MA2321—Analysis in Several Variables
School of Mathematics, Trinity College
Michaelmas Term 2017
Lecture 25 (December 4, 2017)

David R. Wilkins

11.3. The Inverse Function Theorem

The *Inverse Function Theorem* ensures that, for a continuously differentiable function of several real variables, mapping an open set in one Euclidean space into a Euclidean space of the same dimension, the invertibility of the derivative of the function at a given point is sufficient to ensure the local invertibility of that function around the given point, and moreover ensures that the inverse function is also locally a continuously differentiable function.

The proof uses the method of successive approximations, using a convergence criterion for infinite sequences of points in Euclidean space that we established in Proposition 11.4.

Theorem 11.5 (Inverse Function Theorem)

Let $\varphi\colon X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set X in n-dimensional Euclidean space $\mathbb{R}^n$ and mapping X into $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Suppose that the derivative $(D\varphi)_{\mathbf{p}}\colon \mathbb{R}^n \to \mathbb{R}^n$ of the map φ at the point $\mathbf{p}$ is an invertible linear transformation. Then there exists an open set W in $\mathbb{R}^n$ and a continuously differentiable function $\mu\colon W \to X$ that satisfies the following conditions:—

(i) $\mu(W)$ *is an open set in* $\mathbb{R}^n$ *contained in* X, *and* $\mathbf{p} \in \mu(W)$;

(ii) $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ *for all* $\mathbf{y} \in W$.

Proof

We may assume, without loss of generality, that $\mathbf{p} = \mathbf{0}$ and $\varphi(\mathbf{p}) = \mathbf{0}$. Indeed the result in the general case can then be deduced by applying the result in this special case to the function that sends $\mathbf{z}$ to $\varphi(\mathbf{p}+\mathbf{z}) - \varphi(\mathbf{p})$ for all $\mathbf{z} \in \mathbb{R}^n$ for which $\mathbf{p}+\mathbf{z} \in X$. Now $(D\varphi)_{\mathbf{0}} \colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, by assumption. Let $T = (D\varphi)_{\mathbf{0}}^{-1}$, and let $\psi \colon X \to \mathbb{R}^n$ be defined such that

$$\psi(\mathbf{x}) = \mathbf{x} - T(\varphi(\mathbf{x}))$$

for all $\mathbf{x} \in X$.

Now the derivative of any linear transformation at any point is equal to that linear transformation (see Lemma 9.2). It follows from the Chain Rule (Proposition 9.8) that the derivative of the composition function $T \circ \varphi$ at any point $\mathbf{x}$ of X is equal to $T(D\varphi)_{\mathbf{x}}$. It follows that $(D\psi)_{\mathbf{x}} = I - T(D\varphi)_{\mathbf{x}}$ for all $\mathbf{x} \in X$, where I denotes the identity operator on $\mathbb{R}^n$. In particular $(D\psi)_{\mathbf{0}} = I - T(D\varphi)_{\mathbf{0}} = 0$. It then follows from Proposition 9.10 that there exists a positive real number δ such that

$$|\psi(\mathbf{u}) - \psi(\mathbf{v})| \leq \tfrac{1}{2}|\mathbf{u} - \mathbf{v}|$$

whenever $|\mathbf{u}| < \delta$ and $|\mathbf{v}| < \delta$.

Now $\psi(\mathbf{0}) = \mathbf{0}$. It follows from the inequality just proved that $|\psi(\mathbf{x})| \leq \frac{1}{2}|\mathbf{x}|$ whenever $|\mathbf{x}| < \delta$.

Let W be the open set in $\mathbb{R}^n$ defined so that

$$W = \{\mathbf{y} \in \mathbb{R}^n : |T(\mathbf{y})| < \tfrac{1}{2}\delta\},$$

and let $\mu_0, \mu_1, \mu_2, \ldots$ be the infinite sequence of functions from W to $\mathbb{R}^n$ defined so that $\mu_0(\mathbf{y}) = 0$ for all $\mathbf{y} \in W$ and

$$\mu_j(\mathbf{y}) = \mu_{j-1}(\mathbf{y}) + T(\mathbf{y} - \varphi(\mu_{j-1}(\mathbf{y})))$$

for all positive integers j. Now $\varphi(\mathbf{0}) = \mathbf{0}$. It follows that if $\mu_{j-1}(\mathbf{0}) = \mathbf{0}$ for some positive integer j then $\mu_j(\mathbf{0}) = \mathbf{0}$. It then follows by induction on j that $\mu_j(\mathbf{0}) = \mathbf{0}$ for all non-negative integers j.

We shall prove that there is a well-defined function $\mu\colon W \to \mathbb{R}^n$ defined such that $\mu(\mathbf{y}) = \lim\limits_{j \to +\infty} \mu_j(\mathbf{y})$ and that this function μ is a local inverse for φ defined on the open set W that satisfies the required properties.

Let $\mathbf{y} \in W$ and let $\mathbf{x}_j = \mu_j(\mathbf{y})$ for all non-negative integers j. Then $\mathbf{x}_0 = \mathbf{0}$ and

$$\begin{aligned} \mathbf{x}_j &= \mathbf{x}_{j-1} + T(\mathbf{y} - \varphi(\mathbf{x}_{j-1})) \\ &= \psi(\mathbf{x}_{j-1}) + T\mathbf{y} \end{aligned}$$

for all positive integers j. Now we have already shown that $|\psi(\mathbf{x})| \leq \frac{1}{2}|\mathbf{x}|$ whenever $|\mathbf{x}| < \delta$. Also the definition of the open set W ensures that $|T\mathbf{y}| < \frac{1}{2}\delta$. It follows that if $|\mathbf{x}_{j-1}| < \delta$ then

$$|\mathbf{x}_j| \leq |\psi(\mathbf{x}_{j-1})| + |T\mathbf{y}| \leq \tfrac{1}{2}|\mathbf{x}_{j-1}| + |T\mathbf{y}| < \tfrac{1}{2}\delta + |T\mathbf{y}| < \delta.$$

It follows by induction on j that $|\mathbf{x}_j| < \frac{1}{2}\delta + |T\mathbf{y}|$ for all non-negative integers j. Also

$$\begin{aligned} \mathbf{x}_{j+1} - \mathbf{x}_j &= \mathbf{x}_j - \mathbf{x}_{j-1} - T(\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_{j-1})) \\ &= \psi(\mathbf{x}_j) - \psi(\mathbf{x}_{j-1}) \end{aligned}$$

for all positive integers j.

But $|\mathbf{x}_j| < \delta$ and $|\mathbf{x}_{j-1}| < \delta$ and therefore

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| = |\psi(\mathbf{x}_j) - \psi(\mathbf{x}_{j-1})| \leq \tfrac{1}{2}|\mathbf{x}_j - \mathbf{x}_{j-1}|$$

for all positive integers j. It then follows from Lemma 11.4 that the infinite sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is convergent. Now $\mathbf{x}_j = \mu_j(\mathbf{y})$ for all non-negative integers j, where $\mathbf{y}$ is an arbitrary element of the open set W. The convergence result just obtained therefore guarantees that there is a well-defined function $\mu\colon W \to \mathbb{R}^n$ which satisfies

$$\mu(\mathbf{y}) = \lim_{j \to +\infty} \mu_j(\mathbf{y})$$

for all $\mathbf{y} \in W$. Moreover $|\mu_j(\mathbf{y})| < \frac{1}{2}\delta + |T\mathbf{y}|$ for all positive integers j and for all $\mathbf{y} \in W$, and therefore

$$|\mu(\mathbf{y})| \leq \tfrac{1}{2}\delta + |T\mathbf{y}| < \delta$$

for all $\mathbf{y} \in W$.

Next we prove that $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$. Now

$$\begin{aligned} \mu(\mathbf{y}) &= \lim_{j \to +\infty} \mu_j(\mathbf{y}) = \lim_{j \to +\infty} (\mu_{j-1}(\mathbf{y}) + T(\mathbf{y} - \varphi(\mu_{j-1}(\mathbf{y})))) \\ &= \mu(\mathbf{y}) + T(\mathbf{y} - \varphi(\mu(\mathbf{y}))) \end{aligned}$$

It follows that $T(\mathbf{y} - \varphi(\mu(\mathbf{y}))) = \mathbf{0}$. But $T = (D\varphi)_{\mathbf{0}}^{-1}$. It follows that

$$\mathbf{y} - \varphi(\mu(\mathbf{y})) = (D\varphi)_{\mathbf{0}}(T(\mathbf{y} - \varphi(\mu(\mathbf{y})))) = (D\varphi)_{\mathbf{0}}(\mathbf{0}) = \mathbf{0}.$$

Thus $\mathbf{y} = \varphi(\mu(\mathbf{y}))$ for all $\mathbf{y} \in W$. Also $\mu_j(\mathbf{0}) = \mathbf{0}$ for all non-negative integers j, and therefore $\mu(\mathbf{0}) = \mathbf{0}$.

Next we show that if $\mathbf{x} \in \mathbb{R}^n$ satisfies $|\mathbf{x}| < \delta$ and if $\varphi(x) \in W$ then $\mathbf{x} = \mu(\varphi(\mathbf{x}))$. Now $\mathbf{x} = \psi(\mathbf{x}) + T\varphi(\mathbf{x})$ for all $\mathbf{x} \in X$. Also

$$|T\varphi(\mathbf{x})| \leq \|T\|_{\mathrm{op}} \, |\varphi(\mathbf{x})|$$

for all $\mathbf{x} \in X$, where $\|T\|_{\mathrm{op}}$ denotes the operator norm of T (see Lemma 8.1). It follows that

$$\begin{array}{rcl} |\mathbf{x} - \mathbf{z}| & = & |\psi(\mathbf{x}) - \psi(\mathbf{z}) + T(\varphi(\mathbf{x}) - \varphi(\mathbf{z}))| \\ & \leq & |\psi(\mathbf{x}) - \psi(\mathbf{z})| + |T(\varphi(\mathbf{x}) - \varphi(\mathbf{z}))| \\ & \leq & \frac{1}{2}|\mathbf{x} - \mathbf{z}| + \|T\|_{\mathrm{op}} \, |\varphi(\mathbf{x}) - \varphi(\mathbf{z})| \end{array}$$

for all $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ satisfying $|\mathbf{x}| < \delta$ and $|\mathbf{z}| < \delta$. Subtracting $\frac{1}{2}|\mathbf{x} - \mathbf{z}|$ from both sides of the above inequality, and then multiplying by two, we find that

$$|\mathbf{x} - \mathbf{z}| \leq 2\|T\|_{\mathrm{op}} \, |\varphi(\mathbf{x}) - \varphi(\mathbf{z})|.$$

whenever $|\mathbf{x}| < \delta$ and $|\mathbf{z}| < \delta$.

Substituting $\mathbf{z} = \mu(\mathbf{y})$, we find that

$$|\mathbf{x} - \mu(\mathbf{y})| \leq 2\|T\|_{\mathrm{op}}\,|\varphi(\mathbf{x}) - \mathbf{y}|$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x}| < \delta$ and for all $\mathbf{y} \in W$. It follows that if $\mathbf{x} \in X$ satisfies $|\mathbf{x}| < \delta$ and if $\varphi(\mathbf{x}) = \mathbf{y}$ for some $\mathbf{y} \in W$ then $\mathbf{x} = \mu(\mathbf{y})$. The inequality also ensures that

$$|\mu(\mathbf{y}) - \mu(\mathbf{w})| \leq 2\|T\|_{\mathrm{op}}\,|\mathbf{y} - \mathbf{w}|$$

for all $\mathbf{y}, \mathbf{w} \in W$. Thus the function $\mu\colon W \to X$ is Lipschitz continuous. It then follows from Lemma 11.3 that the function μ is continuously differentiable.

Next we prove that $\mu(W)$ is an open subset of X. Now $\mu(W) \subset \varphi^{-1}(W)$ because $\mathbf{y} = \varphi(\mu(\mathbf{y}))$ for all $\mathbf{y} \in W$. We have also proved that $|\mu(\mathbf{y})| < \delta$ for all $\mathbf{y} \in W$. It follows that

$$\mu(W) \subset \varphi^{-1}(W) \cap \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < \delta\}.$$

But we have also shown that if $\mathbf{x} \in X$ satisfies $|\mathbf{x}| < \delta$, and if $\varphi(\mathbf{x}) \in W$ then $\mathbf{x} = \mu(\varphi(\mathbf{x}))$, and therefore $\mathbf{x} \in \mu(W)$. It follows that

$$\mu(W) = \varphi^{-1}(W) \cap \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < \delta\}.$$

Now $\varphi^{-1}(W)$ is an open subset in X, because $\varphi\colon X \to \mathbb{R}^n$ is continuous and W is an open set in $\mathbb{R}^n$ (see Proposition 4.18). It follows that $\mu(W)$ is an intersection of two open sets, and is thus itself an open set. Moreover $\mathbf{0} \in \mu(W)$, because $\mu(\mathbf{0}) = \mathbf{0}$. We can now conclude that $\mu\colon W \to X$ is a local inverse for $\varphi\colon X \to \mathbb{R}^n$.

We have shown that the function $\mu\colon W \to X$ is Lipschitz continuous. It therefore follows from Lemma 11.3 that the function $\mu\colon W \to X$ is continuously differentiable. This completes the proof of the Inverse Function Theorem for continuously differentiable functions whose derivative at a given point is an invertible linear transformation. $\blacksquare$