1 Inverse Function Theorem proof.

1.0.1 $K$ is a contraction map.

For each $y' \in \mathbb{E}^d$, define a function $K : \mathbb{E}^d \to \mathbb{E}^d$ by

$$K(u) \overset{\text{def}}{=} u - Df(x)^{-1}[f(u) - y'],$$

taking into account the hypothesis that $Df(x)$ is nonsingular, hence invertible.

Use $K$ to define a sequence $\{x_n\}$ recursively:

$$x_0 \overset{\text{def}}{=} x; \quad x_{n+1} = K(x_n) = x_n - Df(x)^{-1}[f(x_n) - y'], \quad n = 0, 1, 2, \ldots$$

From the definition of differentiability for functions $f : \mathbb{E}^n \to \mathbb{E}^m$,

$$f(u) = f(v) + Df(v)(u - v) + E(u,v),$$

where the error satisfies $\|E(u,v)\| = o(\|u - v\|)$ as $u \to v$. Substitute this estimate into $K(u)$ and $K(v)$ to show that $K$ is a contraction map near $x$:

$$K(u) - K(v) = u - v - Df(x)^{-1}[f(u) - f(v)]$$

$$= u - v - Df(x)^{-1}\left[Df(v)(u - v) + E(u,v)\right]$$

$$= \left[I - Df(x)^{-1}Df(v)\right](u - v) + Df(x)^{-1}E(u,v).$$

Now $f$ is continuously differentiable near $x$, so $Df(v) \to Df(x)$ as $v \to x$, so $I - Df(x)^{-1}Df(v) \to 0$. (Every coefficient of the matrix tends to 0.) Thus

$$\exists \delta_1 > 0 \quad v \in B(x, \delta_1) \implies \|I - Df(x)^{-1}Df(v)\|_{\text{op}} < \frac{1}{4} \quad (1)$$

Here $\|\cdot\|_{\text{op}}$ is the “operator norm.” For $d \times d$ matrix $A$, $\|A\|_{\text{op}} \overset{\text{def}}{=} \sup\{\|Ax\|/\|x\| : x \in \mathbb{E}^d, \ x \neq 0\}$. It is a continuous function of the coefficients of $A$ and thus $\|A\|_{\text{op}} \to 0$ as $A \to 0$. It satisfies $\|Ax\| \leq \|A\|_{\text{op}} \|x\|$ for all $x \in \mathbb{E}^d$. 


Likewise, since $Df(x)^{-1}$ is bounded and $\|E(u,v)\| = o(\|u - v\|)$,

\[
(\exists \delta > 0) \quad u, v \in B(x, \delta) \implies \|Df(x)^{-1}E(u,v)\| \leq \frac{1}{4}\|u - v\|. 
\tag{2}
\]

Put $\delta = \min\{\delta_1, \delta_2\}$ to conclude that if $u, v \in B(x, \delta)$, then

\[
\|K(u) - K(v)\| = \| [I - Df(x)^{-1}Df(v)] (u - v) + Df(x)^{-1}E(u,v) \|
\leq \| [I - Df(x)^{-1}Df(v)] \| u - v \| + \| Df(x)^{-1}E(u,v) \|
\leq \frac{1}{4}\|u - v\| + \frac{1}{4}\|u - v\| = \frac{1}{2}\|u - v\|.
\]

If $\{x_n\}$ stays in this ball $B(x, \delta)$, then by the contraction mapping theorem it will converge to a unique fixed point.

1.0.2 $K$ preserves a neighborhood of $x$.

To show that $\{x_n\}$ stays sufficiently near $x$ if $y'$ is sufficiently near $y$, use

\[
f(u) = y + Df(x)(u - x) + E(u,x),
\]

where $\|E(u,x)\| = o(\|u - x\|)$ as $u \to x$, to estimate

\[
K(u) - x = u - Df(x)^{-1} [f(u) - y'] - x
= u - x - Df(x)^{-1} [y' - y' + Df(x)(u-x) + E(u,x)]
= -Df(x)^{-1}(y - y') - Df(x)^{-1}E(u,x).
\]

But $Df(x)^{-1}$ is bounded, so there exists $\epsilon > 0$ such that

\[
y' \in B(y, \epsilon) \implies \|Df(x)^{-1}(y - y')\| < \frac{1}{2}\delta,
\]

where $\delta = \min\{\delta_1, \delta_2\}$ from Eq.1 and Eq.2. Then Eq.2 implies

\[
u \in B(x, \delta) \implies \|Df(x)^{-1}E(u,x)\| \leq \frac{1}{4}\|u - x\| \leq \frac{1}{4}\delta.
\]

Hence, if $y' \in B(y, \epsilon)$, then

\[
u \in B(x, \delta) \implies \|K(u) - x\| \leq \frac{1}{2}\delta + \frac{1}{4}\delta < \delta.
\]

Conclude that $x_n \in B(x, \delta) \implies x_{n+1} = K(x_n) \in B(x, \delta)$, so that for all $y' \in B(y, \epsilon),

\[
x_o \in B(x, \delta) \implies \{x_n\} \subset B(x, \delta).
\]
1.0.3 Sequence \( \{x_n\} \) converges to a unique fixed point.

Suppose that \( y' \in B(y, \epsilon) \) so that by the previous results, \( K : B(x, \delta) \to B(x, \delta) \) is a contraction map satisfying

\[
u, v \in B(x, \delta) \implies \|K(u) - K(v)\| \leq \frac{1}{2}\|u - v\|.
\]

Then for any \( x_0 \in B(x, \delta) \), the sequence defined by \( x_{n+1} = K(x_n) \) will satisfy

\[
\|x_{N+1} - x_N\| = \|K(x_N) - K(x_{N-1})\| \leq \frac{1}{2}\|x_N - x_{N-1}\|,
\]

for \( N = 0, 1, 2, \ldots \). Repeat this estimate \( N \) times to get the inequality

\[
\|x_{N+1} - x_N\| \leq \left(\frac{1}{2}\right)^N \|x_1 - x_0\| < 2\delta \left(\frac{1}{2}\right)^N,
\]

since \( x_1, x_0 \in B(x, \delta) \). But this implies that \( \{x_n\} \) is a Cauchy sequence:

\[
\|x_{N+n} - x_N\| \leq \sum_{i=0}^{n-1} \|x_{N+i+1} - x_{N+i}\| < \sum_{i=0}^{n-1} 2\delta \left(\frac{1}{2}\right)^{N+i} < \frac{4\delta}{2^N}.
\]

Metric space \( E^d \) is complete, so \( \{x_n\} \) converges to a point \( x' \in E^d \) that satisfies \( K(x') = x' \). This limit is unique in \( B(x, \delta) \), since for any other point \( x'' \in B(x, \delta) \) with \( x'' = K(x'') \), the contraction property implies

\[
\|x'' - x'\| = \|K(x'') - K(x')\| \leq \frac{1}{2}\|x'' - x'\|,
\]

which forces \( \|x'' - x'\| = 0 \) and thus \( x'' = x' \).

Since \( x' = K(x') = x' - Df(x)^{-1}[f(x') - y'] \), conclude that this unique fixed point satisfies \( f(x') = y' \). Thus it defines a function

\[
g(y') = x' \in B(x, \delta), \quad y' \in B(y, \epsilon).
\]

This \( g \) is the inverse function for \( f \).

1.0.4 The inverse map is differentiable.

Since \( y' = f \circ g(y') \) at all \( y' \) in an open neighborhood of \( y = f(x) \), apply the chain rule to compute

\[
I = D[f \circ g](y') = Df(g(y'))Dg(y') = Dg(y')Df(x'),
\]

for \( x' = g(y') \). Neither \( d \times d \) factor matrix is singular since their product is the nonsingular identity matrix. It follows that \( Df(x') \) is invertible with

\[
Dg(y') = Df(x')^{-1}.
\]

This completes the proof of the Inverse Function Theorem. \( \square \)