Inverse vs Implicit function theorems - MATH 402/502 - Spring 2017

Instructor: C. Pereyra

On Tuesday April 25, I stated the Inverse Function and Implicit Function Theorems for you. I gave you some intuition for both theorems (steaming from considering linearizations of the problems, see below). I briefly discussed a few examples on where an implicit function theorem could be useful:

(1) The inverse function problem can be turned into an implicit function theorem (more in the notes).
(2) An exact differential equation can be turned into an implicit function problem so existence and uniqueness of a solution is a direct implication of these theorems.
(3) Geometric regions can be defined via systems of equations, being able to locally describe the region as the graph of a function is very important.

Tao presents a proof of the Inverse function theorem, and deduces from it the implicit function theorem (a less general version than ours, $m = 1$). As it turns out these two theorems are equivalent in the sense that one could have chosen to prove the general Implicit Function Theorem ($0 < m \in \mathbb{N}$) and deduce the Inverse Function Theorem from it (we need $m = 2n$), I did not have time on Tuesday to show you this. I did show you how to get from the inverse function theorem the general implicit function theorem. In these notes I present both arguments in detail.

Inverse Funtion Theorem

The inverse function theorem gives conditions on a differentiable function so that locally near a base point we can guarantee the existence of an inverse function that is differentiable at the image of the base point, furthermore we have a formula for this derivative: the derivative of the function at the image of the base point is the reciprocal of the derivative of the function at the base point. (See Tao's Section 6.7.)

Theorem 0.1 (Inverse Funtion Theorem)**.** *Let E be an open subset of $\mathbb{R}^n$, and let $f : E \to \mathbb{R}^n$ be a continuously differentiable function on E. Assume $x_0 \in E$ (the base point) and $f'(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then there exists an open set $U \subset E$ containing x_0, and an open set $V \subset \mathbb{R}^n$ containing $f(x_0)$ (the image of the base point), such that f is a bijection from U to V. In particular there is an inverse map $f^{-1} : V \to U$. Moreover f^{-1} is differentiable at y_0 and*

$$(f^{-1})'(y_0) = (f'(x_0))^{-1}. \tag{0.1}$$

When $n = 1$ this is the familiar one-variable inverse function theorem (Theorem 10.4.2) that we discussed in Math 401.

Heuristics: To understand where formula (2.2) comes from, it is illuminating to consider the linear approximation y in $\mathbb{R}^n$ to $f(x)$ near the base point x_0

$$y = y_0 + f'(x_0)(x - x_0).$$

We can solve for x in terms of y provided $f'(x_0)$ is invertible, which is our driving hypothesis,

$$x = x_0 + (f'(x_0))^{-1}(y - y_0).$$

This time we expect x to be a linear approximation to $f^{-1}(y)$ near y_0, in which case we will conclude that $(f^{-1})'(y_0) = (f'(x_0))^{-1}$, the formula in the theorem.

1. Implicit Function Theorem

The implicit function theorem gives sufficent conditions on a function F so that the equation $F(x,y)=0$ can be solved for y in terms of x (or solved for x in terms of y) locally near a base point (x_0,y_0) that satisfies the same equation $F(x_0,y_0)=0$.

Here is a slightly different version of the implicit function theorem I stated in class on Tuesday, it is designed to be able to deduce the inverse function theorem from it.

Theorem 1.1 (Implicit Function Theorem I)**.** *Let m,n be positive integers. Let A be an open subset of $\mathbb{R}^{n+m}$, and let $F:A\to\mathbb{R}^m$ be a continuously differentiable function on A. Let $(x_0,y_0)\in A$ such that $F(x_0,y_0)=0$. Assume that $D_YF(x_0,y_0)$ is invertible*[1]*. Then there are open sets $U\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^m$ such that $x_0\in U$, $y_0\in V$, and there is a function $g:U\to V$ differentiable at x_0 such that $(x,g(x))\in A$ and $F(x,g(x))=0$ for all $x\in U$. Moreover*

$$g'(x_0)=-\big(D_YF(x_0,y_0)\big)^{-1}D_XF(x_0,y_0). \tag{1.1}$$

Heuristics: To understand where formula (1.1) comes from, it is illuminating to use a linear approximation z in $\mathbb{R}^m$ to $F(x,y)$ near the base point (x_0,y_0) with $z_0=F(x_0,y_0)=0$,

$$z=z_0+F'(x_0,y_0)(x-x_0,y-y_0)\qquad\text{where}\qquad z\sim F(x,y).$$

Remember we are solving the equation $z=F(x,y)=0$ so this becomes

$$0=[D_XF(x_0,y_0),D_YF(x_0,y_0)](x-x_0,y-y_0)^t=D_XF(x_0,y_0)(x-x_0)+D_YF(x_0,y_0)(y-y_0).$$

We can solve the linear equation for y in terms of x provided $D_YF(x_0,y_0)$ is invertible which is the assumption to get,

$$y=y_0-(D_YF(x_0,y_0))^{-1}D_XF(x_0,y_0)(x-x_0).$$

We expect y to be a linear approximation to $g(x)$ for x near x_0, and hence $g'(x_0)=-(D_YF(x_0,y_0))^{-1}D_XF(x_0,y_0)$ as expected.

Proof of Inverse Funtion Theorem given Implicit Function Theorem I. We are given $f:E\to\mathbb{R}^n$ a continuously differentiable function on the open $E\subset\mathbb{R}^n$, $x_0\in E$ (the base point), let $y_0=f(x_0)$, and we are given that $f'(x_0):\mathbb{R}^n\to\mathbb{R}^n$ is invertible. Let us define $F:A\to\mathbb{R}^n$ where $A=E\times\mathbb{R}^n\subset\mathbb{R}^{n+n}=\mathbb{R}^{2n}$ by

$$F(x,y)=y-f(x),\qquad F(x_0,y_0)=0$$

We wish to apply the implicit function theorem to this function but we want to write x in terms of y, so the roles of x and y are interchanged in the statement of the implicit function theorem above and $m=n$, we must verify that $D_XF(x_0,y_0)$ is invertible. A calculation shows that

$$F'(x_0,y_0)=[-f'(x_0),I_{n\times n}],$$

where $I_{n\times n}$ is the $n\times n$ identity matrix. Hence $D_XF(x_0,y_0)=-f'(x_0)$, and is invertible by assumption on f, and $D_YF(x_0,y_0)=I_n$ the $n\times n$ identity matrix. We can apply the theorem, there are open sets $U\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^n$ such that $x_0\in U$, $y_0\in V$ and a a function $g:V\to U$ differentiable at y_0 such that $(g(y),y)\in A$ and $F(g(y),y)=0$ for all $y\in V$. Moreover g is differentiable at y_0 and $g'(y_0)=-\big(D_XF(x_0,y_0)\big)^{-1}D_YF(x_0,y_0)$. Notice that

[1]Note that $F'(x_0,y_0)=[D_XF(x_0,y_0),D_YF(x_0,y_0)]$ and $D_XF(x_0,y_0)$ is an $m\times n$ matrix, $D_YF(x_0,y_0)$ is an $m\times m$ matrix.

because of our choice of function F, $0 = F(g(y), y) = y - f(g(y))$, that is we conclude that $y = f(g(y))$ for all $y \in V$, that is $g = f^{-1}$ on V and

$$(f^{-1})'(y_0) = -\big(-f'(x_0)\big)^{-1} I_{n\times n} = (f'(x_0))^{-1}.$$

□

2. Inverse implies Implicit

Here is the version of the implicit function theorem that I stated on Tuesday. It can be deduced as a corollary of the inverse function theorem. Here I am including all the details, whereas on Tuesday I just gave you the general idea, without bothering with the pesky details.

Theorem 2.1 (Implicit Function Theorem II). *Let m, n be positive integers. Let E be an open subset of $\mathbb{R}^{n+m}$, and let $F : E \to \mathbb{R}^m$ be a continuously differentiable function on E. Let $(x_0, y_0) \in E$ such that $F(x_0, y_0) = 0$. Assume that $D_Y F(x_0, y_0)$ is invertible*[2]*. Then there are open sets $W \subset \mathbb{R}^n$ and $U \subset E \subset \mathbb{R}^{n+m}$ such that $x_0 \in W$, $(x_0, y_0) \in U$, and there is a function $G : W \to \mathbb{R}^m$ differentiable at x_0 such that*

$$\{(x, y) \in U : F(x, y) = 0\} = \{(x, G(x)) : x \in W\}. \tag{2.1}$$

Moreover

$$G'(x_0) = -\big(D_Y F(x_0, y_0)\big)^{-1} D_X F(x_0, y_0).$$

Proof of Implicit Funtion Theorem given Inverse Function Theorem. We are given a continuously differentiable function $F : E \to \mathbb{R}^m$, E open subset of $\mathbb{R}^{n+m}$, $(x_0, y_0) \in E$ such that $F(x_0, y_0) = 0$, and we are given that $D_Y F(x_0, y_0)$ is invertible. Define a new continuously differentiable function $f : E \to \mathbb{R}^{n+m}$ by

$$f(x, y) =: (x, F(x, y)),$$

to which we want to apply the inverse function theorem at the base point $z_0 = (x_0, y_0)$, and at its image point under f, $f(z_0) = (x_0, F(x_0, y_0)) = (x_0, 0)$. We must first verify that $f'(z_0)$ is invertible. The $(n+m)\times(n+m)$ matrix representation of the linear transformation $f'(z_0)$ is given by the following block matrix

$$f'(z_0) = \left[\begin{array}{c|c} I_{n\times n} & 0_{n\times m} \\ \hline D_X F(x_0, y_0) & D_Y F(x_0, y_0) \end{array}\right],$$

where $I_{n\times n}$ is the $n \times n$ identity matrix and $0_{n\times m}$ is the $n \times m$ zero matrix. One can operate with block matrices like if we had a 2×2 lower triangular matrix with no zeros in the

[2]Note that $F'(x_0, y_0) = [D_X F(x_0, y_0), D_Y F(x_0, y_0)]$ and $D_X F(x_0, y_0)$ is an $m \times n$ matrix, $D_Y F(x_0, y_0)$ is an $m \times m$ matrix.

diagonal[3] and verify that

$$\big(f'(z_0)\big)^{-1} = \left[\begin{array}{c|c} I_{n\times n} & 0_{n\times m} \\ \hline -(D_YF(x_0,y_0))^{-1}D_XF(x_0,y_0) & (D_YF(x_0,y_0))^{-1} \end{array}\right],$$

So $f'(z_0)$ is invertible and the inverse function theorem ensures that there exists an open set $U \subset E \subset \mathbb{R}^{n+m}$ containing z_0, and an open set $V \subset \mathbb{R}^{n+m}$ containing $f(z_0) = (x_0, F(x_0,y_0)) = (x_0, 0)$ (the image of the base point), such that f is a bijection from U to V. In particular there is an inverse map $f^{-1} : V \to U$. Moreover f^{-1} is differentiable at $f(z_0)$ and

$$(f^{-1})'(f(z_0)) = (f^{-1})'(x_0, 0) = \big(f'(x_0,y_0)\big)^{-1}. \tag{2.2}$$

Let us write f^{-1} in coordinates, $f^{-1} = (h, g)$ where $h = (h_1, h_2, \dots, h_n)$ and $g = (g_1, g_2, \dots, g_m)$, note that the domain of each of the g_i and h_i is V. Since by definition of f,

$$(h(x,y), F(h(x,y), g(x,y))) = f(f^{-1}(x,y)) = (x,y)$$

then $h(x,y) = x$ and $F(h(x,y), g(x,y)) = F(x, g(x,y)) = y$. Also g is differentiable at $(x_0, 0) \in V$. We now define $W \subset \mathbb{R}^n$ and function $G : W \to \mathbb{R}^m$ as follows

$$W := \{x \in \mathbb{R}^n : (x, 0) \in V\}, \qquad G(x) := g(x, 0).$$

Note that $x_0 \in W$, and since V is open in $\mathbb{R}^{n+m}$, W is also open in $\mathbb{R}^n$, finally G so defined is differentiable at x_0. We now prove (2.1), namely

$$A := \{(x,y) \in U : F(x,y) = 0\} = \{(x, G(x)) : x \in W\} =: B.$$

Assume $(x,y) \in U$ and $F(x,y) = 0$ then $f(x,y) = (x,0) \in V$ therefore $x \in W$. Applying f^{-1}, we see that $(x,y) = f^{-1}(x,0)$ in particular $y = g(x,0) = G(x)$, therefore $A \subset B$. All the steps can be reversed since f is a bijection from U to V to conclude that $B \subset A$.

By previous discussion $F(x, G(x)) = 0$ for all $x \in W$, F is differentiable at $(x_0, G(x_0)) = (x_0, y_0)$ and G is differentiable at x_0, the formula for the derivative follows by the chain rule:

$$D_XF(x_0,y_0) + D_YF(x_0,y_0)G'(x_0) = 0,$$

and by simple algebra we conclude that $G'(x_0) = -\big(D_YF(x_0,y_0)\big)^{-1} D_XF(x_0,y_0)$. □

Remark 2.2. Tao only proves the implicit function theorem when $m = 1$, but basically is the same argument I just presented, see Section 6.8 in Book II. With only the case $m = 1$ he couldn't go from implicit to inverse function theorem except in the case $n = m = 1$, because one needs the implicit function theorem from $\mathbb{R}^{2n}$ into $\mathbb{R}^n$, that is the case $n = m$.

Remark 2.3. Final comment, under the hypothesis that the functions are continuously differentiable one gets more differentiability than just at the base point, because the hypothesis of invertibility of the matrices $f'(x_0)$ or $D_YF(x_0,y_0)$ will persist in a neighborhood of x_0 or (x_0, y_0) respectively. The Theorems can be streghtened to say that f^{-1} is continuously differentiable on a possibly smaller open set V, and G is continuously differentiable on a possibly smaller open set W.

[3] In this case $1 \neq 0$ and $c \neq 0$: $A = \begin{bmatrix} 1 & 0 \\ b & c \end{bmatrix}$, $\qquad A^{-1} = \begin{bmatrix} 1 & 0 \\ -c^{-1}b & c^{-1} \end{bmatrix}$.