

The Implicit Function Theorem

Case 1: A linear equation with $m = n = 1$ (We'll say what m and n are shortly.)

Suppose we know that x and y must always satisfy the equation

$$ax + by = c. \tag{1}$$

Let's write the expression on the left-hand side of the equation as a function: $F(x, y) = ax + by$, so the equation is $F(x, y) = c$. [See Figure 1]

We want to know the answers to two questions:

(i) Can we say that y is a function of x , say $y = f(x)$ — *i.e.*, that for every value of x there will be one and only one value of y that satisfies the equation (1)? (We say that the function f , if it exists, gives y as an *explicit* function of x ; and if the function f exists, we say that the equation in (1) defines y *implicitly* as a function of x , or as an *implicit* function of x .)

(ii) If the answer to (i) is Yes, can we say anything about the derivative of the function f — does the derivative exist (*i.e.*, is f differentiable) and if so, can we determine the value of $f'(x)$?

Of course, if we're able to simply “solve for y in terms of x ”, as we can obviously do in this case, then we have the explicit function f and we can differentiate it. But we want to know the answers to (i) and (ii) *in general*, not just for a specific function — and not just for linear functions.

Focusing on the linear case for now, if a and b have the same sign, both positive or both negative, then the diagram looks like Figure 1. In this case it's clear that $y = \frac{c}{b} - \frac{a}{b}x$: we have $f(x) = \frac{c}{b} - \frac{a}{b}x$ and $f'(x) = -\frac{a}{b}$. Similarly, if a and b have opposite signs then the diagram looks like Figure 2, and we again have $y = f(x) = \frac{c}{b} - \frac{a}{b}x$ and $f'(x) = -\frac{a}{b}$.

If $a = 0$ and $b \neq 0$ then we have Figure 3, and $y = f(x) = \frac{c}{b}$ and $f'(x) = 0$: $f(x)$ is still $\frac{c}{b} - \frac{a}{b}x$, and $f'(x)$ is still $-\frac{a}{b}$. Everything is still good.

But if $a \neq 0$ and $b = 0$ then we have Figure 4: $x = \frac{c}{a}$ and y cannot be expressed as a function of x .

When we generalize to nonlinear functions we're going to express everything in terms of F and its derivatives instead of a and b . Let's use the shorthand notation F_x and F_y for the partial derivatives of F :

$$F_x = \frac{\partial F}{\partial x} \quad \text{and} \quad F_y = \frac{\partial F}{\partial y}.$$

We can summarize Case 1 as follows:

$$\text{If } F_y \neq 0 \text{ then equation (1) defines a function } y = f(x), \text{ and } f'(x) = -\frac{F_x}{F_y}. \tag{2}$$

Note that we haven't said where the derivatives F_x and F_y are to be evaluated. Because F is linear, that's not necessary: the derivatives are the same everywhere, namely $F_x = a$ and $F_y = b$.

Case 2: A nonlinear equation with $m = n = 1$

Now suppose we know that x and y must always satisfy the equation

$$F(x, y) = c, \tag{3}$$

where $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ is differentiable but nonlinear, and let (\bar{x}, \bar{y}) be a point that satisfies (3). Write F_x and F_y for the partial derivatives of F at (\bar{x}, \bar{y}) :

$$F_x = \frac{\partial F}{\partial x}(\bar{x}, \bar{y}) \quad \text{and} \quad F_y = \frac{\partial F}{\partial y}(\bar{x}, \bar{y}).$$

Now the diagram looks like Figure 5, where we have the nonlinear level curve of F through (\bar{x}, \bar{y}) and also the tangent to the curve at that point. The equation of the tangent is

$$F_x x + F_y y = F_x \bar{x} + F_y \bar{y},$$

or

$$F_x \Delta x + F_y \Delta y = 0. \tag{4}$$

We've "linearized" the function F , which we know is a good approximation to F in a neighborhood of (\bar{x}, \bar{y}) , or of $(\Delta x, \Delta y) = (0, 0)$. The function F behaves like its linear approximation in this neighborhood, and the linear approximation behaves exactly like the linear function in Case 1. So we have

$$\text{If } F_y \neq 0 \text{ then the equation (3) defines a function } y = f(x), \text{ and } f'(x) = -\frac{F_x}{F_y}. \tag{5}$$

In fact, if $F_y \neq 0$ then equation (4) can be written as $\frac{\Delta y}{\Delta x} = -\frac{F_x}{F_y}$, the exact value of $f'(\bar{x})$ — the slope of the tangent line is exactly the derivative of f .

Figure 6 shows what goes wrong when $F_y = 0$ — *i.e.*, when $\frac{\partial F}{\partial y}(x, y) = 0$ at (\bar{x}, \bar{y}) .

Example: Indifference curves

We have a utility function $u(x, y)$ and we want to know about the indifference curve through a bundle (\bar{x}, \bar{y}) , *i.e.*, the set of bundles (x, y) that satisfy $u(x, y) = c$, where $c = u(\bar{x}, \bar{y})$. Is the indifference curve describable by a function $y = f(x)$, and if so, what is the derivative $f'(\bar{x})$ — *i.e.*, the slope of the curve at (\bar{x}, \bar{y}) , which is also the negative of the *MRS* at (\bar{x}, \bar{y}) ? The Implicit Function Theorem tells us that if $\frac{\partial u}{\partial y} \neq 0$ at (\bar{x}, \bar{y}) as in Figure 7, then the answer to the first question is Yes and the *MRS* is $\frac{u_x}{u_y}$, *i.e.*,

$$MRS(\bar{x}, \bar{y}) = -f'(\bar{x}) = \frac{\frac{\partial u}{\partial x}(\bar{x}, \bar{y})}{\frac{\partial u}{\partial y}(\bar{x}, \bar{y})}.$$

Figure 8 shows what happens if $u_y = 0$ — *i.e.*, if $\frac{\partial u}{\partial y}(x, y) = 0$ at (\bar{x}, \bar{y}) .

Case 3: A nonlinear equation, and m and n are arbitrary:

The Implicit Function Theorem: Let $F : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a C^1 -function and let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a point in $\mathbb{R}^m \times \mathbb{R}^n$. Let $\mathbf{c} = F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \mathbb{R}^n$. If the derivative of F with respect to \mathbf{y} is nonsingular — *i.e.*, if the $n \times n$ matrix

$$\left[\frac{\partial F_k}{\partial y_i} \right]_{k,i=1}^n$$

is nonsingular at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ — then there is a C^1 -function $f : \mathfrak{N} \rightarrow \mathbb{R}^n$ on a neighborhood \mathfrak{N} of $\bar{\mathbf{x}}$ that satisfies

- (a) $f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$, *i.e.*, $F(\bar{\mathbf{x}}, f(\bar{\mathbf{x}})) = \mathbf{c}$,
- (b) $\forall \mathbf{x} \in \mathfrak{N} : F(\mathbf{x}, f(\mathbf{x})) = \mathbf{c}$, and
- (c) $\left[\frac{\partial f_i}{\partial x_j} \right] = - \left[\frac{\partial F_k}{\partial y_i} \right]^{-1} \left[\frac{\partial F_k}{\partial x_j} \right]$, where the partial derivatives are evaluated at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. ■

In economics the Implicit Function Theorem is applied ubiquitously to optimization problems and their solution functions. The first-order conditions for an optimization problem comprise a system of n equations involving an n -tuple of decision variables $\mathbf{x} = (x_1, \dots, x_n)$ and an m -tuple of parameters $\theta = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$. Let's write these first-order-condition equations as $F(\mathbf{x}, \theta) = \mathbf{0}$. The main object of interest for us is the **solution function**, which gives the decision variables $\mathbf{x} = (x_1, \dots, x_n)$ as a function of the parameters $\theta = (\theta_1, \dots, \theta_m)$: $\mathbf{x} = f(\theta)$. We want to know the answers to the two questions posed in Case 1:

- (i) Does the equation $F(\mathbf{x}, \theta) = \mathbf{0}$ actually implicitly define a solution function $\mathbf{x} = f(\theta)$?
- (ii) If the answer to (i) is Yes, can we say anything about the derivatives of the function f , which tell us how the decision variables will be affected by changes in the parameters?

So let's rewrite the Implicit Function Theorem in terms of parameters $\theta = (\theta_1, \dots, \theta_m)$ and decision variables $\mathbf{x} = (x_1, \dots, x_n)$ — replacing \mathbf{x} above with $\theta \in \mathbb{R}^m$ and replacing \mathbf{y} with $\mathbf{x} \in \mathbb{R}^n$.

The Implicit Function Theorem: Let $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a C^1 -function and let $(\bar{\mathbf{x}}, \bar{\theta}) \in \mathbb{R}^n \times \mathbb{R}^m$ be a point at which $F(\bar{\mathbf{x}}, \bar{\theta}) = \mathbf{0} \in \mathbb{R}^n$. If the derivative of F with respect to \mathbf{x} is nonsingular — *i.e.*, if the $n \times n$ matrix

$$\left[\frac{\partial F_k}{\partial x_i} \right]_{k,i=1}^n$$

is nonsingular at $(\bar{\mathbf{x}}, \bar{\theta})$ — then there is a C^1 -function $f : \mathfrak{N} \rightarrow \mathbb{R}^n$ on a neighborhood \mathfrak{N} of $\bar{\theta}$ that satisfies

- (a) $f(\bar{\theta}) = \bar{\mathbf{x}}$, *i.e.*, $F(f(\bar{\theta}), \bar{\theta}) = \mathbf{0}$,
- (b) $\forall \theta \in \mathfrak{N} : F(f(\theta), \theta) = \mathbf{0}$, and
- (c) $\left[\frac{\partial f_i}{\partial \theta_j} \right] = - \left[\frac{\partial F_k}{\partial x_i} \right]^{-1} \left[\frac{\partial F_k}{\partial \theta_j} \right]$, where the partial derivatives are evaluated at $(\bar{\mathbf{x}}, \bar{\theta})$.

Case 4: A linear equation and m and n are arbitrary

This is obviously a special case of Case 3. But it's useful to see how the general case works when everything is linear, and of course the general (nonlinear) case is actually addressed by turning it into the linear case.

Suppose we have a linear function $F : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by $F(\mathbf{x}, \mathbf{y}) = A\mathbf{x} + B\mathbf{y}$, where A is an $n \times m$ matrix and B is an $n \times n$ matrix. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a point in the domain of F , *i.e.* in $\mathbb{R}^m \times \mathbb{R}^n$, and let $\mathbf{c} = F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. If B is nonsingular, then the points (\mathbf{x}, \mathbf{y}) that satisfy the equation $F(\mathbf{x}, \mathbf{y}) = \mathbf{c}$ are exactly the points that satisfy

$$\mathbf{y} = B^{-1}\mathbf{c} - B^{-1}A\mathbf{x}, \quad (8)$$

which we verify as follows:

$$\begin{aligned} F(\mathbf{x}, \mathbf{y}) &= A\mathbf{x} + B\mathbf{y} \\ &= A\mathbf{x} + B(B^{-1}\mathbf{c} - B^{-1}A\mathbf{x}) \\ &= A\mathbf{x} + BB^{-1}\mathbf{c} - BB^{-1}A\mathbf{x} \\ &= A\mathbf{x} + \mathbf{c} - A\mathbf{x} \\ &= \mathbf{c}. \end{aligned}$$

Therefore if we define $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ as $f(\mathbf{x}) = B^{-1}\mathbf{c} - B^{-1}A\mathbf{x}$, we have

- (a) $f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$, *i.e.*, $F(\bar{\mathbf{x}}, f(\bar{\mathbf{x}})) = \mathbf{c}$,
- (b) $\forall \mathbf{x} \in \mathbb{R}^m : F(\mathbf{x}, f(\mathbf{x})) = \mathbf{c}$, and
- (c) $\left[\frac{\partial y_i}{\partial x_j} \right] = \left[\frac{\partial f_i}{\partial x_j} \right] = - \left[\frac{\partial F_k}{\partial y_i} \right]^{-1} \left[\frac{\partial F_k}{\partial x_j} \right] = -B^{-1}A$, which is $n \times m$.

Note the similarity of the expression for $f(\mathbf{x})$ here to the expression for $f(x)$ in Case 1: here A replaces a and B^{-1} replaces $1/b$. So the function f in Case 1 is the special case in which $n = m = 1$.

Also note that, as in Case 1, because the function F is linear, its derivatives are the same over all of $\mathbb{R}^m \times \mathbb{R}^n$, so we don't need to say where they're being evaluated. Therefore the linear function f defined implicitly by the equation $F(\mathbf{x}, \mathbf{y}) = \mathbf{c} = F(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is the same for any $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, and the differential, which gives $\Delta \mathbf{y}$ as a linear function of $\Delta \mathbf{x}$,

$$\Delta \mathbf{y} = -B^{-1}A\Delta \mathbf{x}. \quad (9)$$

holds *exactly*, not just approximately, and holds for all $\Delta \mathbf{x} \in \mathbb{R}^m$, not just "small" $\Delta \mathbf{x}$.

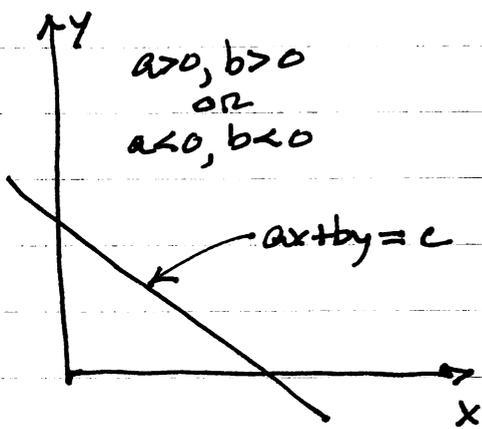


FIGURE 1

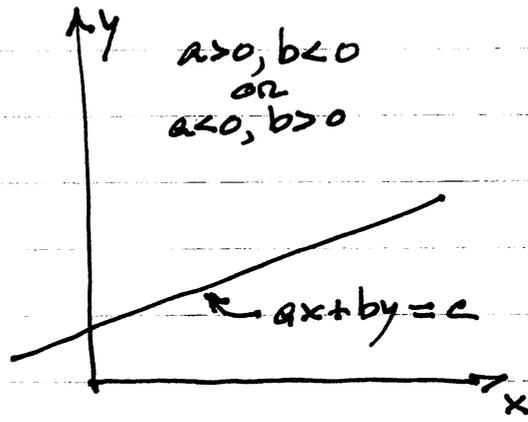


FIGURE 2

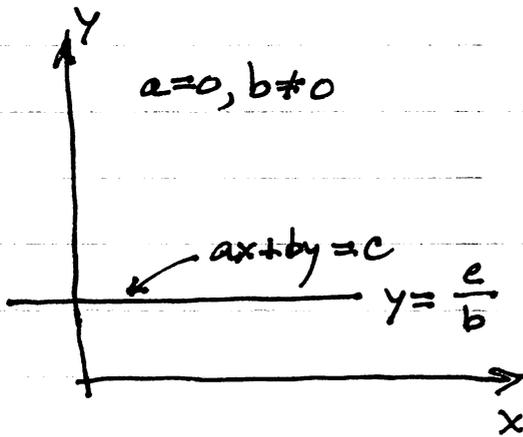


FIGURE 3

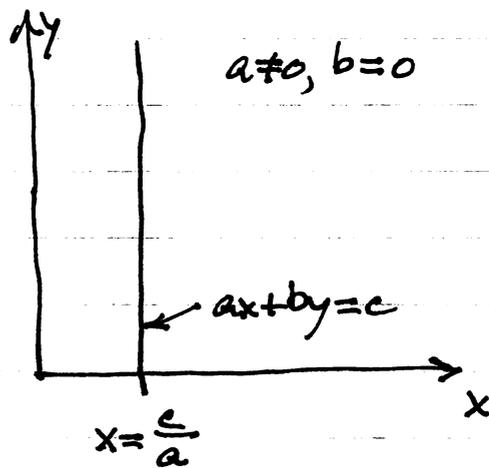


FIGURE 4

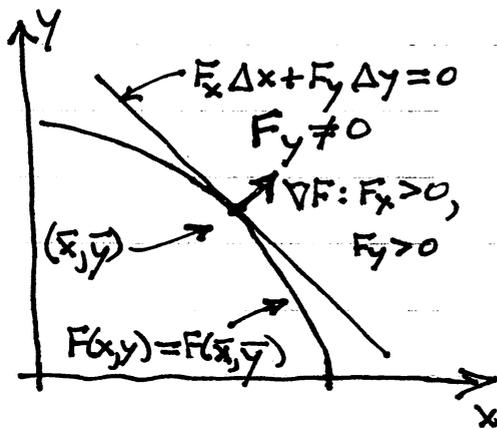


FIGURE 5

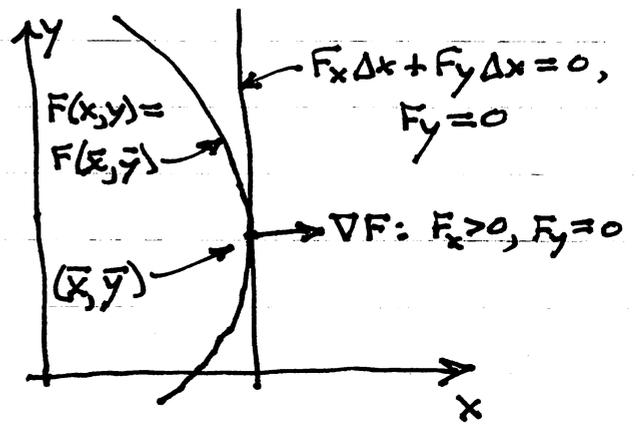


FIGURE 6

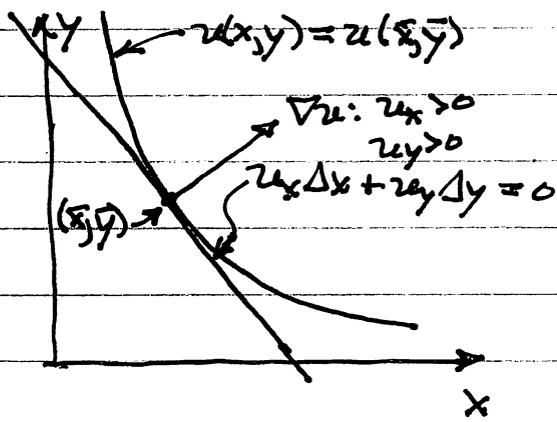


FIGURE 7

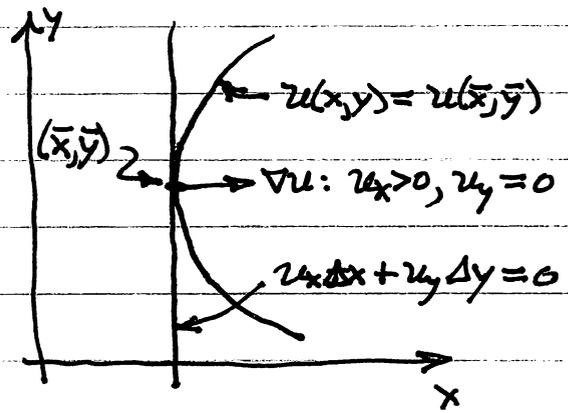


FIGURE 8