Real Analysis:

Differential Calculus

1. One-Variable Calculus: Differentiability of Functions

- **Slope of a Linear Function:**

  The slope of a linear function f measures how much $f(x)$ changes for each unit increase in x.

  – It measures the *rate of change* of the function f.

    - Linear functions have the same rate of change no matter where we start.

- The view of the *slope* of a linear function as its *rate of change* (the *marginal effect*) plays a key role in economic analysis:

  – marginal cost;

  – marginal utility;

  – marginal product of labour.

1.1 Slope of Nonlinear Functions

- How do we measure the marginal effects of nonlinear functions?

- Suppose that $y = f(x)$ is a nonlinear function and we are at the point $(x_0, f(x_0))$ on the graph of f, as in the following figure.
  - We want to measure the rate of change of f when $x = x_0$.
    - A natural solution to this problem is to draw the *tangent line* to the graph of f at x_0 as pictured in the figure.

- Since *the tangent line very closely approximates the graph of f around* $(x_0, f(x_0))$, it is a good proxy for the graph of f itself.
  - The slope of the tangent line should be a good measure for the slope of the nonlinear function at x_0.

- For nonlinear f, the slope of tangent line varies from point to point.

$$f'(x_0) \qquad \text{or} \qquad \frac{df}{dx}(x_0).$$

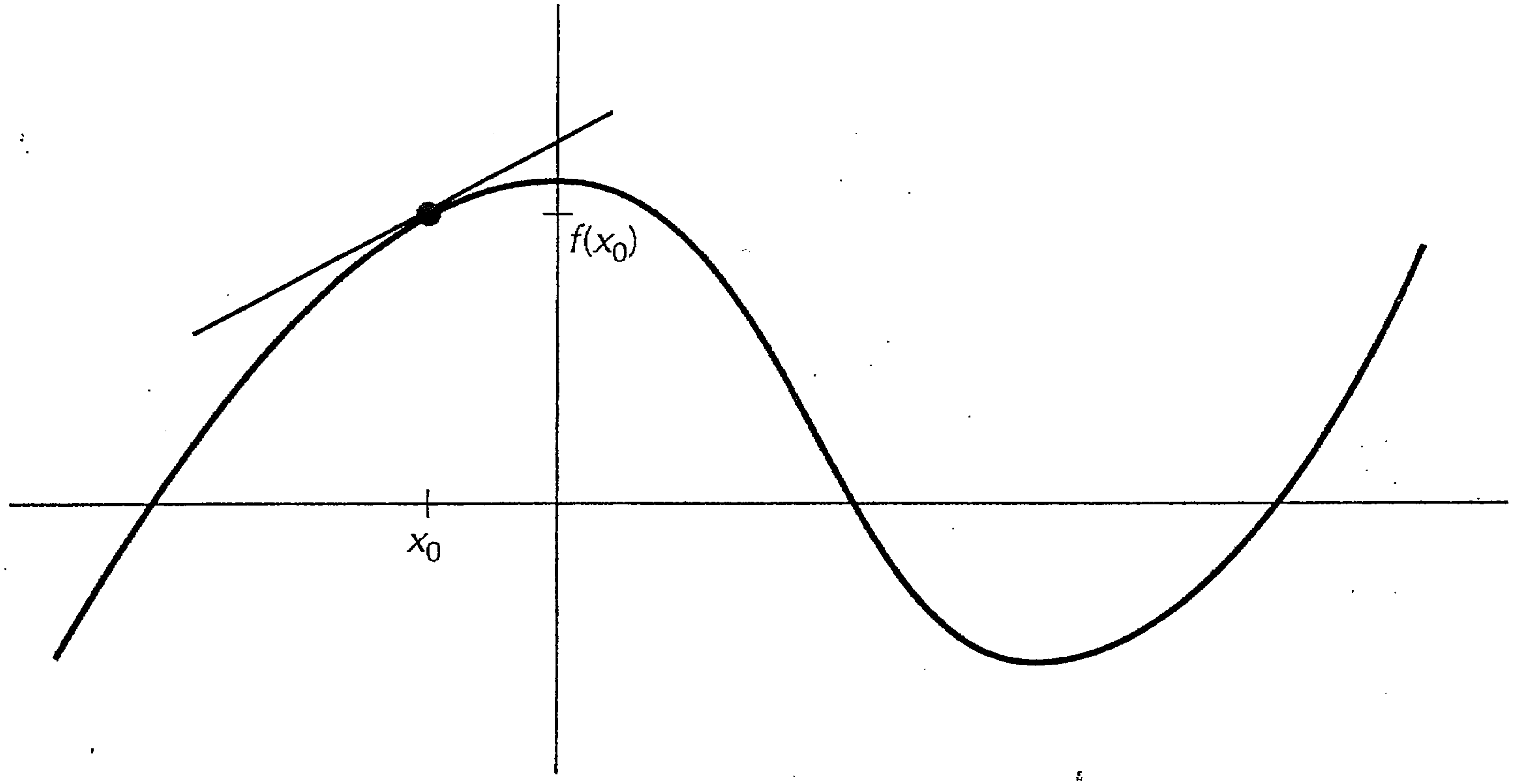


Figure 2.8 *The graph of a nonlinear function.*

- We use the notion of the tangent line approximation to a graph in our daily lives.
  - Contractors who plan to build a large mall or power plant or farmers who want too subdivide large plots of land will generally assume that they are working on a flat plane,
    - even though they know that they are actually working on the rather round earth surface.
  - In effect, they are working with the tangent plane to the earth and the computations that they make on it will be exact to 10 or 20 decimal places – easily close enough for their purposes.

- We define the *slope of a nonlinear function* f at a point $(x_0, f(x_0))$ on its graph as the *slope of the tangent line to the graph of* f at that point.
  - We call the *slope of the tangent line to the graph of* f at $(x_0, f(x_0))$ the **derivative** of f at x_0, and we write it as

$$f'(x_0) \text{ or } \frac{df}{dx}(x_0).$$

    - The latter notation comes from the fact that the slope is the change in f divided by the change in x, or $\dfrac{\Delta f}{\Delta x}$.

- Since the derivative is such an important concept, we need an analytical definition to work with.
  - This problem is best handled using a limiting process.

- Recall that a line segment joining two points on a graph is called a *secant line*.

- Back off a bit from the point $(x_0, f(x_0))$ on the graph of f to the point $(x_0 + h_1, f(x_0 + h_1))$, where h_1 is some small number.
  - Draw the secant line ℓ_1 to the graph joining these two points, as in the following figure.

- By choosing the second point closer and closer to $(x_0, f(x_0))$, we will be drawing better and better approximations to the desired tangent line.

- Choose h_2 closer to zero than h_1 and draw the secant line ℓ_2 joining $(x_0, f(x_0))$ and $(x_0 + h_2, f(x_0 + h_2))$.

- Continue in this way choosing a sequence $\{h_n\}$ of small numbers which converges monotonically to 0.
  - For each n, draw the secant line through the two *distinct* points on the graph $(x_0, f(x_0))$ and $(x_0 + h_n, f(x_0 + h_n))$.

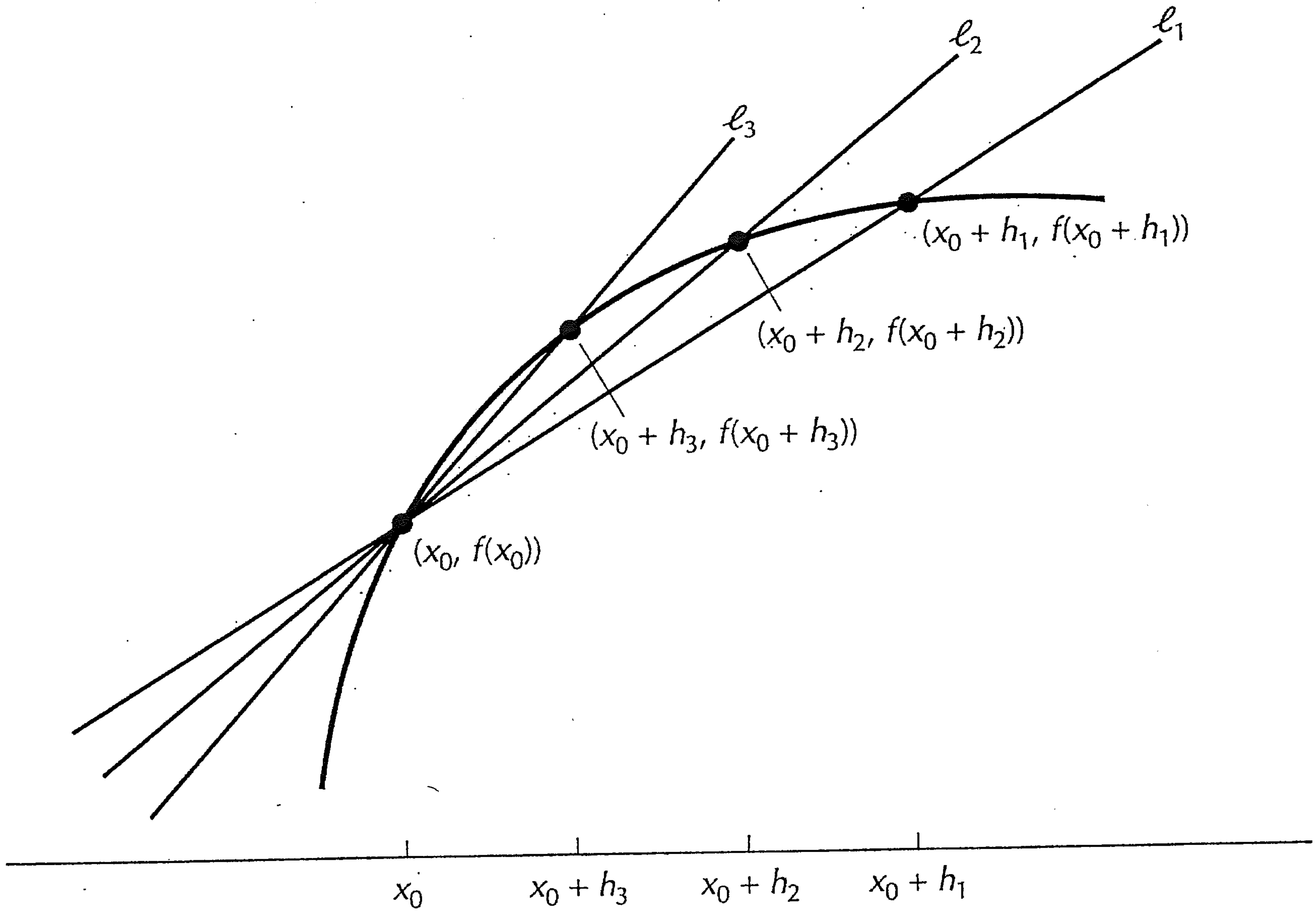


Figure 2.10 *Approximating the tangent line by a sequence of secant lines.*

- The secant lines $\{\ell_n\}$ geometrically approach the tangent line to the graph of f at $(x_0, f(x_0))$, and their slopes approach the slope of the tangent line.

- As ℓ_n passes through the two points $(x_0, f(x_0))$ and $(x_0 + h_n, f(x_0 + h_n))$, its slope is
$$\frac{f(x_0 + h_n) - f(x_0)}{(x_0 + h_n) - x_0} = \frac{f(x_0 + h_n) - f(x_0)}{h_n}.$$

- Hence the slope of the tangent line is the limit of this process as h_n converges to 0.

- **Definition:**

  Let $(x_0, f(x_0))$ be a point on the graph of $y = f(x)$. The **derivative** of f at x_0, written as $f'(x_0)$ or $\dfrac{df}{dx}(x_0)$ or $\dfrac{dy}{dx}(x_0)$, is the slope of the tangent line to the graph of f at $(x_0, f(x_0))$.

  – Analytically, $f'(x_0) = \lim\limits_{h \to 0} \dfrac{f(x_0 + h) - f(x_0)}{h}$, if this limit exists.

  – When the limit does exist, we say that the function f is **differentiable** at x_0 with derivative $f'(x_0)$.

1.2 Differentiability and Continuity

- The continuity of f is a *necessary* condition for its differentiability, but *not sufficient*.

- **Theorem 1:**

  Let f be defined on an open interval I containing the point x_0. If f is differentiable at x_0, then f is continuous at x_0.

  – Proof: To be discussed in class.

- **Counter-example:** Consider the function $f(x) = |x|$.

  – Show that this function is *continuous* at $x = 0$.

  – Show that this function is *not differentiable* at $x = 0$.

1.3 Higher-Order Derivatives

- **Continuously Differentiable Functions**:

  If f is a differentiable function, its derivative $f'(x)$ is another function of x. If $f'(x)$ is a continuous function of x, we say that the original function f is *continuously differentiable*, or C^1 for short.

  – Geometrically, the function f' will be continuous if the tangent line to the graph of f at $(x, f(x))$ changes continuously as x changes.

  – Example: Every polynomial is a continuous function. Since the derivative of a polynomial is a polynomial of one less degree, it is also continuous.

  $\Rightarrow$ Every polynomial is a C^1 function.

- **Second Derivative:**

  Let f be a C^1 function on $\mathbb{R}$. Since its derivative $f'(x)$ is a continuous function on $\mathbb{R}$, we can ask whether or not the function f' has a derivative at a point x_0.

  – The derivative of $f'(x)$ at x_0 is called the second derivative of f at x_0 and is written as

  $$f''(x_0) \text{ or } \frac{d}{dx}\left(\frac{df}{dx}\right)(x_0) = \frac{d^2 f}{dx^2}(x_0).$$

  – The second derivative of f at x_0 is defined as

  $$f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h) - f'(x_0)}{h}, \text{ if this limit exists.}$$

- **Twice Continuously Differentiable Functions:**

  If f has a second derivative everywhere, then f'' is a well-defined function of x. If f'' is a continuous function of x, then we say that f is *twice continuously differentiable*, or C^2 for short.

  – Example: Every polynomial is a C^2 function.

- This process continues for all positive integers.

  – If $f(x)$ has derivatives of order $1, 2, \ldots, k$ and if the kth derivative of f – written as $f^{[k]}(x)$, or $\dfrac{d^k f}{dx^k}(x)$ – is itself a continuous function, we say that f is C^k.

- If f has a continuous derivative of every order, that is, if f is C^k for every positive integer k, then we say that f is C^∞ or "infinitely differentiable".

  – Example: All polynomials are C^∞ functions.

1.4 Approximation by Differentials

- The definition of the derivative leads us naturally to the construction of the *linear approximation* of a function.

- Recall that for a linear function $f(x) = mx + b$, the derivative $f'(x) = m$ gives the slope of the graph of f and measures the rate of change or *marginal change* of f:

  increase in the value of f for every unit increase in the value of x.

- Let us carry over this marginal analysis to nonlinear functions.

  – After all, this was one of the main reasons for defining the derivative of such an f.

- In formulating the analytic definition of the derivative of f, we used the fact that the slope of the tangent line to the graph at $(x_0, f(x_0))$ is well approximated by the slope of the secant line through $(x_0, f(x_0))$ and a nearby point $(x_0 + h, f(x_0 + h))$ on the graph.
  – In symbols, $\dfrac{f(x_0 + h) - f(x_0)}{h} \approx f'(x_0)$, for small h, where $\approx$ means "is well approximated by" or "is close in value to".

- If we set $h = 1$, then the above relationship becomes

$$f(x_0 + 1) - f(x_0) \approx f'(x_0);$$

  – in words, the derivative of f at x_0, $f'(x_0)$, is a good approximation to the *marginal change* of f at x_0.
    - Of course, the less curved the graph of f at x_0, the better is the approximation.

- Example: Consider the production function $F(L) = \frac{1}{2}\sqrt{L}$.
  - Suppose that the firm is currently using 100 units of labour.
  - Marginal product of labour: $F'(100) = \frac{1}{4}(100)^{-\frac{1}{2}} = \frac{1}{40} = 0.025$;
  - The actual increase in output is: $F(101) - F(100) = 0.02494...$, pretty close to 0.025.
- What if the change in the amount of x is not exactly one unit?
  - Substituting Δx, the exact change in x, for h, we get

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \approx f'(x_0)\,, \text{ implying that}$$

$$\Delta y \equiv f(x_0 + \Delta x) - f(x_0) \approx f'(x_0)\Delta x,$$

$$\text{or} \qquad f(x_0 + \Delta x) \approx f(x_0) + f'(x_0)\Delta x.$$

  - Here we write Δy for the exact change in $y = f(x)$ when x changes by Δx.
    - Once again, the less curved the graph and/or the smaller the change Δx in x, the better the approximation.

- The expression $f\left(x_0+\Delta x\right) \approx f\left(x_0\right)+f'\left(x_0\right)\Delta x$ gives an effective way of approximating $f(x)$ for x close to some x_0.

- Example: Consider again the production function $y = F(x) = \frac{1}{2}\sqrt{x}$.

  Suppose the firm cuts its labour force from 900 to 896 units. Let us estimate the change in output, Δy and the new output at $x = 896$.

  – Note that $F'(x) = \frac{1}{4}x^{-\frac{1}{2}}$. Substituting $x_0 = 900$ and $\Delta x = -4$,

    ◦ $\Delta y \approx F'\left(x_0\right)\Delta x = \frac{1}{4}\left(900\right)^{-\frac{1}{2}}\left(-4\right) = -\dfrac{1}{30}$, that is, output will decrease by approximately $\dfrac{1}{30}$ units;

    ◦ New output, $F\left(x_0+\Delta x\right) \approx F\left(x_0\right)+F'\left(x_0\right)\Delta x = 15-\dfrac{1}{30} = 14.9666...$

  – The actual new output is: $F\left(896\right) = \frac{1}{2}\sqrt{896} = 14.9663...$;

    ◦ once again the approximation by the derivative is a good one.

- The equations

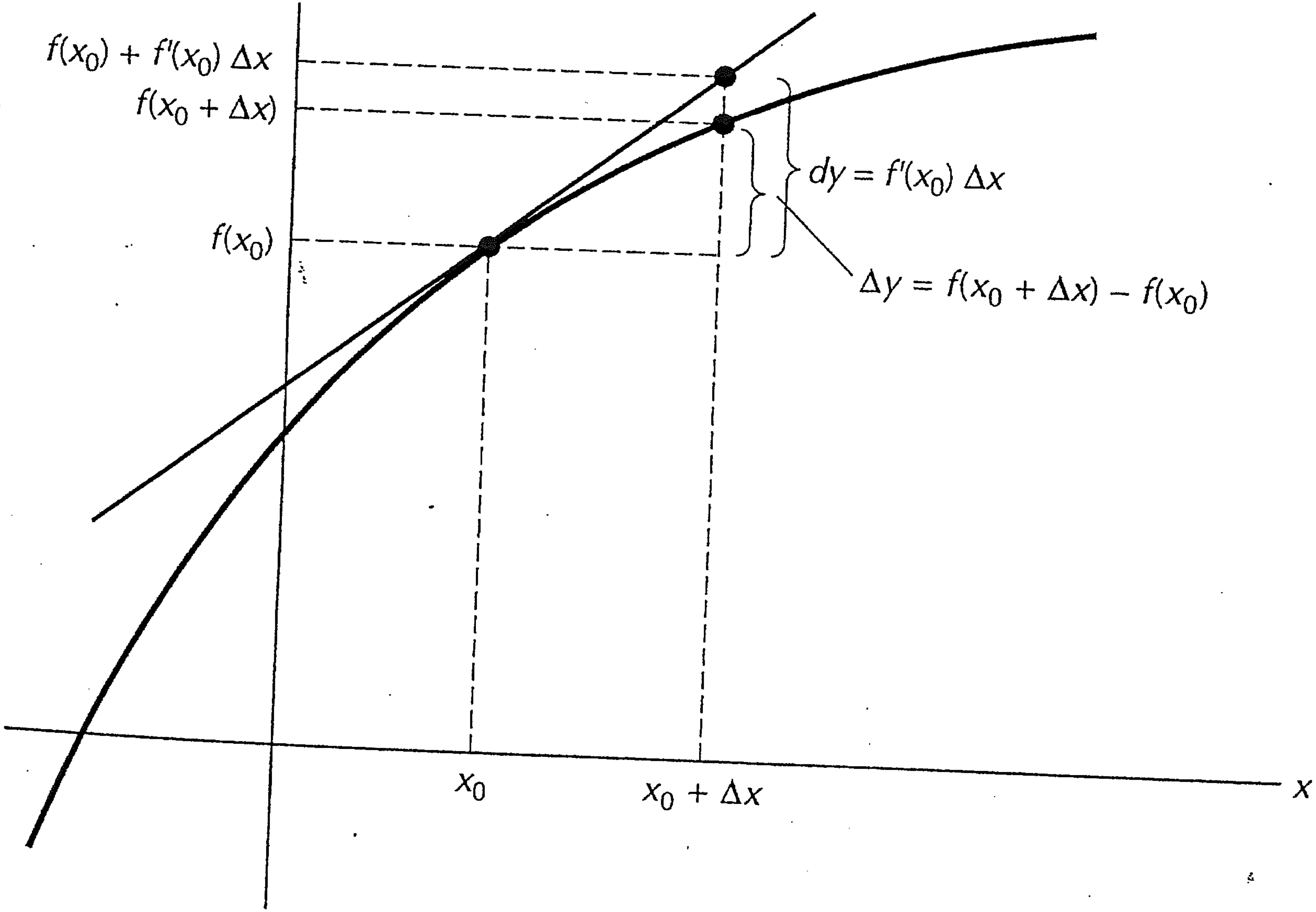
$$\Delta y \equiv f(x_0 + \Delta x) - f(x_0) \approx f'(x_0)\,\Delta x,$$

$$\text{or} \qquad f(x_0 + \Delta x) \approx f(x_0) + f'(x_0)\,\Delta x,$$

are merely analytic representations of the geometric fact that the tangent line ℓ to the graph of $y = f(x)$ at $(x_0, f(x_0))$ is a good approximation to the graph itself for x near x_0.

- As the following figure indicates,
  - the left-hand sides of the equations pertain to movement along the graph of f,
  - the right-hand sides pertain to movement along the tangent line ℓ,
    - because the equation of the tangent line, the line through the point $(x_0, f(x_0))$ with slope $f'(x_0)$, is

$$y = f(x_0) + f'(x_0)(x - x_0) = f(x_0) + f'(x_0)\,\Delta x.$$



Comparing dy and Δy.

Figure 2.15

- Write Δy for the *actual* change in y as x changes by Δx, that is, for the change along the graph of f, as in the figure.
- Write dy for change in y along the tangent line ℓ as x changes by Δx.
- Then the above equation can be written as

$$\Delta y \approx dy = f'\left(x_0\right) \Delta x.$$

- We usually write dx instead of Δx while working with changes along the tangent line, even though Δx is equal to dx.
- **Differentials:** The increments dy and dx along the tangent line ℓ are called *differentials*.
- We sometimes write the differential df in place of dy.
- The equation of differentials for the variation along the tangent line

$$df = f'\left(x_0\right) dx \qquad \text{or} \qquad dy = f'\left(x_0\right) dx$$

gives added weight to the notation $\dfrac{df}{dx}$ for the derivative $f'(x)$.

2. Calculus of Several Variables: Partial Derivatives

- To apply calculus to the study of functions of several variables, we take the simplest approach.
  - We change one variable at a time, keeping all the other variables constant.
  - Since we are not looking at the total variation of f but just the partial variation – the variation brought about by the change in only one variable, say x_i – the corresponding derivative is called the *partial derivative* of f with respect to x_i.
    - It is denoted by $\dfrac{\partial f}{\partial x_i}$; other common notations include f_i, f_{x_i}, and $D_i f$.
- Recall that the derivative of a function f of one variable at x_0 is

$$\frac{df}{dx}(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

- The partial derivative with respect to x_i of a function of several variables, $f(x_1, x_2, ..., x_n)$, at the vector $x_0 = \left(x_1^0, x_2^0, ..., x_n^0\right)$ is defined in a similar manner.

- **Definition:**

  Let A be an open set in $\mathbb{R}^n$, $x_0 = \left(x_1^0, x_2^0, ..., x_n^0\right) \in A$, and $f : A \to \mathbb{R}$. Then

$$\frac{\partial f}{\partial x_i}\left(x_1^0, x_2^0, ..., x_n^0\right) = \lim_{h \to 0} \frac{f\left(x_1^0, x_2^0, ..., x_i^0 + h, ..., x_n^0\right) - f\left(x_1^0, x_2^0, ..., x_i^0, ..., x_n^0\right)}{h},$$

  if this limit exists, where $i = 1, 2, ..., n$.

- Note that only ith variable changes; others are treated as constants.
  - This means, of course, that we can compute partial derivatives just like ordinary derivatives of a function of one variable.
    - That is, if $f(x_1, x_2, ..., x_n)$ is given by some formula involving $(x_1, x_2, ..., x_n)$, then we find $D_i f(x)$ by differentiating the function whose value at x_i is given by the formula when all x_j (for $j \neq i$) are thought of as constants.

- Example: Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x_1, x_2) = x_1^3 + 3x_2^3 + 2x_1x_2$. Then
  - $D_1 f(x) = 3x_1^2 + 2x_2$,
  - $D_2 f(x) = 9x_2^2 + 2x_1$.

2.1 Geometric Interpretation of Partial Derivatives

- Note that we can represent a function $z = f(x, y)$ geometrically by drawing its graph in $\mathbb{R}^3$.
- When we study $\dfrac{\partial f}{\partial x}(a, b)$, we are holding y constant at b and looking at variations in x around $x = a$.
  - In terms of the graph, we are looking at f only on the two-dimensional slice $\{y = b\}$ in $\mathbb{R}^3$ as in the following figure.
  - On this slice, the graph of f is a curve – the graph of the function of *one* variable $x \mapsto f(x, b)$.
    - The partial derivative $\dfrac{\partial f}{\partial x}(a, b)$ is the slope of the tangent line to this graph on this slice, line ℓ in the figure.
- Similarly, $\dfrac{\partial f}{\partial y}(a, b)$ is the slope of the tangent line to the curve which is the intersection of the graph of f with the slice $\{x = a\}$, as illustrated in the next figure.

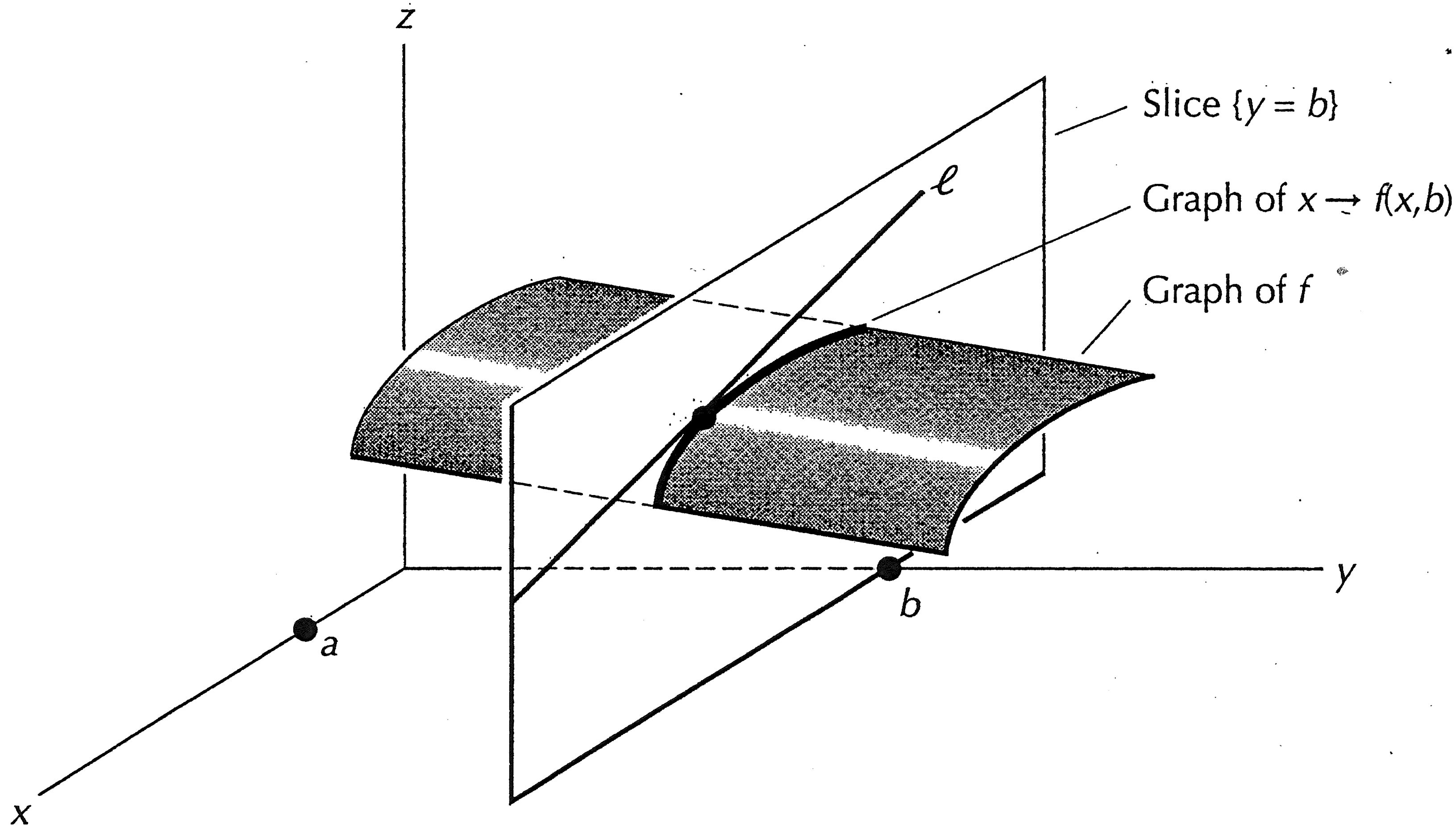


Figure 14.1 *The graph of $x \longmapsto f(x, b)$ on the slice $\{y = b\}$.*

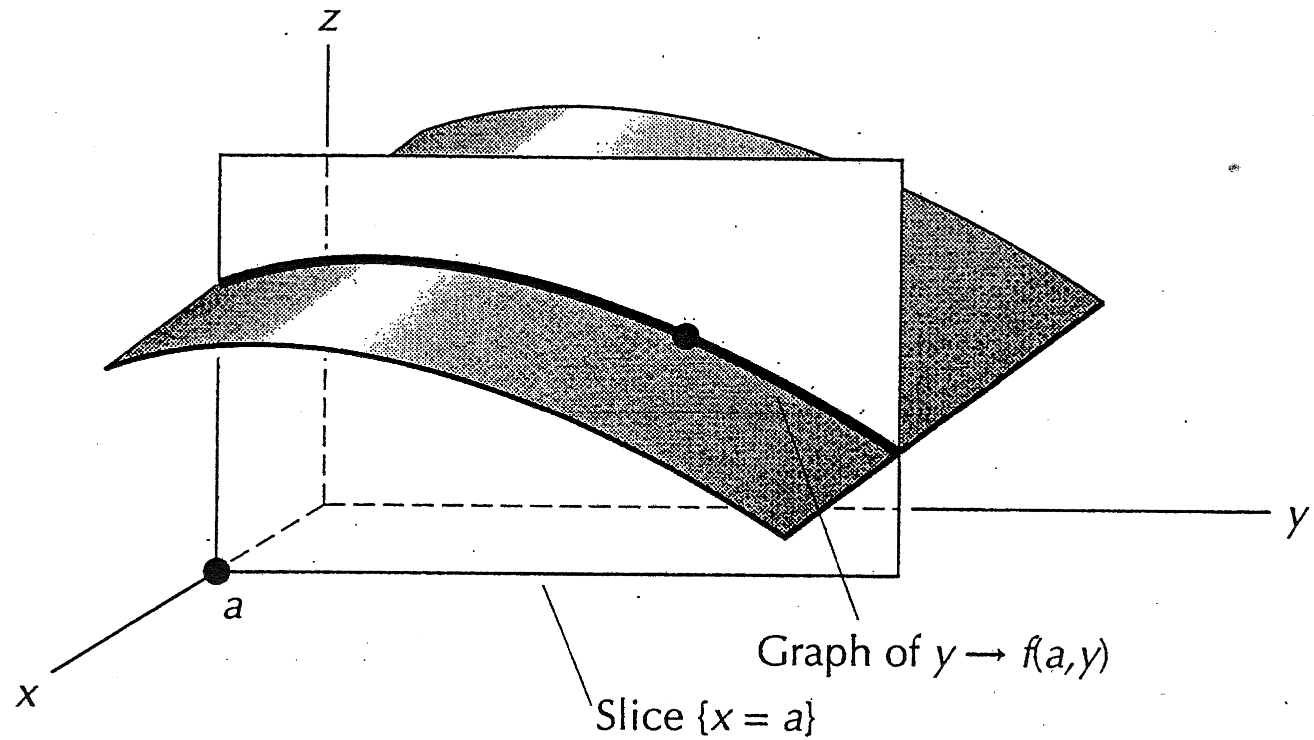


Figure 14.2

The graph of $y \longmapsto f(a, y)$ on the slice $\{x = a\}$.

2.2 The Total Derivative

- Suppose we are interested in the behaviour of a function $F(x, y)$ in the neighbourhood of a given point (x^*, y^*).
  - Calculus of one variable and the concept of partial derivative tell us that
    - if we hold y fixed at y^* and change x^* to $x^* + \Delta x$, then

$$F\left(x^* + \Delta x, y^*\right) - F(x^*, y^*) \approx \frac{\partial F}{\partial x}(x^*, y^*) \cdot \Delta x;$$

    - if we hold x fixed at x^* and change y^* to $y^* + \Delta y$, then

$$F\left(x^*, y^* + \Delta y\right) - F(x^*, y^*) \approx \frac{\partial F}{\partial y}(x^*, y^*) \cdot \Delta y.$$

- What if we allow both x and y to vary simultaneously?
  - Since we are working in the realm of linear approximations, it is natural that the effect of the combined change is roughly the *sum* of the effects of the one-variable

changes:

$$F\left(x^{*}+\Delta x, y^{*}+\Delta y\right)-F(x^{*}, y^{*}) \approx \frac{\partial F}{\partial x}(x^{*}, y^{*}) \cdot \Delta x+\frac{\partial F}{\partial y}(x^{*}, y^{*}) \cdot \Delta y.$$

– We sometimes use the above expression in the following form:

$$F\left(x^{*}+\Delta x, y^{*}+\Delta y\right) \approx F(x^{*}, y^{*})+\frac{\partial F}{\partial x}(x^{*}, y^{*}) \cdot \Delta x+\frac{\partial F}{\partial y}(x^{*}, y^{*}) \cdot \Delta y.$$

- Example: Consider a production function $Q = F(K, L)$ = $4K^{\frac{3}{4}}L^{\frac{1}{4}}$ around the point $(K^{*}, L^{*}) = (10000, 625)$.
  – $Q = F(10000, 625) = 20,000;$ $\quad \dfrac{\partial Q}{\partial K}(10000, 625) = 1.5;$ $\quad \dfrac{\partial Q}{\partial L}(10000, 625) = 8.$
  – If L is held constant and K increased by ΔK , Q will increase by approximately $1.5 \cdot \Delta K$.
    - For an increase in K by 10 units, we estimate

$$Q(10010, 625) = 20,000 + 1.5 \cdot 10 = 20,015,$$

a good approximation to $Q(10010, 625) = 20,014.998$ ... .

– Similarly, a 2-unit decrease in L should induce a $2 \cdot 8 = 16$-unit decrease in Q.

- Consequently, for a 2-unit decrease in L we estimate

$$Q(10000, 623) = 20,000 - 8 \cdot 2 = 19,984,$$

a good approximation to $Q(10000, 623) = 19,983.981$ ... .

– Finally, if we want to consider the effect of both changes, we would use the above expression to estimate

$$\begin{aligned} F(10010, 623) &\approx F(10000, 625) + \frac{\partial F}{\partial K}(10000, 625) \cdot 10 + \frac{\partial F}{\partial L}(10000, 625) \cdot (-2) \\ &= 20,000 + (1.5 \cdot 10) + (8 \cdot (-2)) \\ &= 19,999, \end{aligned}$$

which compares well with the exact value $19,998.967...$.

2.2.1 Geometric Interpretation of Total Derivative

- What is the geometric significance of the approximation

$$F\left(x^{*}+\Delta x, y^{*}+\Delta y\right) \approx F(x^{*}, y^{*})+\frac{\partial F}{\partial x}(x^{*}, y^{*}) \cdot \Delta x+\frac{\partial F}{\partial y}(x^{*}, y^{*}) \cdot \Delta y?$$

- For a function of one variable, the corresponding approximation is

$$f\left(x^{*}+h\right) \approx f\left(x^{*}\right)+f^{\prime}\left(x^{*}\right) \cdot h.$$

  - As discussed above, the right-hand side is the equation of the tangent line to the graph of f at x^{*}.
  - That is, this equation states that the tangent line to the graph of f at x^{*} is a good approximation to the graph itself in the vicinity of $(x^{*}, f(x^{*}))$, as illustrated in the following figure.

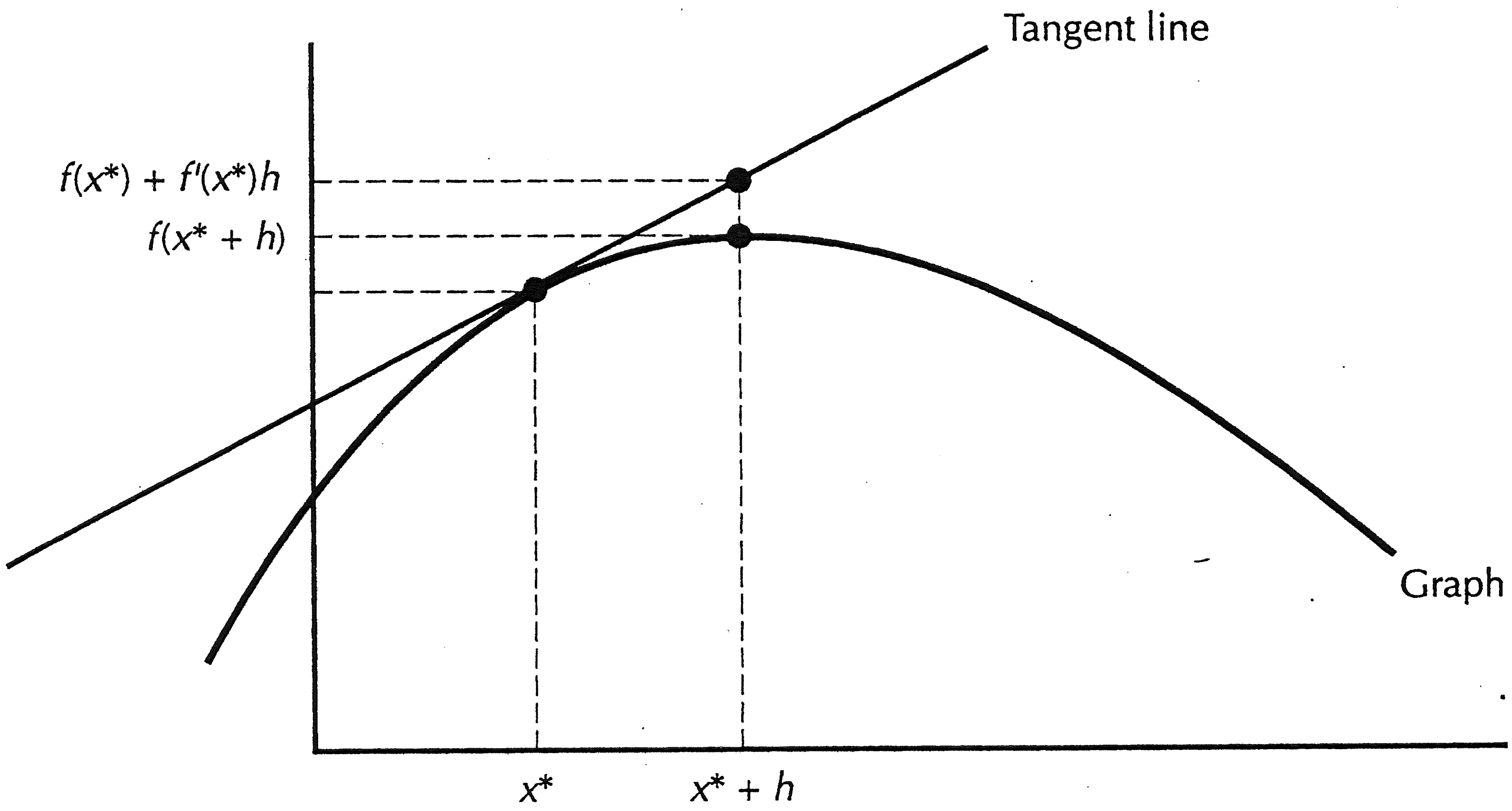


Figure 14.4

The geometry of expression (8).

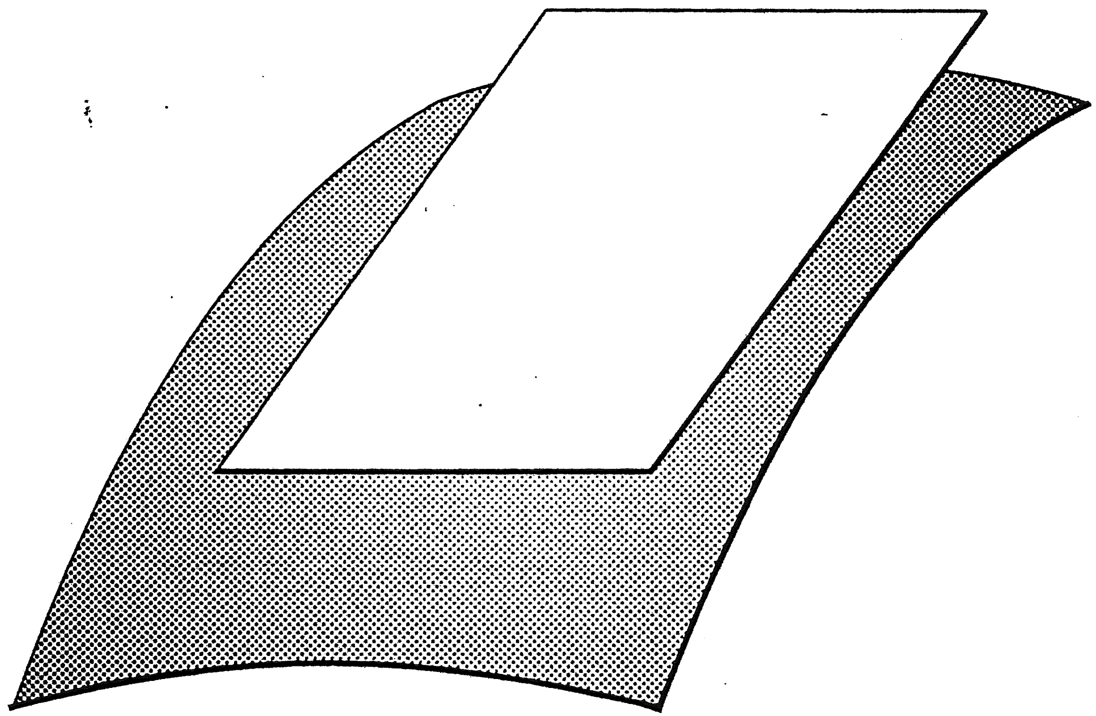
- For a function $z = F(x, y)$ of two variables, the analogue of the tangent line is the *tangent plane* to the graph, as illustrated in the following figure.
  - That is, the expression

$$F\left(x^{*}+\Delta x, y^{*}+\Delta y\right) \approx F(x^{*}, y^{*})+\frac{\partial F}{\partial x}(x^{*}, y^{*}) \cdot \Delta x+\frac{\partial F}{\partial y}(x^{*}, y^{*}) \cdot \Delta y$$

  says that the tangent plane to the graph of F at $(x^{*}, y^{*}, F(x^{*}, y^{*}))$ is a good approximation to the graph itself in the vicinity of $(x^{*}, y^{*}, F(x^{*}, y^{*}))$.

$\Rightarrow$ The change $F(x^{*}+s, y^{*}+t)-F(x^{*}, y^{*})$ can be well approximated by the *linear mapping*

$$(s, t) \mapsto \frac{\partial F}{\partial x}(x^{*}, y^{*}) \cdot s+\frac{\partial F}{\partial y}(x^{*}, y^{*}) \cdot t.$$



The tangent plane to the graph of F.

Figure 14.5

- Thus, we consider the vector $\left(\frac{\partial F}{\partial x}(x^*, y^*), \frac{\partial F}{\partial y}(x^*, y^*)\right)$ as representing the linear approximation of F around (x^*, y^*).
  - In this sense, we call this linear map and the vector which represents it the **derivative** of F at (x^*, y^*) and write it as

$$DF(x^*, y^*) = \nabla F(x^*, y^*) = \left(\frac{\partial F}{\partial x}(x^*, y^*), \frac{\partial F}{\partial y}(x^*, y^*)\right).$$

  - This vector is also called the **gradient vector** of F at (x^*, y^*).

- It is rather natural to form a vector whose entries are all the partial derivatives of F and call it the derivative of F.
  - But, it is important to realize that more is happening here since the linear mapping which this gradient vector represents is the appropriate **linear approximation** of F at (x^*, y^*).
    - As mentioned above, we use this approximation all the time when we use *linear* mathematics for major construction projects on our *round* earth.

- We use the notations dx, dy, and dF when we are working on the tangent plane to the graph of F at (x^*, y^*).
  - These variations on the tangent plane are called **differentials**.

- Using the differentials, the above expression is written as

$$dF = \frac{\partial F}{\partial x}(x^*, y^*) \cdot dx + \frac{\partial F}{\partial y}(x^*, y^*) \cdot dy.$$

  - This expression for dF in terms of dx and dy is called the **total differential** of F at (x^*, y^*).

2.2.2 Functions of More than Two Variables

- The observations and analytical expressions for functions on $\mathbb{R}^2$ carry over in a natural way to functions on $\mathbb{R}^n$.

- If we are studying a function $F(x_1, x_2, ..., x_n)$ of n variables in a neighbourhood of some selected point $x^* = (x_1^*, x_2^*, ..., x_n^*)$, then

$$F\left(x_1^* + \Delta x_1,\ ...,\ x_n^* + \Delta x_n\right)$$
$$\approx F\left(x_1^*, x_2^*, ..., x_n^*\right) + \frac{\partial F}{\partial x_1}(x^*) \cdot \Delta x_1 + \ ... \ + \frac{\partial F}{\partial x_n}(x^*) \cdot \Delta x_n.$$

  – The right-hand side is the representation of the (n-dimensional) *tangent hyperplane* to the graph of F.

- The above equation says that the linear mapping

$$(h_1, h_2, ..., h_n) \mapsto \frac{\partial F}{\partial x_1}(x^*) \cdot h_1 + \ ... \ + \frac{\partial F}{\partial x_n}(x^*) \cdot h_n$$

  is a good approximation to the actual change in F.

- We call this linear map and the vector which represents it the **derivative** of F at x^* and write it as

$$DF(x^*) = \nabla F(x^*) = \left(\frac{\partial F}{\partial x_1}(x^*), \; ..., \; \frac{\partial F}{\partial x_n}(x^*) \right).$$

  – This vector is also called the **gradient vector** of F at x^*.

- We frequently use the **differentials** dF, dx_1, dx_2, ..., dx_n to denote changes on the tangent hyperplane.

- The above expression says that in the vicinity of the point x^*, the tangent hyperplane to the graph of F is a good approximation to the graph itself in that the actual change $\Delta F = F(x^* + \Delta x) - F(x^*)$ is well approximated by the **total differential**

$$dF = \frac{\partial F}{\partial x_1}(x^*) \cdot dx_1 + \; ... \; + \frac{\partial F}{\partial x_n}(x^*) \cdot dx_n$$

  on the tangent hyperplane with $dx_i = \Delta x_i$ for all i.

2.3 Explicit Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

- Up to now we have been studying functions with only one endogenous variable. But functions with several endogenous variables arise naturally in economic models.

- Most real-world firms produce more than one product. To model their production, we need a production function for each product.
  - If the firm uses three inputs to produce two outputs, we need two separate production functions: $q_1 = f^1(x_1, x_2, x_3)$, and $q_2 = f^2(x_1, x_2, x_3)$.
  - We can write $q = (q_1, q_2)$ as an output vector for this firm and summarize the firm's activities by a function $F = (f^1, f^2)$:
$$q = (q_1, q_2) = \left(f^1(x_1, x_2, x_3), f^2(x_1, x_2, x_3)\right) \equiv F(x_1, x_2, x_3).$$
  - The domain of F lies in $\mathbb{R}^3$, and its target space is $\mathbb{R}^2$. We write $F : \mathbb{R}^3 \to \mathbb{R}^2$.

- A firm which uses n inputs to produce m products would have a production function $F : \mathbb{R}^n \to \mathbb{R}^m$.

- When there are m endogenous variables in the system, there should be m separate functions to determine their values:

$$\begin{aligned} y_1 &= f^1(x_1, ..., x_n), \\ y_2 &= f^2(x_1, ..., x_n), \\ &\vdots \\ y_m &= f^m(x_1, ..., x_n). \end{aligned}$$

- Instead of viewing the system as m functions of n variables, we can view it as a single function from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$F(x_1, ..., x_n) = \left(f^1(x_1, ..., x_n), f^2(x_1, ..., x_n), \; ..., \; f^m(x_1, ..., x_n)\right).$$

- On the other hand, if we start with a single function $F : \mathbb{R}^n \to \mathbb{R}^m$ as above, we see that each component of F, $f^i(x_1, ..., x_n)$, is a function from $\mathbb{R}^n$ to $\mathbb{R}$.

- This method is especially useful when we develop and use the theorems of calculus for a function from $\mathbb{R}^n$ to $\mathbb{R}^m$.
  - We apply the usual theory to each component function $f^i : \mathbb{R}^n \to \mathbb{R}$ separately and then combine all the information learned so far back into one big vector or matrix.

- Consider $F = \left(f^1, f^2,\ ...,\ f^m\right) : \mathbb{R}^n \to \mathbb{R}^m$ at a specific point $x^* = (x_1^*, x_2^*, ..., x_n^*)$ in $\mathbb{R}^n$ and use approximation by differentials to estimate the effect of a change at x^* by $\Delta x = (\Delta x_1, \Delta x_2, ..., \Delta x_n)$.

- We first apply the approximation by differentials to each component f^i of F, $i = 1, 2, \ldots, m$:

$$f^1(x^* + \Delta x) - f^1(x^*) \approx \frac{\partial f^1}{\partial x_1}(x^*) \cdot \Delta x_1 + \ ...\ + \frac{\partial f^1}{\partial x_n}(x^*) \cdot \Delta x_n,$$

$$f^2(x^* + \Delta x) - f^2(x^*) \approx \frac{\partial f^2}{\partial x_1}(x^*) \cdot \Delta x_1 + \ ...\ + \frac{\partial f^2}{\partial x_n}(x^*) \cdot \Delta x_n,$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$f^m(x^* + \Delta x) - f^m(x^*) \approx \frac{\partial f^m}{\partial x_1}(x^*) \cdot \Delta x_1 + \ ...\ + \frac{\partial f^m}{\partial x_n}(x^*) \cdot \Delta x_n.$$

- Then we use vector and matrix notation to combine these and get

$$F(x^* + \Delta x) - F(x^*) \approx \begin{pmatrix} \frac{\partial f^1}{\partial x_1}(x^*) & \cdots & \frac{\partial f^1}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x_1}(x^*) & \cdots & \frac{\partial f^m}{\partial x_n}(x^*) \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix}.$$

  – This expression describes the linear approximation of F at x^*.

- We write the matrix on the right-hand side as

$$DF(x^*) = DF_{x^*} = \begin{pmatrix} \frac{\partial f^1}{\partial x_1}(x^*) & \cdots & \frac{\partial f^1}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x_1}(x^*) & \cdots & \frac{\partial f^m}{\partial x_n}(x^*) \end{pmatrix},$$

  and call it the **derivative** or the **Jacobian derivative** of F at x^*.

  – This matrix, also referred to as the **Jacobian matrix**, is the natural generalization of the *gradient vector* of a single endogenous variable to m endogenous variables.

- As we emphasized in the last subsection, it is natural to form a matrix composed of all the first order partial derivatives of the component functions, f^is, and call it the derivative of F.
  - But more is happening here.
    - The above expression says that the linear map which this matrix represents is the effective linear approximation of F around x^*.

- This is the essence of what calculus is all about.

  In studying the behaviour of a nonlinear function $F : \mathbb{R}^n \to \mathbb{R}^m$ in the vicinity of some specific point x^*:

(1) we use derivatives to form the linear approximation $DF(x^*)$,

(2) we use linear algebra to study the behaviour of the linear mapping $DF(x^*)$, and

(3) we use calculus theory to translate information about the linear function $DF(x^*)$ to the corresponding information about the nonlinear function F in a neighbourhood of x^*.

2.4 Higher Order Derivatives

- Let A be an open set in $\mathbb{R}^n$, and $f : A \to \mathbb{R}$.

- Example: Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x_1, x_2) = x_1^3 + 3x_2^3 + 2x_1 x_2$. Then
  - $D_1 f(x) = 3x_1^2 + 2x_2$,
  - $D_2 f(x) = 9x_2^2 + 2x_1$.
  - In this example the gradient vector is

$$\nabla f(x) = \left(3x_1^2 + 2x_2,\ 9x_2^2 + 2x_1\right), \text{ for all } x \in \mathbb{R}^2.$$

- When $f : A \to \mathbb{R}$ has (first-order) partial derivatives at each $x \in A$, we say that f has (first-order) *partial derivatives on* A.

- **Second-Order Partial Derivatives and the Hessian Matrix:**

  – When $f : A \to \mathbb{R}$ has (first-order) partial derivatives at each x on A, these first-order partial derivatives are themselves functions from A to $\mathbb{R}$.

  - If these (first-order) partial derivatives are continuous on A, then we say that f is *continuously differentiable* $\left(C^1\right)$ *on* A.

  - If these functions have (first-order) partial derivatives on A, these partial derivatives are called the *second-order partial derivatives of* f *on* A.

  – To elaborate, if $D_i f\,(x)$ exists for all $x \in A$, we can define the function $D_i f : A \to \mathbb{R}$.

  - If this function has (first-order) partial derivatives on A, then the j-th (first-order) partial derivative of $D_i f$ at x (that is, $D_j\left(D_i f\,(x)\right)$) is a second-order partial derivative of f at x, and is denoted by $D_{ij} f\,(x)$. [Here $i = 1, 2, ..., n$, and $j = 1, 2, ..., n$.]

  - In the example described above,

$$\begin{aligned} D_{11} f\,(x) &= 6x_1, \; D_{22} f\,(x) = 18x_2, \\ D_{12} f\,(x) &= 2 = D_{21} f\,(x)\,. \end{aligned}$$

- We note in this example that the "cross partials" $D_{12}f(x)$ and $D_{21}f(x)$ are equal. This is not a coincidence; it is a more general phenomenon as noted in the following Theorem, known as "Young's Theorem".

- **Theorem 2 (Young's Theorem):**

  Suppose A is an open set in $\mathbb{R}^n$, and f has first and second-order partial derivatives on A. If $D_{ij}f$ and $D_{ji}f$ are continuous on A, then $D_{ij}f(x) = D_{ji}f(x)$, for all $x \in A$.

- When all the hypotheses of Theorem 1 hold for all $i = 1, 2, ..., n$, and $j = 1, 2, ..., n$, we will say that f is *twice continuously differentiable* $\left(C^2\right)$ *on A.*

  – This will be the typical situation in many applications.

- When the first and second-order partial derivatives of $f : A \to \mathbb{R}$ exist on A, the $n \times n$ matrix of second-order partial derivatives of f described below

$$H_f(x) = D^2 f(x) = \begin{pmatrix} D_{11}f(x) & D_{12}f(x) & \cdots & D_{1n}f(x) \\ D_{21}f(x) & D_{22}f(x) & \cdots & D_{2n}f(x) \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}f(x) & D_{n2}f(x) & \cdots & D_{nn}f(x) \end{pmatrix}$$

is called the Hessian matrix of f at $x \in A$, and is denoted by $H_f(x)$ or $D^2 f(x)$.

  – When f is twice continuously differentiable on A, the Hessian matrix of f is symmetric at all $x \in A$.

  – In the example described above, the Hessian matrix of f for all $(x_1, x_2) \in \mathbb{R}^2$ is

$$H_f(x) = \begin{pmatrix} 6x_1 & 2 \\ 2 & 18x_2 \end{pmatrix}.$$

3. Composite Functions and the Chain Rule

- Let $g : A \rightarrow \mathbb{R}^m$ be a function with component functions $g^i : A \rightarrow \mathbb{R}$ ($i = 1, 2, ..., m$) which are defined on an open set $A \subset \mathbb{R}^n$, and let $f : B \rightarrow \mathbb{R}$ be a function defined on an open set $B \subset \mathbb{R}^m$ which contains the set $g(A)$.
  - Then we can define $F : A \rightarrow \mathbb{R}$ by $F(x) = f(g(x)) = f\left(g^1(x), g^2(x), ..., g^m(x)\right)$ for each $x \in A$.
    - This is known as a *composite function* (of f and g).

- The "Chain Rule" of differentiation provides us with a formula for finding the partial derivatives of a composite function, F, in terms of the partial derivatives of the individual functions, f and g.

- **Theorem 3 (Chain Rule):**

Let $g : A \to \mathbb{R}^m$ *be a function with component functions* $g^i : A \to \mathbb{R}$ ($i = 1, 2, ..., m$) *which are continuously differentiable on an open set* $A \subset \mathbb{R}^n$. *Let* $f : B \to \mathbb{R}$ *be a continuously differentiable function on an open set* $B \subset \mathbb{R}^m$ *which contains the set* $g(A)$. *If* $F : A \to \mathbb{R}$ *is defined by* $F(x) = f(g(x))$ *on* A, *and* $a \in A$, *then* F *is differentiable at* a *and we have*, *for* $i = 1, 2, ..., n$,

$$D_i F(a) = \sum_{j=1}^{m} D_j f\left(g^1(a), g^2(a), ..., g^m(a)\right) D_i g^j(a).$$

4. Homogeneous Functions and Euler's Theorem

- A function $f : \mathbb{R}^n_+ \to \mathbb{R}$ is homogeneous of degree r on $\mathbb{R}^n_+$ if for all x in $\mathbb{R}^n_+$, and all $t > 0$, $f(tx) = t^r f(x)$.
  - Example: Consider $f : \mathbb{R}^2_+ \to \mathbb{R}$ given by $f(x_1, x_2) = x_1^a x_2^b$, where $a > 0$ and $b > 0$. Then, if $t > 0$, we have $f(tx_1, tx_2) = (tx_1)^a (tx_2)^b = t^{a+b} x_1^a x_2^b = t^{a+b} f(x_1, x_2)$. So f is homogeneous of degree $(a + b)$.

- **Theorem 4:**

  Suppose $f : \mathbb{R}^n_+ \to \mathbb{R}$ is homogeneous of degree r on $\mathbb{R}^n_+$, and continuously differentiable on $\mathbb{R}^n_{++}$. Then, for each $i = 1, 2, ..., n$, $D_i f$ is homogeneous of degree $r - 1$ on $\mathbb{R}^n_{++}$.

  - Proof: To be discussed in class.

- **Theorem 5 (Euler's Theorem):**

  Suppose $f : \mathbb{R}^n_+ \to \mathbb{R}$ is homogeneous of degree r on $\mathbb{R}^n_+$, and continuously differentiable on $\mathbb{R}^n_{++}$. Then $x \cdot \nabla f(x) = r f(x)$, for all $x \in \mathbb{R}^n_{++}$.

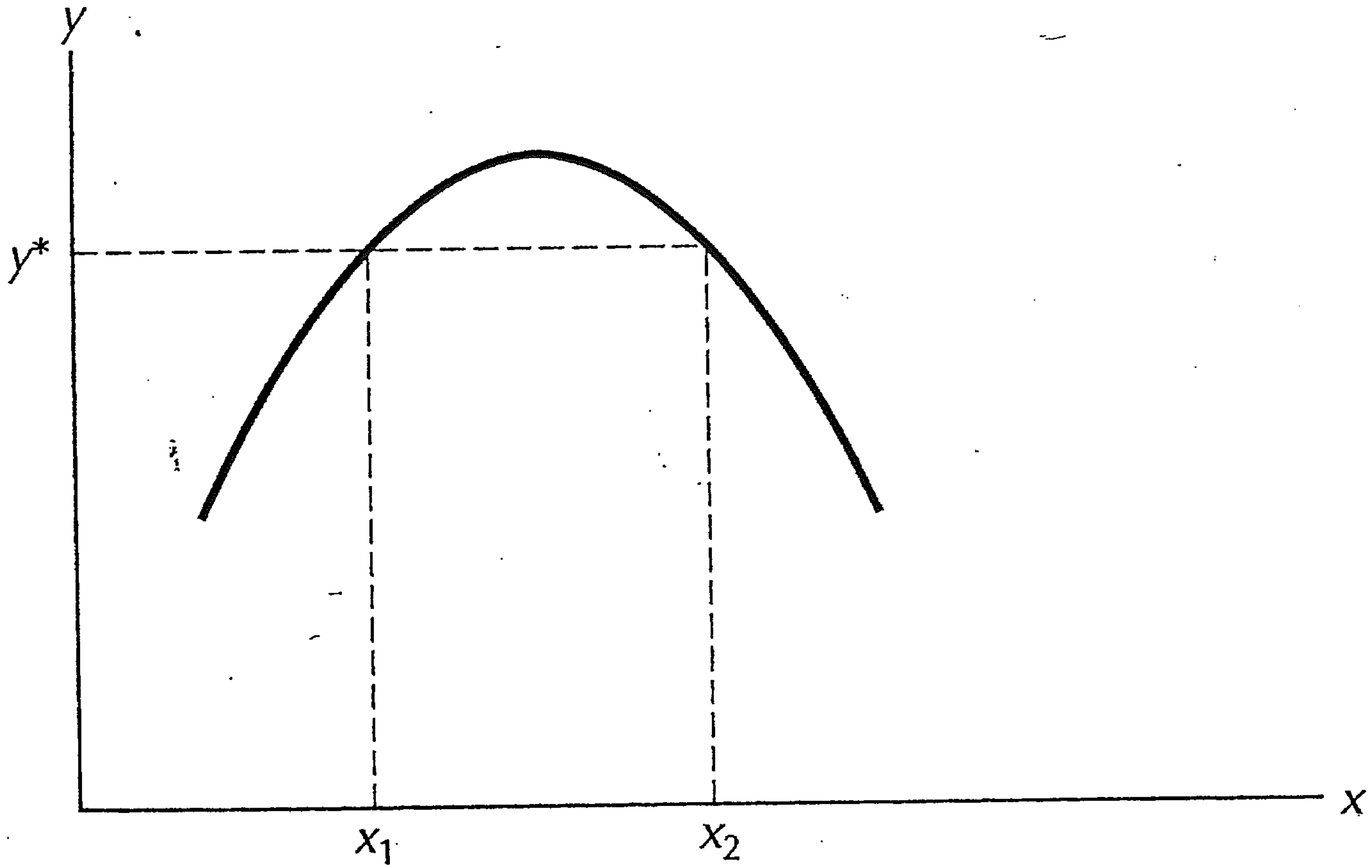
  - Proof: To be discussed in class.

5. Inverse Functions

- Let A be a set in $\mathbb{R}^n$, and let f be a function from A to $\mathbb{R}^n$.
  - f is **one-to-one** on A if whenever x^1, $x^2 \in A$ and $x^1 \neq x^2$, we have $f\left(x^1\right) \neq f\left(x^2\right)$.
  - If there is a function g, from $f(A)$ to A, such that $g(f(x)) = x$ for each $x \in A$, then g is called the **inverse function** of f on $f(A)$.
    - Notation: We often write f^{-1} for the inverse function of f.

- **Example:** Consider $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 2x$.
  - Note that f is one-to-one on $\mathbb{R}$.
  - Also, we can define the function $g : \mathbb{R} \to \mathbb{R}$ by $g(y) = \dfrac{y}{2}$, and note that it has the property $g(f(x)) = x$. Hence g is then the inverse function of f on $\mathbb{R}$.
  - Furthermore, $g'(f(x)) = \dfrac{1}{2} = \dfrac{1}{f'(x)}$, for all $x \in \mathbb{R}$.

- For functions of a single variable, it is easy to look at the graph of the function defined on an interval E and determine whether or not the function is one-to-one on E.
  - As the following figure illustrates, the graph of f cannot turn around; i.e., it cannot have any *local maxima* or *minima* on E.
  - It must be *monotonically increasing* or *monotonically decreasing* on E.
  - The function whose graph is pictured in the following figure is not one-to-one because two points x_1 and x_2 map to the same point y^*.
- Example: Consider the function $f(x) = x^2$.
  - As a function defined on the entire real line $\mathbb{R}$, f is *not* one-to-one. Why?
  - However, if we restrict the domain of f to be $[0, \infty)$, then the restricted f is one-to-one and has a well-defined inverse $g(y) = \sqrt{y}$.



A function is not one-to-one in an interval containing a local max or min.

Figure 4.1

- Example: Consider the function $f(x) = x^3 - 3x$.
  - Look at its graph in the following figure.
  - f is not one-to-one on the entire real line $\mathbb{R}$.
  - f has two local extrema, so it is *not* a monotone function.
  - However, since f is monotone for $x > 1$, its restriction to $(1, \infty)$ is invertible.
- The following theorem summarizes the discussion thus far for functions of a single variable.
- **Theorem 6:**

  A function f defined on an interval E in $\mathbb{R}$ has a well-defined inverse on the interval $f(E)$ if and only if f is monotonically increasing on all of E or monotonically decreasing on all of E.

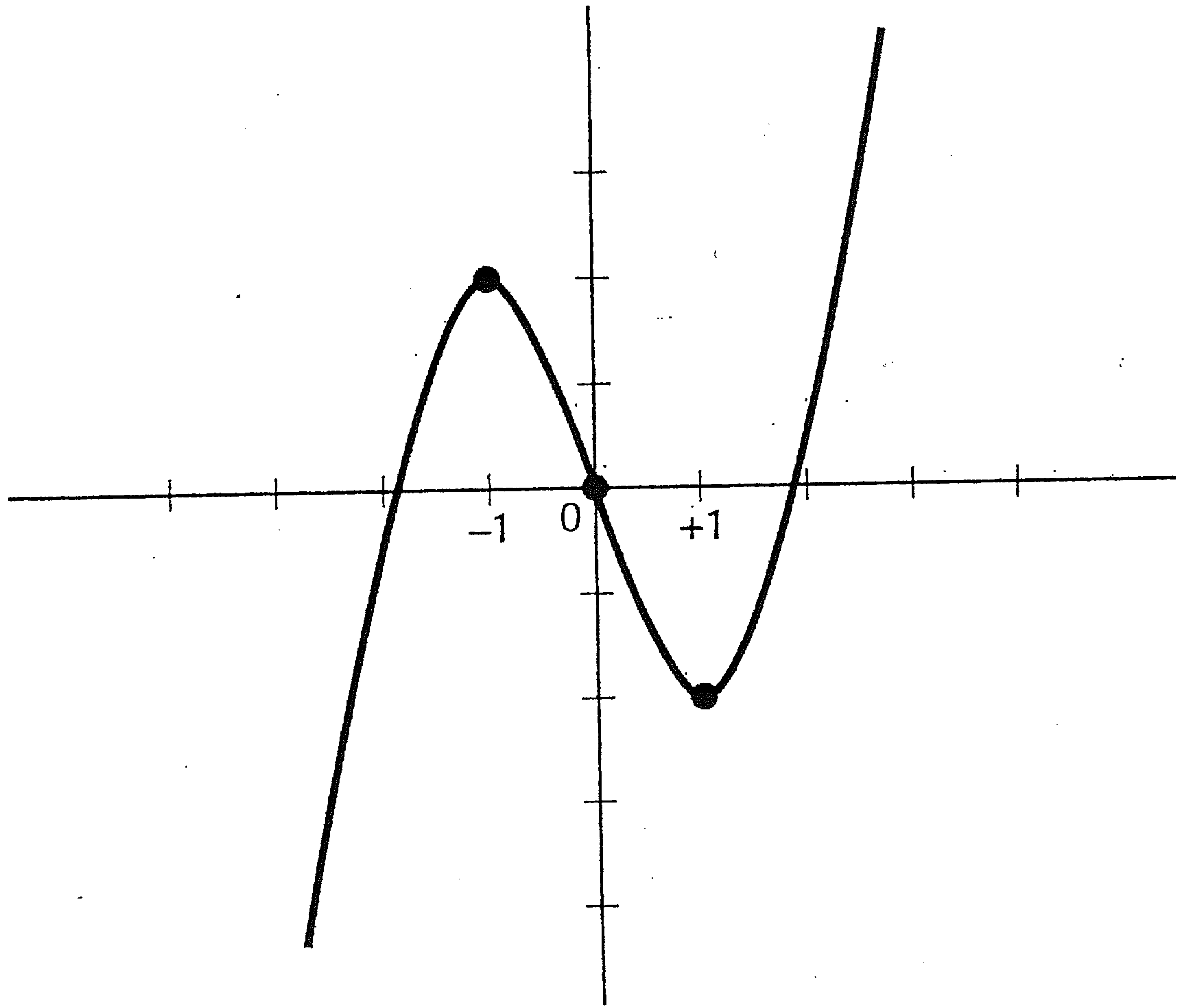


Figure 3.3

The graph of $f(x) = x^3 - 3x$.

- **Calculus Criterion for a Single-Variable Function to be Monotonically Increasing or Decreasing:**
  - f is an *increasing function* if $x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$;
  - f is a *decreasing function* if $x_1 > x_2 \Rightarrow f(x_1) < f(x_2)$.
  - **Theorem 7:**

    Let f be a continuously differentiable function on domain $D \subset \mathbb{R}$.

    If $f' > 0$ on interval $(a, b) \subset D$, then f is increasing on (a, b).

    If $f' < 0$ on interval $(a, b) \subset D$, then f is decreasing on (a, b).

    If f is increasing on (a, b), then $f' \geq 0$ on (a, b).

    If f is decreasing on (a, b), then $f' \leq 0$ on (a, b).

    \- Proof: To be discussed in class.

- Combining Theorems 6 and 7 we get:

  Theorem 8:

  A C^1 function f defined on an interval E in $\mathbb{R}$ is one-to-one and therefore invertible on E if either $f'(x) > 0$ for all $x \in E$ or $f'(x) < 0$ for all $x \in E$.

- Let E be an open interval in $\mathbb{R}$, and $f : E \to \mathbb{R}$ be continuously differentiable on E. Let $a \in E$, and suppose that $f'(a) \neq 0$.

  – Let $\boldsymbol{f'(a) > 0}$.

  $\Rightarrow$ Since f' is continuous, there is an open ball $B_\epsilon(a)$ such that $f'(x) > 0$ for all x in $B_\epsilon(a)$.

  $\Rightarrow$ f is increasing in $B_\epsilon(a)$.

  - Thus, for every $y \in f(B_\epsilon(a))$, there is a unique x in $B_\epsilon(a)$ such that $f(x) = y$.

  - That is, there is a unique function $g : f(B_\epsilon(a)) \to B_\epsilon(a)$ such that $g(f(x)) = x$ for all $x \in B_\epsilon(a)$.

- Thus, g is an inverse function of f on $f(B_\epsilon(a))$.
- We say that g is the inverse of f “locally” around the point $f(a)$.
- [Note that there is no guarantee that the inverse function is defined on the entire set $f(E)$.]

– Let $\boldsymbol{f}'(a) < \mathbf{0}$.

- An inverse function could similarly be defined “locally” around $f(a)$.

- The important restriction to carry out the above kind of analysis is $\boldsymbol{f}'(a) \neq \mathbf{0}$.
  - To illustrate this, consider $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(x) = x^2$ and consider $a = 0$.
    - Clearly f is continuously differentiable on $\mathbb{R}$, but $f'(a) = f'(0) = 0$.
    - Draw the curve for $f(x) = x^2$ and convince yourself that we cannot define a unique inverse function of f even "locally" around $f(0)$.
      - That is, choose any open ball $B_\epsilon(0)$, and consider any point, $y \neq 0$, in the set $f(B_\epsilon(0))$.
      - There will be *two* values $x_1,\ x_2$ in $B_\epsilon(0)$, $x_1 \neq x_2$, such that $f(x_1) = y$ and $f(x_2) = y$.

- **Note:** $f'(a) \neq 0$ *is not a necessary condition to get a unique inverse function of* f.
  - For example, consider $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^3$.
  - Then f is continuously differentiable on $\mathbb{R}$, with $f'(0) = 0$.
  - However f is an increasing function, and clearly has a unique inverse function $g(y) = y^{1/3}$ on $\mathbb{R}$, and hence locally around $f(0)$.

5.1 Derivative of the Inverse Function

- From a geometric point of view, if f maps x_0 to y_0, so that the point (x_0, y_0) is on the graph of f, then f^{-1} maps y_0 back to x_0, and therefore the point (y_0, x_0) is on the graph of f^{-1}.
  - For any point (x, y) on the graph of f, the point (y, x) is on the graph of f^{-1}.
  - This means that the graph of f^{-1} is simply the reflection of the graph of f across the 45^0 line.
    - The following two figures illustrate this phenomenon.
  - Due to the close relationship between the graph of an invertible function f and the graph of its inverse f^{-1}, it is not surprising that there is a close relationship between their derivatives.
  - In particular, if f is C^1 so that its graph has a smoothly varying tangent line, then the graph of f^{-1} also will have a smoothly varying tangent line; that is, f^{-1} will be C^1 too.

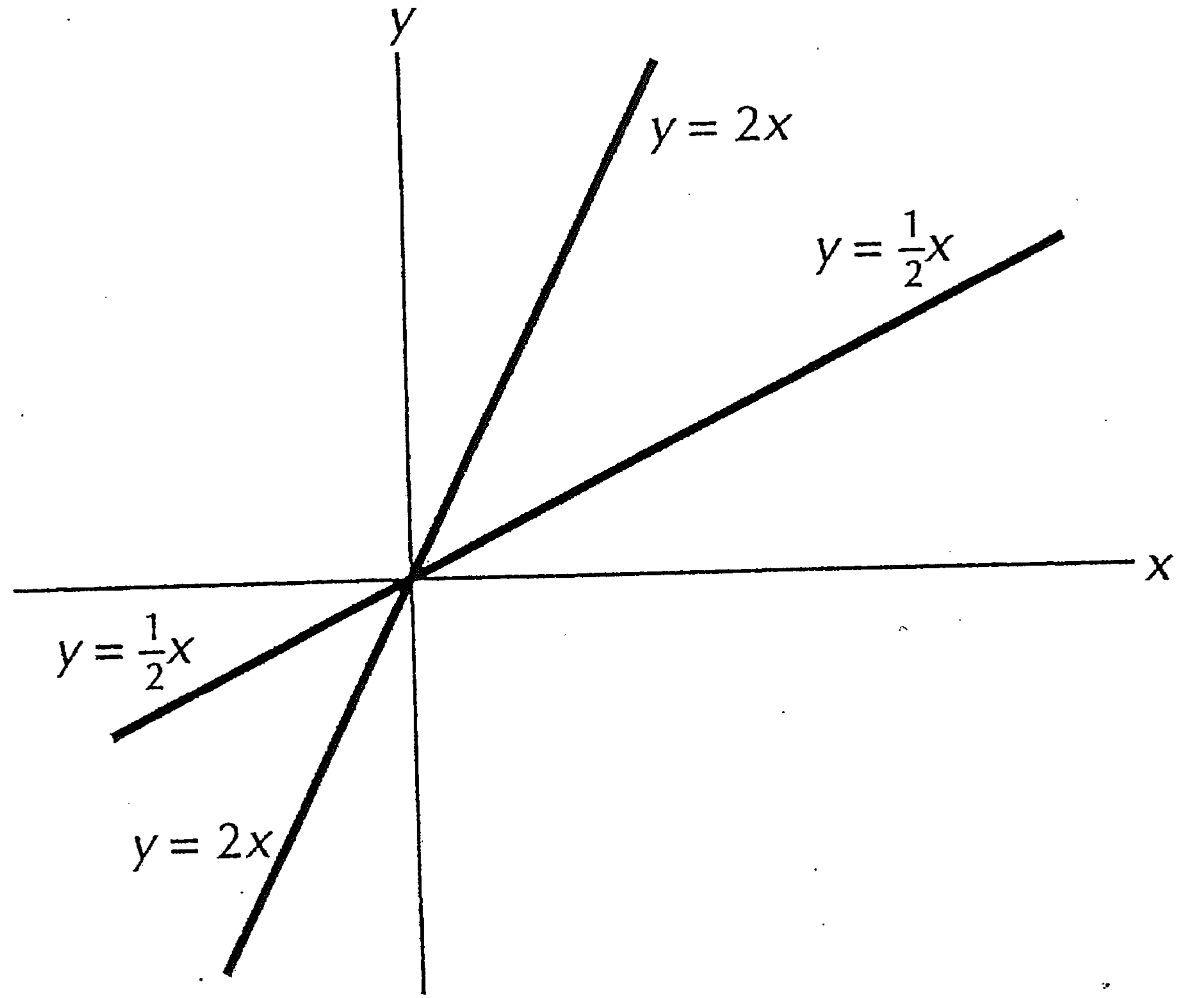


Figure 4.2

The graphs of the functions $y = 2x$ and $y = \frac{1}{2}x$.

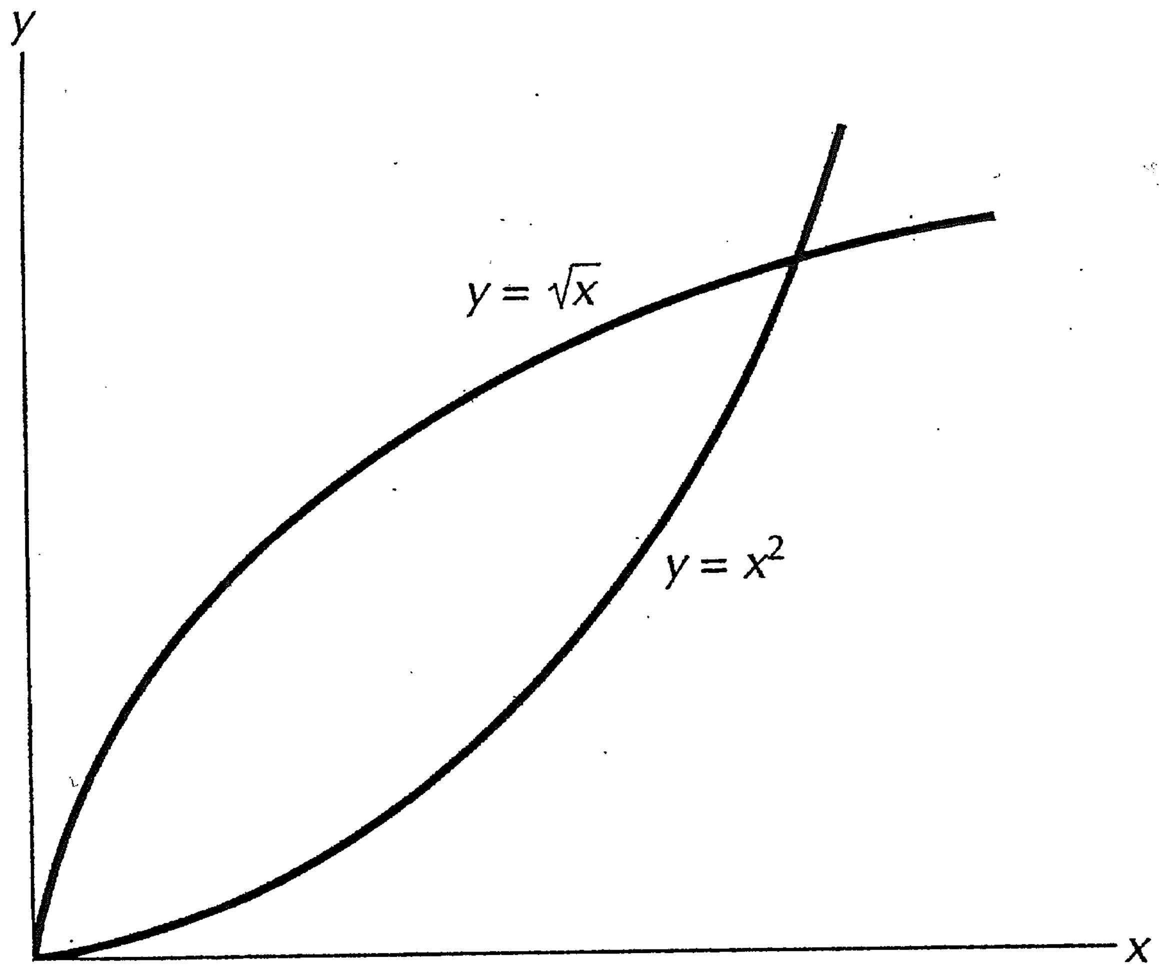


Figure 4.3

The graphs of the functions $y = x^2$ *and* $y = \sqrt{x}$ *for* $x, y \geq 0$.

- The following theorem gives a complete picture for the existence and differentiability of the inverse of a single-variable C^1 function.

- **Theorem 9:**

  Let f be a C^1 function defined on the interval I in $\mathbb{R}$. If $f'(x) \neq 0$ for all $x \in I$, then

(a) f is invertible on I,

(b) its inverse g is a C^1 function on the interval $f(I)$, and

(c) for all y in the domain of the inverse function g,

$$g'(y) = \frac{1}{f'(g(y))}.$$

- Example: Consider the following pair of functions which are inverses of each other:

$$f(x) = \frac{x-1}{x+1}, \text{ and } g(y) = \frac{1+y}{1-y}.$$

– Since $f(2) = \frac{1}{3}$, the inverse g of f maps $\frac{1}{3}$ to 2, that is, $g\left(\frac{1}{3}\right) = 2$.

– Since $f'(x) = \frac{2}{(x+1)^2}$, $f'(2) = \frac{2}{9} \neq 0$, by Theorem 9,

$$g'\left(\frac{1}{3}\right) = \frac{1}{f'(2)} = \frac{9}{2}.$$

– We can check this by computing directly that

$$g'(y) = \frac{2}{(1-y)^2}$$
$$\Rightarrow g'\left(\frac{1}{3}\right) = \frac{2}{4/9} = \frac{9}{2}.$$

- Let A be a set in $\mathbb{R}^n$, and let f be a function from A to $\mathbb{R}^n$.

- **Jacobians:**

  Suppose A is an open set in $\mathbb{R}^n$, and f is a function from A to $\mathbb{R}^n$, with component functions f^1, ..., f^n. If $a \in A$, and the partial derivatives of f^1, ..., f^n exist at a, then the $n \times n$ matrix

$$Df(a) = \begin{pmatrix} D_1 f^1(a) & D_2 f^1(a) & \cdots & D_n f^1(a) \\ D_1 f^2(a) & D_2 f^2(a) & \cdots & D_n f^2(a) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f^n(a) & D_2 f^n(a) & \cdots & D_n f^n(a) \end{pmatrix}$$

  is defined as the **Jacobian matrix** of f at a.

  – The determinant of this matrix, denoted by $J_f(a)$, is called the **Jacobian** of f at a.

- When $A \subset \mathbb{R}^n$, and $f : A \to \mathbb{R}^n$, the counter-part of $f'(a) \neq 0$ is that the Jacobian of f at a, $J_f(a)$, is non-zero.

- **Theorem 10 (Inverse Function Theorem):**

  Let A be an open set of $\mathbb{R}^n$, and $f : A \to \mathbb{R}^n$ be continuously differentiable on A. Suppose $a \in A$ and the Jacobian of f at a, $J_f(a)$, is non-zero. Then there is an open set $X \subset A$ containing a, and an open set $Y \subset \mathbb{R}^n$ containing $f(X)$, and a unique function $g : Y \to X$, such that

(i) *for all $x \in X$, $g(f(x)) = x$, and*

(ii) g *is continuously differentiable on Y* with

$$Dg(f(a)) = [Df(a)]^{-1}.$$

- In order to show that $Dg\left(f\left(a\right)\right)=\left[Df\left(a\right)\right]^{-1}$ we proceed as follows.
  - Under the hypothesis of the Inverse Function Theorem, there is a function $g: Y \rightarrow X$, such that g is continuously differentiable on Y and $g\left(f\left(x\right)\right)=x$ for all $x \in X$.
  - We can define for $x \in X$, $F^1\left(x\right)=g^1\left(f\left(x\right)\right)$ as a composite function of f and g^1.
  - Using the Chain Rule we get

$$D_iF^1\left(x\right)=\sum_{j=1}^{n} D_jg^1\left(f\left(x\right)\right)D_if^j\left(x\right),\ i=1,...,n.$$

  - But since $F^1\left(x\right)=g^1\left(f\left(x\right)\right)=x_1$, we have

$$D_iF^1\left(x\right)=\begin{cases}1 & \text{for } i=1\\ 0 & \text{for } i\neq 1.\end{cases}$$

  - We can repeat these calculations with $F^2\left(x\right)=g^2\left(f\left(x\right)\right)$, and get

$$D_iF^2\left(x\right)=\begin{cases}1 & \text{for } i=2\\ 0 & \text{for } i\neq 2.\end{cases}$$

– The results for $F^3(x), \ldots, F^n(x)$ should now be obvious.

– This information can then be written in familiar matrix multiplication form:

$$I = \begin{pmatrix} D_1 g^1(f(x)) & D_2 g^1(f(x)) & \cdots & D_n g^1(f(x)) \\ D_1 g^2(f(x)) & D_2 g^2(f(x)) & \cdots & D_n g^2(f(x)) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 g^n(f(x)) & D_2 g^n(f(x)) & \cdots & D_n g^n(f(x)) \end{pmatrix} \cdot \begin{pmatrix} D_1 f^1(x) & D_2 f^1(x) & \cdots & D_n f^1(x) \\ D_1 f^2(x) & D_2 f^2(x) & \cdots & D_n f^2(x) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f^n(x) & D_2 f^n(x) & \cdots & D_n f^n(x) \end{pmatrix}.$$

– That is, $I = Dg(f(x)) \cdot Df(x)$.

– Thus, the matrix $Df(a)$ is invertible, and we have $Dg(f(a)) = [Df(a)]^{-1}$.

– It follows that (taking determinants) the Jacobian of g at $f(a)$ is equal to the reciprocal of the Jacobian of f at a:

$$J_g(f(a)) = \frac{1}{J_f(a)}.$$

6. Implicit Functions

- So far we have been working with functions where the endogenous or independent variables are explicit functions of the exogenous or independent variables:

$$y = F(x_1, x_2, ..., x_n).$$

When the variables are separated like this, we say that the endogenous variable (y) is an *explicit function* of the exogenous variables ($x_1, x_2, ..., x_n$).

- Consider the expression

$$G(x_1, x_2, ..., x_n, y) = 0$$

where the exogenous variables ($x_1, x_2, ..., x_n$) are mixed with the endogenous variable (y).

  – If for each ($x_1, x_2, ..., x_n$), the above equation determines a corresponding value of y, we say that the equation defines the endogenous variable y as an **implicit function** of the exogenous variables $x_1, x_2, ..., x_n$.

- An expression like above is often so complicated that one cannot solve it to separate the exogenous variables on one side and the endogenous one on the other.
  - However, we still want to answer the basic question: how does a small change in one of the exogenous variables affect the value of the endogenous variable?
    - In what follows we explain how to answer this question for *implicit functions*.

- Example 1: The equation $4x + 2y - 5 = 0$ express y as an implicit function of x.
  - We can solve the equation and write y as an explicit function of x: $y = 2.5 - 2x$.
- Example 2: A more complex example is the equation: $y^2 - 5xy + 4x^2 = 0$.
  - We substitute any specified value of x into the equation and then solve the resulting quadratic equation for y.
  - Even though this equation is more complex than Example 1, we can still convert the equation into an explicit function (actually, a correspondence) by applying the quadratic formula:

$$y = \frac{5x \pm \sqrt{25x^2 - 16x^2}}{2} = \frac{1}{2}(5x \pm 3x) = \begin{cases} 4x \\ x. \end{cases}$$

- Example 3: Changing one exponent in Example 2 to construct the implicit function $y^5 - 5xy + 4x^2 = 0$ yields an expression which cannot be solved into an explicit function because there is no general formula for solving quintic equations.
  - However, this equation still defines y as a function of x: for example, $x = 0$ implies $y = 0$; $x = 1$ gives $y = 1$, and so on.

- Example 4: Consider a profit-maximizing firm that uses a single input x at a cost of w dollars per unit to produce a single output via a production function $y = f(x)$. If output sells for p dollars a unit, the firm's profit function for any fixed p and w is: $\pi(x) = p \cdot f(x) - w \cdot x$.
  - The profit-maximizing choice of x is determined from setting the derivative of the profit function equal to zero:

$$p \cdot f'(x) - w = 0.$$

  - Think of p and w as exogenous variables. For each choice of p and w, the firm chooses x that satisfies the above equation.
  - There is no reason to limit the models to production functions for which the above equation can be solved explicitly for x in terms of p and w.
  - To study the profit-maximizing behaviour of a general firm, we need to work with the above equation as defining x as an *implicit function* of p and w.
  - We will want to know, for example, how the optimal choice of input x changes as p or w increases.

- The fact that we can write down an implicit function $G(x, y) = c$ does not mean that this equation automatically defines y as a function of x.

- Consider, for example, the simple implicit function

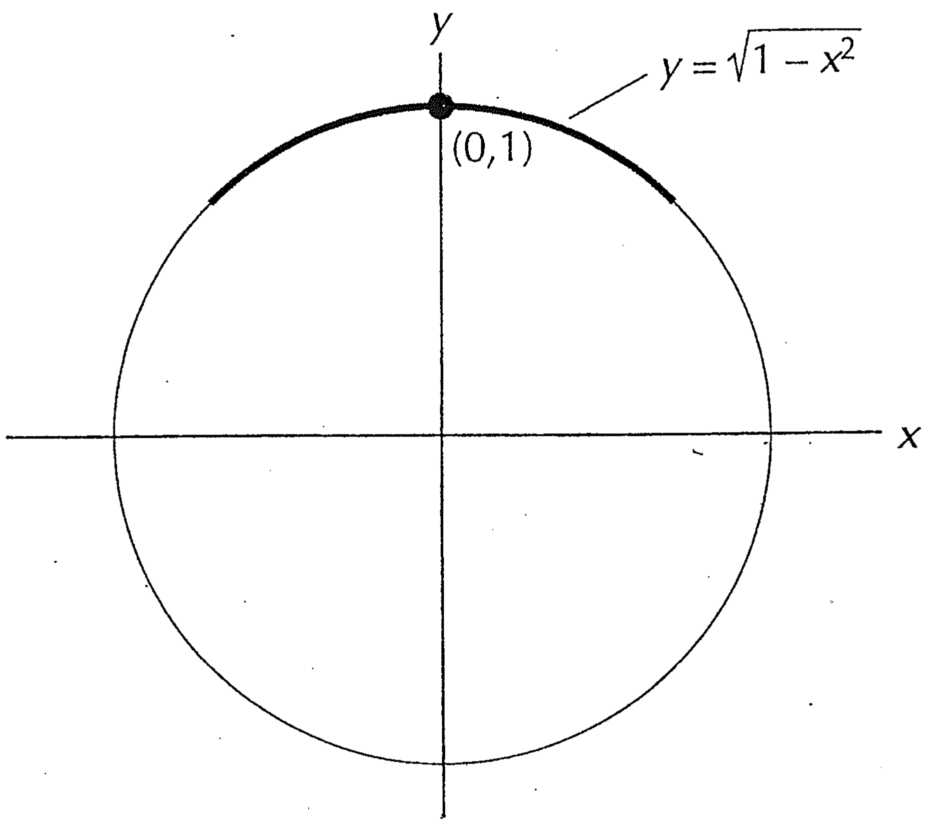
$$x^2 + y^2 = 1. \tag{$*$}$$

  – When $x > 1$, there is no y which satisfies (*).

  – Usually we start with a specific solution (x_0, y_0) of the implicit equation $G(x, y) = c$ and ask if we vary x a little from x_0, can we find a y near the original y_0 that satisfies the equation.

    - For example, if we start with the solution $(x = 0, y = 1)$ of (*) and vary x a little, we can find a unique $y = \sqrt{1 - x^2}$ near $y = 1$ that corresponds to the new x.

    - We can even draw the graph of this explicit relationship, $y = \sqrt{1 - x^2}$, around the point $(0, 1)$, as we do in the following figure.



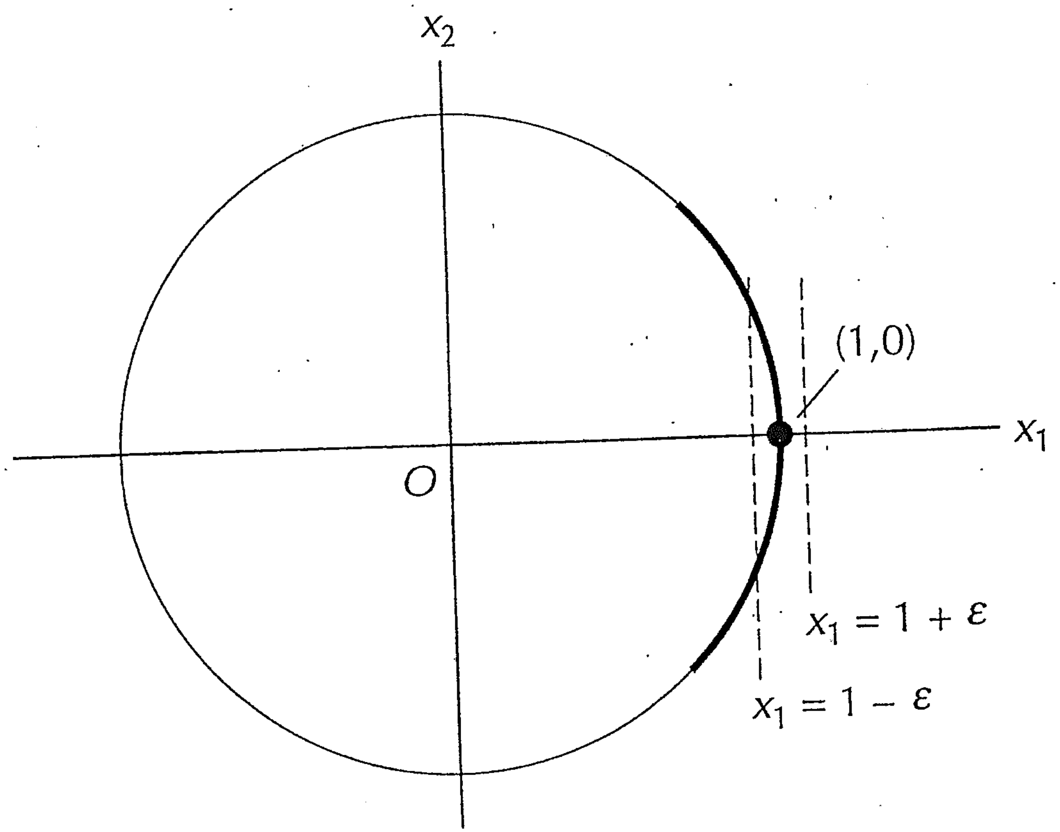
The graph of $x^2 + y^2 = 1$ *near the point* $(0, 1)$.

Figure 15.1

- However, if we start at the solution $(x = 1, y = 0)$ of (*), then *no such functional relationship exists*.
  - Look at the figure in the next page.
    - If we increase x a little to $x = 1 + \varepsilon$, then there is no corresponding y so that $(1 + \varepsilon, y)$ solves (*).
    - If we decrease x a little to $x = 1 - \varepsilon$, then there are two equally good candidates for y *near* $y = 0$, namely

$$y = +\sqrt{2\varepsilon - \varepsilon^2}, \text{ and } y = -\sqrt{2\varepsilon - \varepsilon^2}.$$

    - As the figure illustrates, because the curve $x^2 + y^2 = 1$ is vertical around $(1, 0)$, it does **not** define y as a function of x around $(1, 0)$.



The graph of $x^2 + y^2 = 1$ *near the point* (1, 0).

Figure 15.2

6.1 Implicit Function Theorem: One Exogenous Variable

- For a given implicit function $G(x, y) = c$ and a specified solution point (x_0, y_0), we want the answers to the following two questions:
  - Does $G(x, y) = c$ determine y as a continuous function of x, $y = y(x)$, for x near x_0 and y near y_0?
  - If so, how do changes in x affect the corresponding y's?
- Answers of these two questions are closely related to each other:
  - If the first question has a positive answer, one can use the chain rule to compute $y'(x)$ in terms of $\partial G/\partial x$ and $\partial G/\partial y$.
  - On the other hand, this formula for $y'(x)$ in terms of $\partial G/\partial x$ and $\partial G/\partial y$ leads to the natural criterion for the existence question.
- We suppose that there is a C^1 solution $y = y(x)$ to the equation $G(x, y) = c$, that is,

$$G(x, y(x)) = c. \tag{A}$$

– Use the Chain Rule to differentiate (A) with respect to x at x_0:

$$\frac{\partial G}{\partial x}(x_0, y(x_0)) \cdot \frac{dx}{dx} + \frac{\partial G}{\partial y}(x_0, y(x_0)) \cdot \frac{dy}{dx}(x_0) \;=\; 0,$$

or
$$\frac{\partial G}{\partial x}(x_0, y_0) + \frac{\partial G}{\partial y}(x_0, y_0) \cdot y'(x_0) \;=\; 0.$$

– Solving for $y'(x_0)$ yields

$$y'(x_0) = -\frac{\dfrac{\partial G}{\partial x}(x_0, y_0)}{\dfrac{\partial G}{\partial y}(x_0, y_0)}. \tag{B}$$

– We see from (B) that if the solution $y(x)$ to the equation $G(x, y) = c$ exists and is differentiable, it is *necessary* that $\dfrac{\partial G}{\partial y}(x_0, y_0)$ be *non-zero*.

- As the following fundamental result of mathematical analysis indicates, this necessary condition is also a sufficient condition.

• **Theorem 11(a) (Implicit Function Theorem: one exogenous variable):**

Let $G(x, y)$ *be a* C^1 *function on an open ball around* (x^*, y^*) *in* $\mathbb{R}^2$. *Suppose that* $G(x^*, y^*) = c$. *If* $\frac{\partial G}{\partial y}(x^*, y^*) \neq 0$, *then there exists a* C^1 *function* $y = y(x)$ *defined on an open interval* I *around the point* x^* *such that*:

(a) $G(x, y(x)) \equiv c$ *for all* x *in* I,

(b) $y(x^*) = y^*$, *and*

(c) $y'(x^*) = -\dfrac{\frac{\partial G}{\partial x}(x^*, y^*)}{\frac{\partial G}{\partial y}(x^*, y^*)}$.

- Example: Consider the function $G : \mathbb{R}^2_{++} \to \mathbb{R}$ defined by $G(x, y) = x^\alpha y^\beta$, where α, β are positive constants.
  - At $(x^*, y^*) = (1, 1)$, $G(x^*, y^*) = 1$.
  - $\frac{\partial G}{\partial y}(1, 1) = \beta > 0$, that is, $\frac{\partial G}{\partial y}(1, 1) \neq 0$.
  - Also G is a C^1 function on $\mathbb{R}^2_{++}$.
  - Hence we can invoke the implicit function theorem to obtain a C^1 function $y = y(x)$ defined on an open interval I around the point $x^* = 1$ such that:
    - $x^\alpha (y(x))^\beta = 1$ for all x in I; $y(1) = 1$, and
    - $y'(1) = -\frac{\partial G}{\partial x}(1, 1) \Big/ \frac{\partial G}{\partial y}(1, 1) = -\frac{\alpha}{\beta}$.

- Example: Return now to the equation $x^2 + y^2 = 1$.
  - We saw that this equation does determine y as a function of x around the point $(x = 0, y = 1)$.

    - $\dfrac{\partial G}{\partial y}(0,1) = 2 \neq 0$. So the Implicit Function Theorem assures us that $y(x)$ exists around $(0,1)$.

    - Furthermore, the theorem tells us that $y'(0) = -\dfrac{\dfrac{\partial G}{\partial x}(0,1)}{\dfrac{\partial G}{\partial y}(0,1)} = -\dfrac{0}{2} = 0.$

  - In this case, we have an explicit formula for $y(x)$: $y(x) = \sqrt{1 - x^2}$.

    - We can compute directly from this explicit formula that $y'(0) = -\dfrac{x}{\sqrt{1-x^2}}\Bigg|_{x=0} = 0.$

  - On the other hand, we noted earlier that no function $y(x)$ exists for $x^2 + y^2 = 1$ around $(x = 1, y = 0)$.

    - This is consistent with the theorem since $\dfrac{\partial G}{\partial y} = 2y = 0$ at $(1,0)$.

6.2 Implicit Function Theorem: Several Exogenous Variables

- The Implicit Function Theorem for one exogenous variable and the discussion around it carry over in a straightforward way to the situation where there are many exogenous variables, but still one equation and therefore one endogenous variable:

$$G(x_1, x_2, ..., x_n, y) = c. \tag{C}$$

  - Around a given point $(x_1^*, x_2^*, ..., x_n^*, y^*)$, we want to vary $x = (x_1, x_2, ..., x_n)$ and then find a y-value which corresponds to each such $(x_1, x_2, ..., x_n)$.
  - In this case, we say that the equation (C) defines y as an implicit function of $(x_1, x_2, ..., x_n)$.

- Again, given $G(\cdot)$ and $(x_1^*, x_2^*, ..., x_n^*, y^*)$, we want to know whether this functional relationship exists, and, if yes, how y changes if any of the x_i's change from x_i^*.
  - Since we are working with a function of several variables $(x_1, x_2, ..., x_n)$, we will hold all but one of the x_i's constant and vary one exogenous variable at a time.

– But this puts us back in the two-variable situation that we have been discussing.

- The natural extension of the Implicit Function Theorem for one exogenous variable in this setting is the following.

- **Theorem 11(b) (Implicit Function Theorem: several exogenous variables):**

  Let $G(x_1, x_2, ..., x_k, y)$ *be a* C^1 *function on an open ball around* $(x_1^*, x_2^*, ..., x_k^*, y^*)$ *in* $\mathbb{R}^{k+1}$ *such that* $G(x_1^*, x_2^*, ..., x_k^*, y^*) = c$. *If* $\dfrac{\partial G}{\partial y}(x_1^*, x_2^*, ..., x_k^*, y^*) \neq 0$, *then there exists a* C^1 *function* $y = y(x_1, x_2, ..., x_k)$ *defined on an open ball* B *around* $(x_1^*, x_2^*, ..., x_k^*)$ *such that*:

(a) $G(x_1, x_2, ..., x_k, y(x_1, x_2, ..., x_k)) = c$ *for all* $(x_1, x_2, ..., x_k) \in B$,

(b) $y^* = y(x_1^*, x_2^*, ..., x_k^*)$, *and*

(c) $\dfrac{\partial y}{\partial x_i}(x_1^*, x_2^*, ..., x_k^*) = -\dfrac{\dfrac{\partial G}{\partial x_i}(x_1^*, x_2^*, ..., x_k^*, y^*)}{\dfrac{\partial G}{\partial y}(x_1^*, x_2^*, ..., x_k^*, y^*)}$, *for all* $i = 1, 2, ..., k$.

6.3 System of Implicit Functions

- **Definition:** A set of m equations in $m+n$ unknowns

$$\begin{aligned} G^1(x_1, x_2, ..., x_{m+n}) &= c_1, \\ G^2(x_1, x_2, ..., x_{m+n}) &= c_2, \\ &\vdots \\ G^m(x_1, x_2, ..., x_{m+n}) &= c_m, \end{aligned} \qquad \text{(D)}$$

is called a *system of implicit functions* if there is a partition of the variables into *exogenous variables* and *endogenous variables*, so that if one substitutes into (D) numerical values for the exogenous variables, the resulting system can be solved uniquely for corresponding values of the endogenous variables.

– This is a natural generalization of the single-equation implicit function.

6.3.1 Linear System of Implicit Functions

- From our discussion in linear algebra we know that for *linear* system of implicit functions, in order for each choice of values of the exogenous variables to determine a unique set of values of the endogenous variables it is necessary and sufficient that:

(1) the number of endogenous variables is equal to the number of equations, and

(2) the (square) matrix of coefficients corresponding to the endogenous variables be invertible (have non-zero determinant).

- Example: Consider the linear system of implicit functions:

$$\begin{aligned} 4x + 2y + 2z - r + 3s &= 5, \\ 2x + 0 \cdot y + 2z + 8r - 5s &= 7, \\ 2x + 2y + 0 \cdot z + r - s &= 0. \end{aligned}$$

  – Since there are three equations, we need three endogenous variables, and, therefore, two exogenous variables.

– Let us try to work with y, z and r as endogenous and x and s as exogenous. Putting the exogenous variables on the right side and the endogenous variables on the left, we rewrite the system as

$$\begin{bmatrix} 2 & 2 & -1 \\ 0 & 2 & 8 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ z \\ r \end{bmatrix} = \begin{bmatrix} 5-4x-3s \\ 7-2x+5s \\ -2x+s \end{bmatrix}.$$

– Since the determinant of the coefficient matrix is 40, we can invert it and solve for (y,z,r) explicitly in terms of x and s:

$$\begin{bmatrix} y \\ z \\ r \end{bmatrix} = \begin{bmatrix} 2 & 2 & -1 \\ 0 & 2 & 8 \\ 2 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 5-4x-3s \\ 7-2x+5s \\ -2x+s \end{bmatrix} = \frac{1}{40} \begin{bmatrix} 2 & -2 & 18 \\ 16 & 4 & -16 \\ -4 & 4 & 4 \end{bmatrix} \begin{bmatrix} 5-4x-3s \\ 7-2x+5s \\ -2x+s \end{bmatrix}.$$

– On the other hand, if we want x, y and z to be endogenous, we have to solve the system

$$\begin{bmatrix} 4 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 + r - 3s \\ 7 - 8r + 5s \\ 0 - r + s \end{bmatrix}.$$

– Since the determinant of the coefficient matrix is zero, we know that there are right-hand sides for which the above system cannot be solved for (x, y, z).

- For example, for $r = -5$ and $s = 0$, the system becomes

$$\begin{aligned} 4x + 2y + 2z &= 0, \\ 2x + 0 \cdot y + 2z &= 47, \\ 2x + 2y + 0 \cdot z &= 5. \end{aligned}$$

- Adding the last two equations yields the inconsistent system:

$$\begin{aligned} 4x + 2y + 2z &= 0, \\ 4x + 2y + 2z &= 52. \end{aligned}$$

- Since there is no solution in (x, y, z) for $(r, s) = (-5, 0)$, this partition into exogenous and endogenous variables does not work.

6.3.2 Nonlinear System of Implicit Functions

- The corresponding result for nonlinear systems follows from the usual calculus paradigm:
  - linearize by taking the derivative;
  - apply the linear theorem to this linearized system; and
  - transfer these results back to the original nonlinear system.
- Write the nonlinear system of m equations in $m+n$ unknowns as

$$\begin{aligned} F^1(y_1, y_2, ..., y_m, x_1, x_2, ..., x_n) &= c_1, \\ F^2(y_1, y_2, ..., y_m, x_1, x_2, ..., x_n) &= c_2, \\ \vdots \qquad\qquad & \quad \vdots \\ F^m(y_1, y_2, ..., y_m, x_1, x_2, ..., x_n) &= c_m. \end{aligned} \tag{E}$$

  - Here we want $(y_1, y_2, ..., y_m)$ to be endogenous variables and $(x_1, x_2, ..., x_n)$ to be exogenous variables.

- The linearization of system (E) around the point $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$ is:

$$\begin{aligned} \frac{\partial F^1}{\partial y_1} dy_1 + ... + \frac{\partial F^1}{\partial y_m} dy_m + \frac{\partial F^1}{\partial x_1} dx_1 + ... + \frac{\partial F^1}{\partial x_n} dx_n &= 0, \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \frac{\partial F^m}{\partial y_1} dy_1 + ... + \frac{\partial F^m}{\partial y_m} dy_m + \frac{\partial F^m}{\partial x_1} dx_1 + ... + \frac{\partial F^m}{\partial x_n} dx_n &= 0, \end{aligned} \tag{E'}$$

  where all the partial derivatives are evaluated at $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$.

- The linear system (E') can be solved for ($dy_1, dy_2, ..., dy_m$) in terms of ($dx_1, dx_2, ..., dx_n$) if and only if the coefficient matrix of the dy_i 's

$$\begin{pmatrix} \frac{\partial F^1}{\partial y_1} & \cdots & \frac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial y_1} & \cdots & \frac{\partial F^m}{\partial y_m} \end{pmatrix} \tag{F}$$

  is **invertible** (have **non-zero determinant**) at $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$.

- Since the system is linear, when the coefficient matrix in (F) is invertible, we can use its inverse to solve the system (E') for $(dy_1, dy_2, ..., dy_m)$ in terms of $(dx_1, dx_2, ..., dx_n)$:

$$\begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_m \end{bmatrix} = - \begin{bmatrix} \frac{\partial F^1}{\partial y_1} & \cdots & \frac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F^m}{\partial y_1} & \cdots & \frac{\partial F^m}{\partial y_m} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial F^1}{\partial x_1} dx_1 + ... + \frac{\partial F^1}{\partial x_n} dx_n \\ \vdots \\ \frac{\partial F^m}{\partial x_1} dx_1 + ... + \frac{\partial F^m}{\partial x_n} dx_n \end{bmatrix}. \tag{G}$$

- Since the linear approximation (E') of the original system (E) is a true implicit function of $(dy_1, dy_2, ..., dy_m)$ in terms of $(dx_1, dx_2, ..., dx_n)$, the basic principle of calculus leads us to the following conclusion:

  – if the coefficient matrix in (F) is invertible, then the nonlinear system (E) defines $(y_1, y_2, ..., y_m)$ as implicit functions of $(x_1, x_2, ..., x_n)$, at least in a neighbourhood of $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$.

⇒ The sufficient condition for the existence of implicit functions for the nonlinear system (E) is: *the coefficient matrix in* (F) *is invertible*.

- Furthermore, one can actually use the linear solution (G) of $(dy_1, dy_2, ..., dy_m)$ in terms of $(dx_1, dx_2, ..., dx_n)$ to find the derivatives of the y_i's with respect to the x_j's at $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$.

- To compute $\dfrac{\partial y_k}{\partial x_h}$ for some h and k, recall that this derivative estimates the effect on y_k of a unit increase in x_h ($dx_h = 1$).

- So we set all the dx_j's equal to zero in (E') or (G) except dx_h and then solve (E') or (G) for the corresponding dy_i's.

  – For example, if we use (G), we find

$$\begin{bmatrix} \dfrac{\partial y_1}{\partial x_h} \\ \vdots \\ \dfrac{\partial y_m}{\partial x_h} \end{bmatrix} = - \begin{bmatrix} \dfrac{\partial F^1}{\partial y_1} & \cdots & \dfrac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F^m}{\partial y_1} & \cdots & \dfrac{\partial F^m}{\partial y_m} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\partial F^1}{\partial x_h} \\ \vdots \\ \dfrac{\partial F^m}{\partial x_h} \end{bmatrix}.$$

- Alternatively we can apply Cramer's Rule to (E') and compute

$$\frac{\partial y_k}{\partial x_h} = -\frac{\begin{vmatrix} \dfrac{\partial F^1}{\partial y_1} & \cdots & \dfrac{\partial F^1}{\partial x_h} & \cdots & \dfrac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \dfrac{\partial F^m}{\partial y_1} & \cdots & \dfrac{\partial F^m}{\partial x_h} & \cdots & \dfrac{\partial F^m}{\partial y_m} \end{vmatrix}}{\begin{vmatrix} \dfrac{\partial F^1}{\partial y_1} & \cdots & \dfrac{\partial F^1}{\partial y_k} & \cdots & \dfrac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \dfrac{\partial F^m}{\partial y_1} & \cdots & \dfrac{\partial F^m}{\partial y_k} & \cdots & \dfrac{\partial F^m}{\partial y_m} \end{vmatrix}}.$$

- The following theorem – the most general form of the Implicit Function Theorem – summarizes these conclusions.

- **Theorem 11(c) (Implicit Function Theorem: most general form):**

  Let $F^1, F^2, ..., F^m : \mathbb{R}^{m+n} \to \mathbb{R}$ *be* C^1 *functions. Consider the system of equations*

$$\begin{aligned} F^1(y_1, ..., y_m, x_1, ..., x_n) &= c_1 \\ F^2(y_1, ..., y_m, x_1, ..., x_n) &= c_2 \\ &\vdots \\ F^m(y_1, ..., y_m, x_1, ..., x_n) &= c_m \end{aligned}$$

  as possibly defining $y_1, ..., y_m$ *as implicit functions of* $x_1, ..., x_n$. *Suppose that* $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$ *is a solution to the system of equations. If the determinant of the* $m \times m$ *matrix*

$$\begin{pmatrix} \dfrac{\partial F^1}{\partial y_1} & \cdots & \dfrac{\partial F^1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F^m}{\partial y_1} & \cdots & \dfrac{\partial F^m}{\partial y_m} \end{pmatrix}$$

  evaluated at $(y_1^*, ..., y_m^*, x_1^*, ..., x_n^*)$ *is nonzero, then there exist* C^1 *functions*

$$
\begin{aligned}
y_1 &= f^1(x_1, ..., x_n) \\
y_2 &= f^2(x_1, ..., x_n) \\
&\vdots \\
y_m &= f^m(x_1, ..., x_n)
\end{aligned}
$$

defined on an open ball B *around* $(x_1^*, ..., x_n^*)$ *such that*

$$
\begin{aligned}
F^1\left(f^1(x), ..., f^m(x), x_1, ..., x_n\right) &= c_1 \\
F^2\left(f^1(x), ..., f^m(x), x_1, ..., x_n\right) &= c_2 \\
&\vdots \\
F^m\left(f^1(x), ..., f^m(x), x_1, ..., x_n\right) &= c_m
\end{aligned}
$$

for all $x = (x_1, ..., x_n)$ *in* B *and*

$$
\begin{aligned}
y_1^* &= f^1(x_1^*, ..., x_n^*) \\
y_2^* &= f^2(x_1^*, ..., x_n^*) \\
&\vdots \\
y_m^* &= f^m(x_1^*, ..., x_n^*).
\end{aligned}
$$

Furthermore, one can compute $\dfrac{\partial y_k}{\partial x_h}(y^*, x^*) = \dfrac{\partial f^k}{\partial x_h}(x^*)$ *by setting* $dx_h = 1$ *and* $dx_j = 0$ *for* $j \neq h$ *in*

$$\begin{aligned}
\frac{\partial F^1}{\partial y_1} dy_1 + \cdots + \frac{\partial F^1}{\partial y_m} dy_m + \frac{\partial F^1}{\partial x_1} dx_1 + \cdots + \frac{\partial F^1}{\partial x_n} dx_n \ &= \ 0 \\
&\vdots \\
\frac{\partial F^m}{\partial y_1} dy_1 + \cdots + \frac{\partial F^m}{\partial y_m} dy_m + \frac{\partial F^m}{\partial x_1} dx_1 + \cdots + \frac{\partial F^m}{\partial x_n} dx_n \ &= \ 0
\end{aligned}$$

and solving the resulting system for dy_k.

References

- Must read the following chapters and sections from the textbook:
  - Chapter 2 (pages 10 – 38): One-Variable Calculus: Foundations,
  - Section 4.2 (pages 75 – 81): Inverse Functions and Their Derivatives,
  - Chapter 14 (pages 300 – 333): Calculus of Several Variables,
  - Section 20.1 (pages 483 – 493): Homogeneous Functions,
  - Chapter 15 (pages 334 – 371): Implicit Functions and Their Derivatives.
- This material is based on

1. Bartle, R., *The Elements of Real Analysis*, (chapter 7),
2. Apostol, T., *Mathematical Analysis: A Modern Approach to Advanced Calculus*, (chapters 6, 7).
3. Spivak, M., *Calculus on Manifolds*, (chapter 2).