenous variable y is determined implicitly by the exogenous variables. If we fix the value of an explicit function *f* to a number *c*, the endogenous and exogeneous variables are now related in such a way that if exogeneous variables change the endogeneous variable will change so that the value of the function *f* still remains at *c*.

This is how the isoquant is obtained by fixing a level of output so that the input choices are related implicitly to produce the same level of output.

The following theorem describes the properties of implicit functions.

Implicit Function Theorem (**Simon & Blume** [1994], page 339, Theorem 15.1) Let $f(x, y)$ be a $\boldsymbol{C}^1$ function on a neighborhood about a point (x_0, y_0) in $\boldsymbol{R}^2$. Suppose that $f(x_0, y_0) = c$ and consider the expression

$$f(x, y) = c.$$

If $f_x(x_0, y_0) \neq 0$, then there exists a $\boldsymbol{C}^1$ function $x = x(y)$ defined on a line segment $|y - y_0| < \varepsilon$ for some $\varepsilon > 0$ such that,

a) $f(x(y), y) \equiv c$, for all $|y - y_0| < \varepsilon$.

b) $x(y_0) = x_0$, and

c) $x'(y_0) = -\dfrac{f_y(x_0, y_0)}{f_x(x_0, y_0)}$.

Proof See a sketch of proof in **Simon & Blume** [1994], page 344.❑

This theorem still applies when $f_x(x_0, y_0) = 0$ if $f_y(x_0, y_0) \neq 0$, but now we have to write y as a function of x instead, and the three results are modified correspondingly. That is,

If $f_y(x_0, y_0) \neq 0$, then there exists a $\boldsymbol{C}^1$ function $y = y(x)$ defined on a line segment $|x - x_0| < \varepsilon$ **,** for some $\varepsilon > 0$, such that,

a) $f(x, y(x)) \equiv c$, for all $|x - x_0| < \varepsilon$.

b) $y(x_0) = y_0$, and

c) . $y'(x_0) = -\dfrac{f_x(x_0, y_0)}{f_y(x_0, y_0)}$.

The point (x_0, y_0) is called a ***regular point*** of the $\boldsymbol{C}^1$ function $f(x,y)$ if $f_x(x_0,y_0) \neq 0$ or $f_y(x_0,y_0) \neq 0$.

Example In the simple macroeconomic model,

$$Y = C(Y) + I + G,$$

we can write an implicit function

$$f(Y,I,G) = Y - C(Y) - I - G = 0,$$

where Y is endogeneous while I and G are exogeneous. Applying the Impicity Function Theorem, we obtain

????????.

HW Baldani, p. 136, #5.4.

Implicit Function Theorem (**Sundaram** [1996], Theorem 1.77, page 66) Let $\mathbf{f}: \boldsymbol{S} \to \boldsymbol{R}^k$, $\boldsymbol{S} \subseteq \boldsymbol{R}^n$, be a $\boldsymbol{C}^1$ function on a neighborhood about a point $(\mathbf{x}_0, \mathbf{y}_0) \in \boldsymbol{R}^n$, where $\mathbf{x}_0 \in \boldsymbol{R}^k$, $\mathbf{y}_0 \in \boldsymbol{R}^{n-k}$. Suppose that $\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{c}$, and the gradient of the function $\mathbf{f}$ with respect to the first k variables $\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Then, there is a $\boldsymbol{C}^1$ function $\mathbf{x} = \mathbf{x}(\mathbf{y})$, for some $\varepsilon > 0$ such that

a) $\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{c}$, for all $\|\mathbf{y} - \mathbf{y}_0\| < \varepsilon$,

b) $\mathbf{x}(\mathbf{y}_0) = \mathbf{x}_0$, and

c) $\nabla \mathbf{x}(\mathbf{y}_0) = -[\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)]^{-1} \nabla_{\mathbf{y}} \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)$.

Proof See **Rudin** [1976], Theorem 9.28, page 224.❑

In the theorem above, each of the variables $y_1, y_2, \ldots, y_{n-k}$ in the vector $\mathbf{y}$ can vary. However, we can keep all $y_1, y_2, \ldots, y_{n-k}$ constant and change only y_{n-k}. What would be the gradient $\nabla \mathbf{x}(\mathbf{y}_0)$?

HW Baldani, p. 136, #5.6.

4.10 Level Sets, Tangents and Gradients

Let $f(x,y)$ be a $\boldsymbol{C}^1$ function. A ***level set*** of the function $f:\boldsymbol{S}\to\boldsymbol{R}$, $\boldsymbol{S}\subseteq\boldsymbol{R}^2$, in $\boldsymbol{C}^1$ for a given level c is given by the set $\boldsymbol{L}_f(c)=\{(x,y)|f(x,y)=c\}$. By the Implicit Function Theorem, the slope of the level set at a regular point $(x_0,y_0)\in\boldsymbol{L}_f(c)=\{(x,y)|f(x,y)=c\}$, where $f_y(x_0,y_0)\neq 0$, is given by

$$-\frac{f_x(x_0,y_0)}{f_y(x_0,y_0)}.$$

A vector **v** that has this slope is given by,

$$\mathbf{v}=\begin{bmatrix} f_y(x_0,y_0) \\ -f_x(x_0,y_0)\end{bmatrix}.$$

The gradient of *f* at this point (x_0,y_0) is given by,

$$\nabla f(x_0,y_0)=\begin{bmatrix} f_x(x_0,y_0) \\ f_y(x_0,y_0)\end{bmatrix}.$$

The gradient $\nabla f(x_0,y_0)$ and the vector **v** are at right angle to each other, or orthogonal, as

$$\nabla f(x_0,y_0)^{\mathrm{T}}\mathbf{v}=\begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0)\end{bmatrix}\begin{bmatrix} f_y(x_0,y_0) \\ -f_x(x_0,y_0)\end{bmatrix}=0.$$

Definition Vectors **x** and **y** in $\boldsymbol{R}^n$ are ***orthogonal*** if $\mathbf{x}^{\mathrm{T}}\mathbf{y}=0$.

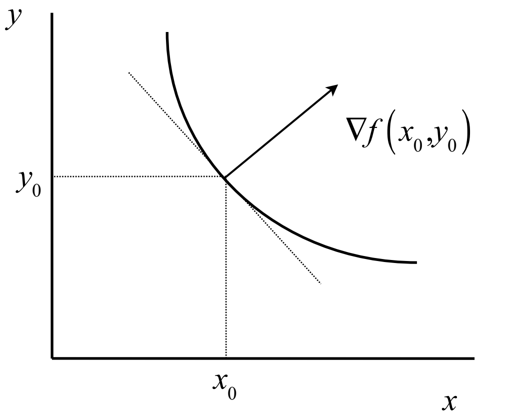

Figure 4.3 The gradient $\nabla f(x_0, y_0)$ and the tangent line of the level set at (x_0, y_0) are orthogonal.

Example Let $f(x,y) = ax + by$. For a fixed number *c*, we have an implicit function $f(x,y) = ax + by = c$, and a level set $\{(x,y) | ax + by = c\}$.

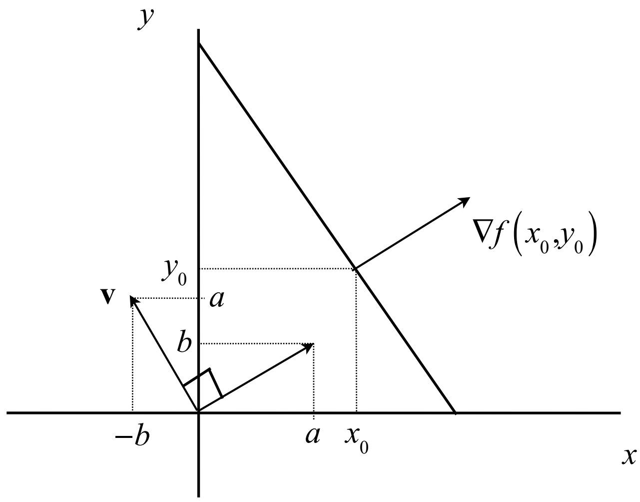

Figure 4.4 The gradient $\nabla f(x_0, y_0)$ and the tangent line of the level set at (x_0, y_0) when the function is linear.

By the Implicit Function Theorem, the slope of the level set at (x_0, y_0), which is regular, is given by

$$-\frac{f_x(x_0,y_0)}{f_y(x_0,y_0)} = -\frac{a}{b}.$$

A vector **v** that has this slope is given by,

$$\mathbf{v} = \begin{bmatrix} -b \\ a \end{bmatrix}.$$

The gradient of *f* at this point (x_0, y_0) is given by,

$$\nabla f(x_0, y_0) = \begin{bmatrix} a \\ b \end{bmatrix}.$$

The gradient $\nabla f(x_0, y_0)$ and the vector **v** are orthogonal, as

$$\nabla f(x_0, y_0)^{\mathrm{T}} \mathbf{v} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} -b \\ a \end{bmatrix} = 0.$$

Note that the gradient $\nabla f(x_0, y_0)$ is the direction that the value of the function f increases, and thus $-\nabla f(x_0, y_0)$ is the direction that it decreases. In Chapter 8 Equality Constrained Optimization, it will be shown that $\nabla f(x_0, y_0)$ is the direction the value of the function *f* increases at the fastest rate.

HW Baldani, p. 137, #5.7.

4.11 Homogeneity

Definition A function $f(\mathbf{x})$ is ***homogeneous of degree*** *k* if

$$f(t\mathbf{x}) = t^k f(\mathbf{x}).$$

- a homogeneous production function is increasing return to scales if $k > 1$.

Theorem The partial derivative $f_j(\mathbf{x})$, $j = 1, 2, \ldots, n$, of a homogeneous function of degree k is also homogeneous of degree $k - 1$.

Proof Let $\mathbf{y} = t\mathbf{x}$. We have

$$f(\mathbf{y}) = t^k f(\mathbf{x})$$
$$\frac{\partial f(\mathbf{y})}{\partial y_j} = t^k \frac{\partial f(\mathbf{x})}{\partial x_j}\frac{dx_j}{dy_j} = t^k \frac{\partial f(\mathbf{x})}{\partial x_j}\frac{1}{t}$$
$$f_j(t\mathbf{x}) = t^{k-1} f_j(\mathbf{x}).$$

Corollary The ratio of partial derivative is homogeneous of degree 0. This makes the expansion path of the firm with homogeneous production linear.

Theorem Euler's Theorem. If f is a homogeneous function of degree k, then $\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{x} = kf(\mathbf{x})$.

Proof Take derivative of $f(t\mathbf{x}) = t^k f(\mathbf{x})$ with respect to t on both sides,

$$\frac{df(t\mathbf{x})}{dt} = kt^{k-1} f(\mathbf{x}).$$

The left-hand side is, with application of the previous theorem,

$$\begin{aligned}\frac{df(t\mathbf{x})}{dt} &= \sum_{j=1}^{n} f_j(t\mathbf{x})\frac{dtx_j}{dt}\\ &= \sum_{j=1}^{n} t^{k-1} f_j(\mathbf{x})x_j\\ &= kt^{k-1} f(\mathbf{x}).\end{aligned}$$

Then we have

$$\sum_{j=1}^{n} f_j(\mathbf{x})x_j = kf(\mathbf{x}).$$

HW Show that

1) A demand function of product 1, as given by $D_1(p_1, p_2, \ldots, p_n, I)$ as a function of its price p_1 and prices of other products $p_2, \ldots, p_n$ and income I, is homogeneous of degree 0.

2) Show that the sum of own price elasticity, cross price elasticity and income elasticity of any product is equal to 0.

HW Baldani, p. 137, #5.8.