A question about Spivak's proof of the implicit function theorem

1 A correct proof of the implicit function theorem

First I'll write what I think is a correct proof of the implicit function theorem.

Notation: Throughout these notes I'll make use of block notation. So for example, if $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, then $\begin{bmatrix} x \\ y \end{bmatrix}$ is a vector in $\mathbb{R}^{n+m}$. If $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$, for example, I'll write $f(x, y)$ rather than the more correct $f(\begin{bmatrix} x \\ y \end{bmatrix})$. And $\frac{\partial f(x,y)}{\partial y}$ is an $m \times m$ matrix. It is the derivative (also called the Jacobian) of function $y \mapsto f(x, y)$ (with x held fixed).

Initial comments: Roughly speaking, the implicit function theorem gives us a guarantee that a system of m equations with m unknowns has a unique solution (under certain assumptions, at least). The equations are allowed to be nonlinear. To be a bit more precise, suppose that $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is continuously differentiable and $f(a, b) = 0$. The implicit function theorem tells us that if the matrix $\frac{\partial f(a,b)}{\partial y}$ is invertible, then for any point x near a, there is a unique point y near b such that $f(x, y) = 0$.

This fact seems plausible from the approximation

$$f(a + \Delta x, b + \Delta y) \approx f(a, b) + \frac{\partial f(a, b)}{\partial x}\Delta x + \frac{\partial f(a, b)}{\partial y}\Delta y.$$

Setting the right hand side equal to 0, we find that we can solve for Δy if $\frac{\partial f(a,b)}{\partial y}$ is invertible. (This intuitive argument is a great example of local linear approximation, which is the key idea of calculus.)

We can use the inverse function theorem to give a short, rigorous proof of the implicit function theorem. The idea behind the proof is to introduce the function $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined by $F(x, y) = \begin{bmatrix} x \\ f(x, y) \end{bmatrix}$. If F happens to be invertible, then for any $x \in \mathbb{R}^n$ the point $\begin{bmatrix} x \\ 0 \end{bmatrix} \in \mathbb{R}^{n+m}$ is the image under F of some point $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{n+m}$. The fact that $F(x, y) = \begin{bmatrix} x \\ 0 \end{bmatrix}$ tells us that $f(x, y) = 0$. To make this proof strategy work, we will need to be careful about the fact that F might only be locally invertible.

Theorem 1 (Implicit Function Theorem)**.** *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a continuously differentiable function. Suppose that $a \in \mathbb{R}^n, b \in \mathbb{R}^m$ and*

$$f(a, b) = 0 \quad \textit{and} \quad \frac{\partial f(a, b)}{\partial y} \textit{ is invertible.}$$

Then there exist open sets $A \ni a$ and $B \ni b$ such that:

1. *For each $x \in A$ there is a unique point $y \in B$ which satisfies $f(x, y) = 0$.*
2. *The function $g : A \to B$ defined implicitly by the equation*

$$f(x, g(x)) = 0 \quad \textit{for all } x \in A \tag{1}$$

is continuously differentiable.

Proof. We will apply the inverse function theorem to the function $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined by

$$F(x, y) = \begin{bmatrix} x \\ f(x, y) \end{bmatrix}$$

for all $x \in \mathbb{R}^n, y \in \mathbb{R}^m$. Note that F is continuously differentiable, and the Jacobian matrix

$$F'(a, b) = \begin{bmatrix} I & 0 \\ \frac{\partial f(a,b)}{\partial x} & \frac{\partial f(a,b)}{\partial y} \end{bmatrix}$$

is invertible. (One way to see this is to check that the null space of $F'(a, b)$ is trivial.) By the inverse function theorem, there exists an open set S containing $\begin{bmatrix} a \\ b \end{bmatrix}$ and an open set W containing $F(a, b) = \begin{bmatrix} a \\ 0 \end{bmatrix}$ such that the restriction of F to S is an invertible mapping from S onto W, and the inverse function $H : W \to S$ is continuously differentiable. This function $H : W \to S$ has the form

$$H(x, y) = \begin{bmatrix} x \\ h(x, y) \end{bmatrix}$$

and the function $h : W \to S$ is continuously differentiable.

We can take S to have the form $S = V \times B$ where $V \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ are open sets. Let $A = \{x \in \mathbb{R}^n \mid \begin{bmatrix} x \\ 0 \end{bmatrix} \in W\}$, which is an open set.

Claim: If $x \in A$, then there exists a unique point $y \in B$ such that $f(x, y) = 0$.

Explanation: Let $x \in A$. Then $\begin{bmatrix} x \\ 0 \end{bmatrix} \in W$, and the point $H(x, 0) = \begin{bmatrix} x \\ h(x, 0) \end{bmatrix}$ is mapped to $\begin{bmatrix} x \\ 0 \end{bmatrix}$ by F. It follows that $f(x, h(x, 0)) = 0$. So we can take $y = h(x, 0)$.

To see that y is unique, suppose that $y_1, y_2 \in B$ and $f(x, y_1) = f(x, y_2) = 0$. Then $F(x, y_1) = F(x, y_2) = \begin{bmatrix} x \\ 0 \end{bmatrix}$. But, the restriction of F to $S = V \times B$ is one-to-one. It follows that $y_1 = y_2$.

In conclusion, if $x \in A$, then $g(x) = h(x, 0)$ is the unique point in B which satisfies $f(x, g(x)) = 0$. The function $g : A \to B$ is continuously differentiable because h is.

□

Comments:

- To see that the matrix $F'(a, b)$ is invertible, let's show that it has a trivial null space. Suppose that

$$\begin{bmatrix} I & 0 \\ \frac{\partial f(a,b)}{\partial x} & \frac{\partial f(a,b)}{\partial y} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

  Then $Iu + 0v = 0 \implies u = 0$. The equation

$$\frac{\partial f(a, b)}{\partial x} u + \frac{\partial f(a, b)}{\partial y} v = 0$$

  then implies that $v = 0$.

- One way to see that the set A is open is to note that A is the preimage of W under the continuous mapping $x \mapsto \begin{bmatrix} x \\ 0 \end{bmatrix}$.

- Is it clear that we can take S to have the form $S = V \times B$ where $V \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ are open sets? If S does not have that form, it is at least true that there are open sets $V \ni a$ and $B \ni b$ such that $V \times B \subset S$. The set $\tilde{W} = F(V \times B)$ is open. But at this point we can rename so that $V \times B$ is now called S, and $\tilde{W}$ is now called W, and we can proceed with the proof. (If we prefer not to do any renaming, we could rephrase the rest of the proof in terms of $\tilde{S}$ and $\tilde{W}$.)

2 Spivak's proof of the implicit function theorem

Spivak's proof of the implicit function theorem is shown in figure 1. Notice that Spivak defines the set A differently than how I defined A above. The way Spivak defines A, I don't think we can assume that if $x \in A$ then $\begin{bmatrix} x \\ 0 \end{bmatrix} \in W$. However, when Spivak says "in other words we can define $g(x) = k(x, 0)$" he seems to be assuming that if $x \in A$ then $\begin{bmatrix} x \\ 0 \end{bmatrix} \in \operatorname{dom} k = W$. **Question:** Is this not an error in Spivak's proof?

For example, we could have $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^2 + y^2 - 25$$

and $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$F(x, y) = \begin{bmatrix} x \\ x^2 + y^2 - 25 \end{bmatrix}.$$

The point $\begin{bmatrix} a \\ b \end{bmatrix}$ could be $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$, and the sets A and B introduced in Spivak's proof could be $A = (2, 5)$ and $B = (3, 5)$. The set $W = \{\begin{bmatrix} x \\ f(x, y) \end{bmatrix} \mid x \in (2, 5), y \in (3, 5)\}$ is visualized in figure 2. We can see that if $x > 4$ then $\begin{bmatrix} x \\ 0 \end{bmatrix} \notin W$. And if $x > 4$ then there is no point $y \in B = (3, 5)$ such that $f(x, y) = 0$. So in this example it is not true that if $x \in A$ then there exists $y \in B$ such that $f(x, y) = 0$.

2-12 Theorem (Implicit Function Theorem). *Suppose $f\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ is continuously differentiable in an open set containing (a,b) and $f(a,b) = 0$. Let M be the $m \times m$ matrix*

$$(D_{n+j}f^i(a,b)) \qquad 1 \leq i, j \leq m.$$

If $\det M \neq 0$, there is an open set $A \subset \mathbf{R}^n$ containing a and an open set $B \subset \mathbf{R}^m$ containing b, with the following property: for each $x \in A$ there is a unique $g(x) \in B$ such that $f(x,g(x)) = 0$. The function g is differentiable.

42 *Calculus on Manifolds*

Proof. Define $F\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$ by $F(x,y) = (x,f(x,y))$. Then $\det F'(a,b) = \det M \neq 0$. By Theorem 2-11 there is an open set $W \subset \mathbf{R}^n \times \mathbf{R}^m$ containing $F(a,b) = (a,0)$ and an open set in $\mathbf{R}^n \times \mathbf{R}^m$ containing (a,b), which we may take to be of the form $A \times B$, such that $F\colon A \times B \to W$ has a differentiable inverse $h\colon W \to A \times B$. Clearly h is of the form $h(x,y) = (x,k(x,y))$ for some differentiable function k (since F is of this form). Let $\pi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ be defined by $\pi(x,y) = y$; then $\pi \circ F = f$. Therefore

$$\begin{aligned} f(x,k(x,y)) &= f \circ h(x,y) = (\pi \circ F) \circ h(x,y) \\ &= \pi \circ (F \circ h)(x,y) = \pi(x,y) = y. \end{aligned}$$

Thus $f(x,k(x,0)) = 0$; in other words we can define $g(x) = k(x,0)$. ∎

Figure 1: Spivak's proof of the implicit function theorem.

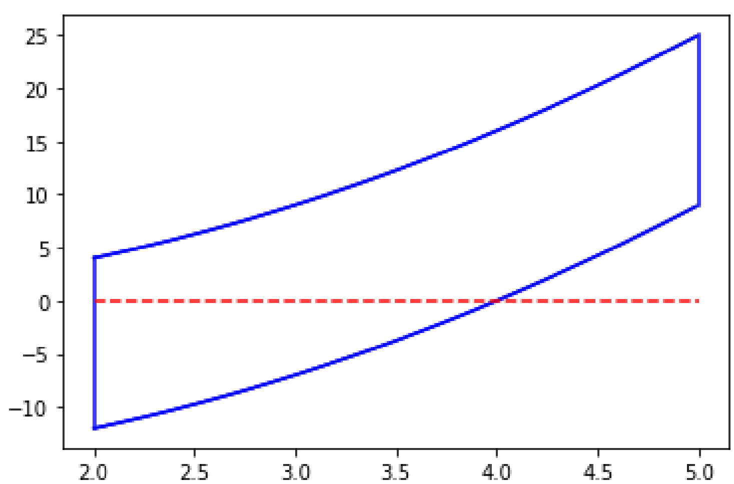

Figure 2: The set W in Spivak's proof when $f(x,y) = x^2 + y^2 - 25$ and $A = (2,5), B = (3,5)$. If $x \in A$ and $x > 4$, there does not exist $y \in B$ such that $f(x,y) = 0$.