Professor John Nachbar
Econ 4111
September 11, 2009

# Inverse Function Theorem

## 1 A Review of the Univariate Case.

Although the univariate Inverse Function Theorem falls under the heading of "results you are supposed to know already," I will review it before stating its multivariate analog.

Let $f : \mathbb{R} \to \mathbb{R}$. If $Df(x) > 0$ on some open interval $U$ then the Mean Value theorem implies that $f$ must be strictly increasing on $U$, which implies that $f$ is invertible on $U$. That is, letting $V = f(U)$, there is a function $f^{-1} : V \to \mathbb{R}$ such that for any $x \in U$, $f^{-1}(f(x)) = x$, and for any $y \in V$, $f(f^{-1}(y)) = y$. Moreover, one can show that the inverse, $f^{-1}$, is continuous. And an analogous result holds if $Df(x) < 0$ on some interval $U$.

For univariate functions, the Inverse Function theorem strengthens the hypothesis "$f$ is differentiable" to "$f$ is $\mathcal{C}^r$," for some positive integer $r$, and in exchange gets a stronger result. If $f$ is $\mathcal{C}^r$ and if, at some point $x^*$, $Df(x^*) \neq 0$, say $Df(x^*) > 0$, then continuity implies that $Df$ is positive on some open set containing $x^*$, so the above argument implies that $f$ is strictly increasing and hence invertible on that neighborhood, and the inverse is continuous. Moreover, and this is harder to show, the inverse is itself differentiable, in fact $\mathcal{C}^r$. That is, by the Inverse Function theorem, if $Df$ is positive at a point then $f$ is invertible on some interval containing that point and the inverse is $\mathcal{C}^r$, instead of merely continuous.

Here are five examples, of increasing subtlety, that illustrate the univariate Inverse Function Theorem.

*Example* 1. Let $f(x) = e^x$, which is $\mathcal{C}^\infty$. Then $Df(x) = e^x > 0$ for all $x$. Hence I can take $U = \mathbb{R}$ and $V = f(\mathbb{R}) = (0, \infty)$. The inverse function is then $f^{-1} : V \to \mathbb{R}$ defined by $f^{-1}(y) = \ln(y)$, which is $\mathcal{C}^\infty$. $\square$

*Example* 2. Let $f(x) = x^2$, which is $\mathcal{C}^\infty$. Then $Df(x) = 2x$ for all $x$. At $x^* = 1$, $Df(1) = 2 > 0$, so the Inverse Function theorem guarantees existence of a local inverse. In fact, I can take $U = (0, \infty)$ and $V = f(U) = (0, \infty)$. The inverse function is then $f^{-1} : V \to \mathbb{R}$ defined by $f^{-1}(y) = \sqrt{y}$, which is $\mathcal{C}^\infty$. $\square$

*Example* 3. Again let $f(x) = x^2$. At $x = 0$, $Df(0) = 0$, so the hypothesis of the Inverse Function theorem is violated. $f$ does not, in fact, have an inverse in any neighborhood of 0. For example, if $U = (-1, 1)$ then $V = (0, 1)$. If $y = 0.01$ then we would have to have both $f^{-1}(0.01) = 0.1$ and $f^{-1}(0.01) = -0.1$. $\square$

*Example* 4. Let $f(x) = x^3$, which is $\mathcal{C}^\infty$. Then $Df(x) = 3x^2$ for all $x$. At $x = 0$, $Df(0) = 0$, so the hypothesis of the Inverse Function theorem is violated. But $f$ *is* invertible. In fact, I can take $U = V = \mathbb{R}$ and the inverse is $f^{-1} : \mathbb{R} \to \mathbb{R}$ defined by $f^{-1}(y) = y^{1/3}$. But $f^{-1}$ is not *differentiable* at 0. □

*Example* 5. Let

$$f(x) = \begin{cases} x + x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

This function is differentiable and $Df(0) = 1 > 0$ so one might think that the Inverse Function theorem would apply. But near 0 the graph of $f$ oscillates around the graph of the identity function $h(x) = x$ with increasing frequency (because of the $\sin(1/x)$ term) but rapidly decreasing amplitude (because of the $x^2$ term). Because of the oscillation, $f$ is not invertible on any neighborhood of 0, even though $Df(0) > 0$. The problem here is that $f$, while differentiable, is not *continuously* differentiable, and in particular $Df$ is not continuous at 0. □

## 2 The Inverse Function Theorem.

Recall that a function $f : X \to Y$ is 1-1 (one to one) iff for any $a, b \in X$, if $a \neq b$ then $f(a) \neq f(b)$. The function is *onto* iff $f(X) = Y$: for any $y \in Y$ there is an $x \in X$ such that $f(x) = y$. It is easy to see that $f$ has an inverse defined on all of $Y$ iff f is 1-1 and onto.

**Theorem 1** (Inverse Function theorem)**.** *Fix $x^* \in \mathbb{R}^N$, let $f : \mathbb{R}^N \to \mathbb{R}^N$ be $\mathcal{C}^r$, where r is a positive integer, let $y^* = f(x^*)$, and suppose $Df(x^*)$ is invertible. Then there are open sets $U, V \subseteq \mathbb{R}^N$, with $x^* \in U$ and $y^* \in V$, such that $Df(x)$ has full rank for all $x \in U$, $f$ maps $U$ 1-1 onto $V$, and hence has an inverse $f^{-1} : V \to U$. Furthermore, $f^{-1}$ is $\mathcal{C}^r$.*

**Proof.** The fact that $U$ can be taken such that $Df(x)$ has full rank for all $x \in U$ follows from the fact that $Df(x^*)$ has full rank, $f$ is continuously differentiable, and the determinant is continuous. The remainder of the theorem is difficult. For a $\mathcal{C}^1$ version, see Rudin (1976). ■

I can use the Chain Rule to computer $Df^{-1}(x)$ even if I cannot derive $f^{-1}$ explicitly. Fix $x \in U$ and let $y = f(x)$. Define $h : V \to V$ by $h(x) = f^{-1}(f(x))$. Then, by the Chain Rule,

$$Dh(x) = Df^{-1}(y)Df(x)$$

On the other hand, for any $x \in U$, $h(x) = x$, hence

$$Dh(x) = I,$$

where $I$ is the $N \times N$ identity matrix. Combining,

$$Df^{-1}(y)Df(x) = I. \tag{1}$$

Since $Df(x)$ has full rank for any $x \in U$, I have,

$$Df^{-1}(y) = [Df(x)]^{-1}. \tag{2}$$

In words, the derivative of the inverse is the inverse of the derivative.

Example 4 shows that $Df(x^*)$ being of full rank is not necessary for existence of an inverse. Equation 1 implies, however, that $Df(x^*)$ being of full rank is *necessary* for $Df^{-1}(x^*)$ to be differentiable. The Inverse Function theorem says, loosely, that this necessary condition is also *sufficient* for both the existence and the differentiability of $f^{-1}$.

You may sometimes see Equation 2 referred to as being the Inverse Function theorem. This is incorrect: Equation 2 is just a routine application of the Chain Rule. The Inverse Function theorem is what allows us to conclude that the inverse function $f^{-1}$ exists and is differentiable.

## References

RUDIN, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill, New York, third edn.