Module MA2321: Analysis in Several Real Variables
Michaelmas Term 2017
Part IV (Section 11)

D. R. Wilkins

Contents

11 The Inverse and Implicit Function Theorems

11.1 Local Invertibility of Differentiable Functions

Definition Let $\varphi: X \to \mathbb{R}^n$ be a continuous function defined over an open set X in $\mathbb{R}^n$ and mapping that open set into $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. A *local inverse* of the map $\varphi: X \to \mathbb{R}^n$ around the point $\mathbf{p}$ is a continuous function $\mu: W \to X$ defined over an open set W in $\mathbb{R}^n$ 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 $\mathbf{y} \in W$.

If there exists a function $\mu: W \to X$ satisfying these conditions, then the function φ is said to be *locally invertible* around the point $\mathbf{p}$.

Lemma 11.1 *Let $\varphi: X \to \mathbb{R}^n$ be a continuous function defined over an open set X in $\mathbb{R}^n$ and mapping that open set into $\mathbb{R}^n$, let $\mathbf{p}$ be a point of X. and let $\mu: W \to X$ be a local inverse for the map ϕ around the point $\mathbf{p}$. Then $\varphi(\mathbf{x}) \in W$ and $\mu(\varphi(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mu(W)$.*

Proof The definition of local inverses ensures that $\mu(W)$ is an open subset of X, $\mathbf{p} \in \mu(W)$ and $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$. Let $\mathbf{x} \in \mu(W)$. Then $\mathbf{x} = \mu(\mathbf{y})$ for some $\mathbf{y} \in W$. But then $\varphi(\mathbf{x}) = \varphi(\mu(\mathbf{y})) = \mathbf{y}$, and therefore $\varphi(\mathbf{x}) \in W$. Moreover $\mu(\varphi(\mathbf{x})) = \mu(\mathbf{y}) = \mathbf{x}$, as required. ∎

Let $\varphi: X \to \mathbb{R}^n$ be a continuous function defined over an open set X in $\mathbb{R}^n$ and mapping that open set into $\mathbb{R}^n$, let $\mathbf{p}$ be a point of X. and let $\mu: W \to X$ be a local inverse for the map ϕ around the point $\mathbf{p}$. Then the function from the open set $\mu(W)$ to the open set W that sends each point $\mathbf{x}$ of $\mu(W)$ to $\varphi(x)$ is invertible, and its inverse is the continuous function from W to $\varphi(W)$ that sends each point $\mathbf{y}$ of W to $\mu(\mathbf{y})$. A function between sets is *bijective* if it has a well-defined inverse. A continuous bijective function whose inverse is also continuous is said to be a *homeomorphism.* We see therefore that the restriction of the map φ to the image $\mu(W)$ of the local inverse $\mu: W \to X$ determines a homeomorphism from the open set $\mu(W)$ to the open set W.

Example The function $\varphi: \mathbb{R}^2 \to \mathbb{R}^2 \setminus \{(0,0)\}$ defined such that

$$\varphi(u, v) = (e^u \cos v, e^u \sin v)$$

for all $u, v \in \mathbb{R}^2$ is locally invertible, though it is not bijective. Indeed, given $(u_0, v_0) \in \mathbb{R}$, let

$$\begin{aligned} W &= \{(r \cos(v_0 + \theta), r \sin(v_0 + \theta)) : \\ &\qquad r, \theta \in \mathbb{R},\ \ r > 0 \text{ and } -\pi < \theta < \pi\}, \end{aligned}$$

and let

$$\mu(r \cos(v_0 + \theta), r \sin(v_0 + \theta)) = (\log r, v_0 + \theta)$$

whenever $r > 0$ and $-\pi < \theta < 1$. Then W is an open set in $\mathbb{R}^2$, the function $\mu: W \to \mathbb{R}^2$ is continuous,

$$\mu(W) = \{(u, v) \in \mathbb{R}^2 : v_0 - \pi < v < v_0 + \pi\},$$

and $\mu(\varphi(u, v)) = (u, v)$ for all $(u, v) \in \mu(W)$.

A continuously differentiable function may have a continuous inverse, but that inverse is not guaranteed to be differentiable, as the following example demonstrates.

Example Let $f: \mathbb{R} \to \mathbb{R}$ be defined so that $f(x) = x^3$ for all real numbers x. The function f is continuously differentiable and has a continuous inverse $f^{-1}: \mathbb{R} \to \mathbb{R}$, where $f^{-1}(x) = \sqrt[3]{x}$ when $x \geq 0$ and $f^{-1}(x) = -\sqrt[3]{-x}$ when $x < 0$. This inverse function is not differentiable at zero.

Lemma 11.2 *Let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set X in $\mathbb{R}^n$. Suppose that φ is locally invertible around some point $\mathbf{p}$ of X. Suppose also that a local inverse to φ around $\mathbf{p}$ is differentiable at the point $\varphi(\mathbf{p})$. Then the derivative $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^n$ of φ at the point $\mathbf{p}$ is an invertible linear operator on $\mathbb{R}^n$. Thus if*

$$\varphi(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n),$$

for all $(x_1, x_2, \ldots, x_n) \in X$, where $y_1, y_2, \ldots, y_n$ are differentiable functions of $x_1, x_2, \ldots, x_n$, and if φ has a differentiable local inverse around the point $\mathbf{p}$, then the Jacobian matrix

$$\begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial y_n}{\partial x_1} & \dfrac{\partial y_n}{\partial x_2} & \cdots & \dfrac{\partial y_n}{\partial x_n} \end{pmatrix}$$

is invertible at the point $\mathbf{p}$.

Proof Let $\mu: W \to X$ be a local inverse of φ around $\mathbf{p}$, where W is an open set in $\mathbb{R}^n$, $\mathbf{p} \in \mu(W)$, $\mu(W) \subset X$ and $\mu(\varphi(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mu(W)$. Suppose that $\mu: W \to X$ is differentiable at $\varphi(\mathbf{p})$. The identity $\mu(\varphi(\mathbf{x})) = \mathbf{x}$ holds throughout the open neighbourhood $\mu(W)$ of point $\mathbf{p}$. Applying the Chain Rule (Proposition 9.8), we find that $(D\mu)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}$ is the identity operator on $\mathbb{R}^n$. It follows that the linear operators $(D\mu)_{\varphi(\mathbf{p})}$ and $(D\varphi)_{\mathbf{p}}$ on $\mathbb{R}^n$ are inverses of one another, and therefore $(D\varphi)_{\mathbf{p}}$ is an invertible linear operator on $\mathbb{R}^n$. The result follows. ∎

Definition A function $\mu: W \to X$ between subsets W and X of Euclidean spaces is said to be *Lipschitz continuous* if there exists a positive constant C such that

$$|\mu(\mathbf{u}) - \mu(\mathbf{v})| \leq C|\mathbf{u} - \mathbf{v}|$$

for all $\mathbf{u}, \mathbf{v} \in W$.

It follows from Corollary 9.11 that a continuously differentiable function is Lipschitz continuous throughout some sufficiently small open neighbourhood of any given point in its domain.

Lemma 11.3 *Let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set X in $\mathbb{R}^n$ that is locally invertible around some point of X and let $\mu: W \to X$ be a local inverse for φ. Suppose that $\varphi: X \to \mathbb{R}^n$ is continuously differentiable and that the local inverse $\mu: W \to X$ is Lipschitz continuous throughout W. Then $\mu: W \to X$ is continuously differentiable throughout W.*

Proof The function $\mu: W \to X$ is Lipschitz continuous, and therefore there exists a positive constant C such that

$$|\mu(\mathbf{y}) - \mu(\mathbf{w})| \leq C\,|\mathbf{y} - \mathbf{w}|$$

for all $\mathbf{y}, \mathbf{w} \in W$. Let $\mathbf{q} \in W$, let $\mathbf{p} = \mu(\mathbf{q})$, and let S be the derivative of φ at $\mathbf{p}$. Then

$$S\mathbf{v} = \lim_{t \to 0} \frac{1}{t}(\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p}))$$

for all $\mathbf{v} \in \mathbb{R}^n$ (see Lemma 9.5). If $|t|$ is sufficiently small then $\mathbf{p} + t\mathbf{v} \in \mu(W)$. It then follows from Lemma 11.1 that

$$t\mathbf{v} = \mu(\varphi(\mathbf{p} + t\mathbf{v})) - \mu(\varphi(\mathbf{p})),$$

and therefore

$$|t||\mathbf{v}| \leq C\,|\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p})|.$$

It follows that

$$|S\mathbf{v}| = \lim_{t\to 0} \frac{1}{|t|}|\varphi(\mathbf{p}+t\mathbf{v}) - \varphi(\mathbf{p})| \geq \frac{1}{C}|\mathbf{v}|$$

for all $\mathbf{v} \in \mathbb{R}^n$, and therefore $S\mathbf{v} \neq \mathbf{0}$ for all non-zero vectors $\mathbf{v}$. It then follows from basic linear algebra that the linear operator S on $\mathbb{R}^n$ is invertible. Moreover $|S^{-1}\mathbf{v}| \leq C|\mathbf{v}|$ for all $\mathbf{v} \in \mathbb{R}^n$.

Now

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|}\,|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - S(\mathbf{x}-\mathbf{p})| = 0,$$

because the function φ is differentiable at $\mathbf{p}$. Also $\mu(\mathbf{y}) \neq \mathbf{p}$ when $\mathbf{y} \neq \mathbf{q}$, because $\mathbf{q} = \varphi(\mathbf{p})$ and $\mathbf{y} = \varphi(\mu(\mathbf{y}))$. The continuity of μ ensures that $\mu(\mathbf{y})$ tends to $\mathbf{p}$ as $\mathbf{y}$ tends to $\mathbf{q}$. It follows that

$$\lim_{\mathbf{y}\to\mathbf{q}} \frac{1}{|\mu(\mathbf{y})-\mathbf{p}|}\,|\mathbf{y} - \mathbf{q} - S(\mu(\mathbf{y}) - \mathbf{p})| = 0$$

(see Proposition 4.16). Now

$$|S^{-1}(\mathbf{y}-\mathbf{q}) - (\mu(\mathbf{y}) - \mathbf{p})| \leq C|\mathbf{y} - \mathbf{q} - S(\mu(\mathbf{y}) - \mathbf{p})|$$

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

$$\frac{1}{|\mathbf{y}-\mathbf{q}|} \leq \frac{C}{|\mathbf{p}-\mu(\mathbf{y})|}$$

for all $\mathbf{y} \in W$ satisfying $\mathbf{y} \neq \mathbf{q}$. It follows that

$$\frac{1}{|\mathbf{y}-\mathbf{q}|}|\mu(\mathbf{y}) - \mathbf{p} - S^{-1}(\mathbf{y}-\mathbf{q})| \leq \frac{C^2}{|\mu(\mathbf{y})-\mathbf{p}|}\,|\mathbf{y} - \mathbf{q} - S(\mu(\mathbf{y}) - \mathbf{p})|.$$

It follows that

$$\lim_{\mathbf{y}\to\mathbf{q}} \frac{1}{|\mathbf{y}-\mathbf{q}|}|\mu(\mathbf{y}) - \mathbf{p} - S^{-1}(\mathbf{y}-\mathbf{q})| = 0$$

(see Proposition 4.9), and therefore the function μ is differentiable at $\mathbf{q}$ with derivative S^{-1}. Thus $(D\mu)_{\mathbf{q}} = (D\varphi)_{\mathbf{p}}^{-1}$ for all $\mathbf{q} \in W$. It follows from this that $(D\mu)_{\mathbf{q}}$ depends continuously on $\mathbf{q}$, and thus the function μ is continuously differentiable on W, as required. ∎

11.2 Convergence of Contractive Sequences

Proposition 11.4 *Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be an infinite sequence of points in n-dimensional Euclidean space $\mathbb{R}^n$, and let λ be a real number satisfying $0 < \lambda < 1$. Suppose that*

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| \leq \lambda|\mathbf{x}_j - \mathbf{x}_{j-1}|$$

for all integers j satisfying $j > 1$. Then the infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is convergent.

Proof We show that an infinite sequence of points in Euclidean space satisfying the stated criterion is a Cauchy sequence and is therefore convergent. Now the infinite sequence satisfies

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| \leq C\lambda^j$$

for all positive integers j, where $C = |\mathbf{x}_2 - \mathbf{x}_1|/\lambda$. Let j and k be positive integers satisfying $j < k$. Then

$$\begin{aligned} |\mathbf{x}_k - \mathbf{x}_j| &= \left|\sum_{s=j}^{k-1}(\mathbf{x}_{s+1} - \mathbf{x}_s)\right| \leq \sum_{s=j}^{k-1} |\mathbf{x}_{s+1} - \mathbf{x}_s| \\ &\leq C\sum_{s=j}^{k-1} \lambda^s = C\lambda^j \frac{1-\lambda^{k-j}}{1-\lambda} < \frac{C\lambda^j}{1-\lambda}. \end{aligned}$$

We now show that the infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is a Cauchy sequence. Let some positive real number ε be given. Then a positive integer N can be chosen large enough to ensure that $C\lambda^N < (1-\lambda)\varepsilon$. Then $|\mathbf{x}_k - \mathbf{x}_j| < \varepsilon$ whenever $j \geq N$ and $k \geq N$. Therefore the given infinite sequence is a Cauchy sequence. Now all Cauchy sequences in $\mathbb{R}^n$ are convergent (see Theorem 2.8). Therefore the given infinite sequence is convergent, as required. ∎

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 $\varphi: X \to \mathbb{R}^n$ 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 $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^n$ 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 $\mu: W \to X$ 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 $(D\varphi)_{\mathbf{0}}: \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, by assumption. Let $T = (D\varphi)_{\mathbf{0}}^{-1}$, and let $\psi: X \to \mathbb{R}^n$ 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 $\mu: W \to \mathbb{R}^n$ defined such that $\mu(\mathbf{y}) = \lim_{j\to+\infty} \mu_j(\mathbf{y})$ 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 $\mu: W \to \mathbb{R}^n$ 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

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

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 $\varphi: X \to \mathbb{R}^n$ 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 $\mu: W \to X$ is a local inverse for $\varphi: X \to \mathbb{R}^n$.

We have shown that the function $\mu: W \to X$ is Lipschitz continuous. It therefore follows from Lemma 11.3 that the function $\mu: W \to X$ 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. ∎

11.4 The Implicit Function Theorem

Theorem 11.6 *Let X be an open set in $\mathbb{R}^n$, let $f_1, f_2, \ldots, f_m$ be a continuously differentiable real-valued functions on X, where $m < n$, let*

$$M = \{\mathbf{x} \in X : f_i(\mathbf{x}) = 0 \textit{ for } i = 1, 2, \ldots, m\},$$

and let $\mathbf{p}$ be a point of M. Suppose that $f_1, f_2, \ldots, f_m$ are zero at $\mathbf{p}$ and that the matrix

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_m} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_m} \end{pmatrix}$$

is invertible at the point $\mathbf{p}$. Then there exists an open neighbourhood U of $\mathbf{p}$ and continuously differentiable functions $h_1, h_2, \ldots, h_m$ of $n - m$ real variables, defined around $(p_{m+1}, \ldots, p_n)$ in $\mathbb{R}^{n-m}$, such that

$$\begin{aligned} M \cap U &= \{(x_1, x_2, \ldots, x_n) \in U : \\ &\qquad x_i = h_i(x_{m+1}, \ldots, x_n) \textit{ for } i = 1, 2, \ldots, m\}. \end{aligned}$$

Proof Let $\varphi: X \to \mathbb{R}^n$ be the continuously differentiable function defined such that

$$\varphi(\mathbf{x}) = \Bigl(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Bigr)$$

for all $\mathbf{x} \in X$. (Thus the ith Cartesian component of the function φ is equal to f_i for $i \leq m$, but is equal to x_i for $m < i \leq n$.) Let J be the Jacobian matrix of φ at the point $\mathbf{p}$, and let $J_{i,j}$ denote the coefficient in the ith row and jth column of J. Then

$$J_{i,j} = \frac{\partial f_i}{\partial x_j}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Also $J_{i,i} = 1$ if $i > m$, and $J_{i,j} = 0$ if $i > m$ and $j \neq i$. The matrix J can therefore be represented in block form as

$$J = \left(\begin{array}{c|c} J_0 & A \\ \hline 0 & I_{n-m} \end{array} \right),$$

where J_0 is the leading $m \times m$ minor of the matrix J, A is an $m \times (n-m)$ minor of the matrix J and I_{n-m} is the identity $(n-m) \times (n-m)$ matrix. It follows from standard properties of determinants that $\det J = \det J_0$. Moreover the hypotheses of the theorem require that $\det J_0 \neq 0$. Therefore $\det J \neq 0$. The derivative $(D\varphi)_{\mathbf{p}}$ of φ at the point $\mathbf{p}$ is represented by the Jacobian matrix J. It follows that $(D\varphi)_{\mathbf{p}}\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation.

The Inverse Function Theorem (Theorem 11.5) now ensures the existence of a local inverse $\mu\colon W \to X$ for the function φ around $\mathbf{p}$. The range $\mu(W)$ of this local inverse is then an open set in X containing the point $\mathbf{p}$, and $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$.

Let $\mathbf{y}$ be a point of W, and let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. Then $\mathbf{y} = \varphi(\mu(\mathbf{y}))$, and therefore $y_i = f_i(\mu(\mathbf{y}))$ for $i = 1, 2, \ldots, m$, and y_i is equal to the ith component of $\mu(\mathbf{y})$ when $m < i \leq n$.

Now $\mathbf{p} \in \mu(W)$. Therefore there exists some point $\mathbf{q}$ of W satisfying $\mu(\mathbf{q}) = \mathbf{p}$. Now $\mathbf{p} \in M$, and therefore $f_i(\mathbf{p}) = 0$ for $i = 1, 2, \ldots, m$. But $q_i = f_i(\mu(\mathbf{q})) = f_i(\mathbf{p})$ when $1 \leq i \leq m$. It follows that $q_i = 0$ when $1 \leq i \leq m$. Also $q_i = p_i$ when $i > m$.

Let g_i denote the ith Cartesian component of the continuously differentiable map $\mu\colon W \to \mathbb{R}^n$ for $i = 1, 2, \ldots, n$. Then $g_i\colon W \to \mathbb{R}$ is a continuously differentiable real-valued function on W for $i = 1, 2, \ldots, n$. If $(y_1, y_2, \ldots, y_n) \in W$ then

$$(y_1, y_2, \ldots, y_n) = \varphi(\mu(y_1, y_2, \ldots, y_n)).$$

It then follows from the definition of the map φ that y_i is the ith Cartesian component of $\mu(y_1, y_2, \ldots, y_n)$ when $i > m$, and thus

$$y_i = g_i(y_1, y_2, \ldots, y_n) \quad \text{when} \quad i > m.$$

Now $\mu(W)$ is an open set, and $\mathbf{p} \in \mu(W)$. It follows that there exists some positive real number δ such that $H(\mathbf{p}, \delta) \subset \mu(W)$. where

$$\begin{aligned} &H(\mathbf{p}, \delta) \\ &\quad = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : p_i - \delta < x_i < p_i + \delta \text{ for } i = 1, 2, \ldots, n\}. \end{aligned}$$

Let

$$\begin{aligned} D &= \{(z_1, z_2, \ldots, z_{n-m}) \in \mathbb{R}^{n-m} : p_{m+j} - \delta < z_j < p_{m+j} + \delta \\ &\qquad \text{for } j = 1, 2, \ldots, n-m\}, \end{aligned}$$

and let $h_i \colon D \to \mathbb{R}$ be defined so that

$$h_i(z_1, z_2, \ldots, z_{n-m}) = g_i(0, 0, \ldots, 0, z_1, z_2, \ldots, z_{n-m})$$

for $i = 1, 2, \ldots, m$.

Let $\mathbf{x} \in H(\mathbf{p}, \delta)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then $\mathbf{x} \in \mu(W)$. It follows from Lemma 11.1 that

$$\begin{aligned} (x_1, x_2, \ldots, x_n) &= \mu(\varphi(\mathbf{x})) \\ &= \mu\Bigl(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Bigr). \end{aligned}$$

On equating Cartesian components we find that

$$x_i = g_i\Bigl(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Bigr).$$

for $i = 1, 2, \ldots, n$.

In particular, if $\mathbf{x} \in H(\mathbf{p}, \delta) \cap M$ then

$$f_1(\mathbf{x}) = f_2(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0,$$

and therefore

$$\begin{aligned} x_i &= g_i\Bigl(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n\Bigr) \\ &= h_i\Bigl(x_{m+1}, \ldots, x_n\Bigr). \end{aligned}$$

for $i = 1, 2, \ldots, m$. It follows that

$$\begin{aligned} M \cap H(\mathbf{p}, \delta) &\subset \{(x_1, x_2, \ldots, x_n) \in H(\mathbf{p}, \delta) : \\ &\qquad x_i = h_i(x_{m+1}, \ldots, x_n) \text{ for } i = 1, 2, \ldots, m\}. \end{aligned}$$

Now let $\mathbf{x}$ be a point of $H(\mathbf{x}, \delta)$ whose Cartesian components $x_1, x_2, \ldots, x_n$ satisfy the equations

$$x_i = h_i(x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$. Then

$$x_i = g_i(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$. Now it was shown earlier that

$$y_i = g_i(y_1, y_2, \ldots, y_n)$$

for all $(y_1, y_2, \ldots, y_n) \in W$ when $i > m$. It follows from this that

$$x_i = g_i(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)$$

when $m < i \leq n$. The functions $g_1, g_2, \ldots, g_n$ are the Cartesian components of the map $\mu: W \to X$. We conclude therefore that

$$(x_1, x_2, \ldots, x_n) = \mu(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n),$$

Applying the function φ to both sides of this equation we see that

$$\begin{aligned} \varphi(x_1, x_2, \ldots, x_n) &= \varphi(\mu(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)) \\ &= (0, 0, \ldots, 0, x_{m+1}, \ldots, x_n). \end{aligned}$$

It then follows from the definition of the map φ that

$$f_i(x_1, x_2, \ldots, x_n) = 0,$$

for $i = 1, 2, \ldots, m$. We have thus shown that if $\mathbf{x}$ is a point of $H(\mathbf{x}, \delta)$ whose Cartesian components $x_1, x_2, \ldots, x_n$ satisfy the equations

$$x_i = h_i(x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$ then $\mathbf{x} \in M$. The converse of this result was proved earlier. The proof of the theorem is therefore completed on taking $U = H(\mathbf{p}, \delta)$. ∎