(d) Let $f(u)=\frac{1}{2}\int_0^1 |u(x)|^2\,dx + h(u)$. Show that on $L^4$, $f$ has a formally non-degenerate critical point at 0, yet this critical point is not isolated.

## 2.5 THE INVERSE AND IMPLICIT FUNCTION THEOREMS

The inverse and implicit function theorems are pillars of nonlinear analysis and geometry, so we give them special attention in this section. Throughout, $\boldsymbol{E}, \boldsymbol{F}, \ldots,$ are assumed to be Banach spaces. In the finite-dimensional case these theorems have a long and complex history; the infinite-dimensional version is due to Hildebrandt and Graves [1927].

We first consider the inverse function theorem. This states that if the linearization of the equation $f(x)=y$ is uniquely invertible then locally so is $f$; i.e. we can uniquely solve $f(x)=y$ for $x$ as a function of $y$. To formulate the theorem, the following terminology is useful.

**2.5.1 Definition.** *A map $f\colon U \subset \boldsymbol{E} \to V \subset \boldsymbol{F}$ ($U, V$ open) is a* ***$C^r$ diffeomorphism*** *if $f$ is of class $C^r$, is a bijection (that is, one-to-one and onto $V$), and $f^{-1}$ is also of class $C^r$.*

**2.5.2 Inverse Mapping Theorem.** *Let $f\colon U \subset \boldsymbol{E} \to \boldsymbol{F}$ be of class $C^r$, $r \geqslant 1$, $x_0 \in U$, and suppose $Df(x_0)$ is a linear isomorphism. Then $f$ is a $C^r$ diffeomorphism of some neighborhood of $x_0$ onto some neighborhood of $f(x_0)$ and $Df^{-1}(y)=[Df(f^{-1}(y))]^{-1}$ for $y$ in this neighborhood of $f(x_0)$.*

Although our immediate interest is the finite-dimensional case, for Banach spaces keep in mind the Banach isomorphism theorem: if $\boldsymbol{T}\colon \boldsymbol{E} \to \boldsymbol{F}$ is linear, bijective, and continuous, then $\boldsymbol{T}^{-1}$ is continuous. (See Box 2.2A.)

***Proof of Theorem 2.5.2.*** We begin by assembling a few standard lemmas. First recall the contraction mapping principle from Section 1.2.

**2.5.3 Lemma.** *Let $M$ be a complete metric space with distance function $d\colon M \times M \to \mathbb{R}$. Let $F\colon M \to M$ and assume there is a constant $\lambda$, $0 \leqslant \lambda < 1$ such that for all $x, y \in M$,*

$$d(F(x), F(y)) \leqslant \lambda d(x, y).$$

*Then $F$ has a unique fixed point $x_0 \in M$; that is $F(x_0)=x_0$.*

This result is the basis of many important existence theorems in analysis. The other fundamental fixed point theorem in analysis is the Schauder fixed point theorem, which states that a continuous map of a compact convex set (in a Banach space, say) to itself, has a fixed point—not necessarily unique however.

**2.5.4 Lemma.** *Let $\mathrm{GL}(\boldsymbol{E}, \boldsymbol{F})$ denote the set of linear isomorphisms from $\boldsymbol{E}$ onto $\boldsymbol{F}$. Then $\mathrm{GL}(\boldsymbol{E}, \boldsymbol{F}) \subset \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F})$ is open.*

***Proof.*** Let

$$\|\alpha\| = \sup_{\substack{e \in E \\ \|e\|=1}} \|\alpha(e)\|$$

be the norm on $\boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F})$ relative to given norms on $\boldsymbol{E}$ and $\boldsymbol{F}$.

We can assume $\boldsymbol{E}=\boldsymbol{F}$. Indeed, if $\varphi_0 \in \mathrm{GL}(\boldsymbol{E}, \boldsymbol{F})$, the map $\psi \mapsto \varphi_0^{-1} \circ \psi$ from $\boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F})$ to $\boldsymbol{L}(\boldsymbol{E}, \boldsymbol{E})$ is continuous and $\mathrm{GL}(\boldsymbol{E}, \boldsymbol{F})$ is the inverse image of $\mathrm{GL}(\boldsymbol{E}, \boldsymbol{E})$.

For $\varphi \in \mathrm{GL}(\boldsymbol{E}, \boldsymbol{E})$, we shall prove that any $\psi$ sufficiently near $\varphi$ is also invertible, which will give the result. More precisely, $\|\psi - \varphi\| < \|\varphi^{-1}\|^{-1}$ implies $\psi \in \mathrm{GL}(\boldsymbol{E}, \boldsymbol{E})$. The key is that $\|\cdot\|$ is an algebra norm. That is, $\|\beta \circ \alpha\| \leqslant \|\beta\|\|\alpha\|$ for $\alpha \in \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{E})$ and $\beta \in \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{E})$ (see Section 2.2). Since $\psi = \varphi \circ (\boldsymbol{I} - \varphi^{-1} \circ (\varphi - \psi))$, $\varphi$ is invertible, and our norm assumption shows that $\|\varphi^{-1} \circ (\varphi - \psi)\| < 1$, is sufficient to show that $\boldsymbol{I} - \xi$ is invertible whenever $\|\xi\| < 1$. ($\boldsymbol{I}$ is the identity operator). Consider the following sequence (called the *Neumann series*):

$$\begin{aligned}
\xi_0 &= \boldsymbol{I} \\
\xi_1 &= \boldsymbol{I} + \xi \\
\xi_2 &= \boldsymbol{I} + \xi + \xi \circ \xi \\
&\ \vdots \\
\xi_n &= \boldsymbol{I} + \xi + \xi \circ \xi + \cdots + (\xi \circ \xi \circ \cdots \circ \xi).
\end{aligned}$$

Using the triangle inequality and the foregoing norm inequality, we can compare this sequence to the sequence of real numbers, $1, 1+\|\xi\|, 1+\|\xi\|+\|\xi\|^2, \ldots,$ which we know is a Cauchy sequence since $\|\xi\| < 1$. Because $\boldsymbol{L}(\boldsymbol{E}, \boldsymbol{E})$ is complete, $\xi_n$ must converge. The limit, say $\rho$, is the inverse of $\boldsymbol{I} - \xi$. Indeed $(\boldsymbol{I} - \xi)\xi_n = \boldsymbol{I} - (\xi \circ \xi \circ \cdots \circ \xi)$, so the result follows. ▼

**2.5.5 Lemma.** *Let $\mathcal{G}\colon \mathrm{GL}(\boldsymbol{E}, \boldsymbol{F}) \to \mathrm{GL}(\boldsymbol{F}, \boldsymbol{E})\colon \varphi \mapsto \varphi^{-1}$. Then $\mathcal{G}$ is of class $C^\infty$ and $D\mathcal{G}(\varphi)\cdot\psi = -\varphi^{-1}\psi\varphi^{-1}$.* (For $D^r\mathcal{G}$, see Box 2.5C.)

***Proof.*** We may assume $\mathrm{GL}(\boldsymbol{E}, \boldsymbol{F}) \neq \varnothing$. If we can show that $D\mathcal{G}(\varphi)\cdot\psi = -\varphi^{-1} \circ \psi \circ \varphi^{-1}$, then it will follow from Leibniz' rule that $\mathcal{G}$ is of class $C^\infty$. Indeed, $D\mathcal{G} = \boldsymbol{B}(\mathcal{G}, \mathcal{G})$ where $\boldsymbol{B} \in \boldsymbol{L}^2(\boldsymbol{L}(\boldsymbol{F}, \boldsymbol{E}); \boldsymbol{L}(\boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F}), \boldsymbol{L}(\boldsymbol{F}, \boldsymbol{E})))$ is defined by $\boldsymbol{B}(\psi_1, \psi_2)(\boldsymbol{A}) = -\psi_1 \circ \boldsymbol{A} \circ \psi_2$, where $\psi_1, \psi_2 \in \boldsymbol{L}(\boldsymbol{F}, \boldsymbol{E})$ and $\boldsymbol{A} \in \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F})$, which shows inductively that if $\mathcal{I}$ is $C^k$ then it is $C^{k+1}$. Since $\psi \mapsto -\varphi^{-1}\psi\varphi^{-1}$ is linear ($\psi \in \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F})$), we must show that

$$\lim_{\psi \to \varphi} \frac{\|\psi^{-1} - (\varphi^{-1} - \varphi^{-1}\psi\varphi^{-1} + \varphi^{-1}\varphi\varphi^{-1})\|}{\|\psi - \varphi\|} = 0.$$

Note that

$$\begin{aligned}\psi^{-1} - (\varphi^{-1} - \varphi^{-1}\psi\varphi^{-1} + \varphi^{-1}\varphi\varphi^{-1}) &= \psi^{-1} - 2\varphi^{-1} + \varphi^{-1}\psi\varphi^{-1}\\ &= \psi^{-1}(\psi - \varphi)\varphi^{-1}(\psi - \varphi)\varphi^{-1}.\end{aligned}$$

Again, using $\|\beta \circ \alpha\| \leqslant \|\alpha\|\|\beta\|$ for $\alpha \in \boldsymbol{L}(\boldsymbol{E}, \boldsymbol{F}), \beta \in \boldsymbol{L}(\boldsymbol{F}, \boldsymbol{G})$,

$$\|\psi^{-1}(\psi - \varphi)\varphi^{-1}(\psi - \varphi)\varphi^{-1}\| \leqslant \|\psi^{-1}\|\|\psi - \varphi\|^2\|\varphi^{-1}\|^2.$$

With this inequality, the limit is clearly zero. ▼

To prove the theorem it is useful to note that it is enough to prove it under the simplifying assumptions that $x_0 = \mathbf{0}$, $f(x_0) = \mathbf{0}$, $\boldsymbol{E} = \boldsymbol{F}$, and $Df(\mathbf{0})$ is the identity. (Indeed, replace $f$ by $h(x) = Df(x_0)^{-1} \circ [f(x + x_0) - f(x_0)]$.)

Now let $g(x) = x - f(x)$ so $Dg(\mathbf{0}) = \mathbf{0}$. Choose $r > 0$ such that $\|x\| \leqslant r$ implies $\|Dg(x)\| \leqslant \frac{1}{2}$, which is possible by continuity of $Dg$. Thus by the mean value inequality, $\|x\| \leqslant r$ implies $\|g(x)\| \leqslant r/2$. Let $\overline{D}_\varepsilon(\mathbf{0}) = \{x \in \boldsymbol{F} \mid \|x\| \leqslant \varepsilon\}$. For $y \in \overline{D}_{r/2}(\mathbf{0})$, let $g_y(x) = y + x - f(x)$. By the mean value inequality, if

$$y \in \overline{D}_{r/2}(\mathbf{0}) \quad \text{and} \quad x_1, x_2 \in \overline{D}_r(\mathbf{0}),$$

then

$$\text{(i)} \qquad \|g_y(x)\| \leqslant \|y\| + \|g(x)\| \leqslant r$$

and

$$\text{(ii)} \qquad \|g_y(x_1) - g_y(x_2)\| \leqslant \tfrac{1}{2}\|x_1 - x_2\|.$$

Thus by Lemma 2.5.3, $g_y$ has a unique fixed point $x$ in $\overline{B}_r(\mathbf{0})$. This point $x$ is the unique solution of $f(x) = y$. Thus $f$ has an inverse

$$f^{-1}: V_0 = D_{r/2}(\mathbf{0}) \to U_0 = f^{-1}(D_{r/2}(\mathbf{0})) \subset D_r(\mathbf{0}).$$

From (ii) we have (iii) $\|f^{-1}(y_1) - f^{-1}(y_2)\| \leqslant 2\|y_1 - y_2\|$, so $f^{-1}$ is continuous.

From Lemma 2.5.4 we can choose $r$ small enough so that $Df(x)^{-1}$ will exist for $x \in D_r(\mathbf{0})$. Moreover, by continuity, $\|Df(x)^{-1}\| \leqslant M$ for some $M$ and all $x \in D_r(\mathbf{0})$ can be assumed as well. If $y_1, y_2 \in D_{r/2}(\mathbf{0})$, $x_1 = f^{-1}(y_1)$, and $x_2 = f^{-1}(y_2)$, then

$$\begin{aligned}&\|f^{-1}(y_1) - f^{-1}(y_2) - Df(x_2)^{-1}(y_1 - y_2)\|\\ &\qquad = \|x_1 - x_2 - Df(x_2)^{-1}[f(x_1) - f(x_2)]\|\\ &\qquad = \|Df(x_2)^{-1}\{Df(x_2)\cdot(x_1 - x_2) - f(x_1) + f(x_2)\}\|\\ &\qquad \leqslant M\|f(x_1) - f(x_2) - Df(x_2)(x_1 - x_2)\|.\end{aligned}$$

This, together with (iii), shows that $f^{-1}$ is differentiable with derivative $Df(x)^{-1}$ at $f(x)$; i.e., $D(f^{-1}) = \mathcal{I} \circ Df \circ f^{-1}$ on $V_0 = D_{r/2}(\mathbf{0})$. This formula, the chain rule, and Lemma 2.5.5 show inductively that if $f^{-1}$ is $C^{k-1}$ then $f^{-1}$ is $C^k$ for $1 \leqslant k \leqslant r$. ■

This argument also proves the following: If $f: U \to V$ is a $C^r$ homeomorphism where $U \subset \boldsymbol{E}$, and $V \subset \boldsymbol{F}$ are open sets, and $Df(u) \in \mathrm{GL}(\boldsymbol{E}, \boldsymbol{F})$ for $u \in U$, then $f$ is a $C^r$ diffeomorphism.

### BOX 2.5A THE SIZE OF THE NEIGHBORHOODS IN THE INVERSE MAPPING THEOREM

An analysis of the preceding proof also gives *explicit estimates* on the size of the ball on which $f(x) = y$ is solvable.[†] Such estimates are sometimes useful in applications. The easiest one to use in examples involves estimates on the second derivative.

**2.5.6 Corollary.** *Suppose $f: U \subset \boldsymbol{E} \to \boldsymbol{F}$ is of class $C^r$, $r \geqslant 2$, $x_0 \in U$ and $Df(x_0)$ is an isomorphism. Let*

$$L = \|Df(x_0)\| \quad \textit{and} \quad M = \|Df(x_0)^{-1}\|.$$

*Assume*

$$\|D^2 f(x)\| \leqslant K \quad \text{for} \quad \|x - x_0\| \leqslant R \quad \textit{and} \quad \overline{D}_R(x_0) \subset U.$$

*Let*

$$R_1 = \min\left\{\frac{1}{2KM}, R\right\},$$

$$R_2 = \min\left\{\frac{1}{R_1}, \frac{1}{2M(L + KR_1)}\right\} \quad \textit{and} \quad R_3 = \frac{R_2}{2L}.$$

---

[†] We thank M. Buchner for providing this formulation.