The Inverse Function Theorem

Suppose X and Y are Banach spaces and $f\colon X \to Y$ is C^1 (continuously differentiable). Under what circumstances does a (local) inverse f^{-1} exist? In the simplest (and most important) case, f is affine:

$$y = f(x) = Ax + b$$

where A is a linear operator. (If X and Y are n-dimensional, then b is an n-dimensional vector, and A is (represented by) an $n \times n$ matrix.) The inverse $x = f^{-1}(y)$ exists iff A is invertible, in which case

$$x = f^{-1}(y) = A^{-1}(y - b).$$

Since, for any $x_0 \in X$, $A = f'(x_0)$,

- An *affine* function f has an *affine* inverse function defined *everywhere* provided $f'(x_0)$ is invertible.

The Inverse Function Theorem is analogous:

- A C^1 function f has a C^1 inverse defined in a neighborhood of $f(x_0)$ provided $f'(x_0)$ is invertible.

Exercise 1: Let $X = R^2 = Y$. Denote vectors in X by $\begin{bmatrix} x \\ y \end{bmatrix}$ and vectors in Y by $\begin{bmatrix} u \\ v \end{bmatrix}$. (The confusing notation, in which $y \notin Y$, makes the intermediate results look familiar.) Let

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x^2 - y^2 \\ 2xy \end{bmatrix}.$$

Find $f'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$. When is $f'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$ invertible, and what is its inverse? (Solutions to the exercises are in the last section of the paper.)

Numerical Methods: Joel[1] has already shown us how to solve $y = f(x)$ for x using Newton's method. For a fixed y, Newton's method approximates $\hat{x} = f^{-1}(y)$, solving the linearization of f for $\hat{x}$:

$$y = f(\hat{x}) \approx f(x) + f'(x)(\hat{x} - x) \qquad \text{so} \qquad \hat{x} \approx x + f'(x)^{-1}(y - f(x)).$$

Denote the Newton operator by

$$F_y(x) = x + f'(x)^{-1}(y - f(x)) \qquad (F_y(x)).$$

Newton's Method iteratively refines a starting value x_0, generating the sequence

$$x_{n+1} = F_y(x_n). \qquad (N)$$

Exercise 2: Find the Newton operator for the function in Exercise 1. Compute the first 2 iterates of Newton's method for $f^{-1}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$ starting from $(x_0 =)\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

The Chord Method: Each iteration of Newton's method (N) requires the computation and inversion of the derivative $f'(x_n)$. The inversion can be costly when n is large, and numerical analysts use clever

[1] Joel H. Shapiro, *Fixated by Fixed Points*

techniques to avoid it. The easiest technique is to invert $f'(x_0)$ once and for all, and use $f'(x_0)^{-1}$ in place of $f'(x_n)^{-1}$ in the Newton's Method. The corresponding operator is

$$G_y(x) = x + f'(x_0)^{-1}\left(y - f(x)\right) \qquad (G_y(x)),$$

and iteratively refining x_0 using the operator G_y is called the **Chord Method**. In other words, the Chord Method is the modification of Newton's Method (N) that replaces $f'(x_n)^{-1}$ with $f'(x_0)^{-1}$. Note in particular that, since $f'(x_0)$ is invertible,

$$x \text{ is a fixed point of } G_y(\cdot) \iff f(x) = y. \qquad (1\text{-}1)$$

Exercise 3: Find the Chord Method operator and compute the two iterations of the method corresponding to the Newton iterates from Exercise 2.

The Chord Method has theoretical value, too: $G_y(x)$ is differentiable with

$$G_y'(x) = I - f'(x_0)^{-1}f'(x) = f'(x_0)^{-1}\left[f'(x_0) - f'(x)\right]. \qquad (G_y')$$

In order for the Newton's Method operator $F_y()$ to be differentiable, f must be twice differentiable — a substantial disadvantage for proving the Inverse Function Theorem.

N.B.: The derivative of G_y is independent of y because the Chord Method fixes the inverse $f'(x_0)^{-1}$, so $f'(x_0)^{-1}y$ is a constant.

The Plan: Prove the Inverse Function Theorem by proving that the Chord Method is a contraction mapping. Informally: a modified Newton's Method plus the Contraction Mapping Theorem imply the Inverse Function Theorem.

Contraction Estimates: Analysts almost always use the derivative to obtain contraction estimates. Roughly speaking, the estimates are of the form

$$\left\|g(u_1) - g(u_0)\right\| \le \sup \left\|g'(\cdot)\right\| \left\|u_1 - u_0\right\|.$$

In English: a difference in $g(u)$ is not larger than the largest derivative times the difference in u. In many ways, the point of requiring that a function have a derivative is to provide this estimate.

In one dimension, the estimate follows immediately from the (differential) Mean Value Theorem, and the sup norm is evaluated over the points between u_0 and u_1. In many dimensions, the correct notion of "between" is convexity.

Exercise 4: Suppose K is convex, u_0 and u_1 belong to K, and $g \in C^1(K)$ with $\|g'(u)\|$ uniformly bounded on K. Show that

$$\left\|g(u_1) - g(u_0)\right\|_Y \le \sup_{u\in K} \|g'(u)\|_{B(X,Y)} \left\|u_1 - u_0\right\|_X.$$

Exercise 4 and Equation (G_y') imply the following

Lemma: If $f \in C^1(X, Y)$ and $x_0 \in X$ with $f'(x_0)$ invertible, then for any $\rho \in (0, 1)$, there exists a $\delta > 0$ such that

$$u_0, u_1 \in B_\delta(x_0) \Longrightarrow \left\|G_y(u_1) - G_y(u_0)\right\|_X < \rho \left\|u_1 - u_0\right\|_X.$$

($B_\delta(x_0)$ is the open ball of radius δ centered at $x_0 \in X$.) In English: G_y maps points in $B_\delta(x_0)$ closer to each other by at least a factor of ρ.

N.B.: While the estimate makes G_y look like a contraction on $B_\delta(x_0)$, analysts usually require a contraction to map $B_\delta(x_0)$ *into* itself. This requirement forces analysts to use unwieldy language such as "G_y is Lipschitz continuous on $B_\delta(x_0)$ with Lipschitz constant ρ".

Proof: Use the continuity of f' (as an operator-valued function) to choose $\delta > 0$ so that

$$u_0, u_1 \in B_\delta(x_0) \Longrightarrow \|f'(u_1) - f'(u_0)\|_{B(X,Y)} < \frac{\rho}{\left\| f'(x_0)^{-1} \right\|_{B(Y,X)}}. \qquad (\delta)$$

Then Exercise 4 and Equation (G'_y) imply

$$\begin{aligned}
\left\|G_y(u_1) - G_y(u_0)\right\|_X &\le \sup_{u \in B_\delta(x_0)} \left\|G'_y(u)\right\|_{B(X)} \left\|u_1 - u_0\right\|_X \\
&\le \sup_{u \in B_\delta(x_0)} \left\| f'(x_0)^{-1} \left(f'(u) - f'(x_0)\right)\right\|_{B(X)} \left\|u_1 - u_0\right\|_X \\
&\le \sup_{u \in B_\delta(x_0)} \left\| f'(x_0)^{-1} \right\|_{B(Y,X)} \left\|f'(u) - f'(x_0)\right\|_{B(X,Y)} \left\|u_1 - u_0\right\|_X \\
&< \rho \left\|u_1 - u_0\right\|_X .
\end{aligned}$$

////

The Lemma implies that, for any y, G_y has at most one fixed point in $B_\delta(x_0)$. Observation (1-1) therefore implies that f is one-to-one on $B_\delta(x_0)$.

Range Control: The Contraction Mapping Theorem requires that G_y map some closed set into itself, and the contraction estimate of the last lemma does not guarantee $G_y\left(\overline{B_\delta(x_0)}\right) \subset \overline{B_\delta(x_0)}$. Indeed, the lemma is independent of y, so it would be optimistic to hope to control the range of $G_y(B_\delta(x_0)$ for arbitrary y using the lemma alone.

Suppose $x_1 \in B_\delta(x_0)$ and denote $f(x_1) = y_1$. Then $G_{y_1}(x_1) = x_1$. Choose δ_1 so that $\overline{B_{\delta_1}(x_1)} \subset B_\delta(x_0)$. If $u \in \overline{B_{\delta_1}(x_1)}$, then the Lemma and the definition of G_y guarantee

$$\begin{aligned}
\left\|G_y(u) - x_1\right\|_X &= \left\|G_y(u) - G_{y_1}(x_1)\right\|_X \\
&\le \left\|G_y(u) - G_y(x_1)\right\|_X + \left\|G_y(x_1) - G_{y_1}(x_1)\right\|_X \\
&\le \rho \left\|u - x_1\right\|_X + \left\| f'(x_0)^{-1} (y - y_1)\right\|_X \\
&\le \rho\delta_1 + \left\| f'(x_0)^{-1} \right\|_{B(Y,X)} \left\|y - y_1\right\|_Y
\end{aligned}$$

Consequently,

$$G_y(u) \in \overline{B_{\delta_1}(x_1)} \quad \text{provided} \quad \left\|y - y_1\right\|_Y \le \frac{1-\rho}{\left\| f'(x_0)^{-1} \right\|_{B(Y,X)}} \delta_1.$$

In English: If y is not too far from $f(x_1)$, then G_y is a contraction *into* $B_{\delta_1}(x_1)$. In particular, the Contraction Mapping Theorem guarantees that every y near $f(x_1)$ has an inverse $f^{-1}(y)$ near x_1. f therefore maps open subsets of $B_\delta(x_0)$ onto open subsets of Y. Consequently,

Theorem: If $f \in C^1(X,Y)$ and $f'(x_0)$ is invertible, then f is an open mapping of a neighborhood of x_0 onto a neighborhood of $f(x_0)$. In particular, f^{-1} exists and is continuous in a neighborhood of $f(x_0)$.

Differentiability: The Inverse Function Theorem guarantees that the inverse function is not only continuous, but is continuously differentiable. The Chain Rule says that *if* f^{-1} is differentiable, then

$$f^{-1}\left(f(x)\right) = x \quad \text{so} \quad f^{-1}\left(f(x)\right)' = f^{-1\prime}\left(f(x)\right) f'(x) = I,$$

the identity mapping. Consequently, *if* f^{-1} is differentiable and we relabel $f(x) = y$, then

$$f^{-1\prime}\left(y\right) = f'(x)^{-1}$$

(The most difficult part of the computation is keeping straight the names of the arguments of the functions.)

The Chain Rule determines what the derivative must be, if it exists, so it suffices to prove

1. That the candidate for the derivative exists — meaning, $f'(x)$ is invertible if $x \in B_\delta(x_0)$, and
2. That the candidate satisfies the definition of the derivative:

$$\frac{f^{-1}(y+k) - f^{-1}(y) - f'(x)^{-1}k}{\|k\|} \to 0 \quad \text{as } \|k\| \to 0.$$

(Again, $x \in B_\delta(x_0)$ and $y = f(x)$.)

Item 1 is true because "operators close to invertible operators are invertible". The topological language is "the invertible operators are an open subset of $B(X,Y)$". The precise statement is

Exercise 5: If $T \in B(X)$ is invertible and $\left\|I - T\right\|_{B(X)} \le \rho < 1$, then T is invertible and

$$\left\|T^{-1}\right\|_{B(X)} \le \frac{1}{1-\rho}$$

(Suggestion: if T were a real number and $I = 1$, then the geometric series

$$\frac{1}{T} = \frac{1}{1-(1-T)} = 1 + (1-T) + (1-T)^2 + \cdots$$

would be dominated by $\frac{1}{1-\rho}$.)

The definition of δ implies that if $x \in B_\delta(x_0)$ then

$$\begin{aligned}\left\|I - f'(x_0)^{-1}f'(x)\right\|_{B(X)} &= \left\|f'(x_0)^{-1}\left(f'(x_0) - f'(x)\right)\right\|_{B(X)} \\ &\le \left\|f'(x_0)^{-1}\right\|_{B(Y,X)} \left\|f'(x_0) - f'(x)\right\|_{B(X,Y)} < \rho,\end{aligned}$$

so Exercise 5 implies that $f'(x_0)^{-1}f'(x)$ is invertible with

$$\left\|\left(f'(x_0)^{-1}f'(x)\right)^{-1}\right\|_{B(X)} = \left\|f'(x)^{-1}f'(x_0)\right\|_{B(X)} \le \frac{1}{1-\rho}$$

Item 1 is therefore proven, but it is useful to estimate the norm of $f'(x)^{-1}$ as well:

$$\begin{aligned}\left\|f'(x)^{-1}\right\|_{B(Y,X)} &= \left\|f'(x)^{-1}f'(x_0)f'(x_0)^{-1}\right\|_{B(Y,X)} \\ &\le \left\|f'(x)^{-1}f'(x_0)\right\|_{B(X)} \left\|f'(x_0)^{-1}\right\|_{B(Y,X)} \\ &\le \frac{1}{1-\rho}\left\|f'(x_0)^{-1}\right\|_{B(Y,X)}.\end{aligned}$$

The proof of Item 2 consists largely of giving helpful names to the various terms. Specifically, suppose $x \in B_\delta(x_0)$ and call $y = f(x) \in f(B_\delta(x_0))$. Suppose $y + k = f(x+h) \in f(B_\delta(x_0))$. The choice of variables means

$$\begin{aligned} f^{-1}(y+k) - f^{-1}(y) - f'(x)^{-1}k &= (x+h) - x - f'(x)^{-1}k \\ &= h - f'(x)^{-1}k \\ &= f'(x)^{-1}\left(f'(x)h - k\right) \\ &= -f'(x)^{-1}\left(f(x+h) - f(x) - f'(x)h\right). \end{aligned}$$

According to the lemma, then,

$$\left\| f^{-1}(y+k) - f^{-1}(y) - f'(x)^{-1}k \right\|_X \le \frac{\left\| f'(x_0)^{-1} \right\|_{B(Y,X)}}{1-\rho} \left\| f(x+h) - f(x) - f'(x)h \right\|_Y$$

The right side is $o(\|h\|_X)$ — but we require that the left side be $o(\|k\|_Y)$, not $o(\|h\|_X)$. To estimate $\|h\|_X$ in terms of $\|k\|_Y$, use the nomenclature and Exercise 4:

$$\begin{aligned} \|h\|_X &= \left\| f^{-1}(y+k) - f^{-1}(y) \right\|_X \\ &\le \sup \left\| {f^{-1}}'(\cdot) \right\|_{B(Y,X)} \|k\|_Y \\ &\le \frac{1}{1-\rho} \left\| f'(x_0)^{-1} \right\|_{B(Y,X)} \|k\|_Y . \end{aligned}$$

In particular, $h \to 0$ if $k \to 0$. The cumulative estimate is therefore

$$\frac{\left\| f^{-1}(y+k) - f^{-1}(y) - f'(x)^{-1}k \right\|_X}{\|k\|_Y} \le \frac{\left\| f'(x_0)^{-1} \right\|_{B(Y,X)}^2}{1-\rho} \frac{\left\| f(x+h) - f(x) - f'(x)h \right\|_Y}{\|h\|_X}.$$

Since the right side goes to zero with $\|h\|_X$ and since $\|h\|_X$ goes to zero with $\|k\|_Y$, the left side goes to zero with $\|k\|_Y$, and the the inverse is differentiable.

The derivative of the inverse, $f'\left(f^{-1}(y)\right)^{-1}$, is continuous because $f^{-1}(y)$ and (by the result following Exercise 5), inversion is continuous. The Inverse Function Theorem is proven.

Solutions and Comments:

Solution 1: The derivative of f is (the operator represented by) the Jacobian matrix

$$f'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x & -2y \\ 2y & 2x \end{bmatrix} = 2\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$$

because it is the linear operator (acting on $\begin{bmatrix} h \\ k \end{bmatrix}$) in the expansion

$$f\left(\begin{bmatrix} x+h \\ y+k \end{bmatrix}\right) - f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2xh + h^2 - 2yk - k^2 \\ 2hy + 2xk + 2hk \end{bmatrix} = \underbrace{2\begin{bmatrix} x & -y \\ y & x \end{bmatrix}}_{f'} \begin{bmatrix} h \\ k \end{bmatrix} + \underbrace{\begin{bmatrix} h^2 - k^2 \\ 2hk \end{bmatrix}}_{=o\left(\sqrt{h^2+k^2}\right)}.$$

The Jacobian matrix is invertible whenever the determinant

$$\left| 2\begin{bmatrix} x & -y \\ y & x \end{bmatrix} \right| = 4\left(x^2 + y^2\right) \ne 0.$$

The Inverse Function Theorem therefore says that for any $f\left(\begin{bmatrix}x\\y\end{bmatrix}\right)$ except $f\left(\begin{bmatrix}0\\0\end{bmatrix}\right)=\begin{bmatrix}0\\0\end{bmatrix}$, there is a neighborhood on which a continuously differentiable f^{-1} exists.

N.B.: Exercise 1 has a complex analogue: identify the complex number $z=x+iy$ with the vector $\begin{bmatrix}x\\y\end{bmatrix}$ in $X=R^2=Y$. Then

$$f(x)=z^2=(x+iy)^2=\left(x^2-y^2\right)+i(2xy)\quad\text{is identified with}\quad f\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x^2-y^2\\2xy\end{bmatrix}=\begin{bmatrix}u\\v\end{bmatrix}.$$

Exercise 1's interpretation for complex numbers is: for any complex $z\neq 0$ there is a neighborhood of $w=z^2$ in which a complex $\sqrt{w}$ exists. Note that *complex* differentiability does not immediately follow from the inverse function theorem as stated — the Cauchy-Riemann equations require a separate computation. Also note that there are 2 square roots, one "positive" and one "negative" — the "sign" in the neighborhood of w is determined by the z whose image is $z^2=w$.

Solution 2: The inverse of the $f'\left(\begin{bmatrix}x\\y\end{bmatrix}\right)$ found in Solution 1 is

$$f'\left(\begin{bmatrix}x\\y\end{bmatrix}\right)^{-1}=\frac{1}{2\left(x^2+y^2\right)}\begin{bmatrix}x&y\\-y&x\end{bmatrix}.$$

The Newton operator for f is therefore

$$\begin{aligned}F_{\begin{bmatrix}u\\v\end{bmatrix}}\left(\begin{bmatrix}x\\y\end{bmatrix}\right)&=\begin{bmatrix}x\\y\end{bmatrix}+\frac{1}{2(x^2+y^2)}\begin{bmatrix}x&y\\-y&x\end{bmatrix}\left(\begin{bmatrix}u\\v\end{bmatrix}-\begin{bmatrix}x^2-y^2\\2xy\end{bmatrix}\right)\\&=\frac{1}{2}\begin{bmatrix}x\\y\end{bmatrix}+\frac{1}{2(x^2+y^2)}\begin{bmatrix}x&y\\-y&x\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}.\end{aligned}$$

Newton's method for $\begin{bmatrix}u\\v\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}$ refines the starting point $\begin{bmatrix}x_0\\y_0\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$ to

$$\begin{bmatrix}x_1\\y_1\end{bmatrix}=F_{\begin{bmatrix}0\\1\end{bmatrix}}\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=\frac{1}{2}\begin{bmatrix}1\\0\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}.$$

A second iteration of Newton's method refines the guess to

$$\begin{bmatrix}x_2\\y_2\end{bmatrix}=F_{\begin{bmatrix}0\\1\end{bmatrix}}\left(\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\right)=\frac{1}{4}\begin{bmatrix}1\\1\end{bmatrix}+\frac{1}{2\sqrt{\frac{1}{2}}}\frac{1}{2}\begin{bmatrix}1&1\\-1&1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}=\frac{2+\sqrt{2}}{4}\begin{bmatrix}1\\1\end{bmatrix}.$$

Evidently, the Newton iterates all lie in the span of $\begin{bmatrix}1\\1\end{bmatrix}$, so $x_n=y_n$. If they converge (and they do), then

$$f\left(\begin{bmatrix}x_n\\y_n\end{bmatrix}\right)=\begin{bmatrix}x_n^2-y_n^2\\2x_ny_n\end{bmatrix}=\begin{bmatrix}x_n^2-x_n^2\\2x_n^2\end{bmatrix}\to\begin{bmatrix}u\\v\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix},$$

so the inverse must be $f^{-1}\left(\begin{bmatrix}0\\1\end{bmatrix}\right)=\begin{bmatrix}\frac{\sqrt{2}}{2}\\\frac{\sqrt{2}}{2}\end{bmatrix}$.

N.B.: Since f represents the complex function $w=z^2$, the two iterations above represent the Newton steps for finding $f^{-1}(w)=\sqrt{i}=\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}i$.

Solution 3: The Chord operator's

$$f'\left(\begin{bmatrix}1\\0\end{bmatrix}\right)^{-1} = \frac{1}{2\left(1^2+0^2\right)}\begin{bmatrix}1 & 0\\ -0 & 1\end{bmatrix} = \frac{1}{2}I$$

is a special case of the first computation in Solution 2. The chord operator is therefore

$$\begin{aligned} G_{\begin{bmatrix}u\\v\end{bmatrix}}\left(\begin{bmatrix}x\\y\end{bmatrix}\right) &= \begin{bmatrix}x\\y\end{bmatrix} + \frac{1}{2}I\left(\begin{bmatrix}u\\v\end{bmatrix} - \begin{bmatrix}x^2-y^2\\2xy\end{bmatrix}\right) \\ &= \begin{bmatrix}x-\frac{1}{2}x^2+\frac{1}{2}y^2\\ y-xy\end{bmatrix} + \frac{1}{2}\begin{bmatrix}u\\v\end{bmatrix}. \end{aligned}$$

The first two iterates of the Chord Method are therefore

$$\begin{aligned} \begin{bmatrix}x_1\\y_1\end{bmatrix} &= G_{\begin{bmatrix}0\\1\end{bmatrix}}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}1-\frac{1}{2}1^2+\frac{1}{2}0^2\\ 0-1\cdot 0\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}, \text{ and} \\ \begin{bmatrix}x_2\\y_2\end{bmatrix} &= G_{\begin{bmatrix}0\\1\end{bmatrix}}\left(\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}\frac{1}{2}-\frac{1}{8}+\frac{1}{8}\\ \frac{1}{2}-\frac{1}{4}\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{4}\begin{bmatrix}2\\3\end{bmatrix} \end{aligned}$$

Solution 4: Parameterize the line segment from u_0 to u_1 by $u(t) = (1-t)u_0 + t\,u_1$. Then the Chain Rule says

$$\frac{d}{dt}g(u(t)) = g'(u(t))\,u'(t) = g'(u(t))\,(u_1-u_0)\,.$$

Note that $g'(u(\cdot))$ is a one-parameter family of operators, and notice that these operators operate on the constant vector (u_1-u_0). The (vector-valued version of the 1-dimensional) Fundamental Theorem of Calculus implies

$$g(u_1)-g(u_0) = \int_0^1 g'(u(t))\,(u_1-u_0)\;dt.$$

In effect, the (many-dimensional) increment in g has been reduced to a 1-dimensional increment along the parameterized line segment. The norm of the left side is dominated by

$$\left\|g(u_1)-g(u_0)\right\|_Y \le \int_0^1 \left\|g'(u(t))\right\|_{B(X,Y)}\,\left\|u_1-u_0\right\|_X \le \sup_{u\in K}\left\|g'(u)\right\|_{B(X,Y)}\,\left\|u_1-u_0\right\|_X$$

Solution 5: Products and sums of operators in $B(X)$ are again operators in $B(X)$. It therefore makes sense to define the partial sums

$$S_n = I + (I-T) + (I-T)^2 + \cdots + (I-T)^{n-1}.$$

The partial sums are Cauchy in $B(X)$ for the same reason that convergent geometric series are Cauchy in R or C: if $m<n$ then

$$\begin{aligned} \|S_n - S_m\|_{B(X)} &= \left\|(I-T)^m + (I-T)^{m+1} + \cdots + (I-T)^{n-1}\right\|_{B(X)} \\ &\le \|(I-T)\|_{B(X)}^m + \|(I-T)\|_{B(X)}^{m+1} + \cdots + \|(I-T)\|_{B(X)}^{n-1} \\ &\le \rho^m + \rho^{m+1} + \cdots + \rho^{n-1} \\ &= \frac{1-\rho^{n-m}}{1-\rho}\rho^m < \frac{1}{1-\rho}\rho^m. \end{aligned}$$

If M is sufficiently large, then $M \le m < n$ implies that $\|S_n - S_m\|_{B(X)}$ is small. Since $B(X)$ is complete, the Cauchy sequence converges, say to S. Then

$$\begin{aligned} ST &= \lim_{n\to\infty} S_n T \\ &= \lim_{n\to\infty} S_n \big(I - (I-T)\big) \\ &= \lim_{n\to\infty} I - (I-T)^n = I \end{aligned}$$

because $\|(I-T)^n\|_{B(X)} \le \rho^n \to 0$ as $n \to \infty$. Since $TS = I$ by an analogous argument, $S = T^{-1}$.

////