1 Overview.

If a function $f : \mathbb{R} \to \mathbb{R}$ is $C^1$ and if its derivative is strictly positive at some $x^* \in \mathbb{R}$, then, by continuity of the derivative, there is an open interval $U$ containing $x^*$ such that the derivative is strictly positive for any $x \in U$. The Mean Value Theorem then implies that $f$ is strictly increasing on $U$, and hence that $f$ maps $U$ 1-1 onto $f(U)$. Let $V = f(U)$. Hence $f : U \to V$ is invertible. A similar argument applies if the derivative is strictly negative.

The Inverse Function Theorem generalizes and strengthens the previous observation. If $f : \mathbb{R}^N \to \mathbb{R}^N$ is $C^r$ and if the matrix $Df(x^*)$ is invertible for some $x^* \in \mathbb{R}^N$, then there is an open set, $U \subseteq \mathbb{R}^N$, with $x^* \in U$, such that $f$ maps $U$ 1-1 onto $V = f(U)$. Hence $f : U \to V$ is invertible. Moreover, $V$ is open, the inverse function $f^{-1} : V \to U$ is $C^r$, and for any $x \in U$, setting $y = f(x)$,

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

The last fact allows us to compute $Df^{-1}(y)$ even when we can’t derive $f^{-1}$ analytically.

Here are four examples in $\mathbb{R}$.

**Example 1.** Let $f(x) = e^x$, which is $C^\infty$. 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 $C^\infty$.

Suppose, for example, that $x = 1$ then $y = e$. Since $Df^{-1}(y) = 1/y$, $Df^{-1}(e) = 1/e$. On the other hand, $Df(x) = e^x$, so $Df(1) = e$. Hence $Df^{-1}(y) = 1/Df(x)$, as claimed above. □

**Example 2.** Let $f(x) = x^2$, which is $C^\infty$. 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 $C^\infty$. □

**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 $f^{-1}(0.01) = -0.1$. □
Example 4. Let \( f(x) = x^3 \), which is \( C^\infty \). 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. \( \square \)

An illuminating, but more abstract, way to view the Inverse Function Theorem is the following. A linear isomorphism is a linear function \( h : \mathbb{R}^N \to \mathbb{R}^N \), \( h(x) = Ax \), such that the \( N \times N \) matrix \( A \) is invertible. A \( C^r \) diffeomorphism is a \( C^r \) function \( f : U \to V \), where both \( U \) and \( V \) are open, such that \( f \) maps \( U \) 1-1 onto \( V \) and the inverse function \( f^{-1} : V \to U \) is also \( C^r \). One can think of diffeomorphism as a generalization of linear isomorphism. The Inverse Function Theorem then says that if the linear approximation to \( f \) at \( x^* \) is an isomorphism then \( f \) is a diffeomorphism on an open set containing \( x^* \).

My proofs below follow those in Rudin (1976), Lang (1988), and Spivak (1965).

2 Some Necessary Preliminaries

The argument sketched above for \( N = 1 \) relied on the Mean Value Theorem. Unfortunately, the Mean Value Theorem does not strictly generalize to \( N > 1 \), as the following example illustrates.

Example 5. Define \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) by

\[
 f(x_1, x_2) = (e^{x_1} \cos(x_2), e^{x_1} \sin(x_2)).
\]

Note that \( f(0,0) = f(0,2\pi) \). Therefore, if the Mean Value Theorem held in \( \mathbb{R}^2 \), there would have to be an \( x^* \) for which \( Df(x^*) \) is the zero matrix. But

\[
 Df(x) = \begin{bmatrix}
 e^{x_1} \cos(x_2) & -e^{x_1} \sin(x_2) \\
 e^{x_1} \sin(x_2) & e^{x_1} \cos(x_2)
\end{bmatrix},
\]

which has full rank for all \( x \). \( \square \)

Fortunately, there is an implication of the Mean Value Theorem that does generalize and that is strong enough to imply invertibility of \( f \). Consider first the case where \( N = 1 \). Suppose that there is a number \( c \) such that \( |Df(x)| \leq c \) for all \( x \). By the Mean Value Theorem, for any \( x_1, x_0 \in \mathbb{R} \), there is an \( x \in \mathbb{R} \) such that \( f(x_1) - f(x_0) = Df(x)(x_1 - x_0) \), hence \( |f(x_1) - f(x_0)| = |Df(x)||x_1 - x_0| \), hence,

\[
 |f(x_1) - f(x_0)| \leq c|x_1 - x_0|.
\]

Theorem 1 below, generalizes this inequality to \( N \geq 1 \).

2The affine approximation to \( f \) at \( x^* \) is

\[
 H(x) = Df(x^*)(x - x^*) + f(x^*).
\]

The linear approximation is the linear portion of this, namely \( h(x) = Df(x^*)x \).
To state Theorem 1, I need to define a norm for matrices. I use the same norm that I used when defining what it means for a function to be $C^1$. Explicitly, for any $N \times N$ matrix $A$, define the norm of $A$ to be

$$\|A\| = \max_{x \in S} \|Ax\|,$$

where $S = \{x \in \mathbb{R}^N : \|x\| \leq 1\}$ is the solid unit ball centered at the origin in $\mathbb{R}^N$. Since $S$ is compact and the norm on $\mathbb{R}^N$ is continuous, $\|A\|$ is well defined and one can verify that this norm does, in fact, satisfy all the required properties of a norm.

For later reference, note that for any $x \neq 0$, since $x/\|x\| = 1$,

$${|Ax|} = \left| A \frac{x}{\|x\|} \right| \|x\| \leq \|A\| \|x\|. \tag{1}$$

I can now state and prove the generalized Mean Value Theorem.

**Theorem 1** (Generalized Mean Value Theorem). Suppose that $U \subseteq \mathbb{R}^N$ is convex and open, that $f : U \to \mathbb{R}^N$ is differentiable, and that there is a number $c > 0$ such that for all $x \in U$,

$$\|Df(x)\| \leq c.$$ 

Then for all $x_1, x_0 \in U$,

$$\|f(x_1) - f(x_0)\| \leq c\|x_1 - x_0\|.$$ 

**Proof.** Take any $x_1, x_0 \in U$. Define $g : \mathbb{R} \to \mathbb{R}^N$ by

$$g(t) = (1 - t)x_0 + tx_1.$$ 

Since $U$ is open and convex, there is an $\varepsilon > 0$ sufficiently small such that $g(t) \in U$, for $t \in (-\varepsilon, 1 + \varepsilon)$.

Define $h : (-\varepsilon, 1 + \varepsilon) \to \mathbb{R}^N$ by $h(t) = f(g(t))$. For ease of notation, let $x_t = g(t)$. Then, by the Chain Rule,

$$Dh(t) = Df(g(t))Dg(t) = Df(x_t)(x_1 - x_0).$$

Hence, by inequality 1,

$$\|Dh(t)\| \leq \|Df(x_t)\|\|x_1 - x_0\| \leq c\|x_1 - x_0\|.$$

Note that this holds for all $t \in (-\varepsilon, 1 + \varepsilon)$.

To complete the proof, I argue that there is a $t \in (0, 1)$ such that $\|f(x_1) - f(x_0)\| \leq \|Dh(t)\|$. To see this, define $\phi : (-\varepsilon, 1 + \varepsilon) \to \mathbb{R}$ by

$$\phi(t) = h(t) \cdot (h(1) - h(0)).$$

3
By the standard Mean Value Theorem on \( \mathbb{R} \), there is a \( t^* \in (0,1) \) such that
\[
\phi(1) - \phi(0) = D\phi(t^*),
\]
or
\[
\phi(1) - \phi(0) = Dh(t^*) \cdot (h(1) - h(0)).
\]
On the other hand, by direct substitution,
\[
\phi(1) - \phi(0) = Dh(t^*) \cdot (h(1) - h(0))
\]
Combining and applying the Schwartz inequality yields,
\[
\|h(1) - h(0)\|^2 \leq \|Dh(t^*)\|\|h(1) - h(0)\|.
\]
Dividing through by \( \|h(1) - h(0)\| \) yields
\[
\|h(1) - h(0)\| \leq \|Dh(t^*)\|.
\]
(if \( \|f(\hat{x}) - f(x)\| = 0 \), then the inequality holds trivially). Finally, the result follows by noting that \( h(1) = f(x_1) \) and \( h(0) = f(x_0) \). ■

The Inverse Function Theorem also requires that both \( U \) and \( V = f(U) \) are open. For \( N = 1 \), this can be established by using, among other things, the fact that the image of an interval under a continuous function is also an interval. This fact, however, has no natural analog in \( \mathbb{R}^N \). Therefore, the proof below instead establishes openness using the Contraction Mapping Theorem, which is discussed in the notes on fixed points. Briefly, let \((X,d)\) be a metric space and let \( \phi : X \to X \). \( \phi \) is a contraction iff there is an \( c \in (0,1) \) such that for all \( x, \hat{x} \in X \),
\[
d(\phi(x),\phi(\hat{x})) \leq c \, d(x,\hat{x}).
\]
A point \( x^* \in X \) is a fixed point of \( \phi \) iff \( \phi(x^*) = x^* \).

**Theorem 2** (Contraction Mapping Theorem). Let \((X,d)\) be a complete metric space and let \( \phi : X \to X \) be a contraction. Then \( \phi \) has a unique fixed point.

### 3 Statement and Proof of the Inverse Function Theorem

**Theorem 3** (Inverse Function Theorem). Let \( f : \mathbb{R}^N \to \mathbb{R}^N \) be \( C^r \), where \( r \) is a positive integer. For \( x^* \in \mathbb{R}^N \), suppose that \( Df(x^*) \) is invertible. Then there are open sets \( U, V \subseteq \mathbb{R}^N \), with \( x^* \in U \), such that \( f \) maps \( U \) 1-1 onto \( V = f(U) \). Hence \( f : U \to V \) is invertible. Moreover, the inverse function \( f^{-1} : V \to U \) is \( C^r \), and for any \( y \in V \), setting \( x = f^{-1}(y) \),
\[
Df^{-1}(y) = [Df(x)]^{-1}.
\]
Proof. Let \( y^* = f(x^*) \). For ease of notation, assume that \( Df(x^*) = I \), the \( N \times N \) identity matrix. For the more general case, define \( F(x) = [Df(x^*)]^{-1}f(x) \). Then \( DF(x^*) = I \). Apply the argument below to \( F \) to get a \( C^r \) inverse \( F^{-1} \), and note that \( f^{-1}(y) = F^{-1}([Df(x^*)]^{-1}y) \), which is \( C^r \) since \( F^{-1} \) is \( C^r \). In this sense, the case \( Df(x^*) = I \) is without loss of generality. For a similar reason, one can simplify notation by assuming \( x^* = y^* = 0 \), but I will not do so.

For any \( y \in \mathbb{R}^N \), define \( \phi_y : \mathbb{R}^N \to \mathbb{R}^N \) by

\[
\phi_y(x) = x - f(x) + y.
\]

For any \( y \), \( \phi_y(x) = x \) iff \( f(x) = y \). Below, I use the Contraction Mapping Theorem on \( \phi_y \) to find \( x \in f^{-1}(y) \).

Note next that for any \( y \), \( D\phi_y(x) = I - Df(x) \). Then, for any \( y \), \( \|D\phi_y(x^*)\| = 0 \). Since \( Df \) is continuous, there is an open set \( U \subseteq \mathbb{R}^N \) such that for any \( x \in U \), for any \( y \),

\[
\|D\phi_y(x)\| < \frac{1}{2}.
\]

(It is important that the right-hand side be smaller than 1; it is not important that it be, in particular, \( 1/2 \).) This implies that, for any \( x \in U \), \( Df(x) \) is invertible, since if \( Df(x) \) is singular then there is a \( z \in \mathbb{R}^N \) of norm 1 such that \( Df(x)z = 0 \), but then \( \|D\phi_y(x)\| \geq \|D\phi_y(x)z\| = \|z\| = 1 > 1/2 \). By Theorem 1, for any \( x, \hat{x} \in U \), and for any \( y \),

\[
\|\phi_y(\hat{x}) - \phi_y(x)\| \leq \frac{1}{2}\|\hat{x} - x\|. \tag{3}
\]

Moreover, again since \( Df \) is continuous, since the inverse operation on invertible matrices is continuous, and since \( \|[Df(x^*)]^{-1}\| = \|I\| = 1 \), I can take \( U \) such that for all \( x \in U \),

\[
\|[Df(x)]^{-1}\| \leq 2. \tag{4}
\]

(The fact that I take the bound to be 2, rather than, say, 1000, is not important; all that matters is that the bound exists.)

Let \( V = f(U) \). Inequality 3 implies that \( f \) must be 1-1 on \( U \): if \( f(\hat{x}) = f(x) \), then \( \phi_y(\hat{x}) - \phi_y(x) = \hat{x} - x \), which contradicts inequality 3. Since \( f \) maps \( U \) 1-1 onto \( V \), the inverse function \( f^{-1} : V \to U \) is well defined.

I must show that \( V \) is open. (Since \( U \) is essentially arbitrary, this will also establish that \( f \) is an open mapping: the image of any open set is open.) Take any \( \hat{y} \in V \). I must show that there is an open ball around \( \hat{y} \) contained in \( V \). Let \( \hat{x} = f^{-1}(\hat{y}) \); since \( f \) is 1-1, \( \hat{x} \) is well defined. Take any \( \delta \) such that \( \delta > 0 \) such that \( K = N_\delta(\hat{x}) \subseteq O \). Take the open ball around \( \hat{y} \) to be \( N_\delta(\hat{y}) \). I need to show that this open ball is a subset of \( V \). Take any \( y \in N_\delta(\hat{y}) \). I must show that \( y \in V \).
Inequality 3 implies that for any \( x \in K \),
\[
\|\phi_y(x) - \hat{x}\| \leq \|\phi_y(x) - \phi_y(\hat{x})\| + \|\phi_y(\hat{x}) - \hat{x}\|
\leq \frac{1}{2} \|x - \hat{x}\| + \|y - \hat{y}\|
\leq \delta.
\]
Thus \( \phi_y \) maps \( K \) into itself. Since \( K \) is complete, and since by inequality 3, \( \phi_y \) is a contraction, the Contraction Mapping Theorem implies that there is an \( x \in K \) such that \( \phi_y(x) = x \), hence there is an \( x \in K \) such that \( f(x) = y \). Since \( K \subseteq U \), this implies \( y \in V \), as was to be shown.

The last step is to show that \( f^{-1} \) is \( C' \). I show first that \( f^{-1} \) is continuous. Fix any \( y \in V \), take any other \( \hat{y} \in V \), and set \( x = f^{-1}(y) \), \( \hat{x} = f^{-1}(\hat{y}) \). Then
\[
\|f^{-1}(\hat{y}) - f^{-1}(y)\| = \|\hat{x} - x\|
= \|\phi_{y^*}(\hat{x}) + \hat{y} - \phi_{y^*}(x) - y\|
\leq \|\phi_{y^*}(\hat{x}) - \phi_{y^*}(x)\| + \|\hat{y} - y\|
\leq \frac{1}{2} \|\hat{x} - x\| + \|\hat{y} - y\|,
\]
where the last inequality comes from inequality 3. Rearranging, and substituting \( x = f^{-1}(y) \), \( \hat{x} = f^{-1}(\hat{y}) \),
\[
\|f^{-1}(\hat{y}) - f^{-1}(y)\| \leq 2\|\hat{y} - y\|, \tag{5}
\]
which implies that \( f^{-1} \) is continuous.

I next claim that \( f^{-1} \) is differentiable on \( V \), and that, moreover, for any \( y \in V \), setting \( x = f^{-1}(y) \),
\[
Df^{-1}(y) = [Df(x)]^{-1}.
\]
Consider any \( \hat{y} \in V \), \( \hat{y} \neq y \). Set \( \hat{x} = f^{-1}(\hat{y}) \). Then, since
\[
f^{-1}(\hat{y}) - f^{-1}(y) - [Df(x)]^{-1}(\hat{y} - y) = x - \hat{x} - [Df(x)]^{-1}(f(\hat{x}) - f(x))
= -[Df(x)]^{-1}(f(\hat{x}) - f(x) - Df(x)(x - \hat{x})),
\]
and since \( \|[Df(x)]^{-1}\| \leq 2 \) (by inequality 4),
\[
\|f^{-1}(\hat{y}) - f^{-1}(y) - [Df(x)]^{-1}(\hat{y} - y)\| \leq 2\|f(\hat{x}) - f(x) - Df(x)(x - \hat{x})\|.
\]
Therefore, to show that
\[
\lim_{\hat{y} \to y} \frac{\|f^{-1}(\hat{y}) - f^{-1}(y) - [Df(x)]^{-1}(\hat{y} - y)\|}{\|\hat{y} - y\|} = 0,
\]

it suffices to show that
\[
\lim_{\hat{y} \to y} \frac{\|f(\hat{x}) - f(x) - Df(x)(x - \hat{x})\|}{\|\hat{y} - y\|} = 0.
\]
But since $f$ is differentiable,

$$\lim_{\hat{x} \to x} \frac{\|f(\hat{x}) - f(x) - Df(x)(x - \hat{x})\|}{\|\hat{x} - x\|} = 0.$$  

By continuity of $f^{-1}$, $\hat{y} \to y$ implies $\hat{x} \to x$. In addition, inequality 5 establishes that,

$$\|\hat{x} - x\| \leq 2.$$  

Hence,

$$\lim_{\hat{y} \to y} \frac{\|f(\hat{x}) - f(x) - Df(x)(x - \hat{x})\|}{\|\hat{y} - y\|} = \lim_{\hat{y} \to y} \frac{\|f(\hat{x}) - f(x) - Df(x)(x - \hat{x})\|}{\|\hat{x} - x\|} \cdot \frac{\|\hat{x} - x\|}{\|\hat{y} - y\|} = 0,$$

as was to be proved.

Since

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

for any $y \in V$, since $Df$ is $C^{r-1}$ (since $f$ is $C^r$), and since the inverse operation on invertible matrices is smooth, this establishes that $Df^{-1}$ is $C^{r-1}$. Hence $f^{-1}$ is $C^r$, as was to be shown. □

**Remark 1.** Equation 2 is consistent with the Chain Rule. Explicitly, suppose that we are simply told that $f^{-1}$ exists and is differentiable. Define $h : V \to V$ by $h(x) = f^{-1}(f(x))$. Then, by the Chain Rule,

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

On the other hand, since $h(x) = x$,

$$Dh(x) = I,$$

where $I$ is the $N \times N$ identity matrix. Combining, and setting $y = f(x)$,

$$Df^{-1}(y)Df(x) = I. \quad (6)$$

Rearranging this yields equation 2. □

**Remark 2.** Example 4 shows that $Df(x^*)$ being of full rank is not necessary for existence of an inverse. Equation 2 implies, however, that $Df(x^*)$ being of full rank is necessary for $Df^{-1}(x^*)$ to be differentiable. The Inverse Function Theorem thus says that this necessary condition for a *differentiable* local inverse is also sufficient. □

**Remark 3.** The above proof of the Inverse Function Theorem generalizes easily to infinite dimensional spaces. See Lang (1988). A common alternative proof, used, for example, in Spivak (1965), replaces the contraction mapping argument with an argument based on existence of maxima for continuous functions defined on compact sets. This approach relies on Heine-Borel, which does not generalize. □
References

