Proof of the Inverse Function Theorem For clarity, we break up the proof into a number of steps.

Step 1 *Simplification to a special case.*

We prove the theorem when $\mathbf{D}f(x_0)$ is the identity transformation. To see that this is sufficient to prove the general case, let $T = \mathbf{D}f(x_0)$. Then T^{-1} exists and is linear, so $\mathbf{D}(T^{-1})(f(x_0)) = T^{-1}$, and by the chain rule,

$$\begin{aligned}\mathbf{D}(T^{-1} \circ f)(x_0) &= \mathbf{D}(T^{-1})(f(x_0)) \circ \mathbf{D}f(x_0) = T^{-1} \circ \mathbf{D}f(x_0) \\ &= \text{identity transformation.}\end{aligned}$$

If the theorem is true for $T^{-1} \circ f$, then the theorem is also true for f. Indeed, if g is an inverse for $T^{-1} \circ f$, the inverse for f will be $g \circ T^{-1}$.

We can make one further simplifying assumption, namely, that $x_0 = 0$ and $f(x_0) = 0$. To see this, let us suppose we have proved the theorem for the special case $x_0 = 0$ and $f(x_0) = 0$. To prove the general case from this, let $h(x) = f(x + x_0) - f(x_0)$. Then $h(0) = 0$ and $\mathbf{D}h(0) = \mathbf{D}f(x_0)$, so that $\mathbf{D}h(0)$ is invertible. If h has an inverse near $x = 0$, then the required inverse for f near x_0 is given by

$$f^{-1}(y) = h^{-1}(y - f(x_0)) + x_0.$$

In summary, step 1 demonstrates that *it is sufficient to prove the theorem under the assumptions that $x_0 = 0$, $f(x_0) = 0$, and $\mathbf{D}f(0)$ is the identity.* These will be assumed in the remainder of the discussion.

Step 2 *Application of the contraction mapping principle to get a local inverse.*

Keeping the preceding remarks in mind, what we would like are two neighborhoods of 0 such that given any y from the first neighborhood of 0, there is a unique x from the second neighborhood such that $f(x) = y$. To show this, consider the function g_y defined by $g_y(x) = y + x - f(x)$. If for some closed neighborhood of zero this is a contraction mapping, then it has a unique fixed point, say, x, and so $x = y + x - f(x)$, or x is the unique point belonging to the neighborhood such that $f(x) = y$. To construct this neighborhood, define $g(x) = x - f(x)$; then $\mathbf{D}g(0) = 0$. Assume g to be of class C^p, with $p \geq 1$. This means in particular that

$\mathbf{D}g$ is a continuous function, and so, by continuity at 0, there exists an $r > 0$ such that $\|x\| < r$ implies $\|\mathbf{D}g_i(x)\| < 1/2n$, where $g = (g_1, \ldots, g_n)$. By the mean value theorem, given $x \in \mathbf{D}(0, r)$, there are points $c_1, c_2, \ldots, c_n$ in $\mathbf{D}(0, r)$ such that $g_i(x) = g_i(x) - g_i(0) = \mathbf{D}g_i(c_i)(x - 0) = \mathbf{D}g_i(c_i)(x)$. Therefore,

$$\|g(x)\| \le \sum_{i=1}^{n} |g_i(x)| = \sum_{i=1}^{n} |\mathbf{D}g_i(c_i)(x)| \le \sum_{i=1}^{n} \|\mathbf{D}g_i(c_i)\|\ \|x\| < \frac{\|x\|}{2} < \frac{r}{2},$$

using the Cauchy-Schwarz inequality on $\mathbf{D}g_i(c_i)(x) = \langle \nabla g_i(c_i), x \rangle$.

This shows that g maps the closed r ball $B(0, r)$ into the closed $r/2$ ball $B(0, r/2)$. Now let y be any member of $B(0, r/2)$. The mapping g_y takes $B(0, r)$ into $B(0, r)$; to see this, note that $\|y\| \le r/2$ and $x \in B(0, r)$ implies

$$\|g_y(x)\| = \|y + g(x)\| \le \|y\| + \|g(x)\| < r/2 + r/2 = r.$$

Let x_1 and x_2 be any two points in $B(0, r)$. Then $\|g_y(x_1) - g_y(x_2)\| = \|g(x_1) - g(x_2)\|$, and by the mean value theorem, as before, $\|g(x_1) - g(x_2)\| \le (1/2)\|x_1 - x_2\|$, and so g_y is a contraction map (with constant $K = 1/2$). The contraction mapping principle implies that there is a unique fixed point $x \in B(0, r)$ for g_y, and as we observed before, this implies $f(x) = y$. This means that f has an inverse $f^{-1} : B(0, r/2) \subset \mathbb{R}^n \to B(0, r) \subset \mathbb{R}^n$.

Step 3 *The inverse is continuous.*

Let x_1 and $x_2 \in B(0, r)$. Recalling the definition of g, we get

$$\|x_1 - x_2\| \le \|f(x_1) - f(x_2)\| + \|g(x_1) - g(x_2)\| \le \|f(x_1) - f(x_2)\| + \frac{1}{2}\|x_1 - x_2\|,$$

and hence $\|x_1 - x_2\| \le 2\|f(x_1) - f(x_2)\|$. Thus, if y_1 and $y_2 \in B(0, 1/2)$, then $\|f^{-1}(y_1) - f^{-1}(y_2)\| \le 2\|y_1 - y_2\|$, and so f^{-1} is continuous.

Step 4 *For suitably small r, the inverse is differentiable on* $\mathbf{D}(0, r/2)$.

We were given that $\mathbf{D}f(0)$ is invertible and that $\mathbf{D}f : A \subset \mathbb{R}^n \to \mathbb{R}^{n^2}$ is continuous, and we have shown that $GL(\mathbb{R}^n, \mathbb{R}^n)$ is open in $L(\mathbb{R}^n, \mathbb{R}^n)$. These facts show that for all x in some neighborhood around 0, $[\mathbf{D}f(x)]^{-1}$ exists. If this neighborhood does not contain $\mathbf{D}(0, r/2)$, r is restricted further until this is the case. Hence, we can assume that $[\mathbf{D}f(x)]^{-1}$ exists for all $x \in \mathbf{D}(0, r/2)$. Moreover, we can assume $\|[\mathbf{D}f(x)]^{-1}y\| \le M\|y\|$ for all $x \in \mathbf{D}(0, r/2)$ and $y \in \mathbb{R}^n$, by continuity of $[\mathbf{D}f(x)]^{-1}$ (see Worked Example **4.4,** Chapter 4).

For y_1 and $y_2 \in \mathbf{D}(0, r/2)$, $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$,

$$\frac{\|f^{-1}(y_1) - f^{-1}(y_2) - [\mathbf{D}f(x_2)]^{-1} \cdot (y_1 - y_2)\|}{\|y_1 - y_2\|}$$

$$= \frac{\|x_1 - x_2 - [\mathbf{D}f(x_2)]^{-1} \cdot (f(x_1) - f(x_2))\|}{\|f(x_1) - f(x_2)\|}$$

$$= \left[\frac{\|x_1 - x_2\|}{\|f(x_1) - f(x_2)\|}\right] \frac{\|\{[\mathbf{D}f(x_2)]^{-1}\}\{\mathbf{D}f(x_2)(x_1 - x_2) - (f(x_1) - f}{\|x_1 - x_2\|}$$

Using $\|x_1 - x_2\| \leq 2\|f(x_1) - f(x_2)\|$ and $\|\mathbf{D}f(x_2)^{-1}y\| \leq M\|y\|$ shov
expression is

$$\leq 2M \frac{\|\mathbf{D}f(x_2)(x_1 - x_2) - (f(x_1) - f(x_2))\|}{\|x_1 - x_2\|}.$$

The last expression has a limit zero as $\|x_1 - x_2\| \to 0$, by the differen
f at x_2. This shows that f^{-1} is differentiable at y_2 with derivative [D
$[\mathbf{D}f(f^{-1}(y_2))]^{-1}$.

In the theorem, we set $W = \mathbf{D}(0, r/2)$ and $U = f^{-1}(W)$, both open

Step 5 $f^{-1} : \mathbf{D}(0, r/2) \to \mathbb{R}^n$ *is of class* C^p.

From step 4, it follows that the inverse mapping $f^{-1} : \mathbf{D}(0, r/2)$
differentiable on $\mathbf{D}(0, r/2)$ and that $\mathbf{D}f^{-1}(y) = [\mathbf{D}f(f^{-1}(y))]^{-1}$. We
shown that $f^{-1} : \mathbf{D}(0, r/2) \to \mathbb{R}^n$ is continuous: $\mathbf{D}f$ is continuous by as
and the inversion mapping from $GL(\mathbb{R}^n, \mathbb{R}^n)$ (the set of invertible lir
from $\mathbb{R}^n$ to $\mathbb{R}^n$) to $GL(\mathbb{R}^n, \mathbb{R}^n)$ is continuous and, in fact, C^∞ by lemn
implies that $\mathbf{D}f^{-1}$ is a continuous map from $\mathbf{D}(0, r/2)$ into $L(\mathbb{R}^n, \mathbb{R}^n$
f^{-1} is of class C^1.

Again look at $\mathbf{D}f^{-1}(y) = [\mathbf{D}f(f^{-1}(y))]^{-1}$, and observe that since
class C^1, $\mathbf{D}f$ is of class C^{p-1}, and since the inversion is C^∞, $\mathbf{D}f^{-1}$ i:
C^1. Hence, f^{-1} is of class C^2. Continuing in this way by induction, \
conclude that f^{-1} is of class C^p. ■