7 Multivariable calculus

We now specialize to Euclidean spaces and consider the calculus of functions from $(\mathbb{R}^n, d_2)$, or an open subset thereof, to $(\mathbb{R}^m, d_2)$. As we make frequent use of the Euclidean norm, we denote it simply by $\|\cdot\|$, instead of $\|\cdot\|_2$, as we have been careful to do until now.

Much of what we prove holds for any norm—or indeed, for any normed vector space—but our focus is on Euclidean space.

7.1 The derivative

Recall that the derivative of the function $f : (a, b) \to \mathbb{R}$ at the point $c \in (a, b)$ describes how the function f changes near c. The function f is differentiable at the point $c \in (a, b)$ if there is a number $f'(c) \in \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(c+h) - f(c)}{h} = f'(c).$$

Unfortunately, this definition of the derivative as the value of a limit is inherently one-dimensional. However, a slightly different perspective does generalize. Note that if $f'(c)$ exists, then its defining limit can be rewritten as

$$\lim_{h\to 0} \frac{f(c+h) - f(c) - f'(c)\cdot h}{h} = 0.$$

This is then equivalent to

$$\lim_{h\to 0} \frac{|f(c+h) - f(c) - f'(c)\cdot h|}{|h|} = 0.$$

It is this version of the derivative that we generalize.

Definition 7.1.1. Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, and that $f : U \to \mathbb{R}^m$. Then f is *differentiable at* c if there is a linear transformation $Df(c) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(c+h) - f(c) - Df(c)(h)\|}{\|h\|} = 0,$$

and in this case $Df(c)$ is the *derivative of* f *at* c. If f is differentiable at every point $c \in U$, then f is called *differentiable* (on U).

For every point $c \in \mathbb{R}^n$, we have a derivative $Df(c)$, which is by definition a linear transformation, thus $Df(c)(h)$ represents this linear transformation, at the point c, applied to the vector h.

We prove that derivatives are unique shortly (Proposition 7.1.4), which justifies our calling this *the* derivative.

Therefore, if the function f in the above definition is differentiable at c, then when $\|h\|$ is small, we have

$$f(c+h) \approx f(c) + Df(c)(h).$$

Let us consider an example.

Example 7.1.2. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x_1, x_2) = (x_1^2, x_1 x_2)$. Let $c = (c_1, c_2)$ be fixed and define A to be the matrix

$$A = \begin{pmatrix} 2c_1 & 0 \\ c_2 & c_1 \end{pmatrix}.$$

Given a vector $h = (h_1, h_2) \in \mathbb{R}^2$, we compute that

$$\left\| f\begin{pmatrix} c_1 + h_1 \\ c_2 + h_2 \end{pmatrix} - f\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \begin{pmatrix} 2c_1 & 0 \\ c_2 & c_1 \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \right\| = \left\| \begin{pmatrix} h_1^2 \\ h_1 h_2 \end{pmatrix} \right\| = |h_1|\|h\|.$$

It follows that f is differentiable at (c_1, c_2) and that $Df(c) = L_A$, where L_A denotes the linear transformation defined by $L_A(x) = Ax$.

Throughout this chapter, when it is clear from the context, we frequently write column vectors as row vectors, and we suppress the "extra parentheses" on functions. Strictly speaking, our $f(x_1, x_2)$ in this example should be

$$f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) \quad \text{or} \quad f\left(\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T\right).$$

We leave the proof of the following result to the reader as Exercise 7.1.2.

Proposition 7.1.3 (Differentiability implies continuity)**.** *Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, and that $f : U \to \mathbb{R}^m$. If f is differentiable at c, then f is continuous at c.*

We now verify that derivatives are unique.

Proposition 7.1.4 (Uniqueness of derivatives)**.** *Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, and that $f : U \to \mathbb{R}^m$. If $S, T : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations and both*

$$\lim_{h \to 0} \frac{\|f(c+h) - f(c) - S(h)\|}{\|h\|} = 0$$

and

$$\lim_{h \to 0} \frac{\|f(c+h) - f(c) - T(h)\|}{\|h\|} = 0,$$

then $S = T$.

Proof. Let $\epsilon > 0$ be arbitrary. We can find a $\delta \in (0,1)$ such that for every $h \in \mathbb{R}^n$ with $0 < \|h\| < \delta$, we have $c + h \in U$, and both

That we can find a $\delta > 0$ such that $c + h \in U$ whenever $\|h\| < \delta$ follows from the fact that U is open and $c \in U$.

$$d_2\left(f(c+h) - f(c), S(h)\right) < \epsilon\|h\|$$
$$d_2\left(f(c+h) - f(c), T(h)\right) < \epsilon\|h\|.$$

It follows that by the triangle inequality that for every such h,

$$d_2(S(h), T(h)) < 2\epsilon\|h\|.$$

Now let $x \in \mathbb{R}^n$ be arbitrary. Setting $\lambda = \delta/(\|x\| + 1) > 0$, we have

$$\|\lambda x\| = \frac{\delta\|x\|}{\|x\| + 1} < \delta < 1,$$

so

$$d_2(S(\lambda x), T(\lambda x)) < 2\epsilon\|\lambda x\| < 2\epsilon.$$

This holds for all $\epsilon > 0$, so it follows that we must have $S(\lambda x) = T(\lambda x)$. Since S and T are both linear transformations and $\lambda > 0$, this implies that $S(x) = T(x)$. Finally, since $x \in \mathbb{R}^n$ was arbitrary, we have that $S = T$, as desired. □

Exercises

Exercise 7.1.1. Show, directly from the definition, that the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x_1, x_2) = (x_1^2 - x_2^2, 2x_1x_2)$$

is differentiable (on all of $\mathbb{R}^2$) and compute its derivative at each point. At which points c does the derivative $Df(c)$ fail to be invertible?

Exercise 7.1.2. Prove Proposition 7.1.3 (differentiability implies continuity).

Exercise 7.1.3. Suppose that A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. Show that the function $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by

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

is differentiable and compute $Df(c)$ at every point $c \in \mathbb{R}^n$.

7.2 Directional and partial derivatives

In the previous section we defined the (total) derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ as the multidimensional analogue of the single variable derivative. There is another way we might think of extending the single variable derivative to multivariable functions, which is to pick a particular line in $\mathbb{R}^n$ and take a derivative of our function along this line.

This is the notion of a directional derivative, but first we need to make an observation. Suppose that we have an open set $U \subseteq \mathbb{R}^n$, a point $c \in U$, and a unit vector $u \in \mathbb{R}^n$. Because U is open and $c \in U$, there is some $\delta > 0$ such that

$$N_\delta(c) \subseteq U.$$

This implies that

$$\{c + tu : t \in (-\delta, \delta)\} \subseteq U,$$

so U contains an open line segment centered at c.

Definition 7.2.1. Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, that $u \in \mathbb{R}^n$ is a unit vector, and that $f : U \to \mathbb{R}^m$. The *derivative of f in the direction u at the point c* is defined to be

$$D_u f(c) = \lim_{t \to 0} \frac{f(c + tu) - f(c)}{t},$$

if this limit exists.

The limit in the definition of directional derivatives is to be taken in the Euclidean space $(\mathbb{R}^m, d_2)$.

Our only result of this section relates these directional derivatives to the total derivative.

Proposition 7.2.2. *Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, that $u \in \mathbb{R}^n$ is a unit vector, and that $f : U \to \mathbb{R}^m$. If f is differentiable at the point c, then f is also differentiable in the direction u at the point c, and*

$$D_u f(c) = Df(c)(u).$$

The equality in this proposition states that the derivative in the direction u is equal to the (total) derivative of f at the point c evaluated at u.

Proof. We would like to show that

$$\lim_{t \to 0} \frac{f(c + tu) - f(c)}{t} = Df(c)(u).$$

This is a limit in the Euclidean space $(\mathbb{R}^m, d_2)$, and it holds if and only if

$$\lim_{t \to 0} \left\| \frac{f(c + tu) - f(c)}{t} - Df(c)(u) \right\| = 0.$$

Recall that we denote the Euclidean norm by $\| \cdot \|$ throughout this chapter.

We are given that f is differentiable at the point c, and thus we have

$$\lim_{h \to 0} \frac{\|f(c + h) - f(c) - Df(c)(h)\|}{\|h\|} = 0.$$

If we take $h = tu$ in the above limit, we have $\|h\| = |t|$ because u is a unit vector, and thus we see that

$$\lim_{t\to 0} \frac{\|f(c+tu) - f(c) - Df(c)(tu)\|}{|t|} = 0.$$

Finally, we have $Df(c)(tu) = tDf(c)(u)$ because $Df(c)$ is linear, and thus the limit above implies that the limit that we seek to prove, completing the proof. $\square$

When u is one of the standard basis vectors of $\mathbb{R}^n$, we obtain the familiar partial derivatives.

Definition 7.2.3. Suppose that $U \subseteq \mathbb{R}^n$ is open, that $c \in U$, and that $f : U \to \mathbb{R}^m$. The *partial derivative of* f with respect to x_j at c is the directional derivative of f in the direction e_j at c and is denoted by $D_j f(c)$.

In the case where $m = 1$, so that $f : \mathbb{R}^n \to \mathbb{R}$, it is customary to write

$$\frac{\partial f}{\partial x_j}(c)$$

instead of $D_j f(c)$. Of course if $f : \mathbb{R}^n \to \mathbb{R}^m$, then we can express f as a column vector of functions with codomain $\mathbb{R}$,

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

and then we have

$$D_j f(c) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(c) \\ \frac{\partial f_2}{\partial x_j}(c) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(c) \end{pmatrix}.$$

Exercises

Exercise 7.2.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. For a point $c \in \mathbb{R}^n$, find the unit vector u that maximizes $D_u f(c)$. This direction is the *direction of maximum increase of f at c.* Compare with Exercise **??**.

Exercise 7.2.2. Suppose $f : \mathbb{R}^3 \to \mathbb{R}^2$ is differentiable at 0 and

$$Df(0) = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Find the directional derivative in the direction of the vector

$$v = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}.$$

Exercise 7.2.3. Suppose that $U \subseteq \mathbb{R}^n$ is open and that $f : U \to \mathbb{R}$. Define what it means for f to have a *local minimum* at a point $c \in U$, and prove that if f has a local minimum at c and f is differentiable at c, then $Df(c) = 0$.

The statement that $Df(c) = 0$ in Exercise 7.2.3 means that $Df(c)$ is the zero transformation (the transformation that maps every vector in $\mathbb{R}^n$ to $0 \in \mathbb{R}$). One way to prove this would be to show that $Df(c)(u) = 0 \in \mathbb{R}$ for every unit vector $u \in \mathbb{R}^n$.

Exercise 7.2.4. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \text{ and} \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Show that the partial derivatives of f exist at the origin, even though f is not continuous there.

Exercise 7.2.5. Suppose that $U \subseteq \mathbb{R}^2$ is open and that $f : U \to \mathbb{R}$. Prove that if the partial derivatives of f exist and are bounded, then f is continuous.

Exercise 7.2.6. Suppose that $U \subseteq \mathbb{R}^n$ is open and connected and that the function $f : U \to \mathbb{R}$ is differentiable and satisfies $Df(c) = 0$ for all $c \in U$. Prove that f is a constant function.

To start Exercise 7.2.6, one might want to prove that if $N_\delta(c) \subseteq U$, then f is constant on $N_\delta(c)$. To do this, consider the single variable function defined by $f(c + tu)$ for $t \in (-\delta, \delta)$.

7.3 The chain rule

In this section we prove the multivariable chain rule. First though, we investigate what it should say. Suppose that we have two functions,

$$\begin{aligned} f &: \mathbb{R}^n \to \mathbb{R}^m, \\ g &: \mathbb{R}^m \to \mathbb{R}^p. \end{aligned}$$

If f is differentiable at the point $c \in \mathbb{R}^n$ and g is differentiable at the point $d = f(c) \in \mathbb{R}^m$, then the chain rule tells us what the derivative of their composition $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ is at the point c.

Since f is differentiable at c, for small $h \in \mathbb{R}^n$ we have

$$f(c+h) - f(c) \approx Df(c)(h).$$

Similarly, since g is differentiable at $d = f(c)$, for small $k \in \mathbb{R}^m$,

$$g(d+k) - g(d) \approx Dg(d)(k).$$

In order to motivate the formula in the chain rule, suppose momentarily that f and g are both linear functions, and so the approximations above are actually equalities. Then we would have

$$\begin{aligned} g(f(c+h)) &= g\left(f(c) + Df(c)(h)\right) \\ &= g\left(f(c)\right) + Dg(f(c))\Big(Df(c)(h)\Big) \\ &= g(f(c)) + \Big(Dg(f(c)) \circ Df(c)\Big)(h). \end{aligned}$$

Thus if f and g are both linear, then the derivative of the composition $g \circ f$ is given by the composition $Dg(f(c)) \circ Df(c)$.

Since differentiability means that functions can be locally approximated by a linear transformations, we might hope that this conclusion holds in general. We just need to quantify and control the approximations involved.

Theorem 7.3.1 (Chain rule)**.** *Suppose that $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open and that $f : U \to V$ and $g : V \to \mathbb{R}^p$. If f is differentiable at $c \in U$ and g is differentiable at $f(c) \in V$, then $g \circ f$ is differentiable at c and*

$$D(g \circ f)(c) = Dg(f(c)) \circ Df(c).$$

Proof. Define $d = f(c)$ and $k = f(c+h) - f(c)$, so $d + k = f(c+h)$. Ultimately, we seek to show that

$$\frac{\Big\| g(d+k) - g(d) - \Big(Dg(d) \circ Df(c)\Big)(h) \Big\|}{\|h\|} \to 0$$

as $h \to 0$. Thus we need to bound the numerator of this fraction in terms of $\|h\|$. By the triangle inequality, this numerator is at most

$$\Big\|g(d+k) - g(d) - Dg(d)(k)\Big\| + \Big\|Dg(d)(k) - \Big(Dg(d) \circ Df(c)\Big)(h)\Big\|.$$

We bound these two terms separately. Let $\epsilon \in (0,1)$ be arbitrary.

First, because g is differentiable at d, there is some $\delta > 0$ such that if $\|k\| \in (0, \delta)$, then

$$\Big\|g(d+k) - g(d) - Dg(d)(k)\Big\| < \epsilon\|k\|.$$

Since g is differentiable at d, there is a linear transformation $Dg(d)$ such that

$$\lim_{k \to 0} \frac{\|g(d+k) - g(d) - Dg(d)(k)\|}{\|k\|} = 0$$

This bounds one of the terms we are interested in, but we want to bound it in terms of $\|h\|$, not $\|k\|$.

Because f is differentiable at c, there is some $\gamma > 0$ such that if $\|h\| \in (0, \gamma)$, then

$$\|\underbrace{f(c+h) - f(c)}_{k} - Df(c)(h)\| < \epsilon\|h\| < \|h\|.$$

We can obtain a bound on $\|k\|$ from this inequality. By the triangle inequality, we have

We have $\epsilon\|h\| < \|h\|$ here because $\epsilon < 1$.

$$\|k\| < \|Df(c)(h)\| + \|h\| \leq \Big(1 + \|Df(c)\|_{\text{op}}\Big)\|h\|.$$

From this inequality we obtain two things. First we see that if

The $\|Df(c)\|_{\text{op}}$ here is just a constant, so we have bounded $\|k\|$ by a constant times $\|h\|$.

$$\|h\| < \min\left\{\gamma,\ \frac{\delta}{1 + \|Df(c)\|_{\text{op}}}\right\},$$

then we have both $\|h\| < \gamma$ and $\|k\| < \delta$, and thus our inequalities above will hold. Secondly, we see that for such vectors h, we have

$$\Big\|g(d+k) - g(d) - Dg(d)(k)\Big\| < \epsilon\Big(1 + \|Df(c)\|_{\text{op}}\Big)\|h\|,$$

which gives us the bound in terms of $\|h\|$ that we need.

It remains to bound the quantity

$$\Big\|Dg(d)(k) - \Big(Dg(d) \circ Df(c)\Big)(h)\Big\|.$$

From the linearity of $Dg(d)$, we see that

$$\begin{aligned}\Big\|Dg(d)(k) - \Big(Dg(d) \circ Df(c)\Big)(h)\Big\| &= \Big\|Dg(d)\Big(k - Df(c)(h)\Big)\Big\| \\ &\leq \|Dg(d)\|_{\text{op}}\,\|k - Df(c)(h)\| \\ &< \epsilon\|Dg(d)\|_{\text{op}}\|h\|,\end{aligned}$$

provided that $\|h\| < \gamma$. This bounds the second quantity we sought to bound.

Combining our two bounds, we see that for

$$\|h\| < \min\left\{ \gamma, \frac{\delta}{1 + \|Df(c)\|_{\text{op}}} \right\},$$

we have

$$\left\| g(d+k) - g(d) - \Big(Dg(d) \circ Df(c) \Big)(h) \right\| < \epsilon \Big(1 + \|Df(c)\|_{\text{op}} + \|Dg(d)\|_{\text{op}} \Big) \|h\|.$$

Since $\epsilon > 0$ was arbitrary, this implies that

$$\lim_{h \to 0} \frac{\left\| g(d+k) - g(d) - \Big(Dg(f(c)) \circ Df(c) \Big)(h) \right\|}{\|h\|} = 0,$$

proving the theorem. □

Exercises

Exercise 7.3.1. Suppose that $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable. Write out the chain rule for $g \circ f$ explicitly in terms of the partial derivatives of f and g.

Exercise 7.3.2. Suppose that $f, g : \mathbb{R}^2 \to \mathbb{R}^2$ are differentiable. Write out the chain rule for $g \circ f$ explicitly in terms of the partial derivatives of f and g.

7.4 The contraction mapping theorem

Definition 7.4.1. Let (X, d_X) and (Y, d_Y) be metric spaces. The function $f : X \to Y$ is *Lipschitz continuous* if there is a constant $c \in (0, \infty)$ such that for all $x_1, x_2 \in X$, we have

$$d_Y(f(x_1), f(x_2)) \leq c d_X(x_1, x_2).$$

Any constant $c \in (0, \infty)$ that satisfies this condition is called a *Lipschitz constant for f*. If f has a Lipschitz constant $c \in (0, 1]$, then f is called a *contraction*. If f has a Lipschitz constant $c \in (0, 1)$, then f is called a *strict contraction*.

For example, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is any linear transformation, then it is Lipschitz continuous with Lipschitz constant $\|T\|_{\text{op}}$, because

$$\|T(x_1) - T(x_2)\| = \|T(x_1 - x_2)\| \leq \|T\|_{\text{op}} \|x_1 - x_2\|$$

for all $x_1, x_2 \in X$.

Specializing to the real line (with the standard metric), we see that the function defined by $f(x) = |x|$ is Lipschitz continuous with Lipschitz constant 1. Thus it is a contraction. It is not a strict contraction, however, and it is not everywhere differentiable. The function defined by $f(x) = x + 2$ is also a contraction but not a strict contraction. The function defined by $f(x) = x/2$ is Lipschitz continuous with Lipschitz constant $1/2$, so it is a strict contraction.

The proof of the following result is left to the reader in Exercise 7.4.1.

Proposition 7.4.2. *Let (X, d_X) and (Y, d_Y) be metric spaces. If $f : X \to Y$ is Lipschitz continuous, then it is uniformly continuous.*

We are particularly interested in points that are mapped to themselves by contractions. Obviously, for there to be any such point, our function must map a metric space onto itself.

Definition 7.4.3. The point $x \in X$ is a *fixed point* of the mapping $f : X \to X$ if $f(x) = x$.

When strict contractions have fixed points, they are unique.

Proposition 7.4.4. *Let (X, d) be a metric space and $f : X \to X$ be a strict contraction. Then f has at most one fixed point.*

Proof. Let $f : X \to X$ be a strict contraction and let $c \in (0, 1)$ be a Lipschitz constant for f. If $x, y \in X$ are both fixed points, then on the one hand we have

$$d(f(x), f(y)) = d(x, y),$$

because $f(x) = x$ and $f(y) = y$, while on the other hand we have

$$d(f(x), f(y)) \leq cd(x, y).$$

We can only have $d(x, y) \leq cd(x, y)$ for $c \in (0, 1)$ if $d(x, y) = 0$, which implies that $x = y$, as desired. □

The result this section is named for shows that if a metric space is complete, then every strict contraction has a unique fixed point.

Theorem 7.4.5 (Contraction mapping theorem)**.** *Let (X, d) be a nonempty complete metric space and $f : X \to X$ be a strict contraction. Then f has a unique fixed point.*

Proof. By our previous result, it suffices to find a fixed point of f. Let $f : X \to X$ be a strict contraction and let $c \in (0, 1)$ be a Lipschitz constant for f. Choose $x_0 \in X$ arbitrarily, and define a sequence (x_n) recursively, letting

We may choose a point $x_0 \in X$ because we have assumed that X is nonempty.

$$x_{n+1} = f(x_n)$$

for each $n \geq 0$. For all $n \geq 1$, we have

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq cd(x_n, x_{n-1}).$$

It follows by induction that

$$d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$$

for all $n \geq 0$. We can use this inequality to show that the sequence (x_n) is Cauchy. Indeed, for $m \geq n$, we have

The first inequality here comes from the triangle inequality for the metric d.

$$\begin{aligned} d(x_m, x_n) &\leq \sum_{k=n}^{m-1} d(x_{k+1}, x_k) \\ &\leq (c^n + c^{n+1} + \cdots + c^{m-1})\, d(x_1, x_0) \\ &\leq c^n \frac{d(x_1, x_0)}{1 - c}. \end{aligned}$$

Since $d(x_1, x_0)/(1 - c)$ can be viewed as a constant and $c < 1$, this inequality implies that (x_n) is Cauchy. Because we have assumed that (X, d) is complete, $\lim x_n = x$ for some $x \in X$. Finally, because f is continuous (Proposition 7.4.2), we have that

$$f(x) = \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} x_{n+1} = x,$$

proving that x is a fixed point, as desired. □

Exercises

Exercise 7.4.1. Prove Propposition 7.4.1: Lipschitz continuity implies uniform continuity.

Exercise 7.4.2. Define the mapping $f : [0,1] \to [0,1]$ by $f(x) = x - x^2$, with the standard metric on $[0,1]$. Show that f is a contraction but not a strict contraction.

Exercise 7.4.3. What does the contraction mapping theorem say when (X,d) is a discrete metric space? Provide a simple proof of this specialization.

7.5 Continuous differentiability

Suppose that $U \subseteq \mathbb{R}^n$ and that $f : U \to \mathbb{R}^m$. We have defined the (total) derivative of f as a function

$$Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m).$$

Here we are interested in what it means for this derivative to be continuous.

Definition 7.5.1. Suppose that $U \subseteq \mathbb{R}^n$ is open and that $f : U \to \mathbb{R}^m$. We say that f is *continuously differentiable* if $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ exists and is continuous at every point of U.

In the definition above, we interpret $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ as a metric space with the metric induced by the operator norm. Thus for Df to be continuous simply means that for every point $c \in U$ and every $\epsilon > 0$, there is a $\delta > 0$ (possibly depending on c) such that $\|Df(x) - Df(c)\|_{\text{op}} < \epsilon$ whenever $x \in U$ satisfies $\|x - c\| < \delta$.

On the other hand, if we represent f in terms of its components as

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix},$$

then if $Df(x)$ exists, its matrix representation is given by

$$\begin{pmatrix} D_1 f_1(x) & \cdots & D_n f_1(x) \\ \vdots & & \vdots \\ D_1 f_m(x) & \cdots & D_n f_m(x) \end{pmatrix}.$$

Thus it is natural to ask how the continuity of Df relates to the continuity of the partial derivatives $D_k f_i$.

Our first result shows that the continuity of Df implies the continuity of every directional derivative of f. Recall that we have proved previously that the directional derivative $D_u f(x)$ is equal to $Df(x)(u)$.

Proposition 7.5.2. *Suppose that $U \subseteq \mathbb{R}^n$ is open and that $f : U \to \mathbb{R}^m$ is continuously differentiable. Then for every unit vector $u \in \mathbb{R}^n$, the directional derivative*

$$D_u f : U \to \mathbb{R}^m$$

is continuous.

The converse of Proposition 7.5.2 is true too. In fact if all of the partials $D_k f_i$ are continuous, then Df is continuous. It is a little more work, though, and we only need this direction to establish the inverse function theorem.

Proof. Suppose that f is continuously differentiable and let $c \in U$ and $\epsilon > 0$ be given. There is some $\delta > 0$ such that

$$\|Df(x) - Df(c)\|_{\text{op}} < \epsilon$$

whenever $x \in U$ satisfies $\|x - c\| < \delta$. Thus for these vectors x, we have

$$\begin{aligned}\|D_u f(x) - D_u f(c)\| &= \|Df(x)(u) - Df(c)(u)\| \\ &= \|\left(Df(x) - Df(c)\right)(u)\| \\ &\le \|Df(x) - Df(c)\|_{\text{op}} \|u\| \\ &< \epsilon,\end{aligned}$$

proving the result. □

Our next result requires an additional hypothesis on our domain.

Definition 7.5.3. A subset $U \subseteq \mathbb{R}^n$ is said to be *convex* if for every $x, y \in U$ and every $t \in [0, 1]$, the linear combination $ty + (1-t)x$ also lies in U.

Note that neighborhoods are convex.

Proposition 7.5.4. *Suppose that $U \subseteq \mathbb{R}^n$ is open and convex and that $f : U \to \mathbb{R}^m$ is differentiable. If the derivative Df satisfies $\|Df(x)\|_{\text{op}} \le M$ for all $x \in U$, then f is Lipschitz continuous on U with constant M, meaning that*

$$d_2(f(y), f(x)) \le M\, d_2(y, x)$$

for all $x, y \in U$.

The condition that Df satisfy $\|Df(x)\|_{\text{op}} \le M$ for all $x \in U$ is equivalent to Df being bounded, as a mapping from U to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Proof. Let $x, y \in U$ be arbitrary. Since U is convex, we know that

$$ty + (1-t)x \in U$$

for all $t \in [0, 1]$, but actually since U is open, there is some $\delta > 0$ so that this holds for all $t \in (-\delta, 1+\delta)$. Define the function $s : (-\delta, 1+\delta) \to U$ by

$$s(t) = ty + (1-t)x,$$

and let $g(t) = f \circ s : (-\delta, 1+\delta) \to \mathbb{R}^m$. Since both f and s are differentiable, the chain rule implies that g is differentiable and that

$$Dg(t) = Df(s(t))s'(t) = Df(s(t))(y - x).$$

Thus

$$\|Dg(t)\| \le \|Df(s(t))\|_{\text{op}} \|y - x\| \le M\|y - x\|$$

for all $t \in (-\delta, 1+\delta)$.

From the first fundamental theorem of calculus applied to the components of g, it follows that

$$f(y) - f(x) = g(1) - g(0) = \int_0^1 Dg(t)\, dt.$$

We would like to claim that

$$\|f(y) - f(x)\| \leq \int_0^1 \|Dg(t)\|\, dt \leq M\|y - x\|,$$

which would complete the proof if we knew we were allowed to move the norm inside the integral like this. That can be justified, but here is another way to get what we want.

Viewing x and y as constants, define the function $h : [0, 1] \to \mathbb{R}$ by

$$h(t) = \langle f(y) - f(x), g(t) \rangle.$$

We have

$$\begin{aligned} h(1) - h(0) &= \langle f(y) - f(x), g(1) \rangle - \langle f(y) - f(x), g(0) \rangle \\ &= \langle f(y) - f(x), f(y) \rangle - \langle f(y) - f(x), f(x) \rangle \\ &= \langle f(y) - f(x), f(y) - f(x) \rangle \\ &= \|f(y) - f(x)\|^2. \end{aligned}$$

We also see that h is differentiable, and

$$h'(t) = \langle f(y) - f(x), Dg(t) \rangle.$$

Therefore by the single-variable mean value theorem,

$$h(1) - h(0) = \langle f(y) - f(x), Dg(\xi) \rangle$$

for some point $\xi \in (0, 1)$. By the Cauchy–Schwarz inequality and our hypotheses,

$$\begin{aligned} \langle f(y) - f(x), Dg(\xi) \rangle &\leq \|f(y) - f(x)\| \|Dg(\xi)\| \\ &\leq \|f(y) - f(x)\| \cdot M\|y - x\|. \end{aligned}$$

It follows that

$$\|f(y) - f(x)\|^2 = |h(1) - h(0)|^2 \leq \|f(y) - f(x)\| \cdot M\|y - x\|,$$

proving the result. □

As a consequence of this result, we see that if f is continuously differentiable and if c is a point such that $Df(c) = 0$, then in some neighborhood of c, f acts as a strict contraction. This follows because the continuity of Df implies that there is a neighborhood $N_\delta(c)$ of c on which $\|Df(x)\|_{\text{op}} < 1/2$, and then our previous result shows that on this neighborhood, f is Lipschitz with constant $1/2$.

7.6 The inverse function theorem

In this section we prove a single theorem.

Theorem 7.6.1 (The inverse function theorem). *Suppose that $U \subseteq \mathbb{R}^n$ is open and that $f : U \to \mathbb{R}^n$. If f is continuously differentiable on U and $Df(c)$ is invertible for some point $c \in U$, then there is an open set $V \subseteq U$ containing the point c such that*

(a) $f|_V : V \to f(V)$ is a bijection,
(b) $f(V)$ is open,
(c) $(f|_V)^{-1}$ is continuously differentiable, and
(d) $D(f|_V)^{-1}(f(c)) = (Df(c))^{-1}$.

Before we begin the proof, note that the final part of the theorem—the computation of the derivative of f^{-1} at the point $f(c)$—follows routinely from the other parts of the theorem and the chain rule. If we know that f^{-1} exists and is differentiable near $f(c)$, then since $f^{-1} \circ f$ is the identity, the chain rule says that

$$D(f^{-1})(f(x)) \circ Df(x)$$

is equal to the derivative of the identity (which is also the identity), and thus

$$D(f^{-1})(f(x)) = (Df(x))^{-1}.$$

To keep our proof of the other parts of the theorem as clean as possible, we additionally assume that $c = 0$ and that $Df(c)$ is the identity transformation. The first assumption is clearly harmless; we can simply shift the function to make this true. To see that the second assumption is also harmless, suppose that $Df(c) = T$. Our assumptions imply that T is invertible, so T^{-1} exists and (as with all linear transformations) the mapping $x \mapsto T^{-1}x$ is continuous and is its own derivative. Therefore the function $x \mapsto T^{-1}f(x)$ satisfies all of the hypotheses of the theorem and has derivative equal to the identity.

We are now ready to begin the proof.

Proof of the inverse function theorem. As explained above, we assume that $c = 0$ and that $Df(0) = I$, where I is the identity transformation. We define the function g by $g(x) = x - f(x)$. Because f is continuously differentiable, so is g, and we have

$$Dg(0) = I - I = 0.$$

Thus $\|Dg(0)\|_{\text{op}} = 0$. Since g is continuously differentiable and U is open, there is some open neighborhood of 0 on which $\|Dg(x)\|_{\text{op}} < 1/2$.

Here, $Dg(0) = 0$ means the derivative of g at the origin is the zero linear transformation (that sends every vector to the zero vector).

In proving that f is one-to-one on V, we use only the easier of our two contraction results, which does not guarantee the existence of a fixed point, only that any such fixed point must be unique. We use the more substantial contraction mapping theorem, which guarantees a unique fixed point if the domain is complete, to prove that $f(V)$ is open.

Let $V \subseteq U$ denote this neighborhood. By Proposition 7.5.4, g is Lipschitz continuous on V with constant $1/2$, or in other words, g is a strict contraction on V. Also since $\|Dg(x)\|_{\text{op}} < 1/2$ for $x \in V$ and $f(x) = x - g(x)$, it follows from Proposition **??** that the linear transformation $Df(x) = I - Dg(x)$ is invertible for all $x \in V$.

Proposition **??** states that if A is an $n \times n$ matrix and $\|A\|_{\text{op}} < 1$, then $I_n - A$ is invertible.

To prove (a), it suffices to prove that $f|_V : V \to f(V)$ is an injection, since every function is surjective on its range. For any point $y \in \mathbb{R}^n$, define a function $g_y : V \to \mathbb{R}^n$ by

$$g_y(x) = y + x - f(x).$$

Note that $g_y(x) = x$ if and only if $f(x) = y$. Since we are treating y as a constant, we have $Dg_y(x) = Dg(x)$ for all $x \in V$, so every one of these functions is a strict contraction on V. Because every strict contraction has at most one fixed point (Proposition 7.4.4), for every point $y \in \mathbb{R}^n$, there is at most one point $x \in V$ such that $g_y(x) = x$. This means that there is at most one point $x \in V$ with $f(x) = y$, and so f is injective on V, which is enough to prove part (a) of the theorem.

Now we would like to show that $f(V)$ is open. Let $y_0 \in f(V)$ be arbitrary. Since $y_0 \in f(V)$, there is some $x_0 \in V$ such that $f(x_0) = y_0$. Since V is open, there is some radius $r > 0$ such that

Here $\overline{N_r(x_0)}$ is the closure of the (open) neighborhood $N_r(x_0)$.

$$\overline{N_r(x_0)} \subseteq V.$$

We would like to show that $N_{r/2}(y_0) \subseteq f(V)$, which will prove that $f(V)$ is open. To this end, fix $y \in N_{r/2}(y_0)$, so $\|y - y_0\| < r/2$, and let $x \in \overline{N_r(x_0)}$ be arbitrary. By the triangle inequality, we have

$$\|g_y(x) - x_0\| \leq \|g_y(x) - g_y(x_0)\| + \|g_y(x_0) - x_0\|.$$

We know that g_y is Lipschitz continuous on V with constant $1/2$, so

$$\|g_y(x) - g_y(x_0)\| \leq \frac{1}{2}\|x - x_0\| \leq \frac{r}{2}.$$

We also have

$$\|g_y(x_0) - x_0\| = \|y + x_0 - f(x_0) - x_0\| = \|y - y_0\| < \frac{r}{2}.$$

Together, these two inequalities show that

$$\|g_y(x) - x_0\| < r$$

for all $x \in \overline{N_r(x_0)}$. This means that $g_y(x) \in \overline{N_r(x_0)}$ whenever $x \in \overline{N_r(x_0)}$, so g_y is a strict contraction on $\overline{N_r(x_0)}$. Since $\overline{N_r(x_0)} \subseteq \mathbb{R}^n$ is closed, it is complete. Therefore the contraction mapping theorem (Theorem 7.4.5) implies that g_y has a unique fixed point. Thus there is some point $x \in \overline{N_r(x_0)}$ for which $g_y(x) = x$, which shows that $y \in f(V)$, completing our proof that $f(V)$ is open.

Proposition **??** says that a closed subset of a complete metric space is complete.

It remains only to show that $(f|_V)^{-1}$ is continuously differentiable. Committing a slight abuse of notation, we denote $(f|_V)^{-1}$ simply by f^{-1} for the rest of the proof. The harder part of this is simply showing that f^{-1} is differentiable. Choose $y, y+k \in f(V)$. Thus there exist $x, x+h \in V$ for which $y = f(x)$ and $y+k = f(x+h)$. Let $T = Df(x)$; we know that T is invertible because we showed at the beginning of the proof that $Df(x)$ is invertible for all $x \in V$. We would like to show that the derivative of f^{-1} exists at the point y and that it is equal to T^{-1}. To this end we compute that

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

Thus

$$\begin{aligned} &\|f^{-1}(y+k) - f^{-1}(y) - T^{-1}(k)\| \\ &\qquad\qquad \leq \|T^{-1}\|_{\text{op}} \|f(x+h) - f(x) - T(h)\|. \end{aligned}$$

This inequality relates the numerators in the definitions of $Df(x)$ and $D(f^{-1})(y)$. We only need to bound the denominators. For this we have

$$\begin{aligned} g(x+h) - g(x) &= x + h - f(x+h) - x - f(x) \\ &= h - k, \end{aligned}$$

and since g_y is Lipschitz continuous on V with constant $1/2$, the above calculation implies that

$$\|h - k\| \leq \frac{1}{2}\|(x+h) - x\| = \frac{1}{2}\|h\|.$$

From the triangle inequality and the inequality above, it follows that

$$\|h\| = \|h - k + k\| \leq \|h - k\| + \|k\| \leq \frac{1}{2}\|h\| + \|k\|,$$

and thus

$$\|k\| \geq \frac{1}{2}\|h\|.$$

Putting everything together, we see that

$$\frac{\|f^{-1}(y+k) - f^{-1}(y) - T^{-1}(k)\|}{\|k\|} \leq 2\|T^{-1}\|_{\text{op}} \frac{\|f(x+h) - f(x) - T(h)\|}{\|h\|},$$

and from this it follows that f^{-1} is differentiable at $y = f(x)$.

We have now established that f^{-1} is differentiable on $f(V)$, and thus as we remarked before the proof, we have

$$D(f^{-1})(f(x)) = (Df(x))^{-1},$$

or putting this in terms of y and $f^{-1}(y)$,

$$D(f^{-1})(y) = (Df(f^{-1}(y)))^{-1}.$$

The righthand side here is the composition of three functions: f^{-1} followed by Df followed by inversion. We have just shown that f^{-1} is differentiable, so it must be continuous. We have assumed that Df is continuous (because f is continuously differentiable). Finally, we have proved earlier that inversion is continuous (Proposition **??**). Therefore, since the composition of continuous functions is continuous, $D(f^{-1})$ is continuous. This completes the proof of part (c) of the theorem, and as we have already shown how part (d) follows from first three parts, the proof of the theorem is complete. □

Exercises

Exercise 7.6.1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f(x, y) = (xy, x - y).$$

Verify that the inverse function theorem applies at the point $(1, 1)$ and find an open set V containing $(1, 1)$ on which f is injective. Then find $(f|_V)^{-1}$.

For Exercise 7.6.1, note that if $xy = ab$ and $x - y = a - b$, then

$$x^2 + y^2 = a^2 + b^2,$$

and hence both (x, y) and (a, b) are points of intersection of the same line and circle.

Exercise 7.6.2. In real coordinates, the complex function $z \mapsto z^2$ takes the form $f : \mathbb{R}^2 \to \mathbb{R}^2$ where $f(x, y) = (x^2 - y^2, 2xy)$. Prove that if $(a, b) \neq (0, 0)$, then there is an open set V containing (a, b) such that $f(V)$ is open and $f|_V : V \to f(V)$ is a bijection.

Exercise 7.6.3. Let f be the function described in Exercise 7.6.2. For both of the points $(1, 0)$ and $(-1, 0)$, find a set V that satisfies those conditions and compute $(f|_V)^{-1}$.

Exercise 7.6.4. Define the mapping $f : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$f(\rho, \theta, \phi) = \begin{pmatrix} \rho\cos(\theta)\sin(\phi) \\ \rho\sin(\theta)\sin(\phi) \\ \rho\cos(\phi) \end{pmatrix}.$$

What does the inverse function theorem say about f?

Fixing $\rho = 1$ in the function of Exercise 7.6.4 gives a mapping $g : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$g(\theta, \phi) = \begin{pmatrix} \cos(\theta)\sin(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\phi) \end{pmatrix}.$$

What is the image of g?

7.7 The implicit function theorem

THE IMPLICIT FUNCTION THEOREM tells us when a *relation* can be used to define a *function*. The quintessential example is the defining relation of the unit circle,

$$x^2 + y^2 = 1.$$

Where does this relation define y as a function of x? Certainly not near the point $(1, 0)$, because in any open set containing this point, the points on the unit circle fail the vertical line test. The same is true near the point $(-1, 0)$. On the other hand, near the point $(1/2, \sqrt{3}/2)$, the relation can be used to define y as a function of x, namely

$$y(x) = \sqrt{1 - x^2}.$$

In this case we are able to give an explicit closed form expression for $y(x)$, but that is not generally possible, which is why this is an *implicit definition* of y.

By defining

$$f(x, y) = x^2 + y^2 - 1,$$

we can view the unit circle as the set

$$\{(x, y) : f(x, y) = 0\}.$$

We observed above that there is some $\epsilon > 0$ so that we can define a function $y : (1/2 - \epsilon, 1/2 + \epsilon) \to \mathbb{R}$ that satisfies $y(1/2) = \sqrt{3}/2$ and

$$f(x, y(x)) = 0$$

for all $x \in (1/2 - \epsilon, 1/2 + \epsilon)$.

What is the difference between the point $(1/2, \sqrt{3}/2)$ where we can define such a function $y(x)$ and the point $(1, 0)$ where we cannot define one? As we will see, the difference between these two points is that

$$D_y f(1/2, \sqrt{3}/2) = \sqrt{3} \neq 0 \quad \text{while} \quad D_y f(0, 0) = 0.$$

In the general situation, we have a relation $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$, and we view f as being a function of two vectors, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$, so

$$f(x, y) = f(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^m.$$

Letting $f(x, y) = (f_1(x, y), \ldots, f_m(x, y))$, we see that the matrix representation of Df at a given point in $\mathbb{R}^{n+m}$ is

$$\left(\begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} & \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{array}\right).$$

We denote by $D_y f$ the linear transformation in $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ whose matrix representation is given by the rightmost m columns of this matrix.

Theorem 7.7.1 (The implicit function theorem)**.** *Suppose that the function $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is continuously differentiable and fix a point $(a, b) = (a_1, \ldots, a_n, b_1, \ldots, b_m)$ with $f(a, b) = 0$.*

If $D_y f(a, b)$ is invertible, then we can find a neighborhood $N_r(a) \subseteq \mathbb{R}^n$ so that there there is a unique continuous function $g : N_r(a) \to \mathbb{R}^m$ satisfying $g(a) = b$ and $f(x, g(x)) = 0$ for all $x \in N_r(a)$. Moreover, $N_r(a)$ can be chosen in such a way that this unique continuous function g is in fact continuously differentiable.

There is nothing special about $f(a, b)$ equaling $0 \in \mathbb{R}^m$. The implicit function theorem holds for any *level set*, where $f(a, b) = c$ for a constant $c \in \mathbb{R}^m$. (Simply replace $f(x, y)$ by $f(x, y) - c$ in the statement of the theorem.)

Proof. We prove the theorem by applying the inverse function theorem to the function $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined by

$$F(x, y) = (x, f(x, y)),$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Note that

$$F(a, b) = (a, f(a, b)) = (a, 0).$$

Our hypotheses imply that f is differentiable. It follows that F is differentiable as well and that the matrix representation of the derivative of F is given by the block matrix

$$\begin{pmatrix} I_n & 0 \\ D_x f & D_y f \end{pmatrix}.$$

Indeed, our hypotheses tell us that f is continuously differentiable, and it follows that F is also continuously differentiable (Exercise 7.7.4).

From the matrix representation of $DF(a, b)$ and the fact that $D_y f(a, b)$ is invertible, we see that $DF(a, b)$ is invertible (Exercise **??**). Therefore F satisfies all of the hypotheses of the inverse function theorem, so there is an open set $V \subseteq \mathbb{R}^{n+m}$ containing the point (a, b) such that

(a) $F|_V : V \to F(V)$ is a bijection,
(b) $F(V)$ is open,
(c) $(F|_V)^{-1}$ is continuously differentiable, and
(d) $D(F|_V)^{-1}(F(a, b)) = D(F|_V)^{-1}(a, 0) = (DF(a, b))^{-1}$.

Since $F(V)$ is open and contains the point $F(a, b) = (a, 0)$, there is some radius $r > 0$ so that

$$N_r((a, 0)) \subseteq F(V).$$

This implies that for every $x \in N_r(a) \subseteq \mathbb{R}^n$, we have $(x, 0) \in F(V)$. This means that for each such x, there is some vector $y_x \in \mathbb{R}^m$ with $(x, y_x) \in V$ and $f(x, y_x) = 0$. Indeed, since $F|_V$ is a bijection between V and $F(V)$, if $x \in N_r(a)$, then this vector y_x is unique (otherwise, $(x, 0)$ would have multiple preimages).

Therefore, we can define the function

$$g : N_r(a) \to \mathbb{R}^m$$

by $g(x) = y_x$, where y_x is the unique vector in $\mathbb{R}^m$ satisfying $(x, y_x) \in V$ and $F(x, y_x) = (x, 0)$. We can express the function g symbolically as

$$g(x) = \text{proj}_y \left((F|_V)^{-1}(x, 0) \right),$$

where proj_y denotes the projection of a vector in $\mathbb{R}^{n+m}$ onto its last m coordinates. This expression shows that g is the composition of three continuously differentiable functions: the inclusion mapping $x \mapsto (x, 0)$, the function $(F|_V)^{-1}$, and finally the projection mapping $(x, y) \mapsto y$. It follows that g is continuously differentiable.

It remains only to establish the uniqueness of g. Let $h : N_r(a) \to \mathbb{R}^m$ be any continuous function that satisfies $h(a) = b$ and $f(x, h(x)) = 0$ for all $x \in N_r(a)$. Our proof shows that if $(x, h(x)) \in V$, then since $F|_V$ is a bijection between V and $F(V)$, $h(x)$ must equal $g(x)$. However, we need to worry that we might have $(x, h(x)) \notin V$.

For an example of the uniqueness issues we are considering here, think back to the unit circle. Say that we choose $(a, b) = (1/2, \sqrt{3}/2)$ and $N_r(a) = (1/3, 2/3)$. (So $r = 1/6$.) Then there are a *lot* of choices of *some* function $g : N_r(a) \to \mathbb{R}^m$. We must have $g(1/2) = \sqrt{3}/2$, but then for every point $x \in N_r(a) \setminus \{a\}$, we may choose either $g(x) = \sqrt{1 - x^2}$ or $g(x) = -\sqrt{1 - x^2}$. However, there is only one way to choose a *continuous* function g that satisfies the desired conditions.

Define the set

$$A = \{x \in N_r(a) : g(x) = h(x)\} = (g - h)^{-1}(\{0\}).$$

The set A is closed in the space $(N_r(a), d_2)$, because it is the preimage of a closed set under a continuous function (Exercise **??**). The set A is also open in the space $(N_r(a), d_2)$, because if $h(x_0) = g(x_0)$, then that means that $(x_0, h(x_0)) = (x_0, g(x_0)) \in V$, and since V is open and h is continuous, we must then have $(x, h(x)) \in V$ for all x in some neighborhood of x_0, which implies that $h(x) = g(x)$ for these values of x.

This shows that A is clopen in the space $(N_r(a), d_2)$, but we know that $(N_r(a), d_2)$ is connected (Exercise **??**), so its only clopen subsets are itself and the empty set (Proposition **??**). Clearly $A \neq \emptyset$, because $a \in A$, so it must be that $A = N_r(a)$, which implies that g and h are identical and completes the proof. □

Exercises

Exercise 7.7.1. Using the conclusion of Exercise 7.7.5 to show that $yz = \log(x + z) - \log(3)$ defines z as a function of (x, y) near the point $(2, 0, 1)$, and find $\partial z / \partial x$ at this point.

Exercise 7.7.2. Consider the *folium of Descartes*, described implicitly as

$$f(x, y) = x^3 + y^3 - 3axy = 0,$$

for a fixed parameter $a > 0$. Show that the implicit function theorem says that f is locally the graph of a function for any point $(x_0, y_0) \neq (0, 0)$ with $f(x_0, y_0) = 0$.

Exercise 7.7.2 shows that it is possible (in principle) to solve for y as a function of x *or* x as a function of y near any point on the folium of Descartes, except possibly at the origin. (And a plot shows that this relation is certainly not the graph of a function near the origin.)

Exercise 7.7.3. The *parabolic folium* is described implicitly as

$$x^3 = a(x^2 - y^2) + bxy$$

for fixed parameters $a, b > 0$. What does the implicit function theorem say about this curve?

Exercise 7.7.4. Prove the fact used in the proof of the implicit function theorem that if $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is continuously differentiable, then the function $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined by $F(x, y) = (x, f(x, y))$ is also continuously differentiable (where here, as in that proof, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$).

Exercise 7.7.5. Suppose that $f = f(x, y, z) : \mathbb{R}^3 \to \mathbb{R}$ is continuously differentiable. Show that if $f(a, b, c) = 0$ and $\frac{\partial f}{\partial z}(a, b, c) \neq 0$, then the relation $f(x, y, z) = 0$ defines $z = g(x, y)$ near the point (a, b, c). Show further that

$$\frac{\partial g}{\partial x}(a, b) = -\frac{\frac{\partial f}{\partial x}(a, b, c)}{\frac{\partial f}{\partial z}(a, b, c)}.$$

Exercise 7.7.5 is implicit differentiation. Note that the necessary assumption that $\frac{\partial f}{\partial z}(a, b, c) \neq 0$ is sufficient to establish that, at least locally, z is indeed a function of x and y. (An issue that is not directly addressed in most calculus texts.)

Exercise 7.7.6. The point $(x, y, u, v, w) = (1, 1, 1, 1, -1) \in \mathbb{R}^5$ satisfies the system of equations

$$\begin{aligned} u^5 - xv^2 + y + w &= 0 \\ v^5 - yu^2 + x + w &= 0 \\ w^4 + y^5 - x^4 - 1 &= 0. \end{aligned}$$

Explain why there is an open set $U \subseteq \mathbb{R}^2$ containing the point $(x, y) = (1, 1)$ and continuously differentiable functions

$$u(x, y), v(x, y), w(x, y) : U \to \mathbb{R}$$

such that $u(1, 1) = 1 = v(1, 1)$, $w(1, 1) = -1$, and so that the point

$$(x, y, u(x, y), v(x, y), w(x, y))$$

satisfies the system of equations for all $(x, y) \in U$.

Exercise 7.7.7. What can you say about solving the system

$$\begin{aligned} x^2 - y^2 + 2u^3 + v^2 &= 3 \\ 2xy + y^2 - u^2 + 3v^4 &= 5 \end{aligned}$$

for (u, v) in terms of (x, y) near the point $(x, y, u, v) = (1, 1, 1, 1)$?

Index