The Contraction Mapping Theorem and the Implicit and Inverse Function Theorems

The Contraction Mapping Theorem

Theorem (The Contraction Mapping Theorem) *Let* $B_a = \{\ \vec{x} \in \mathbb{R}^d \ \big|\ \|\vec{x}\| < a\ \}$ *denote the open ball of radius a centred on the origin in* $\mathbb{R}^d$. *If the function*

$$\vec{g} : B_a \to \mathbb{R}^d$$

obeys

(H1) *there is a constant* $G < 1$ *such that* $\|\vec{g}(\vec{x}) - \vec{g}(\vec{y})\| \le G\, \|\vec{x} - \vec{y}\|$ *for all* $\vec{x}, \vec{y} \in B_a$

(H2) $\|\vec{g}(\vec{0})\| < (1 - G)a$

then the equation

$$\vec{x} = \vec{g}(\vec{x})$$

has exactly one solution.

Discussion of hypothesis (H1): Hypothesis (H1) is responsible for the word "Contraction" in the name of the theorem. Because $G < 1$ (and it is crucial that G is strictly smaller than 1) the distance between the images $\vec{g}(\vec{x})$ and $\vec{g}(\vec{y})$ of $\vec{x}$ and $\vec{y}$ is strictly smaller than the original distance between $\vec{x}$ and $\vec{y}$. Thus the function g contracts distances. Note that, when the dimension $d = 1$ and the function g is C^1,

$$|g(x) - g(y)| = \Big|\int_x^y g'(t)\, dt\Big| \le \Big|\int_x^y |g'(t)|\, dt\Big| \le \Big|\int_x^y \sup_{t' \in B_a} |g'(t')|\, dt\Big| = |x - y| \sup_{t' \in B_a} |g'(t')|$$

For a once continuously differentiable function, the smallest G that one can pick and still have $|g(x) - g(y)| \le G|x - y|$ for all x, y is $G = \sup_{t' \in B_a} |g'(t')|$. In this case (H1) comes down to the requirement that there exist a constant $G < 1$ such that $|g'(t)| \le G < 1$ for all $t' \in B_a$. For dimensions $d > 1$, one has a whole matrix $\mathcal{G}(\vec{x}) = \big[\frac{\partial g_i}{\partial x_j}(\vec{x})\big]_{1 \le i,j \le d}$ of first partial derivatives. There is a measure of the size of this matrix, called the norm of the matrix and denoted $\big\|\mathcal{G}(\vec{x})\big\|$ such that

$$\|\vec{g}(\vec{x}) - \vec{g}(\vec{y})\| \le \|\vec{x} - \vec{y}\| \sup_{\vec{t} \in B_a} \big\|\mathcal{G}(\vec{t})\big\|$$

Once again (H1) comes down to $\big\|\mathcal{G}(\vec{t})\big\| \le G < 1$ for all $\vec{t} \in B_a$. Roughly speaking, (H1) forces the derivative of $\vec{g}$ to be sufficiently small, which forces the derivative of $\vec{x} - \vec{g}(\vec{x})$ to be bounded away from zero.

If we were to relax (H1) to $G \le 1$, the theorem would fail. For example, $g(x) = x$ obeys $|g(x) - g(y)| = |x - y|$ for all x and y. So G would be one in this case. But every x obeys $g(x) = x$, so the solution is certainly not unique.

Discussion of hypothesis (H2): If $\vec g$ only takes values that are outside of B_a, then $\vec x = \vec g(\vec x)$ cannot possibly have any solutions. So there has to be a requirement that $\vec g(\vec x)$ lies in B_a for at least some values of $\vec x \in B_a$. Our hypotheses are actually somewhat stronger than this:

$$\|\vec g(\vec x)\| = \|\vec g(\vec x) - \vec g(\vec 0) + \vec g(\vec 0)\| \le \|\vec g(\vec x) - \vec g(\vec 0)\| + \|\vec g(\vec 0)\| \le G\|\vec x - \vec 0\| + (1-G)a$$

by (H1) and (H2). So, for all $\vec x$ in B_a, that is, all $\vec x$ with $\|\vec x\| < a$, $\|\vec g(\vec x)\| < Ga + (1-G)a = a$. With our hypotheses $\vec g : B_a \to B_a$. Roughly speaking, (H2) requires that $\vec g(\vec x)$ be sufficiently small for at least one $\vec x$.

If we were to relax (H2) to $\|\vec g(\vec 0)\| \le (1-G)a$, the theorem would fail. For example, let $d = 1$, pick any $a > 0$, $0 < G < 1$ and define $g : B_a \to \mathbb{R}$ by $g(x) = a(1-G) + Gx$. Then $g'(x) = G$ for all x and $g(0) = a(1-G)$. For this g,

$$g(x) = x \quad \Longleftrightarrow \quad a(1-G) + Gx = x \quad \Longleftrightarrow \quad a(1-G) = (1-G)x \quad \Longleftrightarrow \quad x = a$$

As $x = a$ is not in the domain of definition of g, there is no solution.

Proof that there is at most one solution: Suppose that $\vec x^*$ and $\vec y^*$ are two solutions. Then

$$\begin{aligned} \vec x^* = \vec g(\vec x^*),\ \vec y^* = \vec g(\vec y^*) \quad &\Longrightarrow \quad \|\vec x^* - \vec y^*\| = \|\vec g(\vec x^*) - \vec g(\vec y^*)\| \\ &\overset{(H1)}{\Longrightarrow} \quad \|\vec x^* - \vec y^*\| \le G\|\vec x^* - \vec y^*\| \\ &\Longrightarrow \quad (1-G)\|\vec x^* - \vec y^*\| = 0 \end{aligned}$$

As $G < 1$, $1 - G$ is nonzero and $\|\vec x^* - \vec y^*\|$ must be zero. That is, $\vec x^*$ and $\vec y^*$ must be the same.

Proof that there is at least one solution: Set

$$\vec x_0 = 0 \quad \vec x_1 = \vec g(\vec x_0) \quad \vec x_2 = \vec g(\vec x_1) \quad \cdots \quad \vec x_n = \vec g(\vec x_{n-1}) \quad \cdots$$

We showed in "Significance of hypothesis (H2)" that $\vec g(\vec x)$ is in B_a for all $\vec x$ in B_a. So $\vec x_0,\ \vec x_1,\ \vec x_2,\ \cdots$ are all in B_a. So the definition $\vec x_n = \vec g(\vec x_{n-1})$ is legitimate. We shall show that the sequence $\vec x_0,\ \vec x_1,\ \vec x_2,\ \cdots$ converges to some vector $\vec x^* \in B_a$. Since $\vec g$ is continuous, this vector will obey

$$\vec x^* = \lim_{n\to\infty} \vec x_n = \lim_{n\to\infty} \vec g(\vec x_{n-1}) = \vec g\big(\lim_{n\to\infty} \vec x_{n-1}\big) = \vec g(\vec x^*)$$

In other words, $\vec{x}^*$ is a solution of $\vec{x} = \vec{g}(\vec{x})$.

To prove that the sequence converges, we first observe that, applying (H1) numerous times,

$$\begin{aligned}
\|\vec{x}_m - \vec{x}_{m-1}\| &= \big\|\vec{g}(\vec{x}_{m-1}) - \vec{g}(\vec{x}_{m-2})\big\| \\
\le G\big\|\vec{x}_{m-1} - \vec{x}_{m-2}\big\| &= G\big\|\vec{g}(\vec{x}_{m-2}) - \vec{g}(\vec{x}_{m-3})\big\| \\
\le G^2\big\|\vec{x}_{m-2} - \vec{x}_{m-3}\big\| &= G^2\big\|\vec{g}(\vec{x}_{m-3}) - \vec{g}(\vec{x}_{m-4})\big\| \\
&\vdots \\
\le G^{m-1}\big\|\vec{x}_1 - \vec{x}_0\big\| &= G^{m-1}\|\vec{g}(\vec{0})\|
\end{aligned}$$

Remember that $G < 1$. So the distance $\|\vec{x}_m - \vec{x}_{m-1}\|$ between the $(m-1)^{\rm st}$ and $m^{\rm th}$ entries in the sequence gets really small for m large. As

$$\vec{x}_n = \vec{x}_0 + \big(\vec{x}_1 - \vec{x}_0\big) + \big(\vec{x}_2 - \vec{x}_1\big) + \cdots + \big(\vec{x}_n - \vec{x}_{n-1}\big) = \sum_{m=1}^{n} \big(\vec{x}_m - \vec{x}_{m-1}\big)$$

(recall that $\vec{x}_0 = \vec{0}$) it suffices to prove that $\sum\limits_{m=1}^{n} \big(\vec{x}_m - \vec{x}_{m-1}\big)$ converges as $n \to \infty$. To do so it suffices to prove that $\sum\limits_{m=1}^{n} \big\|\vec{x}_m - \vec{x}_{m-1}\big\|$ converges as $n \to \infty$, which we do now.

$$\sum_{m=1}^{n} \big\|\vec{x}_m - \vec{x}_{m-1}\big\| \le \sum_{m=1}^{n} G^{m-1}\|\vec{g}(\vec{0})\| = \frac{1-G^n}{1-G}\|\vec{g}(\vec{0})\|$$

As n tends to ∞, G^n converges to zero (because $0 \le G < 1$) and $\frac{1-G^n}{1-G}\|\vec{g}(\vec{0})\|$ converges to $\frac{1}{1-G}\|\vec{g}(\vec{0})\|$. Hence $\vec{x}_n$ converges to some $\vec{x}^*$ as $n \to \infty$. As

$$\|\vec{x}^*\| \le \sum_{m=1}^{\infty} \big\|\vec{x}_m - \vec{x}_{m-1}\big\| \le \tfrac{1}{1-G}\|\vec{g}(\vec{0})\| < \tfrac{1}{1-G}(1-G)a = a$$

$\vec{x}^*$ is in B_a. ∎

Generalization: The same argument proves the following generalization:

Let X be a complete metric space, with metric d, and $g : X \to X$. If there is a constant $0 \le G < 1$ such that

$$d\big(g(x), g(y)\big) \le G\, d(x,y) \qquad \text{for all } x, y \in X$$

then there exists a unique $x \in X$ obeying $g(x) = x$.

Aliases: The "contraction mapping theorem" is also known as the "Banach fixed point theorem" and the "contraction mapping principle".

The Implicit Function Theorem

As an application of the contraction mapping theorem, we now prove the implicit function theorem. Consider some C^∞ function $\vec f(\vec x,\vec y)$ with $\vec x$ running over $\mathbb{R}^n$, $\vec y$ running over $\mathbb{R}^d$ and $\vec f$ taking values in $\mathbb{R}^d$. Suppose that we have one point $(\vec x_0,\vec y_0)$ on the surface $\vec f(\vec x,\vec y)=0$. In other words, suppose that $\vec f(\vec x_0,\vec y_0)=0$. And suppose that we wish to solve $\vec f(\vec x,\vec y)=0$ for $\vec y$ as a function of $\vec x$ near $(\vec x_0,\vec y_0)$. First observe that for each fixed $\vec x$, $\vec f(\vec x,\vec y)=0$ is a system of d equations in d unknowns. So at least the number of unknowns matches the number of equations. By way of motivation, let's expand the equations in powers of $\vec x-\vec x_0$ and $\vec y-\vec y_0$. The $i^{\rm th}$ equation (with $1\le i\le d$) is then

$$0=f_i(\vec x,\vec y)=f_i(\vec x_0,\vec y_0)+\sum_{j=1}^n \tfrac{\partial f_i}{\partial x_j}(\vec x_0,\vec y_0)(\vec x-\vec x_0)_j+\sum_{j=1}^d \tfrac{\partial f_i}{\partial y_j}(\vec x_0,\vec y_0)(\vec y-\vec y_0)_j+\text{h.o.}$$

where h.o. denotes terms of degree at least two. Equivalently

$$A(\vec y-\vec y_0)=\vec b$$

where A denotes the $d\times d$ matrix $\big[\tfrac{\partial f_i}{\partial y_j}(\vec x_0,\vec y_0)\big]_{1\le i,j\le d}$ of first partial $\vec y$ derivatives of $\vec f$ at $(\vec x_0,\vec y_0)$ and

$$b_i=-\sum_{j=1}^n \tfrac{\partial f_i}{\partial x_j}(\vec x_0,\vec y_0)(\vec x-\vec x_0)_j-\text{h.o.}$$

For $(\vec x,\vec y)$ very close to $(\vec x_0,\vec y_0)$ the higher order contributions h.o. will be very small. If we approximate by dropping h.o. completely, then the right hand side $\vec b$ becomes a constant (remember that are trying to solve for $\vec y$ when $\vec x$ is viewed as a constant) and there is a unique solution if and only if A has an inverse. The unique solution is then $\vec y=\vec y_0+A^{-1}\vec b$.

Now return to the problem of solving $\vec f(\vec x,\vec y)=0$, without making any approximations. Assume that the matrix A exists and has an inverse. When $d=1$, A is invertible if and only if $\frac{\partial f}{\partial y}(x_0,\vec y_0)\ne 0$. For $d>1$, A is invertible if and only if 0 is not an eigenvalue of A or, equivalently, if and only if $\det A\ne 0$. In any event, assuming that A^{-1} exists,

$$\vec f(\vec x,\vec y)=0 \quad\iff\quad A^{-1}\vec f(\vec x,\vec y)=0 \quad\iff\quad \vec y-\vec y_0=\vec y-\vec y_0-A^{-1}\vec f(\vec x,\vec y)$$

(If you expand in powers of $\vec x-\vec x_0$ and $\vec y-\vec y_0$, you'll see that the right hand side is exactly $A^{-1}\vec b$, including the higer order contributions.) This re-expresses our equation in a form to which we may apply the contraction mapping theorem. Precisely, rename $\vec y-\vec y_0=\vec z$ and define $\vec g(\vec x,\vec z)=\vec z-A^{-1}\vec f(\vec x,\vec z+\vec y_0)$. Then

$$\vec f(\vec x,\vec y)=0 \quad\iff\quad \vec y=\vec y_0+\vec z \text{ and } \vec g(\vec x,\vec z)=\vec z$$

Fix any $\vec{x}$ sufficiently near $\vec{x}_0$. Then $\vec{g}(\vec{x},\vec{z})$ is a function of $\vec{z}$ only and one may apply the contraction mapping theorem to it.

We must of course check that the hypotheses are satisfied. Observe first, that when $\vec{z}=\vec{0}$ and $\vec{x}=\vec{x}_0$, the matrix $\big[\frac{\partial g_i}{\partial z_j}(\vec{x}_0,\vec{0})\big]_{1\le i,j\le d}$ of first derivatives of $\vec{g}$ is exactly $\mathbb{1}-A^{-1}A$, where $\mathbb{1}$ is the identity matrix. The identity $\mathbb{1}$ arises from differentiating the term $\vec{z}$ in $\vec{g}(\vec{x}_0,\vec{z})=\vec{z}-A^{-1}\vec{f}(\vec{x}_0,\vec{z}+\vec{y}_0)$ and $-A^{-1}A$ arises from differentiating $-A^{-1}\vec{f}(\vec{x}_0,\vec{z}+\vec{y}_0)$. So $\big[\frac{\partial g_i}{\partial z_j}(\vec{x}_0,\vec{0})\big]_{1\le i,j\le d}$ is exactly the zero matrix. For $(\vec{x},\vec{z})$ sufficiently close to $(\vec{x}_0,\vec{0})$, the matrix $\big[\frac{\partial g_i}{\partial z_j}(\vec{x},\vec{z})\big]_{1\le i,j\le d}$ will, by continuity, be small enough that (H1) is satisfied. This is because, for any $\vec{u}$, $\vec{v}\in\mathbb{R}^d$, and any $1\le i\le d$,

$$g_i(\vec{x},\vec{u})-g_i(\vec{x},\vec{v})=\int_0^1 \tfrac{d}{dt}g_i\big(\vec{x},t\vec{u}+(1-t)\vec{v}\big)\,dt=\sum_{j=1}^d\int_0^1 \big(u_j-v_j\big)\tfrac{\partial g_i}{\partial z_j}\big(\vec{x},t\vec{u}+(1-t)\vec{v}\big)\,dt$$

so that

$$\big|g_i(\vec{x},\vec{u})-g_i(\vec{x},\vec{v})\big|\le d\;\|\vec{u}-\vec{v}\|\max_{\substack{0\le t\le 1\\ 1\le j\le d}}\big|\tfrac{\partial g_i}{\partial z_j}\big(\vec{x},t\vec{u}+(1-t)\vec{v}\big)\big|$$

and

$$\big\|\vec{g}(\vec{x},\vec{u})-\vec{g}(\vec{x},\vec{v})\big\|\le\Gamma\|\vec{u}-\vec{v}\|\qquad\text{with}\qquad\Gamma=d^2\max_{\substack{0\le t\le 1\\ 1\le j\le d}}\big|\tfrac{\partial g_i}{\partial z_j}\big(\vec{x},t\vec{u}+(1-t)\vec{v}\big)\big|$$

By continuity, we may choose $a>0$ small enough that $\Gamma\le\frac{1}{2}$ whenever $\|\vec{x}-\vec{x}_0\|$, $\|\vec{u}\|$ and $\|\vec{v}\|$ are all smaller than a. Also observe that $\vec{g}(\vec{x}_0,\vec{0})=-A^{-1}\vec{f}(\vec{x}_0,\vec{y}_0)=\vec{0}$. So, once again, by continuity, we may choose $0<a'<a$ so that $\big\|\vec{g}(\vec{x},\vec{0})\big\|<\frac{1}{2}a$ whenever $\|\vec{x}-\vec{x}_0\|<a'$.

We conclude from the contraction mapping theorem that, assuming A is invertible, there exist $a,a'>0$ such that, for each $\vec{x}$ obeying $\|\vec{x}-\vec{x}_0\|<a'$, the system of equations $\vec{f}(\vec{x},\vec{y})=0$ has exactly one solution, $\vec{y}(\vec{x})$, obeying $\|\vec{y}(\vec{x})-\vec{y}_0\|<a$. That's the existence and uniqueness part of the

Theorem (Implicit Function Theorem) *Let $n,d\in\mathbb{N}$ and let $U\subset\mathbb{R}^{n+d}$ be an open set. Let $\vec{f}:U\to\mathbb{R}^d$ be C^∞ with $\vec{f}(\vec{x}_0,\vec{y}_0)=0$ for some $\vec{x}_0\in\mathbb{R}^n$, $\vec{y}_0\in\mathbb{R}^d$ with $(\vec{x}_0,\vec{y}_0)\in U$. Assume that* $\det\big[\frac{\partial f_i}{\partial y_j}(\vec{x}_0,\vec{y}_0)\big]_{1\le i,j\le d}\ne 0$. *Then there exist open sets $V\subset\mathbb{R}^d$ and $W\subset\mathbb{R}^n$ with $\vec{x}_0\in W$ and $\vec{y}_0\in V$ such that*

$$\textit{for each } \vec{x}\in W,\ \textit{there is a unique } \vec{y}\in V \textit{ with } \vec{f}(\vec{x},\vec{y})=0.$$

If the $\vec{y}$ above is denoted $\vec{Y}(\vec{x})$, then $\vec{Y}:W\to\mathbb{R}^d$ is C^∞, $\vec{Y}(\vec{x}_0)=\vec{y}_0$ and $\vec{f}\big(\vec{x},\vec{Y}(\vec{x})\big)=0$ for all $\vec{x}\in W$. Furthermore

$$\frac{\partial\vec{Y}}{\partial\vec{x}}(\vec{x})=-\Big[\frac{\partial\vec{f}}{\partial\vec{y}}\big(\vec{x},\vec{Y}(\vec{x})\big)\Big]^{-1}\frac{\partial\vec{f}}{\partial\vec{x}}\big(\vec{x},\vec{Y}(\vec{x})\big)\tag{1}$$

where $\frac{\partial \vec{Y}}{\partial \vec{x}}$ *denotes the* $d\times n$ *matrix* $\left[\frac{\partial Y_i}{\partial x_j}\right]_{\substack{1\le i\le d\\ 1\le j\le n}}$, $\frac{\partial \vec{f}}{\partial \vec{x}}$ *denotes the* $d\times n$ *matrix of first partial derivatives of* $\vec{f}$ *with respect to* $\vec{x}$ *and* $\frac{\partial \vec{f}}{\partial \vec{y}}$ *denotes the* $d\times d$ *matrix of first partial derivatives of* $\vec{f}$ *with respect to* $\vec{y}$.

Proof: We have already proven the existence and uniqueness part of the theorem.

The rest will follow once we know that $\vec{Y}(\vec{x})$ has one continuous derivative, because then differentiating $\vec{f}\big(\vec{x},\vec{Y}(\vec{x})\big)=0$ with respect to $\vec{x}$ gives

$$\tfrac{\partial \vec{f}}{\partial \vec{x}}\big(\vec{x},\vec{Y}(\vec{x})\big)+\tfrac{\partial \vec{f}}{\partial \vec{y}}\big(\vec{x},\vec{Y}(\vec{x})\big)\tfrac{\partial \vec{Y}}{\partial \vec{x}}(\vec{x})=\vec{0}$$

which implies (1). (The inverse of the matrix $\frac{\partial \vec{f}}{\partial \vec{y}}\big(\vec{x},\vec{Y}(\vec{x})\big)$ exists, for all $\vec{x}$ close enough to $\vec{x}_0$, because the determinant of $\frac{\partial \vec{f}}{\partial \vec{y}}(\vec{x},\vec{y})$ is nonzero for all $(\vec{x},\vec{y})$ close enough to $(\vec{x}_0,\vec{y}_0)$, by continuity.) Once we have (1), the existence of, and formulae for, all higher derivatives follow by repeatedly differentiating (1). For example, if we know that $\vec{Y}(\vec{x})$ is C^1, then the right hand side of (1) is C^1, so that $\frac{\partial \vec{Y}}{\partial \vec{x}}(\vec{x})$ is C^1 and $\vec{Y}(\vec{x})$ is C^2.

We have constructed $\vec{Y}(\vec{x})$ as the limit of the sequence of approximations $\vec{Y}_n(\vec{x})$ determined by $\vec{Y}_0(\vec{x})=\vec{y}_0$ and

$$\vec{Y}_{n+1}(\vec{x})=\vec{Y}_n(\vec{x})-A^{-1}\vec{f}\big(\vec{x},\vec{Y}_n(\vec{x})\big) \tag{2}$$

Since $\vec{Y}_0(\vec{x})$ is C^∞ (it's a constant) and $\vec{f}$ is C^∞ by hypothesis, all of the $\vec{Y}_n(\vec{x})$'s are C^∞ by induction and the chain rule. We could prove that $\vec{Y}(\vec{x})$ is C^1 by differentiating (2) to get an inductive formula for $\frac{\partial \vec{Y}_n}{\partial \vec{x}}(\vec{x})$ and then proving that the sequence $\big\{\frac{\partial \vec{Y}_n}{\partial \vec{x}}(\vec{x})\big\}_{n\in\mathbb{N}}$ of derivatives converges uniformly.

Instead, we shall pick any unit vector $\hat{e}\in\mathbb{R}^n$ and prove that the directional derivative of $\vec{Y}(\vec{x})$ in direction $\hat{e}$ exists and is given by formula (1) multiplying the vector $\hat{e}$. Since the right hand side of (1) is continuous in $\vec{x}$, this will prove that $\vec{Y}(\vec{x})$ is C^1. We have $\vec{f}\big(\vec{x}+h\hat{e},\vec{Y}(\vec{x}+h\hat{e})\big)=0$ for all sufficiently small $h\in\mathbb{R}$. Hence

$$\begin{aligned}
0&=\vec{f}\big(\,\vec{x}+h\hat{e}\,,\,\vec{Y}(\vec{x}+h\hat{e})\,\big)-\vec{f}\big(\,\vec{x}\,,\,\vec{Y}(\vec{x})\,\big)\\
&=\vec{f}\big(\,\vec{x}+th\hat{e}\,,\,t\vec{Y}(\vec{x}+h\hat{e})+(1-t)\vec{Y}(\vec{x})\,\big)\Big|_{t=0}^{t=1}\\
&=\int_0^1 \tfrac{d}{dt}\vec{f}\big(\,\vec{x}+th\hat{e}\,,\,t\vec{Y}(\vec{x}+h\hat{e})+(1-t)\vec{Y}(\vec{x})\,\big)\;dt\\
&=h\int_0^1 \tfrac{\partial \vec{f}}{\partial \vec{x}}\hat{e}\;dt+\int_0^1 \tfrac{\partial \vec{f}}{\partial \vec{y}}[\vec{Y}(\vec{x}+h\hat{e})-\vec{Y}(\vec{x})]\;dt
\end{aligned}$$

where the arguments of both $\frac{\partial \vec{f}}{\partial \vec{x}}$ and $\frac{\partial \vec{f}}{\partial \vec{y}}$ are $\big(\,\vec{x}+th\hat{e}\,,\,t\vec{Y}(\vec{x}+h\hat{e})+(1-t)\vec{Y}(\vec{x})\,\big)$. Recall that $\frac{\partial \vec{f}}{\partial \vec{x}}$ is the $d\times n$ matrix $\left[\frac{\partial f_i}{\partial x_j}\right]_{\substack{1\le i\le d\\ 1\le j\le n}}$, $\hat{e}$ is an n component column vector, $\frac{\partial \vec{f}}{\partial \vec{y}}$ is the $d\times d$

matrix $\left[\frac{\partial f_i}{\partial y_j}\right]_{\substack{1\le i\le d\\ 1\le d\le n}}$, and $\vec{Y}$ is a d component column vector. Note that $[\vec{Y}(\vec{x}+h\hat{e})-\vec{Y}(\vec{x})]$ is independent of t and hence can be factored out of the second integral. Dividing by h gives

$$\tfrac{1}{h}[\vec{Y}(\vec{x}+h\hat{e})-\vec{Y}(\vec{x})] = -\left[\int_0^1 \tfrac{\partial \vec{f}}{\partial \vec{y}}\ dt\right]^{-1} \int_0^1 \tfrac{\partial \vec{f}}{\partial \vec{x}}\,\hat{e}\ dt \tag{3}$$

Since

$$\lim_{h\to 0}\big(\,\vec{x}+th\hat{e}\,,\,t\vec{Y}(\vec{x}+h\hat{e})+(1-t)\vec{Y}(\vec{x})\,\big) = \big(\,\vec{x}\,,\,\vec{Y}(\vec{x})\,\big)$$

uniformly in $t\in[0,1]$, the right hand side of (3) — and hence the left hand side of (3) — converges to

$$-\big[\tfrac{\partial \vec{f}}{\partial \vec{y}}\big(\vec{x},\vec{Y}(\vec{x})\big)\big]^{-1}\tfrac{\partial \vec{f}}{\partial \vec{x}}\big(\vec{x},\vec{Y}(\vec{x})\big)\,\hat{e}$$

as $h\to 0$, as desired.

■

The Inverse Function Theorem

As an application of the implicit function theorem, we now prove the inverse function theorem.

Theorem (Inverse Function Theorem) *Let $d \in \mathbb{N}$ and let $U \subset \mathbb{R}^d$ be an open set. Let $\vec{F} : U \to \mathbb{R}^d$ be C^∞ with $\det \left[\frac{\partial F_i}{\partial y_j}(\vec{y}_0)\right]_{1\le i,j\le d} \ne 0$ for some $\vec{y}_0 \in U$. Then there exists an open set $V \subset U$ with $\vec{y}_0 \in V$ such that the restriction $\vec{F}\,\big|V$ of $\vec{F}$ to V maps V one–to–one onto the open set $\vec{F}(V)$ and $\left(\vec{F}\,\big|V\right)^{-1}$ is C^∞. Furthermore, If we denote $\left(\vec{F}\,\big|V\right)^{-1}$ by $\vec{Y}$, then*

$$\frac{\partial \vec{Y}}{\partial \vec{x}}(\vec{x}) = \left[\frac{\partial \vec{F}}{\partial \vec{y}}\big(\vec{Y}(\vec{x})\big)\right]^{-1} \tag{2}$$

Proof: Apply the implicit function theorem with $n = d$, $\vec{f}(\vec{x}, \vec{y}) = \vec{F}(\vec{y}) - \vec{x}$, $\vec{x}_0 = \vec{F}(\vec{y}_0)$ and U replaced by $\mathbb{R}^d \times U$. ■