Definition 17.6.3 (Fixed points). Let $f : X \to X$ be a map, and $x \in X$. We say that x is a *fixed point* of f if $f(x) = x$.

Contractions do not necessarily have any fixed points; for instance, the map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x + 1$ does not. However, it turns out that *strict* contractions always do, at least when X is complete:

Theorem 17.6.4 (Contraction mapping theorem). *Let (X, d) be a metric space, and let $f : X \to X$ be a strict contraction. Then f can have at most one fixed point. Moreover, if we also assume that X is non-empty and complete, then f has exactly one fixed point.*

Proof. See Exercise 17.6.7. □

Remark 17.6.5. The contraction mapping theorem is one example of a *fixed point theorem* - a theorem which guarantees, assuming certain conditions, that a map will have a fixed point. There are a number of other fixed point theorems which are also useful. One amusing one is the so-called *hairy ball theorem*, which (among other things) states that any continuous map $f : S^2 \to S^2$ from the sphere $S^2 := \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 = 1\}$ to itself, must contain either a fixed point, or an anti-fixed point (a point $x \in S^2$ such that $f(x) = -x$). A proof of this theorem can be found in any topology text; it is beyond the scope of this text.

We shall give one consequence of the contraction mapping theorem which is important for our application to the inverse function theorem. Basically, this says that any map f on a ball which is a "small" perturbation of the identity map, remains one-to-one and cannot create any internal holes in the ball.

Lemma 17.6.6. *Let $B(0, r)$ be a ball in $\mathbf{R}^n$ centered at the origin, and let $g : B(0, r) \to \mathbf{R}^n$ be a map such that $g(0) = 0$ and*

$$\|g(x) - g(y)\| \leq \frac{1}{2}\|x - y\|$$

for all $x, y \in B(0, r)$ (here $\|x\|$ denotes the length of x in $\mathbf{R}^n$). Then the function $f : B(0, r) \to \mathbf{R}^n$ defined by $f(x) := x + g(x)$

is one-to-one, and furthermore the image $f(B(0,r))$ of this map contains the ball $B(0,r/2)$.

Proof. We first show that f is one-to-one. Suppose for sake of contradiction that we had two different points $x, y \in B(0,r)$ such that $f(x) = f(y)$. But then we would have $x + g(x) = y + g(y)$, and hence

$$\|g(x) - g(y)\| = \|x - y\|.$$

The only way this can be consistent with our hypothesis $\|g(x) - g(y)\| \leq \frac{1}{2}\|x - y\|$ is if $\|x - y\| = 0$, i.e., if $x = y$, a contradiction. Thus f is one-to-one.

Now we show that $f(B(0,r))$ contains $B(0,r/2)$. Let y be any point in $B(0,r/2)$; our objective is to find a point $x \in B(0,r)$ such that $f(x) = y$, or in other words that $x = y - g(x)$. So the problem is now to find a fixed point of the map $x \mapsto y - g(x)$.

Let $F : B(0,r) \to B(0,r)$ denote the function $F(x) := y - g(x)$. Observe that if $x \in B(0,r)$, then

$$\|F(x)\| \leq \|y\| + \|g(x)\| \leq \frac{r}{2} + \|g(x) - g(0)\| \leq \frac{r}{2} + \frac{1}{2}\|x - 0\| < \frac{r}{2} + \frac{r}{2} =$$

so F does indeed map $B(0,r)$ to itself. Also, for any x, x' in $B(0,r)$ we have

$$\|F(x) - F(x')\| = \|g(x') - g(x)\| \leq \frac{1}{2}\|x' - x\|$$

so F is a strict contraction. By the contraction mapping theorem, F has a fixed point, i.e., there exists an x such that $x = y - g(x)$. But this means that $f(x) = y$, as desired. $\square$

Exercise 17.6.1. Let $f : [a,b] \to \mathbf{R}$ be a differentiable function of one variable such that $|f'(x)| \leq 1$ for all $x \in [a,b]$. Prove that f is a contraction. (Hint: use the mean-value theorem, Corollary 10.2.9.) If in addition $|f'(x)| < 1$ for all $x \in [a,b]$, show that f is a strict contraction.

Exercise 17.6.2. Show that if $f : [a,b] \to \mathbf{R}$ is differentiable and is a contraction, then $|f'(x)| \leq 1$.

Exercise 17.6.3. Give an example of a function $f : [a, b] \to \mathbf{R}$ which is continuously differentiable and which is a strict contraction, but such that $|f'(x)| = 1$ for at least one value of $x \in [a, b]$.

Exercise 17.6.4. Given an example of a function $f : [a, b] \to \mathbf{R}$ which is a strict contraction but which is not differentiable for at least one point x in $[a, b]$.

Exercise 17.6.5. Verify the claims in Examples 17.6.2.

Exercise 17.6.6. Show that every contraction on a metric space X is necessarily continuous.

Exercise 17.6.7. Prove Theorem 17.6.4. (Hint: to prove that there is at most one fixed point, argue by contradiction. To prove that there is at least one fixed point, pick any $x_0 \in X$ and define recursively $x_1 = f(x_0)$, $x_2 = f(x_1)$, $x_3 = f(x_2)$, etc. Prove inductively that $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$, and conclude (using the geometric series formula, Lemma 7.3.3) that the sequence $(x_n)_{n=0}^{\infty}$ is a Cauchy sequence. Then prove that the limit of this sequence is a fixed point of f.)

Exercise 17.6.8. Let (X, d) be a complete metric space, and let $f : X \to X$ and $g : X \to X$ be two strict contractions on X with contraction coefficients c and c' respectively. From Theorem 17.6.4 we know that f has some fixed point x_0, and g has some fixed point y_0. Suppose we know that there is an $\varepsilon > 0$ such that $d(f(x), g(x)) \leq \varepsilon$ for all $x \in X$ (i.e., f and g are within ε of each other in the uniform metric). Show that $d(x_0, y_0) \leq \varepsilon/(1 - \max(c, c'))$. Thus nearby contractions have nearby fixed points.

17.7 The inverse function theorem in several variable calculus

We recall the inverse function theorem in single variable calculus (Theorem 10.4.2), which asserts that if a function $f : \mathbf{R} \to \mathbf{R}$ is invertible, differentiable, and $f'(x_0)$ is non-zero, then f^{-1} is differentiable at $f(x_0)$, and

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.$$

In fact, one can say something even when f is not invertible, as long as we know that f is *continuously* differentiable. If $f'(x_0)$

is non-zero, then $f'(x_0)$ must be either strictly positive or strictly negative, which implies (since we are assuming f' to be continuous) that $f'(x)$ is either strictly positive for x near x_0, or strictly negative for x near x_0. In particular, f must be either strictly increasing near x_0, or strictly decreasing near x_0. In either case, f will become invertible if we restrict the domain and range of f to be sufficiently close to x_0 and to $f(x_0)$ respectively. (The technical terminology for this is that f is *locally invertible near* x_0.)

The requirement that f be continuously differentiable is important; see Exercise 17.7.1.

It turns out that a similar theorem is true for functions $f : \mathbf{R}^n \to \mathbf{R}^n$ from one Euclidean space to the same space. However, the condition that $f'(x_0)$ is non-zero must be replaced with a slightly different one, namely that $f'(x_0)$ is *invertible.* We first remark that the inverse of a linear transformation is also linear:

Lemma 17.7.1. *Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation which is also invertible. Then the inverse transformation $T^{-1} : \mathbf{R}^n \to \mathbf{R}^n$ is also linear.*

Proof. See Exercise 17.7.2. □

We can now prove an important and useful theorem, arguably one of the most important theorems in several variable differential calculus.

Theorem 17.7.2 (Inverse function theorem). *Let E be an open subset of $\mathbf{R}^n$, and let $T : E \to \mathbf{R}^n$ be a function which is continuously differentiable on E. Suppose $x_0 \in E$ is such that the linear transformation $f'(x_0) : \mathbf{R}^n \to \mathbf{R}^n$ is invertible. Then there exists an open set U in E containing x_0, and an open set V in $\mathbf{R}^n$ containing $f(x_0)$, such that f is a bijection from U to V. In particular, there is an inverse map $f^{-1} : V \to U$. Furthermore, this inverse map is differentiable at $f(x_0)$, and*

$$(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}.$$

Proof. We first observe that once we know the inverse map f^{-1} is differentiable, the formula $(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}$ is automatic. This comes from starting with the identity

$$I = f^{-1} \circ f$$

on U, where $I : \mathbf{R}^n \to \mathbf{R}^n$ is the identity map $Ix := x$, and then differentiating both sides using the chain rule at x_0 to obtain

$$I'(x_0) = (f^{-1})'(f(x_0))f'(x_0).$$

Since $I'(x_0) = I$, we thus have $(f^{-1})'(f(x_0)) = (f'(x_0))^{-1}$ as desired.

We remark that this argument shows that if $f(x_0)$ is *not* invertible, then there is no way that an inverse f^{-1} can exist and be differentiable at $f(x_0)$.

Next, we observe that it suffices to prove the theorem under the additional assumption $f(x_0) = 0$. The general case then follows from the special case by replacing f by a new function $\tilde{f}(x) := f(x) - f(x_0)$ and then applying the special case to $\tilde{f}$ (note that V will have to shift by $f(x_0)$). Note that $f^{-1}(y) = \tilde{f}^{-1}(y + f(x_0))$ - why?. Henceforth we will always assume $f(x_0) = 0$.

In a similar manner, one can make the assumption $x_0 = 0$. The general case then follows from this case by replacing f by a new function $\tilde{f}(x) := f(x + x_0)$ and applying the special case to $\tilde{f}$ (note that E and U will have to shift by x_0). Note that $f^{-1}(y) = \tilde{f}^{-1}(y) + x_0$ - why? Henceforth we will always assume $x_0 = 0$. Thus we now have that $f(0) = 0$ and that $f'(0)$ is invertible.

Finally, one can assume that $f'(0) = I$ where $I : \mathbf{R}^n \to \mathbf{R}^n$ is the identity transformation $Ix = x$). The general case then follows from this case by replacing f with a new function $\tilde{f} : E \to \mathbf{R}^n$ defined by $\tilde{f}(x) := f'(0)^{-1}f(x)$, and applying the special case to this case. Note from Lemma 17.7.1 that $f'(0)^{-1}$ is a linear transformation. In particular, we note that $\tilde{f}(0) = 0$ and that

$$\tilde{f}'(0) = f'(0)^{-1}f'(0) = I,$$

so by the special case of the inverse function theorem we know that there exists an open set U' containing 0, and an open set V' containing 0, such that $\tilde{f}$ is a bijection from U' to V', and that $\tilde{f}^{-1} : V' \to U'$ is differentiable at 0 with derivative I. But we have $f(x) = f'(0)\tilde{f}(x)$, and hence f is a bijection from U' to $f'(0)(V')$ (note that $f'(0)$ is also a bijection). Since $f'(0)$ and its inverse are both continuous, $f'(0)(V')$ is open, and it certainly contains 0. Now consider the inverse function $f^{-1} : f'(0)(V') \to U'$. Since $f(x) = f'(0)\tilde{f}(x)$, we see that $f^{-1}(y) = \tilde{f}^{-1}(f'(0)^{-1}y)$ for all $y \in f'(0)(V')$ (why? use the fact that $\tilde{f}$ is a bijection from U' to V'). In particular we see that f^{-1} is differentiable at 0.

So all we have to do now is prove the inverse function theorem in the special case, when $x_0 = 0$, $f(x_0) = 0$, and $f'(x_0) = I$. Let $g : E \to \mathbf{R}^n$ denote the function $f(x) - x$. Then $g(0) = 0$ and $g'(0) = 0$. In particular

$$\frac{\partial g}{\partial x_j}(0) = 0$$

for $j = 1, \ldots, n$. Since g is continuously differentiable, there thus exists a ball $B(0, r)$ in E such that

$$\|\frac{\partial g}{\partial x_j}(x)\| \leq \frac{1}{2n^2}$$

for all $x \in B(0, r)$. (There is nothing particularly special about $\frac{1}{2n^2}$, we just need a nice small number here.) In particular, for any $x \in B(0, r)$ and $v = (v_1, \ldots, v_n)$ we have

$$\|D_v g(x)\| = \|\sum_{j=1}^{n} v_j \frac{\partial g}{\partial x_j}(x)\|$$

$$\leq \sum_{j=1}^{n} |v_j| \|\frac{\partial g}{\partial x_j}(x)\|$$

$$\leq \sum_{j=1}^{n} \|v\| \frac{1}{2n^2} \leq \frac{1}{2n}\|v\|.$$

But now for any $x, y \in B(0,r)$, we have by the fundamental theorem of calculus

$$g(y) - g(x) = \int_0^1 \frac{d}{dt} g(x + t(y-x))\, dt$$

$$= \int_0^1 D_{y-x} g(x + t(y-x))\, dt.$$

By the previous remark, the vectors $D_{y-x}g(x + t(y-x))$ have a magnitude of at most $\frac{1}{2n}\|y-x\|$. Thus every component of these vectors has magnitude at most $\frac{1}{2n}\|y-x\|$. Thus every component of $g(y) - g(x)$ has magnitude at most $\frac{1}{2n}\|y-x\|$, and hence $g(y) - g(x)$ itself has magnitude at most $\frac{1}{2}\|y-x\|$ (actually, it will be substantially less than this, but this bound will be enough for our purposes). In other words, g is a contraction. By Lemma 17.6.6, the map $f = g + I$ is thus one-to-one on $B(0,r)$, and the image $f(B(0,r))$ contains $B(0,r/2)$. In particular we have an inverse map $f^{-1} : B(0,r/2) \to B(0,r)$ defined on $B(0,r/2)$.

Applying the contraction bound with $y = 0$ we obtain in particular that

$$\|g(x)\| \leq \frac{1}{2}\|x\|$$

for all $x \in B(0,r)$, and so by the triangle inequality

$$\frac{1}{2}\|x\| \leq \|f(x)\| \leq \frac{3}{2}\|x\|$$

for all $x \in B(0,r)$.

Now we set $V := B(0,r/2)$ and $U := f^{-1}(B(0,r))$. Then by construction f is a bijection from U to V. V is clearly open, and $U = f^{-1}(V)$ is also open since f is continuous. (Notice that if a set is open relative to $B(0,r)$, then it is open in $\mathbf{R}^n$ as well). Now we want to show that $f^{-1} : V \to U$ is differentiable at 0 with derivative $I^{-1} = I$. In other words, we wish to show that

$$\lim_{x \to 0; x \in V \setminus \{0\}} \frac{\|f^{-1}(x) - f^{-1}(0) - I(x-0)\|}{\|x\|} = 0.$$

Since $f(0) = 0$, we have $f^{-1}(0) = 0$, and the above simplifies to

$$\lim_{x \to 0; x \in V \backslash \{0\}} \frac{\|f^{-1}(x) - x\|}{\|x\|} = 0.$$

Let $(x_n)_{n=1}^{\infty}$ be any sequence in $V \backslash 0$ that converges to 0. By Proposition 14.1.5(b), it suffices to show that

$$\lim_{n \to \infty} \frac{\|f^{-1}(x_n) - x_n\|}{\|x_n\|} = 0.$$

Write $y_n := f^{-1}(x_n)$. Then $y_n \in B(0,r)$ and $x_n = f(y_n)$. In particular we have

$$\frac{1}{2}\|y_n\| \leq \|x_n\| \leq \frac{3}{2}\|y_n\|$$

and so since $\|x_n\|$ goes to 0, $\|y_n\|$ goes to zero also, and their ratio remains bounded. It will thus suffice to show that

$$\lim_{n \to \infty} \frac{\|y_n - f(y_n)\|}{\|y_n\|} = 0.$$

But since y_n is going to 0, and f is differentiable at 0, we have

$$\lim_{n \to \infty} \frac{\|f(y_n) - f(0) - f'(0)(y_n - 0)\|}{\|y_n\|} = 0$$

as desired (since $f(0) = 0$ and $f'(0) = I$). □

The inverse function theorem gives a useful criterion for when a function is (locally) invertible at a point x_0 - all we need is for its derivative $f'(x_0)$ to be invertible (and then we even get further information, for instance we can compute the derivative of f^{-1} at $f(x_0)$). Of course, this begs the question of how one can tell whether the linear transformation $f'(x_0)$ is invertible or not. Recall that we have $f'(x_0) = L_{Df(x_0)}$, so by Lemmas 17.1.13 and 17.1.16 we see that the linear transformation $f'(x_0)$ is invertible if and only if the matrix $Df(x_0)$ is. There are many ways to check whether a matrix such as $Df(x_0)$ is invertible; for instance, one can use determinants, or alternatively Gaussian elimination methods. We will not pursue this matter here, but refer the reader to any linear algebra text.

If $f'(x_0)$ exists but is non-invertible, then the inverse function theorem does not apply. In such a situation it is not possible for f^{-1} to exist and be differentiable at x_0; this was remarked in the above proof. But it is still possible for f to be invertible. For instance, the single-variable function $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^3$ is invertible despite $f'(0)$ not being invertible.

Exercise 17.7.1. Let $f : \mathbf{R} \to \mathbf{R}$ be the function defined by $f(x) := x + x^2 \sin(1/x^4)$ for $x \neq 0$ and $f(0) := 0$. Show that f is differentiable and $f'(0) = 1$, but f is not increasing on any open set containing 0 (Hint: show that the derivative of f can turn negative arbitrarily close to 0. Drawing a graph of f may aid your intuition.)

Exercise 17.7.2. Prove Lemma 17.7.1.

Exercise 17.7.3. Let $f : \mathbf{R}^n \to \mathbf{R}^n$ be a continuously differentiable function such that $f'(x)$ is an invertible linear transformation for every $x \in \mathbf{R}^n$. Show that whenever V is an open set in $\mathbf{R}^n$, that $f(V)$ is also open. (Hint: use the inverse function theorem.)

17.8 The implicit function theorem

Recall (from Exercise 3.5.10) that a function $f : \mathbf{R} \to \mathbf{R}$ gives rise to a *graph*

$$\{(x, f(x)) : x \in \mathbf{R}\}$$

which is a subset of $\mathbf{R}^2$, usually looking like a curve. However, not all curves are graphs, they must obey the *vertical line test*, that for every x there is exactly one y such that (x, y) is in the curve. For instance, the circle $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$ is not a graph, although if one restricts to a semicircle such as $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1, y > 0\}$ then one again obtains a graph. Thus while the entire circle is not a graph, certain local portions of it are. (The portions of the circle near $(1, 0)$ and $(0, 1)$ are not graphs over the variable x, but they are graphs over the variable y).

Similarly, any function $f : \mathbf{R}^n \to \mathbf{R}$ gives rise to a graph $\{(x, f(x)) : x \in \mathbf{R}^n\}$ in $\mathbf{R}^{n+1}$, which in general looks like some sort of n-dimensional surface in $\mathbf{R}^{n+1}$ (the technical term for this is a *hypersurface*). Conversely, one may ask which hypersurfaces are actually graphs of some function, and whether that function is continuous or differentiable.

If the hypersurface is given geometrically, then one can again invoke the vertical line test to work out whether it is a graph or not. But what if the hypersurface is given algebraically, for instance the surface $\{(x, y, z) \in \mathbf{R}^3 : xy + yz + zx = -1\}$? Or more generally, a hypersurface of the form $\{x \in \mathbf{R}^n : f(x) = 0\}$, where $f : \mathbf{R}^n \to \mathbf{R}$ is some function? In this case, it is still possible to say whether the hypersurface is a graph, locally at least, by means of the *implicit function theorem.*

Theorem 17.8.1 (Implicit function theorem). *Let E be an open subset of $\mathbf{R}^n$, let $f : E \to \mathbf{R}$ be continuously differentiable, and let $y = (y_1, \ldots, y_n)$ be a point in E such that $f(y) = 0$ and $\frac{\partial f}{\partial x_n}(y) \neq 0$. Then there exists an open subset U of $\mathbf{R}^{n-1}$ containing $(y_1, \ldots, y_{n-1})$, an open subset V of E containing y, and a function $g : U \to \mathbf{R}$ such that $g(y_1, \ldots, y_{n-1}) = y_n$, and*

$$\{(x_1, \ldots, x_n) \in V : f(x_1, \ldots, x_n) = 0\}$$

$$= \{(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) : (x_1, \ldots, x_{n-1}) \in U\}.$$

In other words, the set $\{x \in V : f(x) = 0\}$ is a graph of a function over U. Moreover, g is differentiable at $(y_1, \ldots, y_{n-1})$, and we have

$$\frac{\partial g}{\partial x_j}(y_1, \ldots, y_{n-1}) = -\frac{\partial f}{\partial x_j}(y) / \frac{\partial f}{\partial x_n}(y) \tag{17.1}$$

for all $1 \leq j \leq n-1$.

Remark 17.8.2. The equation (17.1) is sometimes derived using *implicit differentiation.* Basically, the point is that if you know that

$$f(x_1, \ldots, x_n) = 0$$

then (as long as $\frac{\partial f}{\partial x_n} \neq 0$) the variable x_n is "implicitly" defined in terms of the other $n-1$ variables, and one can differentiate the above identity in, say, the x_j direction using the chain rule to obtain

$$\frac{\partial f}{\partial x_j} + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial x_j} = 0$$

which is (17.1) in disguise (we are using g to represent the implicit function defining x_n in terms of $x_1, \ldots, x_n$). Thus, the implicit function theorem allows one to define a dependence implicitly, by means of a constraint rather than by a direct formula of the form $x_n = g(x_1, \ldots, x_{n-1})$.

Proof. This theorem looks somewhat fearsome, but actually it is a fairly quick consequence of the inverse function theorem. Let $F : E \to \mathbf{R}^n$ be the function

$$F(x_1, \ldots, x_n) := (x_1, \ldots, x_{n-1}, f(x_1, \ldots, x_n)).$$

This function is continuously differentiable. Also note that

$$F(y) = (y_1, \ldots, y_{n-1}, 0)$$

and

$$DF(y) = \left(\frac{\partial f}{\partial x_1}(y)^T, \frac{\partial f}{\partial x_2}(y)^T, \ldots, \frac{\partial f}{\partial x_n}(y)^T\right)$$

$$= \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ \frac{\partial f}{\partial x_1}(y) & \frac{\partial f}{\partial x_2}(y) & \ldots & \frac{\partial f}{\partial x_{n-1}}(y) & \frac{\partial f}{\partial x_n}(y) \end{pmatrix}.$$

Since $\frac{\partial f}{\partial x_n}(y)$ is assumed by hypothesis to be non-zero, this matrix is invertible; this can be seen either by computing the determinant, or using row reduction, or by computing the inverse explicitly, which is

$$DF(y)^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ -\frac{\partial f}{\partial x_1}(y)/a & \frac{\partial f}{\partial x_2}(y)/a & \dots & \frac{\partial f}{\partial x_{n-1}}(y)/a & 1/a \end{pmatrix},$$

where we have written $a = \frac{\partial f}{\partial x_n}(y)$ for short. Thus the inverse function theorem applies, and we can find an open set V in E containing y, and an open set W in $\mathbf{R}^n$ containing $F(y) = (y_1, \dots, y_{n-1}, 0)$, such that F is a bijection from V to W, and that F^{-1} is differentiable at $(y_1, \dots, y_{n-1}, 0)$.

Let us write F^{-1} in co-ordinates as

$$F^{-1}(x) = (h_1(x), h_2(x), \dots, h_n(x))$$

where $x \in W$. Since $F(F^{-1}(x)) = x$, we have $h_j(x_1, \dots, x_n) = x_j$ for all $1 \leq j \leq n-1$ and $x \in W$, and

$$f(x_1, \dots, x_{n-1}, h_n(x_1, \dots, x_n)) = x_n.$$

Also, h_n is differentiable at $(y_1, \dots, y_{n-1}, 0)$ since F^{-1} is.

Now we set $U := \{(x_1, \dots, x_{n-1}) \in \mathbf{R}^{n-1} : (x_1, \dots, x_{n-1}, 0) \in W\}$. Note that U is open and contains $(y_1, \dots, y_{n-1}, 0)$. Now we define $g : U \to \mathbf{R}$ by $g(x_1, \dots, x_{n-1}) := h_n(x_1, \dots, x_{n-1}, 0)$. Then g is differentiable at $(y_1, \dots, y_{n-1})$. Now we prove that

$$\{(x_1, \dots, x_n) \in V : f(x_1, \dots, x_n) = 0\}$$

$$= \{(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) : (x_1, \dots, x_{n-1}) \in U\}.$$

First suppose that $(x_1, \dots, x_n) \in V$ and $f(x_1, \dots, x_n) = 0$. Then we have $F(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, 0)$, which lies in W. Thus $(x_1, \dots, x_{n-1})$ lies in W. Applying F^{-1}, we see that $(x_1, \dots, x_n) = F^{-1}(x_1, \dots, x_{n-1}, 0)$. In particular $x_n = h_n(x_1, \dots, x_{n-1}, 0)$, and hence $x_n = g(x_1, \dots, x_{n-1})$. Thus every element of the left-hand set lies in the right-hand set. The reverse inclusion comes by reversing all the above steps and is left to the reader.

Finally, we show the formula for the partial derivatives of g. From the preceding discussion we have

$$f(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1})) = 0$$

for all $(x_1, \ldots, x_{n-1}) \in U$. Since g is differentiable at $(y_1, \ldots, y_{n-1})$, and f is differentiable at $(y_1, \ldots, y_{n-1}, g(y_1, \ldots, y_{n-1})) = y$, we may use the chain rule, differentiating in x_j, to obtain

$$\frac{\partial f}{\partial x_j}(y) + \frac{\partial f}{\partial x_n}(y)\frac{\partial g}{\partial x_j}(y_1, \ldots, y_{n-1})$$

and the claim follows by simple algebra. □

Example 17.8.3. Consider the surface $S := \{(x, y, z) \in \mathbf{R}^3 : xy + yz + zx = -1\}$, which we rewrite as $\{(x, y, z) \in \mathbf{R}^3 : f(x, y, z) = 0\}$, where $f : \mathbf{R}^3 \to \mathbf{R}$ is the function $f(x, y, z) := xy + yz + zx + 1$. Clearly f is continuously differentiable, and $\frac{\partial f}{\partial z} = y + x$. Thus for any (x_0, y_0, z_0) in S with $y_0 + x_0 \neq 0$, one can write this surface (near (x_0, y_0, z_0)) as a graph of the form $\{(x, y, g(x, y)) : (x, y) \in U\}$ for some open set U containing (x_0, y_0), and some function g which is differentiable at (x_0, y_0). Indeed one can implicitly differentiate to obtain that

$$\frac{\partial g}{\partial x}(x_0, y_0) = -\frac{y_0 + z_0}{y_0 + x_0} \text{ and } \frac{\partial g}{\partial y}(x_0, y_0) = -\frac{x_0 + z_0}{y_0 + x_0}.$$

In the implicit function theorem, if the derivative $\frac{\partial f}{\partial x_n}$ equals zero at some point, then it is unlikely that the set $\{x \in \mathbf{R}^n : f(x) = 0\}$ can be written as a graph of the x_n variable in terms of the other $n - 1$ variables near that point. However, if some other derivative $\frac{\partial f}{\partial x_j}$ is zero, then it would be possible to write the x_j variable in terms of the other $n - 1$ variables, by a variant of the implicit function theorem. Thus as long as the gradient ∇f is not entirely zero, one can write this set $\{x \in \mathbf{R}^n : f(x) = 0\}$ as a graph of *some* variable x_j in terms of the other $n - 1$ variables. (The circle $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 - 1 = 0\}$ is a good example of this; it is not a graph of y in terms of x, or x in terms of y, but near every point it is one of the two. And this is because the gradient of x^2+y^2-1 is never zero on the circle.) However, if ∇f does vanish at some point x_0, then we say that f has a *critical point* at x_0 and the behavior there is much more complicated. For instance, the set $\{(x,y) \in \mathbf{R}^2 : x^2 - y^2 = 0\}$ has a critical point at $(0,0)$ and there the set does not look like a graph of any sort (it is the union of two lines).

Remark 17.8.4. Sets which look like graphs of continuous functions at every point have a name, they are called *manifolds*. Thus $\{x \in \mathbf{R}^n : f(x) = 0\}$ will be a manifold if it contains no critical points of f. The theory of manifolds is very important in modern geometry (especially differential geometry and algebraic geometry), but we will not discuss it here as it is a graduate level topic.