Differential Topology

Notes for Mathematics 738, Spring 2016.

Ralph Howard
Department of Mathematics
University of South Carolina
Columbia, S.C. 29208, USA
howard@math.sc.edu

Contents

1. A Review of Some Linear Algebra. 1
2. The Inverse and Implict Function Theorems. 4
2.1. The Banach Fixed Point Theorem. 4
2.2. Banach Spaces. 6
2.3. Bounded linear maps between Banach spaces. 8
2.4. The derivative of maps between Banach spaces 11
2.5. Preliminary version of the inverse function theorem. 18
2.6. Inverse Function Theorem 21
2.7. Implicit functions in two and three dimensions 23
2.8. The General Finite Dimensional Implicit Function Theorem. 35
2.9. Higher Derivatives for Inverse and Implicit Functions. 37

1. A Review of Some Linear Algebra.

Here we recall facts from linear algebra that will motivate some of what we do with nonlinear maps. I assume you know the definition of a vector space, a subspace of a vector space, the dimension of a vector space, and the span of a subset of a vector space. If U and W are subspaces of a vector space then the set

$$U + W = \{u + w : u \in U, w \in W\}$$

is a subspace and is the smallest subspace containing both U and W. The intersection is also a subspace.

Proposition 1.1. *Let V be a finite dimensional real vector space and U and W subspaces of V. Then*

$$\dim(U + W) + \dim(U \cap W) = \dim(U) + \dim(W).$$

Problem 1.1. Prove this. *Hint:* Let $p = \dim U$, $q = \dim W$, and $k = \dim U \cap W$. Let $v_1, \ldots, v_k$ be a basis for $U \cap W$. Expend this set to a basis $v_1, \ldots, v_k, u_{k+1}, \ldots, u_p$ of U and a basis $v_1, \ldots, v_k, w_{k+1}, \ldots, w_q$. Then show $v_1, \ldots, v_k, u_{k+1}, \ldots, u_p, w_{k+1}, \ldots, w_q$ is a basis for $U + W$. □

If V is a vector space over $\mathbb{R}$, and $a, b \in V$ are distinct, then the ***line*** through a and b is the set

$$\{(1-t)a + tb : t \in \mathbb{R}\}.$$

An ***affine subspace*** of V is a subset A, that contains the line through any two of its points. More explicitly A is an affine subspace of V if and only if for all $a_0, a_1 \in A$ and $t \in \mathbb{R}$ we have $(1-t)a_0 + ta_1 \in A$. As we are not assuming here that a_0 and a_1 are distinct this implies that a one element subset of V is is an affine subspace.

If $a_0, a_1, \ldots, a_k$ are in the real vector space V than an ***affine combination*** of these vectors is a sum of the form

$$\sum_{j=0}^{k} t_j a_j \qquad \text{where} \qquad \sum_{j=0}^{k} t_j = 1.$$

Note that an affine combination of two points a_0 and a_1 is of the form $t_0 a_0 + t_1 a_1$ where $t_0 + t_1 = 1$. Letting $t = t_1$ we have $t_0 a_0 + t_1 a_1 = (1-t)a_0 + ta_1$. Thus the set of all affine combinations of two distinct points is just the line through the points.

Proposition 1.2. *A subset of the real vector space V is an affine subspace of V if and only if it is closed under taking affine combinations of its elements.*

Problem 1.2. Prove this. *Hint:* One direction is more or less clear. If a subset is closed under affine combinations, then it is an affine subspace. Conversely if A is an affine subspace of V, then it is closed under taking the affine combination of any two of its elements. Use this as the base case for an induction showing that A is closed under general affine combinations. □

Proposition 1.3. *Let V be a real vector space.*

(a) A linear subspace of V is also an affine subspace of V.

(b) An affine subspace of V that contains the origin is a linear subspace of V.

Problem 1.3. Prove this. *Hint:* For part (b) note that if A is an affine subspace of V with $0 \in A$, then for any $a_1, a_2 \in A$ and $\alpha, \beta \in \mathbb{R}$ we have $\alpha a_1 + \beta a_0 = (1 - \alpha - \beta)0 + \alpha a_1 + \beta a_0$ and thus $\alpha a_1 + \beta a_0$ is an affine combination of 0, a_1, and a_2. □

Proposition 1.4. *Let U be a linear subspace of V and $a \in V$ then the translate*

$$a + U := \{a + u : u \in U\}$$

is an affine subspace of V.

Problem 1.4. Prove this. □

Proposition 1.5. *Let A be an affine subspace of the real vector space V. Then there is a unique linear subspace U of V such that for any vector $a_0 \in A$*

$$A = a_0 + U.$$

(That is the affine subspaces of V are just the translations of the linear subspaces.)

Problem 1.5. Prove this. *Hint:* To start let $a_0 \in A$, set $U = \{a - a_0 : a \in A\}$ and show U is a linear subspace of V. One way to do this is to show U is an affine subspace that contains the origin. □

If A is an affine subspace of the real vector space V and $A = a_0 + U$ where U is a linear subspace of V then define the ***dimension*** of A to be

$$\dim(A) = \dim(U).$$

The following is a very special case of one of the main theorems we will prove this term.

Proposition 1.6. *Let A and B be affine subspaces of the real vector space V and assume $\dim(A) + \dim(B) < \dim(V)$. Then for every $\varepsilon > 0$ there is a vector $v \in V$ with $\|v\| < \varepsilon$ and $A \cap (v + B) = \varnothing$.*

Problem 1.6. Prove this. *Hint:* If $A \cap B = \varnothing$ when we can use $v = 0$. If $A \cap B \neq \varnothing$, let $a \in A \cap B$. Then there are unique linear subspaces U and W such that $A = a + U$ and $B = a + W$. As $\dim(U) + \dim(W) < \dim(V)$ the subspace $U + W$ is not all of V. Let $x \in V$ be a vector that is not in $U + W$ and let $0 \neq t \in \mathbb{R}$. Show that $A \cap (tx + B) = \varnothing$ for all $t \neq 0$. □

This shows that in some cases it is easy to move one affine subspace away from another. In some cases this is not true.

Proposition 1.7. *Let U and W be linear subspaces of the real vector space V with $U + W = V$. Then for all $a, b \in V$ we have $(a + U) \cap (b + W) \neq \varnothing$.*

Problem 1.7. Prove this. □

The following is a basic factor about linear maps.

Proposition 1.8 (Rank plus Nullity Theorem.)**.** *Let $L\colon V \to W$ be a linear map between finite dimensional vector spaces. Then*

$$\dim(\ker(L)) + \dim(\mathrm{Image}(L)) = \dim(V).$$

Problem 1.8. Prove this. *Hint:* Let $k = \dim(\ker(L))$ and choose a basis $v_1, \ldots, v_k$ of $\ker(K)$. Extend this to a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ of V (where $n = \dim(V)$). Then show $Lv_{k+1}, \ldots, Lv_n$ is a basis of V. □

One of the main ideas in differential topology is to relate the geometry of a map $f\colon M \to N$ between nice spaces so the geometry of preimages $f^{-1}[y] := \{x \in M : f(x) = y\}$ for $y \in N$. One of the main geometric invariants of a space is its dimension. The following is anther special case of a much more general result we will prove later.

Proposition 1.9. *Let $L\colon V \to W$ be a surjective linear map between finite dimensional real vector spaces. Then*

(a) $\dim(W) \le \dim(V)$
(b) *all* $y \in W$ *the preimage* $f^{-1}[y]$ *is an affine subspace of* V *of dimension* $\dim(V) - \dim(W)$.

Problem 1.9. Prove this. □

Recall that if $f\colon X \to Y$ is a map between sets, then $g\colon Y \to X$ is a ***left inverse*** to f if and only if the composition $g \circ f\colon X \to X$ is the identity map on X. This implies f is injective and g is surjective. Likewise g is a ***right inverse*** if and only if the composition $f \circ g\colon Y \to Y$ is the identity map on Y. In this case f is surjective and g is injective.

Proposition 1.10. *Let* $L\colon V \to W$ *be a linear map between finite dimensional real vector spaces.*
(a) If L *is surjective, then there is a linear map* $S\colon W \to V$ *that is a right inverse to* L*. (Note for* $y \in W$ *that* $x = Sy$ *gives a solution to* $Lx = y$*.)*
(b) If L *is injective, there is a linear map* $T\colon W \to V$ *that is a left inverse to* L*.*

Problem 1.10. Prove this. □

2. The Inverse and Implict Function Theorems.

2.1. **The Banach Fixed Point Theorem.** We now give what is one of the more important existence theorems for general nonlinear equations. Let (X, d) be a metric space (say that (X, d) is the plane $\mathbb{R}^2$ with its usual metric), and let $F\colon X \to X$ be a continuous map. Then given $y \in X$ we would like a method for solving the equation $F(x) = y$ for x. Experience has shown that it is often better to rewrite this equation as $f(x) = x$ for a new function f. (In the case $X = \mathbb{R}$ then for any constant $c \neq 0$ let $f(x) = c(F(x)-y)+x$, then the equation $f(x) = x$ has the same solutions as $F(x) = y$.) It turns out that many useful theorems about solving equations are stated as fixed point theorems. I don't have any real insight into why fixed point theorems turn out to easier to handle than other types of theorems on solutions, but this seems to be the case. (In topology there is the Brouwer fixed point theorem which is a very general result about solving n equations in n unknowns.)

We first recall a bit of elementary analysis. Let a, r be real numbers, with $r \neq 1$. To find the sum

$$S = a + ar + \cdots + ar^n = \sum_{k=0}^{n} ar^k.$$

To do this multiple S by $(1-r)$ to get

$$\begin{aligned}(1-r)S &= S - rS \\ &= (a + ar + \cdots + ar^n) - (ar + ar^2 + \cdots + ar^{n+1}) \\ &= a - ar^{n+1}.\end{aligned}$$

Thus the sum of this finite geometric series is

$$S = \frac{a - ar^{n+1}}{1-r}.$$

Thus when $|r| < 1$, so that $\lim_{n\to\infty} r^{n+1} = 0$, we get the convergent infinite sum

$$\sum_{k=0} ar^k = a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1-r}.$$

We will use these facts several without comment in what follows.

Let (X, d) be a metric space. Then a map $f\colon X \to X$ is a ***contraction*** if and only there is a constant $\rho < 1$ so that $d(f(x), f(y)) \leq \rho d(x, y)$. The number ρ is called the ***c***ontraction factor. It is easy to see that any contraction is continuous.

Let Y be any set and $g\colon Y \to Y$. Then the point $y_0 \in Y$ is a fixed point of g if and only if $g(y_0) = y_0$.

Problem 2.1. Let (X, d) be a metric space and $f\colon X \to X$ a contraction with contraction factor ρ. Then show that f has at most one fixed point. *Hint:* Assume that $a, b \in X$ are fixed points of f. Then $d(a, b) = d(f(a), f(b)) \leq \rho d(a, b)$. □

Theorem 2.1 (Banach Fixed Point Theorem)**.** *Let (X, d) be a complete metric space and $f\colon X \to X$ be a contraction with contraction factor $\rho < 1$. Then show that f has a unique fixed point x_* in X. This fixed point can be found by starting with any $x_0 \in X$ and defining a sequence $\{x_k\}_{k=0}^{\infty}$ by recursion*

$$x_1 = f(x_0), \quad x_2 = f(x_1), \quad x_3 = f(x_2), \quad \ldots \quad x_{k+1} = f(x_k), \quad \ldots$$

Then $x_* = \lim_{k\to\infty} x_k$. There is also an estimate on the error of using x_n as an approximation to x_*. This is

$$d(x_n, x_*) \leq \frac{d(x_0, x_1)\rho^n}{1-\rho}. \tag{1}$$

Remark 2.2. This result was in Banach's thesis which appeared in published form in 1922. In his thesis and this paper he also introduced "complete normed linear space" which have since been renamed as Banach spaces. The idea of looking at the sequence x_0, $x_1 = f(x_0)$, $x_2 = f(x_1) \ldots$ is an abstraction of an idea of Picard. □

Problem 2.2. Prove the theorem by doing the following:

(a) For the sequence defined above show that $d(x_k, x_{k+1}) \leq d(x_0, x_1)\rho^n$.
(b) Let $m < n$ then by repeated use of the triangle inequality show

$$d(x_m, x_n) \leq \sum_{k=m}^{n-1} d(x_k, x_{k+1}).$$

(c) Use Part (a) and sum a geometric series to show

$$d(x_m, x_n) \le \frac{d(x_0, x_1)\rho^m - d(x_0, x_1)\rho^n}{1-\rho} \le \frac{d(x_0, x_1)\rho^m}{1-\rho}.$$

(d) Show the sequence $x_0, x_1, x_2, \ldots$ is Cauchy sequence and therefore converges to some point x_* of X. Use Part (c) to show (1) holds.
(e) Show x_* is a fixed point of f. *Hint:* Take the limit as $k \to \infty$ of the equation $x_{k+1} = f(x_k)$.
(f) Show x_* is the only fixed point of f. □

2.2. **Banach Spaces.** Let $\mathbf{X}$ be a (not necessarily finite dimensional) vector space over the real numbers. Then a ***norm***, $|\cdot|_{\mathbf{X}}$, on $\mathbf{X}$ is a function from $\mathbf{X}$ to the real numbers such that

(a) (nonnegative) for all $x \in \mathbf{X}$ the inequality $|x|_{\mathbf{X}} \ge 0$ holds and with $|x|_{\mathbf{X}} = 0$ if and only if $x = 0$.
(b) (triangle inequality) for all $x, y \in \mathbf{X}$

$$|x+y|_{\mathbf{X}} \le |x|_{\mathbf{X}} + |y|_{\mathbf{X}}$$

(c) (homogeneity) If $x \in \mathbf{X}$ and $\lambda \in \mathbb{R}$ then

$$|\lambda x|_{\mathbf{X}} = |\lambda||x|_{\mathbf{X}}.$$

As examples let $\mathbb{R}^n$ is the usual vector space of length n column vectors of real numbers. That is, because it is more compatible with matrix operations, we view elements $x \in \mathbb{R}^n$ as column vectors

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$$

(The reason for using superscripts rather subscripts will become clear later.) To save space will will sometimes write elements of $\mathbb{R}^n$ as $x = (x^1, x^2, \ldots, x^n)$. For each $p \ge 1$ the function

$$|(x^1, x^2, \ldots, x^n)|_{\ell^p} = \left(|x^1|^p + |x^2|^p + \cdots + |x^n|^p\right)^{\frac{1}{p}}$$

is a norm on $\mathbb{R}^n$. In this case the triangle inequality is a special case of the ***Minkowski inequality*** and is not a trivial inequality. If $\mathbf{X} = C[0,1]$ is the vector space of all continuous functions $f\colon [0,1] \to \mathbb{R}$ and we define

$$|f|_{L^\infty} = \max_{x \in [0,1]} |f(x)|$$

then $|\cdot|_{L^\infty}$ makes $C[0,1]$ into a normed space.

Problem 2.3. Prove that $|\cdot|_{L^\infty}$ does make $C[0,1]$ into a normed vector space. Show this is not finite dimensional by showing that the functions p_n defined by $p_n(t) = t^n$ for $n = 0, 1, 2, \ldots$ are linearly independent. □

Anther example is to let $\mathbf{X} = L^1[0,1]$, that is the Lebesgue integrable functions $f\colon [0,1] \to \mathbb{R}$ and use for the norm

$$|f|_{L^1} = \int_{[0,1]} |f|\, dm$$

where m is Lebesgue measure. (Here I am not being altogether honest, $L^1[0,1]$ is really the vector space of equivalence class of integrable functions were we consider to functions equal if they are equal except on a set of measure zero, that is if they are equal almost everywhere.)

Proposition 2.3. *If $\mathbf{X}$ is a normed linear space with norm, $|\cdot|_{\mathbf{X}}$, then $\mathbf{X}$ is a metric space with distance function*

$$d(x,y) := |x-y|_{\mathbf{X}}.$$

Problem 2.4. Prove this. □

We can now define a ***Banach space***. It is a normed vector space $(\mathbf{X}, |\cdot|_{\mathbf{X}})$ that is complete as a metric space. That is all Cauchy sequences in $\mathbf{X}$ converge. (Recall that a sequence $\langle x_n\rangle_{n=1}^{\infty})$ is ***Cauchy*** if and only if for all $\varepsilon > 0$ there is a positive integer N such that $m,n > N$ implies $|x_m - x_n|_{\mathbf{X}} < \varepsilon$.)

In finite dimensions all normed spaces are complete and thus all finite dimensional normed spaces are Banach spaces. For an example of a normed space that is not a Banach space consider $C[0,1]$ with the norm

$$|f|_{L^1} = \int_{[0,1]} |f|\, dm.$$

Since the continuous continuous functions are dense in $L^1[0,1]$, for any discontinuous function f in $L^1[0,1]$ we choose a sequences $\langle f_n\rangle_{n=1}^{\infty}$ of continuous functions with $\lim_{n\to\infty} |f - f_n|\, dm = 0$. Then $\langle f_n\rangle_{n=1}^{\infty}$ will be Cauchy in $C[0,1]$ but not convergent in $C[0,1]$ as the limit function is not in $C[0,1]$.

Proposition 2.4. *Let $\mathbf{X}$ be a finite dimensional normed space with $e_1, \ldots, e_n$ a basis of $\mathbf{X}$. For any $x \in \mathbf{X}$ write $x = \sum_{j=1}^n x^j e_j$ with $x^j \in \mathbb{R}$ and use this basis to define a $|\cdot|_{\ell^1}$ by*

$$|x|_{\ell^1} = \sum_{j=1}^n |x^j|.$$

Then there are constants $C_1, C_2 > 0$ such that

$$C_1|x|_{\ell^1} \le |x|_{\mathbf{X}} \le C_2|x|_{\ell^1}.$$

Thus on a finite dimensional space all norms define the same topology.

Problem 2.5. Prove this. *Hint:* Let $M = \max_{1\le j\le n} |e_j|_{\mathbf{X}}$. Then show

$$|x|_{\mathbf{X}} \le M \sum_{j=1}^n |x_j| = M|x|_{\ell^1},$$

so $C_2 = M$ works. The inequality just given yields for $x, y \in \mathbf{X}$ that

$$|x-y|_{\mathbf{X}} \leq M|x-y|_{\ell^1}.$$

Therefore $|\cdot|_{\mathbf{X}}$ is continuous with respect to the topology defined by $|\cdot|_{\ell^1}$. The topology defined by $|\cdot|_{\ell^1}$ is the usual topology on a finite dimensional vector space and in this topology all the closed bounded sets are compact. (To be a little more explicit the basis $e_1, \ldots, e_n$ gives a linear isomorphism of $\mathbf{X}$ with $\mathbb{R}^n$ and we use this isomorphism to move the topology of $\mathbb{R}^n$ to $\mathbf{X}$.) Use this fact to show $|\cdot|_{\mathbf{X}}$ achieves it minimum, m, on the set $S := \{x : |x|_{\ell^1} = 1\}$. Then $m > 0$ and

$$m \leq |x|_{\mathbf{X}} \qquad \text{whenever} \qquad |x|_{\ell^1} = 1.$$

Now use the homogeneity of the two norms to show

$$m|x|_{\ell^1} \leq |x|_{\mathbf{X}}$$

for all $x \in \mathbf{X}$. Set $C_1 = m$ to complete the proof. □

2.3. **Bounded linear maps between Banach spaces.** Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces with norms $|\cdot|_{\mathbf{X}}$ and $|\cdot|_{\mathbf{Y}}$. Then a linear map $A\colon \mathbf{X} \to \mathbf{Y}$ is ***bounded*** if and only if there is a constant C so that

$$|Ax|_{\mathbf{Y}} \leq C|x|_{\mathbf{X}} \quad \text{for all} \quad x \in \mathbf{X}.$$

The best constant C in this inequality is the ***operator norm*** (which we will usually just call the ***norm***) of A and denoted by $\|A\|$. Thus $\|A\|$ is given by

$$\|A\| := \sup_{0 \neq x \in \mathbf{X}} \frac{|Ax|_{\mathbf{Y}}}{|x|_{\mathbf{X}}}.$$

Problem 2.6. Show that a linear map $L\colon \mathbf{X} \to \mathbf{Y}$ is continuous if and only if it is bounded. *Hint:* If L is bounded, then for all $x_0, x_1 \in \mathbf{X}$, $|Lx_1 - Lx_0|_{\mathbf{Y}} = |L(x_1 - x_0)|_{\mathbf{Y}} \leq \|L\||x_1 - x_0|_X$ and this can be used to show L is continuous. Conversely if L is continuous at 0 then for $\varepsilon = 1$ there is a $\delta > 0$ such that

$$|x - 0|_{\mathbf{X}} < \delta \qquad \text{implies} \qquad |Lx - L0|_{\mathbf{Y}} < \varepsilon = 1.$$

Let $x \neq 0$ then $\frac{\delta}{2|x|_{\mathbf{X}}} x$ satisfies

$$\left| \frac{\delta}{2|x|_{\mathbf{X}}} x \right|_{\mathbf{X}} = \frac{\delta}{2} < \delta$$

and therefore

$$\left| L\left(\frac{\delta}{2|x|_{\mathbf{X}}} x \right) \right|_{\mathbf{Y}} < 1$$

which can be used to show that L is bounded. □

Denote by $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ the set of all bounded linear maps $A\colon \mathbf{X} \to \mathbf{Y}$.

Proposition 2.5. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces with $\mathbf{X}$ finite dimensional. Then every linear map $A\colon \mathbf{X} \to \mathbf{Y}$ is bounded.*

Problem 2.7. Prove this. *Hint:* Let $n = \dim(\mathbf{X})$ and $e_1, \ldots, e_n$ a basis of $\mathbf{X}$. If $x \in \mathbf{X}$ write $x = \sum_{j=1}^n x^j e_j$ with $x^j \in \mathbb{R}$. Let $M = \max_{1 \le j \le n} |Ae_j|_{\mathbf{Y}}$. Then, using the notation of Proposition 2.4, show

$$|Ax|_{\mathbf{X}} \le M \sum_{j=1}^n |x^j| \le \frac{M}{C_1} |x|_{\mathbf{X}},$$

which shows that A is bounded. □

Problem 2.8. (a) Show that the norm $\| \cdot \|$ makes $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ into a normed linear space. That is show if $A, B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and c_1 and c_2 are real numbers then $c_1 A + c_2 B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and

$$\|c_1 A + c_2 B\| \le |c_1| \|A\| + |c_2| \|B\|.$$

(b) Show that the norm $\| \cdot \|$ is complete on $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ and so $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ is a Banach space. *Hint:* This can be done as follows. Let $\{A_k\}_{k=1}^\infty$ be a Cauchy sequence in $\mathbf{X}$. Then $M := \sup_k \|A_k\| < \infty$. Show
  (i) For any $x \in \mathbf{X}$ the sequence $\{A_k x\}_{k=1}^\infty$ is a Cauchy sequence and as $\mathbf{Y}$ is a Banach space this implies $\lim_{k \to \infty} A_k x$ exists.
  (ii) Define a map $A \colon \mathbf{X} \to \mathbf{Y}$ by $Ax := \lim_{k \to \infty} A_k x$. Then show A is linear and $|Ax|_{\mathbf{Y}} \le M |x|_{\mathbf{X}}$ for all $x \in \mathbf{X}$. Thus A is bounded.
  (iii) Let $\varepsilon > 0$ and let N_ε be so that $k, \ell \ge N_\varepsilon$ implies $\|A_k - A_\ell\| \le \varepsilon$ (N_ε exists as $\{A_k\}_{k=1}^\infty$ is Cauchy). Then for any $x \in \mathbf{X}$ and $k \ge N_\varepsilon$ and all $\ell \ge N_\varepsilon$ we have

$$\begin{aligned} |Ax - A_k x|_{\mathbf{Y}} &\le |Ax - A_\ell x|_{\mathbf{Y}} + |(A_\ell - A_k)x|_{\mathbf{Y}} \\ &\le |Ax - A_\ell x|_{\mathbf{Y}} + \varepsilon |x|_{\mathbf{X}} \\ &\xrightarrow{\ell \to \infty} 0 + \varepsilon |x|_{\mathbf{X}} = \varepsilon |x|_{\mathbf{X}}. \end{aligned}$$

  This implies $\|A - A_k\| \le \varepsilon$ for $k \ge N_\varepsilon$ and thus $\lim_{k \to \infty} A_k = A$. This shows any Cauchy sequence in $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ converges.

(c) If $\mathbf{Z}$ is a third Banach space $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and $B \in \mathcal{B}(\mathbf{Y}, \mathbf{Z}))$ then $BA \in \mathcal{B}(\mathbf{X}, \mathbf{Z})$ and $\|BA\| \le \|B\| \|A\|$. In particular if $A \in \mathcal{B}(\mathbf{X}, \mathbf{X})$ then by induction $\|A^k\| \le \|A\|^k$. □

Remark 2.6. Some inequalities involving norms of bounded linear maps will be used repeatedly in what follows without comment. The inequalities in question are

$$|Ax|_{\mathbf{Y}} \le \|A\| |x|_{\mathbf{X}}, \quad \|AB\| \le \|A\| \, \|B\|, \quad \|A^k\| \le \|A\|^k.$$

Of course the various forms of the triangle inequality will also be used. This includes the form $|u - v| \ge |u| - |v|$. □

The linear map $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ is ***invertible*** if and only if there is a $B \in \mathcal{B}(\mathbf{Y}, \mathbf{X})$ so that $AB = I_{\mathbf{Y}}$ and $BA = I_{\mathbf{X}}$ (where $I_{\mathbf{X}}$ is the identity map on $\mathbf{X}$). The map B is called the ***inverse*** of A and is denoted by $B = A^{-1}$.

Proposition 2.7. *Let* $\mathbf{X}$ *be a Banach space and* $A \in \mathcal{B}(\mathbf{X}, \mathbf{X})$ *with* $\|I_{\mathbf{X}} - A\| < 1$. *Then the* A *in invertible and the inverse is given by*

$$A^{-1} = \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^k = I_{\mathbf{X}} + (I_{\mathbf{X}} - A) + (I_{\mathbf{X}} - A)^2 + (I_{\mathbf{X}} - A)^3 + \cdots$$

and satisfies the bound

$$\|A^{-1}\| \leq \frac{1}{1 - \|I_{\mathbf{X}} - A\|}.$$

Moreover if $0 < \rho < 1$ *then*

$$\|A - I_{\mathbf{X}}\|, \|B - I_{\mathbf{X}}\| \leq \rho \quad \textit{implies} \quad \|A^{-1} - B^{-1}\| \leq \frac{1}{(1-\rho)^2} \|A - B\|$$

Proof. Let $B := \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^k$ then $\|(I_{\mathbf{X}} - A)^k\| \leq \|I_X - A\|^k$ and as $\|I_X - A\| < 1$ the geometric series $\sum_{k=0}^{\infty} \|I_X - A\|^k$ converges. Therefore by comparison the series defining B converges and

$$\|B\| \leq \sum_{k=0}^{\infty} \|I_{\mathbf{X}} - A\|^k = \frac{1}{1 - \|I_{\mathbf{X}} - A\|}.$$

Now compute

$$\begin{aligned} AB &= \sum_{k=0}^{\infty} A(I_{\mathbf{X}} - A)^k \\ &= \sum_{k=0}^{\infty} (I_{\mathbf{X}} - (I_{\mathbf{X}} - A))(I_{\mathbf{X}} - A)^k \\ &= \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^k - \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^{k+1} \\ &= (I_{\mathbf{X}} - A)^0 \\ &= I_{\mathbf{X}}. \end{aligned}$$

A similar calculation shows that $BA = I_{\mathbf{X}}$ (or just note A and B commute as all the terms in the sum defining B are polynomials in A). Thus $B = A^{-1}$. (The formula for $B = A^{-1}$ was motivated by the power series $(1-x)^{-1} = \sum_{k=0}^{\infty} x^k$.)

If $\|A - I_{\mathbf{X}}\|, \|B - I_{\mathbf{X}}\| \leq \rho$ then by what we have just done

$$\|A^{-1}\|, \|B^{-1}\| \leq \frac{1}{1-\rho}$$

Therefore

$$\begin{aligned} \|A^{-1} - B^{-1}\| &= \|A^{-1}(B - A)B^{-1}\| \leq \|A^{-1}\| \, \|B^{-1}\| \, \|B - A\| \\ &\leq \frac{1}{(1-\rho)^2} \|A - B\|. \end{aligned}$$

This completes the proof. ∎

The next proposition is a somewhat more general version of the last result. The main point is that the set of invertible operators is an open set and the map $A \mapsto A^{-1}$ is continuous on this set.

Proposition 2.8. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be Banach spaces and let* $A, B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$. *Assume that* A *is invertible. Then if* B *satisfies*

$$\|A - B\| < \frac{1}{\|A^{-1}\|}$$

then B *is also invertible and*

$$\|B^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|A - B\|}, \qquad \|B^{-1} - A^{-1}\| \le \frac{\|A^{-1}\|^2 \|B - A\|}{1 - \|A^{-1}\|\,\|A - B\|}.$$

Therefore the set of invertible maps from $\mathbf{X}$ *to* $\mathbf{Y}$ *is open in* $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ *and the map* $A \mapsto A^{-1}$ *is continuous on this set.*

Proof. This is more or less a corollary to the last result. Write $B = A - (A - B) = A(I_{\mathbf{X}} - A^{-1}(A - B))$. But $\|A^{-1}(A - B)\| \le \|A^{-1}\|\,\|A - B\| < 1$ by assumption. Thus the last proposition gives that $I_{\mathbf{X}} - A^{-1}(A - B)$ is invertible and that

$$\|(I_{\mathbf{X}} - A^{-1}(A - B))^{-1}\| \le \frac{1}{1 - \|A^{-1}(A - B)\|} \le \frac{1}{1 - \|A^{-1}\|\,\|(A - B)\|}.$$

Whence $B = A(I_{\mathbf{X}} - A^{-1}(A - B))$ is the product of invertible maps and whence invertible with $B^{-1} = (I_{\mathbf{X}} - A^{-1}(A - B))^{-1} A^{-1}$. Thus

$$\|B^{-1}\| \le \|(I_{\mathbf{X}} - A^{-1}(A - B))^{-1}\|\,\|A^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|(A - B)\|}$$

which gives the required bound on $\|B^{-1}\|$.

Now $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. Therefore

$$\|A^{-1} - B^{-1}\| \le \|A^{-1}\|\,\|A - B\|\,\|B^{-1}\| \le \frac{\|A^{-1}\|^2 \|B - A\|}{1 - \|A^{-1}\|\,\|A - B\|}.$$

This completes the proof. □

Remark 2.9. When $\mathbf{X}$ is finite dimensional a linear operator $A \colon \mathbf{X} \to \mathbf{X}$ is invertible if and only if $\det(A) \ne 0$. As det is a continuous function the set $\{A : \det(A) \ne 0\}$ is open. Thus in this case that the set of invertible operators is open has an easier proof.

2.4. **The derivative of maps between Banach spaces.** Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces, $U \subseteq \mathbf{X}$ be an open set and $f \colon U \to \mathbf{Y}$ a function. Then f is ***differentiable*** at $a \in U$ if and only if there is a linear map $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ so that

$$f(x) - f(a) = A(x - a) + o(|x - a|_{\mathbf{X}}).$$

More explicitly this means these is a function $x \mapsto \varepsilon(x; a)$ from a neighborhood of a in U so that

$$f(x) - f(a) = A(x - a) + |x - a|_{\mathbf{X}} \varepsilon(x; a) \quad \text{where} \quad \lim_{x \to a} |\varepsilon(x, a)|_{\mathbf{Y}} = 0.$$

When f is differentiable at a the linear map A is unique and called the ***derivative*** of f at a. It will be denoted by $A = f'(a)$. Thus for us the derivative is a bounded linear map $f'(a) \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ rather than a number.

Problem 2.9. Verify the claim above that the A in the definition of the derivative is unique. *Hint:* If A and B are bounded linear maps such that

$$\begin{aligned} f(x) - f(a) &= A(x-a) + |x-a|_{\mathbf{X}}\varepsilon_1(x,a) \\ f(x) - f(a) &= B(x-a) + |x-a|_{\mathbf{X}}\varepsilon_2(x,a) \end{aligned}$$

where $\lim_{x\to a} \varepsilon_1(x,a) = \lim_{x\to a} \varepsilon_2(x,a) = 0$, then

$$(B-A)(x-a) = |x-a|_{\mathbf{X}}(\varepsilon_2(x,a) - \varepsilon_1(x,a)).$$

Let v be any vector in $\mathbf{X}$ and let $x = a + tv$ where $t > 0$. Use the last equation to show

$$t(B-A)v = t|v|_{\mathbf{X}}(\varepsilon_2(a+tv,a) - \varepsilon_1(a+tv,a))$$

divide by t and take the limit at $t \searrow 0$ to show $(B-A)v = 0$. □

Problem 2.10. Show that f is differential at a with $f'(a) = A$ if and only if

$$\lim_{x\to a} \frac{|f(x) - f(a) - A(x-a)|_{\mathbf{Y}}}{|x-a|_{\mathbf{X}}} = 0.$$ □

Problem 2.11. If f is differentiable at a then f is continuous at a. □

To get a feel for what this linear map measures let $v \in \mathbf{X}$, assume that $f\colon U \to \mathbf{Y}$ is differentiable at a and let $c(t) := f(a+tv)$. Then for $t \neq 0$

$$\begin{aligned} \frac{1}{t}(c(t) - c(0)) = \frac{1}{t}(f(a+tv) - f(a)) &= \frac{1}{t}(f'(a)tv + |tv|_{\mathbf{X}}\varepsilon(a+tv;a)) \\ &= f'(a)v + |v|_{\mathbf{X}}\varepsilon(a+tv;a). \end{aligned}$$

But $\lim_{t\to 0} \varepsilon(a+tv;a) = 0$ so this implies that c has a tangent vector at $t = 0$ and that it is given by $c'(0) = f'(a)v$. That is $f'(a)v$ is the "directional derivative" at a of f in the direction v.

Problem 2.12. Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and $U \subset \mathbf{X}$ open. Let $c\colon (a,b) \to U$ be a continuously differentiable map and let $f\colon U \to \mathbf{Y}$ be a map that is differentiable at $c(t_0)$. Then $\gamma(t) := f(c(t))$ is differentiable at t_0 and

$$\left.\frac{d}{dt} f(c(t))\right|_{t=t_0} = \gamma'(t_0) = f'(c(t_0))c'(t_0).$$ □

To make this more concrete let us look at some finite dimensional cases.

Problem 2.13. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a scalar valued function on $\mathbb{R}^n$. Let

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},\ e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

be the standard basis of $\mathbb{R}^n$ and write $x \in \mathbb{R}^n$ as

$$x = \sum_{j=1}^n x^j e_j = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}.$$

Assume that f is differentiable at the point a. Then the derivative $f'(a)$ is a linear map form $\mathbb{R}^n$ to $\mathbb{R}$. (Linear maps from a vector space to the field of scalars are called ***linear functionals***.) Show that the matrix of this linear functional is the row vector

$$\left[\frac{df}{dx^1}, \frac{df}{dx^2}, \dots, \frac{df}{dx^n}\right]$$

where all the components are evaluated at a. *Hint:* Here is one way to see what is going in when $n = 2$. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ and assume that f is differentiable at (a, b). Every linear functional $\lambda\colon \mathbb{R}^2 \to \mathbb{R}$ is of the form $\lambda(x, y) = c_1x + c_2y$ for some $c_1, c_2 \in \mathbb{R}$ and therefore we have $f'(a, b)(x, y) = c_1x + c_2y$. As f is differentiable at (a, b)

$$\begin{aligned} f(x, y) &= f(a, b) + f'(a, b)(x - a, y - b) + \varepsilon(x, y, a, b)|(x - a, y - b)|_{\ell^2} \\ &= f(a, b) + c_1(x - a) + c_2(y - b) + \varepsilon(x, y, a, b)|(x - a, y - b)|_{\ell^2} \end{aligned}$$

where

$$\lim_{(x,y)\to((a,b)} \varepsilon(x, y, a, b) = 0.$$

Let $y = b$ and rearrange:

$$\begin{aligned} \frac{f(x, b) - f(a, b)}{(x - a)} &= c_1 + \frac{\varepsilon(x, y, a, b)|(x - a, 0)|_{\ell^2}}{x - a} \\ &= c_1 \pm \varepsilon(x, y, a, b) \end{aligned}$$

as $|(x - a, 0)|_{\ell^2}/(x - a) = \pm 1$. Taking a limit

$$\begin{aligned} \frac{\partial f}{\partial x}(a, b) &= \lim_{x\to a} \frac{f(x, b) - f(a, b)}{(x - a)} \\ &= \lim_{x\to a}\left(c_1 + \frac{\varepsilon(x, y, a, b)|(x - a, 0)|_{\ell^2}}{x - a}\right) \\ &= c_1 \end{aligned}$$

A similar calculation shows $c_2 = \frac{\partial f}{\partial y}(a, b)$. Therefore

$$f'(a, b)\begin{bmatrix} x \\ y \end{bmatrix} = \frac{\partial f}{\partial x}(a, b)x + \frac{\partial f}{\partial y}(a, b)y = \left[\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b)\right]\begin{bmatrix} x \\ y \end{bmatrix}$$

and thus $f'(a, b)$ is the row vector $\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$ where the components are evaluated at (a, b). $\square$

Problem 2.14. Let $\mathbf{X} = \mathbb{R}^n$ and $\mathbf{Y} = \mathbb{R}^m$ and let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function given in components as

$$f(x) = \begin{bmatrix} f^1(x) \\ f^2(x) \\ \vdots \\ f^m(x) \end{bmatrix}.$$

Let $\dfrac{\partial f}{\partial x^i}$ be the partial derivative of f with respect to x^i. That is

$$\frac{\partial f}{\partial x^i}(x) = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^i}(x) \\ \\ \dfrac{\partial f^2}{\partial x^i}(x) \\ \vdots \\ \dfrac{\partial f^m}{\partial x^i}(x) \end{bmatrix}$$

Assume that f is differentiable at $x = a$. Then show that the matrix A of $f'(a)$ in the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the matrix with columns $\dfrac{\partial f}{\partial x^1}, \dfrac{\partial f}{\partial x^2}, \ldots, \dfrac{\partial f}{\partial x^n}$. That is

$$A = \left[\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \ldots, \frac{\partial f}{\partial x^m}\right] = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^1} & \dfrac{\partial f^1}{\partial x^2} & \cdots & \dfrac{\partial f^1}{\partial x^n} \\ \\ \dfrac{\partial f^2}{\partial x^1} & \dfrac{\partial f^2}{\partial x^2} & \cdots & \dfrac{\partial f^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f^m}{\partial x^1} & \dfrac{\partial f^m}{\partial x^2} & \cdots & \dfrac{\partial f^m}{\partial x^n} \end{bmatrix}$$

where these are all evaluated at $x = a$. *Hint:* As in the last problem let us look at the case where $n = 2$ and assume that f is differentiable at (a, b). Then we have

$$f(x, y) = \begin{bmatrix} f^1(x, y) \\ f^2(x, y) \\ \vdots \\ f^m(x, y) \end{bmatrix}$$

Rather than use the method of the last problem we use Problem 2.12.

$$\begin{aligned}
f'(a,b)e_1 &= \frac{d}{dt}\bigg|_{t=0} f\left((a,b)+te_1\right) \\
&= \frac{d}{dt}\bigg|_{t=0} f\left(a+t,b\right) \\
&= \frac{d}{dt}\bigg|_{t=0} \begin{bmatrix} f^1(a+t,b) \\ f^2(a+t,b) \\ \vdots \\ f^m(a+t,b) \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\partial f^1}{\partial x}(a,b) \\ \dfrac{\partial f^2}{\partial x}(a,b) \\ \vdots \\ \dfrac{\partial f^m}{\partial x}(a,b) \end{bmatrix}.
\end{aligned}$$

Likewise

$$f'(a,b)e_2 = \begin{bmatrix} \dfrac{\partial f^1}{\partial y}(a,b) \\ \dfrac{\partial f^2}{\partial y}(a,b) \\ \vdots \\ \dfrac{\partial f^m}{\partial y}(a,b) \end{bmatrix}$$

The rest is just remembering that the matrix of a linear map between Euclidean spaces has as its columns the images of the standard basis. □

The following gives trivial examples of differentiable maps.

Proposition 2.10. *Let $A\colon \mathbf{X} \to \mathbf{Y}$ be a bounded linear map between Banach spaces and $y_0 \in \mathbf{Y}$. Set $f(x) = Ax + y_0$ then f is differentiable at all points $a \in \mathbf{X}$ and $f'(a) = A$ for all a.*

Proof. $f(x) - f(a) = A(x-a)$ so the definition of differentiable is verified with $\varepsilon(x,a) \equiv 0$. □

The following gives a less trivial example.

Proposition 2.11. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and let $U \subset \mathcal{B}(\mathbf{X},\mathbf{Y})$ be the set of invertible elements (this is an open set by Proposition 2.8). Define a map $f\colon U \to \mathcal{B}(\mathbf{Y},\mathbf{X})$ by*

$$f(X) = X^{-1}.$$

Then f is differentiable and for $A \in U$ the derivative $f'(A)\colon \mathcal{B}(\mathbf{X},\mathbf{Y}) \to \mathcal{B}(\mathbf{Y},\mathbf{X})$ is the linear map whose value on $V \in \mathcal{B}(\mathbf{X},\mathbf{Y})$ is

$$f'(A)V = -A^{-1}VA^{-1}.$$

Proof. Let $L\colon \mathcal{B}(\mathbf{X},\mathbf{Y}) \to \mathcal{B}(\mathbf{Y},\mathbf{X})$ be the linear map $LV := -A^{-1}VA^{-1}$ Then for $X \in U$

$$\begin{aligned} f(X) - f(A) - L(X-A) &= X^{-1} - A^{-1} + A^{-1}(X-A)A^{-1} \\ &= X^{-1}(A-X)A^{-1} + A^{-1}(X-A)A^{-1} \\ &= (-X^{-1} + A^{-1})(X-A)A^{-1} \\ &= X^{-1}(X-A)A^{-1}(X-A)A^{-1}, \end{aligned}$$

so that

$$\|f(X) - f(A) - L((X-A)\| \le \|X^{-1}\|\,\|A^{-1}\|^2\|X-A\|^2.$$

The map $X \mapsto X^{-1}$ is continuous (Proposition 2.8) so $\lim_{X\to A} X^{-1} = A^{-1}$. Thus

$$\lim_{X\to A} \frac{\|f(X) - f(A) - L((X-A)\|}{\|X-A\|} \le \lim_{X\to A} \|X^{-1}\|\,\|A^{-1}\|^2\|X-A\| = 0.$$

The result now follows from Problem 2.10. □

Next we consider the chain rule.

Proposition 2.12. *Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ be Banach spaces $U \subseteq \mathbf{X}$, $V \subseteq \mathbf{Y}$ open sets $f\colon U \to \mathbf{Y}$ and $g\colon V \to \mathbf{Z}$. Let $a \in U$ so that $f(a) \in V$ and assume that f is differentiable at a and g is differentiable at $f(a)$. Then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a))f'(a)$$

Proof. From the definitions $f(x) - f(a) = f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_1(x;a)$ and $g(y) - g(f(a)) = g'(f(a)) + |y - f(a)|_{\mathbf{Y}}\varepsilon_2(x; f(a))$ where $\lim_{x\to a}\varepsilon_1(x;a) = 0$ and $\lim_{y\to f(a)}\varepsilon_2(y; f(a)) = 0$. Then

$$\begin{aligned} g(f(x)) - g(f(a)) &= g'(f(a))(f(x) - f(a)) + |f(x) - f(a)|_{\mathbf{Y}}\varepsilon_2(f(x), f(a)) \\ &= g'(f(a))f'(a)(x-a) + g'(f(a))|x-a|_{\mathbf{X}}\varepsilon_1(x,a) \\ &\qquad + |f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_1(x,a)|_{\mathbf{X}}\varepsilon_2(f(x), f(a)) \\ &= g'(f(a))f'(a)(x-a) + |x-a|_{\mathbf{X}}\Bigg(g'(f(a))\varepsilon_1(x,a) \\ &\qquad + \left|f'(a)\frac{x-a}{|x-a|_{\mathbf{X}}} + \varepsilon_1(x,a)\right|_{\mathbf{X}} \varepsilon_2(f(x); f(a))\Bigg) \\ &= g'(f(a))f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_3(x;a), \end{aligned}$$

where this defines $\varepsilon_3(x,a)$. Then

$$|\varepsilon_3(x,a)|_{\mathbf{X}} \le \|g'(f(a))\||\varepsilon_1(x,a)|_{\mathbf{Y}} + (\|f'(a)\| + |\varepsilon_1(x;a)|_{\mathbf{Y}})|\varepsilon_2(f(x), f(a))|_{\mathbf{Z}}.$$

This (and the continuity of f at a) implies $\lim_{x\to a}\varepsilon_3(x,a)=0$ which completes the proof. $\square$

Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces, $U\subset\mathbf{X}$ open and $f\colon U\to\mathbf{Y}$. Then f is ***continuously differentiable*** on U if and only if f is differentiable at each point $x\in U$ and the map $x\mapsto f'(x)$ is a continuous map from U to $\mathcal{B}(\mathbf{X},\mathbf{Y})$. Or what is the same thing, $f'(x)$ exists for all $x\in U$ and for $a\in U$ we have $\lim_{x\to a}\|f'(x)-f'(a)\|=0$.

If $c\colon[a,b]\to U$ is a continuously differentiable curve and $f\colon U\to\mathbf{Y}$ is a continuously differentiable map, then $\gamma(t):=f(c(t))$ is a continuously differential curve $\gamma\colon[a,b]\to\mathbf{Y}$ and by the chain rule (or Problem (2.12))

$$\gamma'(t)=f'(c(t))c'(t).$$

Using the fundamental theorem of calculus this gives

$$\gamma(b)=\gamma(a)=\int_a^b\gamma'(t)\,dt=\int_a^b f'(c(t))c'(t)\,dt$$

But

$$|\gamma'(t)|_{\mathbf{Y}}=|f'(c(t))c'(t)|_{\mathbf{Y}}\le\|f'(c(t))\||c'(t)|_{\mathbf{X}}.$$

These can be combined to give

$$|\gamma(b)-\gamma(a)|_{\mathbf{Y}}=\left|\int_a^b f'(c(t))c'(t)\,dt\right|_{\mathbf{Y}}\le\int_a^b\|f'(c(t))\||c'(t)|_{\mathbf{X}}\,dt. \tag{2}$$

Recall that a U is ***convex*** if and only if when $x_0,x_1\in U$ and $0\le t\le 1$, then $(1-t)x_0+tx_1\in U$.

Proposition 2.13 (Mean Value inequality)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and let $U\subset\mathbf{X}$ be open and convex. Assume that $f\colon U\to\mathbf{Y}$ is continuously differentiable and that $\|f'(x)\|\le C$ for all $x\in U$. Then*

$$|f(x_1)-f(x_0)|_{\mathbf{Y}}\le C|x_1-x_0|_{\mathbf{X}}$$

for all $x_1,x_0\in U$

Proof. Let $c\colon[0,1]\to U$ be given by $c(t)=(1-t)x_0+tx_1=x_0+t((x_1-x_0)$ (this curves lies in U as $x_0,x_1\in U$ and U is convex). Then $c'(t)=(x_1-x_0)$. Let $\gamma(t):=f(c(t))$. Then $\gamma'(t)=f'(c(t))c'(t)=f'(c(t))(x_1-x_0)$. Putting this into (2) implies

$$\begin{aligned}|f(x_1)-f(x_0)|_{\mathbf{Y}}=|\gamma(1)-\gamma(0)|_{\mathbf{Y}}&\le\int_0^1\|f'(c(t))\||x_1-x_0|_{\mathbf{X}}\,dt\\&\le\int_0^1 C|x_1-x_0|_{\mathbf{X}}\,dt=C|x_1-x_0|_{\mathbf{X}}.\end{aligned}$$

This completes the proof. $\square$

2.5. **Preliminary version of the inverse function theorem.** In a Banach space $\mathbf{X}$ we let

$$B(a,r) := \{x : |x-a|_{\mathbf{X}} < r\}, \qquad \overline{B}(x,r) := \{x : |x-a|_{\mathbf{X}} \le r\}$$

be the open and closed balls of radius r centered at a. In the following theorem and its proof we will always be referring to the same Banach space $\mathbf{X}$, and therefore we simplify notation by shorting $|x|_{\mathbf{X}}$ to $|x|$.

Theorem 2.14. *Let $\mathbf{X}$ be a Banach space and W an open subset of $\mathbf{X}$ that contains 0. Let $f\colon W \to \mathbf{X}$ be a continuously differentiable function with*

$$f(0) = 0, \qquad \text{and} \qquad f'(0) = I$$

where I is the identity map. Then 0 has an open neighborhood $U \subseteq W$ such that the image $V := f[U]$ is open and there is a continuously differentiable map $g\colon V \to U$ inverse to $f\big|_U$ and the derivative of g is given by

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

Problem 2.15. Prove this along the following lines. As motivation note that on some level finding the inverse of f is equivalent to solving $f(x) = y$ for x. We will reduce this to finding x as the fixed point of a contraction. For $y \in \mathbf{X}$ set

$$\varphi_y(x) := x - f(x) + y.$$

(a) Show

$$f(x) = y \qquad \text{if and only if} \qquad \varphi_y(x) = x.$$

Also show that

$$\varphi_y = \varphi_0 + y$$

and therefore all the φ_y's have the same derivative:

$$\varphi_y'(x) = I - f'(x)$$

where I is the identity map on $\mathbf{X}$.

(b) We would like φ_y to be a contraction. This will not necessarily be true on the entire domain of f. So we need to restrict to a smaller set. The map $x \mapsto f'(x)$ is continuous by assumption. Use this to show that $x \mapsto \|I - f'(x)\|$ is continuous and therefore there is a $r > 0$ such that

$$x \in \overline{B}(0,r) \qquad \text{implies} \qquad \|I - f'(x)\| \le \frac{1}{2}.$$

(For the continuity of $x \mapsto \|I - f'(x)\|$ note this is the composition of $x \mapsto I - f'(x)$ on W and the map $A \mapsto \|A\|$ on $\mathcal{B}(\mathbf{X},\mathbf{X})$ and these maps are continuous.) By part (a) this is equivalent to

$$x \in \overline{B}(0,r) \qquad \text{implies} \qquad \|\varphi_y'(x)\| \le \frac{1}{2}.$$

for any $y \in \mathbf{X}$. For future use also note that $\|I - f'(x)\| \le 1/2$ implies (by Proposition 2.7) that for $x \in \overline{B}(0,r)$ the linear map $f'(x)$ has an inverse, $f'(x)^{-1}$ and

$$\|f'(x)^{-1}\| \le 2.$$

(c) Now use the Mean Value Inequality (Proposition 2.13) to show that for all $x_0, x_1 \in \overline{B}(0,r)$ and for any $y \in \mathbf{X}$

$$|\varphi_y(x_1) - \varphi_y(x_0)| \le \frac{1}{2}|x_1 - x_0|.$$

(d) The last part of the problem does not quite show that φ_y is a contraction on $\overline{B}(0,r)$ as φ_y need not map $\overline{B}(0,r)$ into itself. (For example if $|y| > r$ the $\varphi_y(0) \notin \overline{B}(0,r)$.) This is not hard to fix. Note

$$|\varphi_y(x)| = |\varphi_0(x) + y| \le |\varphi_0(x)| + |y|.$$

Now show if $x \in \overline{B}(0,r)$, then

$$|\varphi_0(x)| = |\varphi_0(x) - \varphi_0(0)| \le \frac{1}{2}|x| \le \frac{r}{2}.$$

Thus

$$x \in \overline{B}(0,r),\ y \in \overline{B}(0,r/2) \qquad \text{implies} \qquad \varphi_y(x) \in \overline{B}(0,r).$$

(e) As $\overline{B}(0,r)$ is a closed subset of a complete metric space, it is itself a complete metric space. Use what has been done so far to show that for all $y \in \overline{B}(0,r/2)$ that $\varphi_y \colon \overline{B}(0,r) \to \overline{B}(0,r)$ is a contraction and therefore by the Banach Fixed Point Theorem (Theorem 2.1) φ_y has a unique fixed point in $\overline{B}(0,r)$ and therefore for all $y \in \overline{B}(0,r/2)$ the equation $f(x) = y$ has a unique solution with $x \in \overline{B}(0,r)$.

(f) We now deal with an annoying minor point. We are looking an inverse of f on a open neighborhood of 0, but so far we are working with the closed sets $\overline{B}(0,r)$ and $\overline{B}(0,r/2)$. So show that if $y \in B(0,r/2)$ that $f(x) = y$ has a unique solution with $x \in B(0,r)$. *Hint:* If $y \in B(0,r)$, then $|y| < r/2$. If $f(x) = y$ with $f(x)$, then $x = \varphi_y(x)$ and therefore

$$|x| = |\varphi_y(x)| = |\varphi_0(x) + y| \le |\varphi_0(x)| + |y|$$

now proceed as in Part (d).

(g) We have shown the image of $B(0,r)$ under f contains $B(0,r/2)$. To get an inverse we also need that f is injective. To do this use that $f(x) = x - \varphi_0(x)$ and therefore use the triangle inequality and the reverse triangle inequality (i.e. $|a+b| \ge |a| - |b|$) to show for all $x_0, x_1 \in B(0,r)$ that

$$|x_1 - x_0| - |\varphi_0(x_1) - \varphi_0(x_0| \le |f(x_1) - f(x_0)| \le |x_1 - x_0| + |\varphi_0(x_1) - \varphi_0(x_0)|$$

and therefore

$$\frac{1}{2}|x_1 - x_0| \le |f(x_1) - f(x_0)| \le \frac{3}{2}|x_1 - x_0|.$$

This shows that f is injective on $B(0,r)$. (The upper bound on $|f(x_1) - f(x_0)|$ will be used in showing the inverse of f is continuous.)

(h) Let $U = \left(f\big|_{B(0,r)}\right)^{-1}[B(0,r/2)]$ be the preimage of $B(0,r/2)$ by the restriction of f to $B(0,r)$. Show that U is open and that $f\big|_U \colon U \to B(0,r/2)$ is a bijection. We set $V = B(0,r/2)$ Therefore there is an inverse $g\colon V \to U$ to $f\big|_U$.

(i) Now show that g is continuous by showing for $y_0, y_1 \in B(0,r/2) = V$ that
$$\frac{2}{3}|y_1 - y_0| \le |g(y_1) - g(y_0)| \le 2|y_1 - y_0|.$$
Hint: Let $x_0 = g(y_0)$ and $x_1 = g(y_1)$ and use these values in the inequalities of Part (g).

(j) Show that g is differentiable and its derivative is $g'(y) = f'(g(y))^{-1}$. *Hint:* Let $y, y_0 \in V = B(0,r/2)$ and let $x = g(y)$ and $x_0 = g(y_0)$. As f is differentiable
$$f(x) - f(x_0) = f'(x - x_0)(x - x_0) + \varepsilon(x, x_0)|x - x_0|$$
with
$$\lim_{x \to x_0} \varepsilon(x, x_0) = 0.$$
Using $f(x) = y$, $f(x_0) = y_0$ show
$$\begin{aligned} g(y) - g(y_0) &= f'(g(y_0))^{-1}(y - y_0) - f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))|g(y) - g(y_0)| \\ &= f'(g(y_0))^{-1}(y - y_0) + \rho(y, y_0) \end{aligned}$$
Where this defines $\rho(y, y_0)$. Rewrite $\rho(y, y_0)$ as
$$\begin{aligned} \rho(y, y_0) &= -f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))\frac{|g(y) - g(y_0)|}{|y - y_0|}|y - y_0| \\ &= \varepsilon_1(y, y_0)|y - y_0| \end{aligned}$$
where
$$\varepsilon_1(y, y_0) = -f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))\frac{|g(y) - g(y_0)|}{|y - y_0|}.$$
To show g is differentiable at y_0 it is enough to show
$$\lim_{y \to y_0} \varepsilon_1(y, y_0) = 0.$$
By Parts (b) and (i) we have
$$\|f'(g(y_0))^{-1}\| \le 2, \qquad \frac{|g(y) - g(y_0)|}{|y - y_0|} \le 2.$$
Use these to show
$$|\varepsilon_1(y, y_0)| \le 4|\varepsilon(g(y), g(y_0))|$$
and use the continuity of g to show
$$\lim_{y \to y_0} \varepsilon(g(y), g(y_0)) = 0.$$
Put these pieces together to conclude $\lim_{y \to y_0} \varepsilon_1(y, y_0) = 0$.

(k) Now that g is differentiable, all that remains is to show it is continuously differentiable. Prove this. *Hint:* From the last part of the problem we have that

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

This is the composition of three maps. The first is $y \to g(y)$ form $V = B(0, r/2)$ to U, the second is $x \mapsto f'(x)$ from U to $\mathcal{B}(\mathbf{X}, \mathbf{X})$, and the third is the inverse map $A \mapsto A^{-1}$ on the set of invertible elements of $\mathcal{B}(\mathbf{X}, \mathbf{X})$. All of these maps are continuous (for the continuity of $A \mapsto A^{-1}$ see Proposition 2.8). □

2.6. **Inverse Function Theorem.**

Theorem 2.15. *Let W be an open subset of the Banach space $\mathbf{X}$, and $f\colon W \to \mathbf{Y}$ a continuously differentiable map from W to the Banach space $\mathbf{Y}$. If $a \in W$ and $f'(a)\colon \mathbf{X} \to \mathbf{Y}$ is an invertible linear map, then a has an open neighborhood U and $f(a)$ has an open neighborhood V and there is a continuously differentiable map $g\colon V \to U$ that is the inverse of the restriction $f\big|_U$. To be very explicit*

$$f(g(y)) = y \qquad \textit{and} \qquad g(f(x)) = x$$

for all $x \in U$ and $y \in V$.

Problem 2.16. Prove this. *Hint:* This can be reduced to Theorem 2.14. Let $W_0 = W - a := \{w - a : w \in W\}$. This is a neighborhood of 0 in $\mathbf{X}$. Define $f_0\colon W_0 \to \mathbf{X}$ by

$$f_0(x) = f'(a)\,(f(x+a) - f(a))\,.$$

(a) Show that $f_0(0) = 0$, $f_0'(0) = I$, and

$$f(x) = f'(a) f_0(x - a) + f(a).$$

(b) Now use Theorem 2.14 to show that 0 has open neighborhoods U_0 and $V_0 = f[U_0]$ and a continuously differentiable map $g_0\colon V_0 \to U_0$ such that g_0 is the inverse of the restriction $f\big|_{U_0}$.

(c) Let $U = U_0 + a$ and $W = W_0 + f(a)$ and

$$g(y) = g_0\left(f'(a)^{-1}(y - f(a))\right).$$

Show

$$g_0(y) = g\left(f'(a)y + f(a)\right) - f(a)$$

and then that

$$f(g(y)) = y \qquad \text{and} \qquad g(f(x)) = x$$

for all $x \in U$ and $y \in V$. Thus g is the required inverse to the restriction of f to U. As g_0 is continuous differentiable we see that g is the composition of continuous differentiable functions and therefore g is continuous differentiable.

(d) To get the formula for $g'(y)$ note that
$$f \circ g(y) = f(g(y)) = y = Iy$$
By the chain rule
$$(f \circ g)'(y) = f'(g(y))g'(y)$$
and the derivative of I is I. Thus $f'(g(y))g'(y) = I$. Now solve this for $g'(y)$ to finish the proof. $\square$

A map f between metric spaces is an ***open map*** if and only if f maps open sets to open sets. That is if V is an open subset of the domain of f, then the image $f[V]$ is open.

Proposition 2.16. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and U an open subset of $\mathbf{X}$. If $f\colon U \to \mathbf{Y}$ is a continuous differentiable map such that $f'(x)$ is invertible for all $x \in U$, then f is an open map.*

Problem 2.17. Prove this. $\square$

A map $f\colon M \to N$ between metric spaces is a ***proper map*** if and only if for all compact subsets $K \subseteq M$ the preimage $f^{-1}[K]$ is compact.

Problem 2.18. This problem is to give a feel for what it means for a map to be proper. Show that a map $f\colon \mathbb{R}^n \to \mathbb{R}$ is if and only if
$$\lim_{|x|_{\ell^2} \to \infty} |f(x)| = \infty.$$
More generally if $f\colon \mathbb{R}^m \to \mathbb{R}^n$ then f is proper if and only if
$$\lim_{|x|_{\ell^2} \to \infty} |f(x)|_{\ell^2} = \infty.$$
Thus a map between Euclidean spaces is proper exactly when it maps large points to large points. $\square$

Proposition 2.17. *If $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is a proper map, then the image $f[\mathbb{R}^m]$ is closed in $\mathbb{R}^n$.*

Problem 2.19. Prove this. *Hint:* Let y_0 be a point in the closure of $f[\mathbb{R}^n]$. We need to show that y_0 is in $f[\mathbb{R}^m]$. As y_0 is in the closure of $f[\mathbb{R}^n]$ there is a sequence $y_1, y_2, \ldots \in f[\mathbb{R}^m]$ with $\lim_{k\to\infty} y_k = y_0$. Then $K := \{y_0\} \cup \{y_k : k = 1, 2, \ldots\}$ is a compact subset of $\mathbb{R}^n$. Let $x_k \in \mathbb{R}^m$ with $y_k = f(x_k)$ then $\{x_k : k = 1, 2, \ldots\} \subseteq f^{-1}[K]$. Now use that f is proper to show that $x_1, x_2, \ldots$ is a bounded sequence and therefore has a convergent subsequence. $\square$

Recall that a metric space, M, is ***connected*** if and only if the only subsets C of M that are both open and closed are $\varnothing$ and M.

We can now give our first global result.

Proposition 2.18. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable proper map and assume that $f'(x)$ is invertible for all $x \in \mathbb{R}^n$. Then f is surjective and therefore for all $y \in \mathbb{R}^n$ the equation $f(x) = y$ has a solution.*

Problem 2.20. Prove this. □

Remark 2.19. In the last proposition more can be said. In fact f will be bijective. Showing that f is injective uses that $\mathbb{R}^n$ is simply connected and some results about covering spaces from algebraic topology. □

2.7. **Implicit functions in two and three dimensions.** After this more or less abstract abstract version of the inverse function theorem, let us look at two and three dimensional applications.

2.7.1. *Implict in two dimensions.* The set set of this this. Given an equation such as

$$y^3 + xy + x^5 = 0$$

we would like to be able to understand when we can solve for y in terms of x and get a smooth function. We will only be able to do this locally near a point that satisfies the equation. More explicitly given an equation of the form

$$g(x, y) = c$$

and a point (x_0, y_0) with $g(x_0, y_0) = c$, we want to find a function $y = f(x)$ defined on a neighborhood of x_0 with $f(x_0) = y_0$ and $g(x, f(x)) \equiv c$. As a motivation for one of the hypothesis of the next theorem assume that $g(x, y)$ is linear, that is we wish to solve

$$g(x, y) = ax + by = c$$

for y. We can only do this if $b \neq 0$. In the general case we have the approximation to $g(x, y)$ given by the derivative

$$g(x, y) \approx g(x_0, y_0) + \frac{\partial g}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0)(y - y_0) = c$$

and we can only solve this approximating equation for y when $\frac{\partial g}{\partial x}(x_0, y_0) \neq 0$.

Theorem 2.20. *Let W be an open set in $\mathbb{R}^2$ and $g\colon W \to \mathbb{R}$ a C^1 function. Let $(x_0, y_0) \in W$ and assume*

$$\frac{\partial g}{\partial y}(x_0, y_0) \neq 0.$$

Then there are open neighborhoods U of x_0 and V of y_0 and a C^1 function $f\colon U \to V$ with

$$f(x_0) = y_0 \qquad \text{and} \qquad g(x, f(x)) = g(x_0, y_0)$$

for all $x \in U$.

Problem 2.21. Prove this along the following lines. We will view elements of $\mathbb{R}^2$ as column vectors.

(a) Define a function $F\colon W \to \mathbb{R}^2$ by

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ g(x,y) \end{bmatrix}$$

and show that the derivative of F is the matrix

$$F'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ g_x(x,y) & g_y(x,y) \end{bmatrix}$$

where we have used the abbreviations

$$g_x = \frac{\partial g}{\partial x} \qquad \text{and} \qquad g_y = \frac{\partial g}{\partial y}.$$

(b) Show that the matrix $F'(x_0, y_0)$ has an inverse.

(c) Use the Inverse Function Theorem to find a neighborhood, N, of $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ and a neighborhood, M, of $F\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\right) = \begin{bmatrix} x_0 \\ g(x_0, y_0) \end{bmatrix}$ and a C^1 map $G\colon M \to N$ that is the inverse of the restriction $F\big|_M$.

(d) Let G be given by

$$G\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \xi(x,y) \\ \eta(x,y) \end{bmatrix}.$$

Use that

$$F \circ G\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$$

to show

$$\xi(x,y) = x \qquad \text{and} \qquad g(x, \eta(x,y)) = y.$$

(e) Thus if $f(x) = \eta(x, y_0)$ we have $g(x, f(x)) = y_0$ for all x in the domain of f. Now fiddle around with the details to finish the proof. That is find the neighborhoods U and V etc. $\square$

The following is the justification of the method of implicit differentiation we all learned in our calculus classes.

Corollary 2.21. *Let $f(x)$ be as in the last theorem. Then*

$$f'(x) = -\frac{\dfrac{\partial g}{\partial x}(x, f(x))}{\dfrac{\partial g}{\partial y}(x, f(x))} = -\frac{g_x(x, f(x))}{g_y(x, f(x))}.$$

Problem 2.22. Prove this. $\square$

We define a subset C of $\mathbb{R}^2$ to be a C^1 ***curve*** if and only if for each point $(x_0, y_0) \in C$ is locally either the graph of C^1 function over the x axis or a C^1 function over the y axis. That is there are open neighborhoods U of x_0 and V of y_0 and either a C^1 function $f\colon U \to V$ with

$$C \cap (U \times V) = \{(x, f(x)) : x \in U\}$$

or a C^1 function $h\colon V \to U$ with

$$C \cap (U \times V) = \{(h(y) : y \in V\}.$$

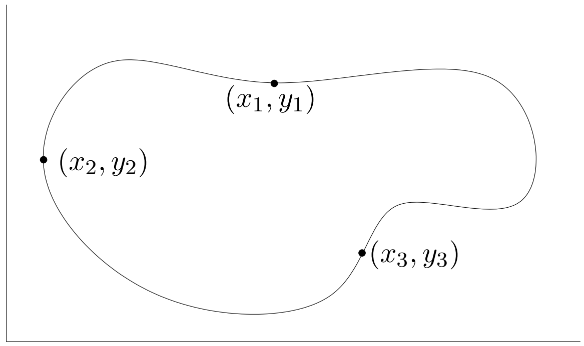


FIGURE 1. A C^1 curve. At (x_1, y_1) the curve is locally a graph over the x-axis. At (x_2, y_2) it is locally a graph over the y-axis. At (x_3, y_3) it is locally a graph over both the x and y-axis.

What Theorem 2.20 shows is that if C is

$$C = \{(x, y) : g(x, y) = c\}$$

where g is a C^1 function defined on an open subset of $\mathbb{R}^2$ that near any point (x_0, y_0) of C where $g_y(x_0, y_0) \neq 0$ that C is locally a C^1 graph over the x-axis. By interchanging the rolls of x and y we see that near any point where $g_x \neq 0$ that C is locally a graph over the y-axis. This yields the following.

Theorem 2.22. *Let $g\colon W \to \mathbb{R}$ be a C^1 function on an open subset W of $\mathbb{R}^2$. If for all $(x, y) \in C$ we have*

$$\nabla g(x, y) \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

then C^1 is a C^1 curve in $\mathbb{R}^2$. □

As an example we look at the unit circle $C := \{(x, y) : x^2 + y^2 = 1\}$. Here $g(x, y) = x^2 + y^2$ and the gradient is

$$\nabla g(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

and the only point where this vanishes is the origin. As the origin is not on C we have that C is a C^1 curve, something you already knew. And it can be seen by noting that at each point it is either a graph over the x-axis of one of the functions $x \mapsto \pm\sqrt{1 - x^2}$ or a graph over the y-axis of one of the functions $y \mapsto \pm\sqrt{1 - y^2}$.

Let us look at an example where we can not solve for x or y explicitly.

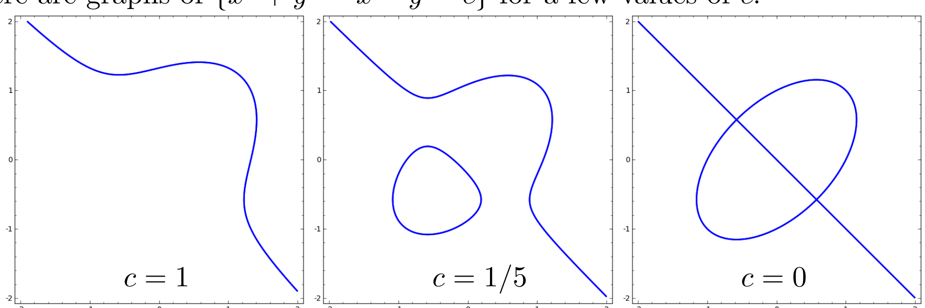
Problem 2.23. Show that the set defined by $x^3 + y^3 - x - y = c$ is a C^1 curve for all

$$c \neq 0, \ \pm c_{\text{crit}}$$

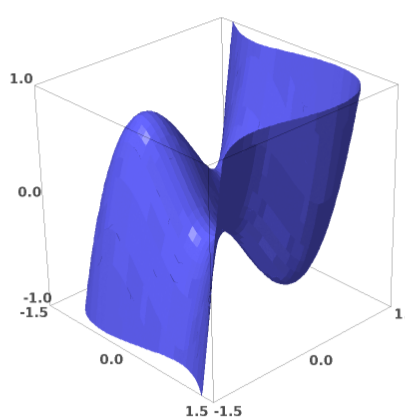
where

$$c_{\text{crit}} = \frac{4}{3\sqrt{3}}.$$

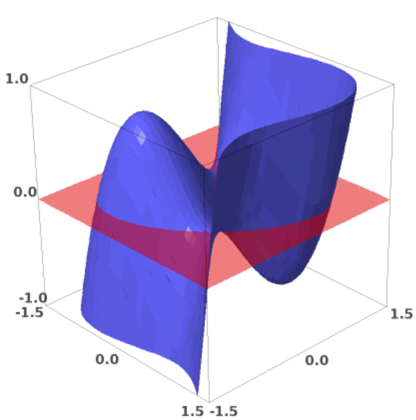
Here are graphs of $\{x^3 + y^3 - x - y = c\}$ for a few values of c.



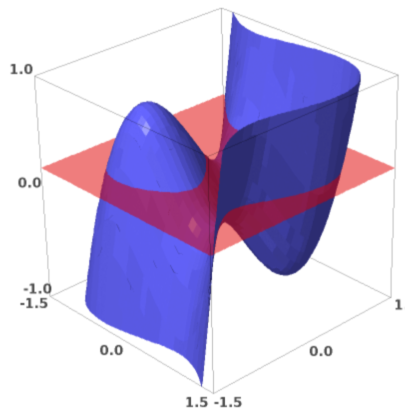
For $c = 0$ it is clear from the picture that it is not a C^1 curve. (Or note in this case the equation becomes $(x + y)(x^2 - xy + y^2 - 1) = 0$ which shows the graph is the union of a line and an ellipse.) What happens when $c = \pm 4/(3\sqrt{3})$? Here are some pictures that may help with this question. In the following c_{crit} is as above and $\varepsilon = .05$.



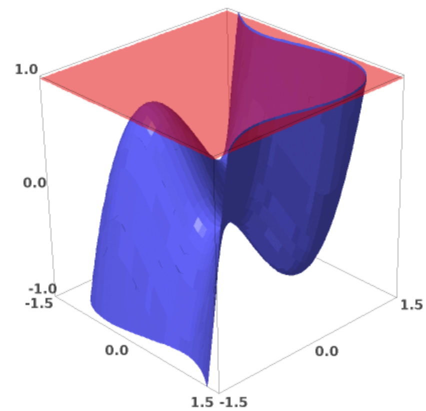
The surface $\{z = x^3 + y^3 - x - y\}$.



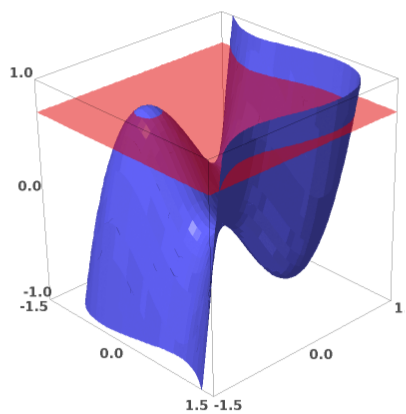
The surface and $\{z = 0\}$.



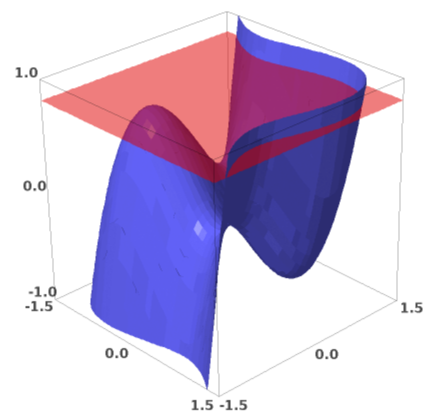
The surface and $\{z = 1/5\}$.



The surface and $\{z = 1\}$.



The surface and $\{z = c_{\text{crit}} - \varepsilon\}$.



The surface and $\{z = c_{\text{crit}} + \varepsilon\}$.

What is the relation between the plane $\{z = c_{\text{crit}}\}$ and the surface $\{z = x^3 + y^3 - x - y\}$? $\square$

Problem 2.24 (Lagrange Multipliers in the plane). Here is anther application to making a subject of vector calculus rigorous, in this case Lagrange multipliers. Let W be an open subset of $\mathbb{R}^2$ and let $C := \{(x, y) : g(x, y) = c\}$ and assume that $\nabla g(x, y) \neq 0$ for all $(x, y) \in C$. Then C is a C^1 curve in $\mathbb{R}^2$. Let $f\colon W \to \mathbb{R}$ be a C^1 function and that the restriction $f\big|_C$ has a local extrema (that is a local maximum or minimum) at (x_0, y_0). Show there there is a scalar λ such that

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0).$$

Hint: As $\nabla g \neq 0$ on C by we conclude by Theorem 2.22 that C is a C^1 curve in $\mathbb{R}^2$. Thus near (x_0, y_0) C is either a C^1 graph over the x-axis or a C^1 graph over the y-axis. Assume that it is a graph over the x-axis, the proof in the case of being a graph over the y-axis being almost identical.

By Theorem 2.20 there is a neighborhood U of x_0 and a neighborhood V of y_0 and a C^1 function $\varphi\colon U \to V$ such that near (x_0, y_0) the set C is just $\{(x, \varphi(x)) : x \in U\}$. As the restriction $f\big|_C$ has an extrema at (x_0, y_0), to be concrete assume it has a local maximum, the function $x \mapsto f(x, \varphi(x))$ has a local maximum at $x = x_0$. Thus its derivative is zero. That is

$$\begin{aligned} \frac{d}{dx} f(x, \varphi(x))\bigg|_{x=x_0} &= f_x(x_0, \varphi(x_0)) + f_y(x_0, \varphi(x_0))\varphi'(x_0) \\ &= f_x(x_0, y_0) + f_y(x_0, y_0)\varphi'(x_0) \\ &= 0 \end{aligned}$$

By Corollary 2.21

$$\varphi'(x_0) = -\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}$$

Using this in what we have just done and clearing of fractions gives

$$f_x(x_0, y_0)g_y(x_0, y_0) - f_y(x_0, y_0)g_x(x_0, y_0) = 0.$$

Which shows the vector $\nabla f(x_0, y_0)$ is orthogonal to the vector

$$\begin{bmatrix} g_y(x_0, y_0) \\ -g_x(x_0, y_0) \end{bmatrix}.$$

But $\nabla g(x_0, y_0)$ is also orthogonal to this vector. Show that this implies that $\nabla f(x_0, y_0)$ is a scalar multiple of $\nabla g(x_0, y_0)$. □

2.7.2. *Implicit functions in three dimensions.* We first look at when an equation

$$g(x, y, z) = c$$

where c is a constant defines z as a function of x and y. As in two dimensions this will only be possible locally. The following is a exact analogue of Theorem 2.20.

Theorem 2.23. *Let W be an open set in $\mathbb{R}^2$ and $g\colon W \to \mathbb{R}$ a C^1 function. Let $(x_0, y_0, z_0) \in W$ and assume*

$$\frac{\partial g}{\partial z}(x_0, y_0, z_0) \neq 0.$$

Then there are open neighborhoods U of (x_0, y_0) in $\mathbb{R}^2$ and V of Z_0 and a C^1 function $f\colon U \to V$ with

$$f(x_0, y_0) = z_0 \qquad \text{and} \qquad g(x, y, f(x, y)) = g(x_0, y_0, z_0)$$

for all $x \in U$.

Problem 2.25. Prove this along the following lines, which in turn follow Problem 2.21. As in the proof Wwe view elements of $\mathbb{R}^3$ as column vectors.

(a) Define a function $F\colon W \to \mathbb{R}^2$ by

$$F\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ g(x,y,z) \end{bmatrix}$$

and show that the derivative of F is the matrix

$$F'\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g_x(x,y,z) & g_y(x,y,z) & g_z(x,y,z) \end{bmatrix}.$$

(b) Show that the matrix $F'(x_0, y_0, z_0)$ has an inverse.

(c) Use the Inverse Function Theorem to find a neighborhood, N, of $\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$ and a neighborhood, M, of $F\left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}\right) = \begin{bmatrix} x_0 \\ y_0 \\ g(x_0,y_0,z_0) \end{bmatrix}$ and a C^1 map $G\colon M \to N$ that is the inverse of the restriction $F\big|_M$.

(d) Let G be given by

$$G\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} \xi(x,y,z) \\ \eta(x,y,z) \\ \zeta(x,y,z) \end{bmatrix}.$$

Use that

$$F \circ G\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

to show

$$\xi(x,y,z) = x \qquad \text{and} \qquad \eta(x,y,z) = y, \qquad g(x,y,\zeta(x,y,z)) = z.$$

(e) Thus if $f(x,y) = \zeta(x,y,z_0)$ we have $g(x,y,f(x,y)) = z_0$ for all x in the domain of f. Now mess around with the details to finish the proof. That is find the neighborhoods U and V etc. □

Corollary 2.24. *Let $f(x,y)$ be as in Theorem 2.23. Then the partial derivatives of f are given by*

$$\frac{\partial f}{\partial x}(x,y) = -\frac{g_x(x,y,f(x,y))}{g_z(x,y,f(x,y))}, \qquad \frac{\partial f}{\partial y}(x,y) = -\frac{g_y(x,y,f(x,y))}{g_z(x,y,f(x,y))}.$$

Problem 2.26. Prove this. □

Let $S \subseteq \mathbb{R}^2$ be a subset of $\mathbb{R}^3$ and let $(x_0, y_0, z_0) \in S$ a point of S. The S is ***locally a C^1 graph over the x-y plane at*** (x_0, y_0, z_0) if and only if there is a neighborhood U of (x_0, y_0) in $\mathbb{R}^2$ and a neighborhood V of z_0 in $\mathbb{R}$ and a C^1 function $f\colon U \to V$ such that

$$S \cap (U \times V) = \{(x,y,f(x,y)) : (x,y) \in U\}.$$

There are analogous definitions of ***locally a*** C^1 ***graph over the*** x-z ***plane at*** (x_0, y_0, z_0) and ***locally a*** C^1 ***graph over the*** y-z ***plane at*** (x_0, y_0, z_0). A subset $S \subseteq \mathbb{R}^3$ is a C^1 ***surface in*** $\mathbb{R}^3$ if and only if at each of its points it is locally a graph over at least one of the coordinate axises. Then Theorem 2.23 implies

Theorem 2.25. *If W is an open subset of $\mathbb{R}^3$ and $g\colon W \to \mathbb{R}$ is a C^1 and $c \in \mathbb{R}$ a constant such that for all (x, y, z) with $g(x, y, z) = c$*

$$\nabla g\,(x, y, z) := \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

then S is a C^1 surface of $\mathbb{R}^3$. □

As a familiar example let

$$g(x, y, z) = x^2 + y^2 - z^2.$$

Then the gradient is

$$\nabla g(x, y, z) = \begin{bmatrix} 2x \\ 2y \\ -2z \end{bmatrix}$$

and the only point where this vanishes is the origin $(0, 0, 0)$. Therefore for all $c \neq 0$ the level set $\{g(x, y, z) = 0\}$ is a C^1 surface in $\mathbb{R}^3$. Figure 2 shows some of the level sets.

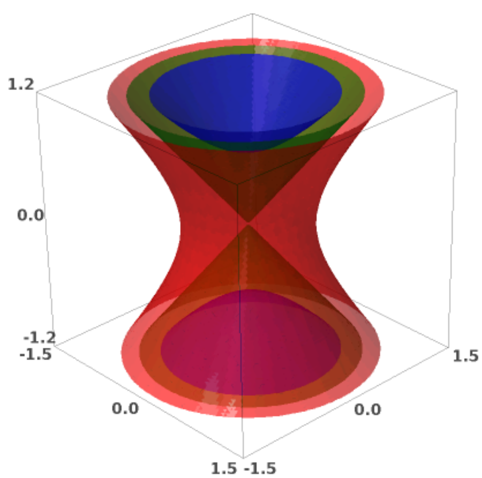


FIGURE 2. The level sets $\{x^2 + y^2 - z^2 = c\}$ for the values $c = -1/2$ (blue), $c = 0$ (green), and $c = 1/2$ (red). Note for $c = 0$ this is a double cone and not a C^1 surface.

Problem 2.27 (Lagrange multipliers on surfaces). Let W be an open set in $\mathbb{R}^3$ and let $g\colon W \to \mathbb{R}$ be a C^1 function. Let c be a constant such that $\nabla g(x, y, z) \neq 0$ at all points of $S := \{(x, y, z) : g(x, y, z) = c\}$. Let $f\colon W \to \mathbb{R}$ be a C^1 function such that the restriction $f\big|_S$ has a local extrema at $(x_0, y_0, z_0) \in S$. Show there is a scalar λ such that

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0).$$

Hint: Proceed as in Problem 2.24 and assume that S is locally a graph over the x-y plane near (x_0, y_0, z_0), say S is of the form $z = \varphi(x, y)$. Use that $(x, y) \mapsto f(x, y, \varphi(x, y))$ has a local extrema at (x_0, y_0) to show

$$f_x g_z - f_z g_x = 0 \qquad \text{and} \qquad f_y g_z - f_z g_y = 0$$

at the point (x_0, y_0, z_0). That is $\nabla f(x_0, y_0, z_0)$ is orthogonal to the two vectors

$$\begin{bmatrix} g_z \\ 0 \\ -g_x \end{bmatrix}, \qquad \begin{bmatrix} 0 \\ g_z \\ -g_y \end{bmatrix}$$

evaluated at (x_0, y_0, z_0). But $\nabla f(x_0, y_0, z_0)$ is also orthogonal to these two vectors. □

In three dimensions there is anther possible for defining a function implicit. We can have two equations

$$\begin{aligned} g(x, y, z) &= c_1 \\ h(x, y, z) &= c_2 \end{aligned}$$

and solve for two of the variables in terms of the remaining one. As motivation first consider the case that the functions are linear so that the system of equations becomes

$$\begin{aligned} g(x, y, z) &= a_1 x + a_2 y + a_3 z = c_1 \\ h(x, y, z) &= b_1 x + b_2 y + b_3 z = c_2 \end{aligned}$$

and we wish to solve for y and z in terms of x. Rewrite this as

$$\begin{aligned} a_2 y + a_3 z &= c_1 - a_1 x \\ b_2 y + b_3 z &= c_2 - a_2 x. \end{aligned}$$

We solve this system when

$$\det \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \neq 0.$$

This is equivalent to the two vectors

$$\begin{bmatrix} a_2 \\ b_2 \end{bmatrix}, \qquad \begin{bmatrix} a_3 \\ b_3 \end{bmatrix}$$

being linearly independent. In the general case get the

$$\begin{aligned} 0 = g(x, y, z) &\approx g(*) + g_x(*)(x - x_0) + g_y(*)(y - y_0) + g_z(*)(z - z_0) \\ 0 = h(x, y, z) &\approx h(*) + h_x(*)(x - x_0) + h_y(*)(y - y_0) + h_z(*)(z - z_0) \end{aligned}$$

where $*$ is the point (x_0, y_0, z_0). This approximating system has a solution when $g_y(*)h_z(*) - g_z(*)h_y(*) \neq 0$.

Theorem 2.26. *Let W be an open set in $\mathbb{R}^2$ and let $g, h\colon W \to \mathbb{R}$ be two C^1 functions. Let $(x_0, y_0, z_0) \in W$ such that the two vectors*

$$\begin{bmatrix} g_y(x_0,y_0,z_0) \\ g_z(x_0,y_0,z_0) \end{bmatrix}, \qquad \begin{bmatrix} h_y(x_0,y_0,z_0) \\ h_z(x_0,y_0,z_0) \end{bmatrix}$$

are linearly independent. (This is often stated as saying $\det \begin{bmatrix} g_y & h_y \\ g_z & h_z \end{bmatrix} \neq 0$ *at the point (x_0, y_0, z_0).) are linearly independent. The there is a neighborhood, U, of x_0 in $\mathbb{R}$ and a neighborhood, V, of (y_0, z_0) in $\mathbb{R}^2$ and a C^1 function $f\colon U \to V$ with $f(x_0) = (y_0, z_0)$ and*

$$\begin{aligned} g(x, f(x)) &= g(x_0, y_0, z_0) \\ h(x, f(x)) &= h(x_0, y_0, z_0) \end{aligned}$$

for all $x \in U$.

Problem 2.28. Prove this along the following lines.

(a) Define a function $F\colon W \to \mathbb{R}^3$ by

$$F\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ g(x,y,z) \\ h(x,y,z) \end{bmatrix}$$

and show the derivative is

$$F'\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ g_x(x,y,z) & g_y(x,y,z) & g_z(x,y,z) \\ h_x(x,y,z) & h_y(x,y,z) & h_z(x,y,z) \end{bmatrix}.$$

(b) Show that $F'\left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}\right)$ is invertible and therefore by the Inverse Function Theorem there are neighborhoods N of (x_0, y_0, z_0) and M of $F(x_0, y_0, z_0)$ and a C^1 function $G\colon M \to N$ that is the inverse of the restriction $F\big|_M$.

(c) Let G be given by

$$G\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} \xi(x,y,z) \\ \eta(x,y,z) \\ \zeta(x,y,z) \end{bmatrix}.$$

Use that

$$F \circ G\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

to show

$$\begin{aligned} \xi(x,y,z) &= x \\ g(x,\eta(x,y,z),\zeta(x,y,z)) &= y \\ h(x,\eta(x,y,z),\zeta(x,y,z)) &= z \end{aligned}$$

and therefore if

$$f(x) = (\eta(x,y_0,z_0),\zeta(x,y_0,z_0))$$

we have

$$\begin{aligned} g(x,f(x)) &= g(\eta(x,y_0,z_0),\zeta(x,y_0,z_0)) = g(x_0,y_0,z_0), \\ h(x,f(x)) &= h(\eta(x,y_0,z_0),\zeta(x,y_0,z_0)) = h(x_0,y_0,z_0). \end{aligned}$$

(d) Now finish the proof. □

We now proceed with what should by now be a familiar pattern. Let $C \subseteq \mathbb{R}^3$ and $(x_0,y_0,z_0) \in C$. Then C is ***locally a graph over the x-axis near*** (x_0,y_0,z_0) if and only if there is a neighborhood U of x_0 in $\mathbb{R}$ and a neighborhood V of (y_0,z_0) in R^2 and a C^1 function $f\colon U \to V$ with $f(x_0) = (y_0,z_0)$ and

$$C \cap (U \times V) = \{(x,f(x)) : x \in U\}.$$

The definitions of C ***locally a graph over the y-axis near*** (x_0,y_0,z_0) and ***locally a graph over the z-axis near*** (x_0,y_0,z_0) are as expected. The set C is a C^1 ***curve in*** $\mathbb{R}^3$ if and only if at each of its points it is locally a graph over at least one of the coordinate axises.

Lemma 2.27. *Let*

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \qquad w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

be linearly independent vectors in $\mathbb{R}^3$. Then at least one of the pairs of vectors

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad \textit{or} \quad \begin{bmatrix} v_1 \\ v_3 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_3 \end{bmatrix} \quad \textit{or} \quad \begin{bmatrix} v_2 \\ v_3 \end{bmatrix}, \begin{bmatrix} w_2 \\ w_3 \end{bmatrix}$$

is linearly independent.

Problem 2.29. Prove this. □

Theorem 2.28. *Let W be an open set in $\mathbb{R}^3$ and $g,h\colon W \to \mathbb{R}$ be C^1 functions. Set*

$$C := \{(x,y,z) : g(x,y,z) = c_1, h(x,y,z) = c_2\}$$

where c_1 and c_2 are constant. Assume that for all $(x,y,z) \in C$ that the gradients $\nabla g(x,y,z)$ and $\nabla h(x,y,z)$ are linearly independent. Then C is a C^1 curve in $\mathbb{R}^3$.

Problem 2.30. Prove this. *Hint:* By the last lemma at least one of the pairs

$$\begin{bmatrix} g_x \\ g_y \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} g_x \\ g_z \end{bmatrix}, \begin{bmatrix} h_x \\ h_z \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} g_y \\ g_z \end{bmatrix}, \begin{bmatrix} h_y \\ h_z \end{bmatrix}$$

are linearly independent. To take one of these cases assume

$$\begin{bmatrix} g_y \\ g_z \end{bmatrix}, \begin{bmatrix} h_y \\ h_z \end{bmatrix}$$

are linearly independent and use Theorem 2.26 to conclude C is locally a graph over the x-axis. □

Problem 2.31. Show that for all $r > 0$ that the intersection of the sphere defined by $x^2+y^2+z^2 = r^2$ and the double cone defined by $x^2+y^2-z^2 = 0$ is a C^1 curve. Is it connected? Draw a picture and give a geometric description of the intersection. □

Proposition 2.29. *With the set up of Theorem 2.26 let the function* $f\colon U \to V \subseteq$ *be given by*

$$f(x) = (f_1(x), f_2(x)).$$

Then

$$f_1'(x) = -\frac{g_x h_z - g_z h_x}{g_y h_z - g_z h_y}, \qquad f_2'(x) = -\frac{g_y h_x - g_x h_y}{g_y h_z - g_z h_y}$$

where all the partial derivatives are evaluated at $(x, f_1(x), f_2(x))$.

Problem 2.32. Prove this. *Hint:* From Theorem 2.26

$$\begin{aligned} g(x, f_1(x), f_2(x)) &= g(x_0, y_0, z_0) \\ h(x, f_1(x), f_2)) &= h(x_0, y_0, z_0) \end{aligned}$$

and taking the derivatives of these with respect to x gives

$$\begin{aligned} g_x + g_y f_1'(x) + g_z f_2'(x) &= 0 \\ h_x + h_y f_1'(x) + h_z f_2'(x) &= 0 \end{aligned}$$

with the partial derivatives evaluated at $(x, f_1(x), f_2(x))$. □

Problem 2.33 (Lagrange multipliers yet again). Let $W \subset \mathbb{R}^3$ be an open set and $g, h\colon W \to \mathbb{R}$ C^1 functions. Let $c_1, c_2 \in \mathbb{R}$ and set

$$C = \{(x, y, z) : g(x, y, z) = c_1, h(x, y, z) = c_2\}.$$

Assume that for all points of C that ∇g and ∇h are linearly independent. Let $f\colon W \to \mathbb{R}$ be C^1 and assume that the restriction $f\big|_C$ has a local extrema at $(x_0, y_0, z_0) \in C$. Show there are scalars λ_1 and λ_2 such that

$$\nabla f = \lambda_1 \nabla g + \lambda_2 \nabla h$$

at the points (x_0, y_0, z_0). *Hint:* By Theorem 2.28 C is C^1 curve in $\mathbb{R}^2$ and so at the point (x_0, y_0, z_0) it is the graph over at least one of the coordinate axises. Assume it is a graph over the x-axis. Then with the notation of Problem 2.32 there are C^1 functions $\varphi_1(x)$ and $\varphi_2(x)$ defined in a neighborhood of x_0 such that

$$\begin{aligned} g(x, \varphi_1(x), \varphi_2(x)) &= c_1 \\ h(x, \varphi_1(x), \varphi_2(x)) &= c_2. \end{aligned}$$

Talking the derivative with respect to x gives

$$\begin{aligned} g_x + g_y\varphi_1'(x) + g_z\varphi_2'(x) &= 0 \\ h_x + h_y\varphi_1'(x) + h_z\varphi_2'(x) &= 0 \end{aligned}$$

which shows that both ∇g and ∇h are orthogonal to the vector

$$v(x) := \begin{bmatrix} 1 \\ \varphi_1'(x) \\ \varphi_2'(x) \end{bmatrix}.$$

As ∇g and ∇h are linearly independent they are a basis of $v(x)^\perp$. The function $x \mapsto f(x, \varphi_1(x), \varphi_2(x))$ has a local extrema at x_0 and therefore

$$\frac{d}{dx} f(x, \varphi_1(x), \varphi_2(x))\bigg|_{x=x_0} = f_x + f_y\varphi_1'(x_0) + f_z\varphi_2'(x_0) = 0$$

with the partial derivatives evaluated at $(x_0, \varphi_1(x_0), \varphi_2(x_0)) = (x_0, y_0, z_0)$. Thus $\nabla f(x_0, y_0, z_0)$ is orthogonal to $v(x_0)$. $\square$

2.8. **The General Finite Dimensional Implicit Function Theorem.** Let m and n be positive integers. Then we write elements in the product space $\mathbb{R}^m \times \mathbb{R}^n$ as (x, y) where $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. We will be considering functions $g \colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ and for a constant vector $c \in \mathbb{R}^n$ trying to solve

$$g(x, y) = c$$

for y as a function of x. To get motivate some of the hypothesis of the theorem we write this in coordinates

$$g(x^1, \dots, x^m, y^1, \dots, y^n) = \begin{bmatrix} g^1(x^1, \dots, x^m, y^1, \dots, y^n) \\ g^2(x^1, \dots, x^m, y^1, \dots, y^n) \\ \vdots \\ g^n(x^1, \dots, x^m, y^1, \dots, y^n) \end{bmatrix}$$

the vector equation $g(x, y) = c$ is the same as the system

$$\begin{aligned} g^1(x^1, \dots, x^m, y^1, \dots, y^n) &= c^1 \\ g^2(x^1, \dots, x^m, y^1, \dots, y^n) &= c^2 \\ \vdots \qquad\qquad & \quad \vdots \\ g^n(x^1, \dots, x^m, y^1, \dots, y^n) &= c^n \end{aligned}$$

of n equation in $m+n$ unknowns. If these equations were linear, say

$$\begin{aligned}
g^1 &= a^1_1x^1 + a^1_2x^2 + \cdots + a^1_mx^m + b^1_1y^1 + b^1_2y^2 + \cdots + b^1_ny^n = c^1\\
g^2 &= a^2_1x^1 + a^2_2x^2 + \cdots + a^2_mx^m + b^2_1y^2 + b^2_2y^2 + \cdots + b^2_ny^n = c^2\\
&\qquad\qquad\qquad\qquad\vdots \qquad\qquad\qquad\qquad\qquad \vdots\\
g^n &= a^n_1x^1 + a^n_2x^2 + \cdots + a^n_mx^m + b^n_1y^n + b^n_2y^n + \cdots + b^n_ny^n = c^n
\end{aligned}$$

where the a^i_j's and b^ℓ_k's are constants. Rewriting as

$$\begin{aligned}
b^1_1y^1 + b^1_2y^2 + \cdots + b^1_ny^n &= c^1 - a^1_1x^1 - a^1_2x^2 - \cdots - a^1_mx^m\\
b^2_1y^2 + b^2_2y^2 + \cdots + b^2_ny^n &= c^2 - a^2_1x^1 - a^2_2x^2 - \cdots - a^2_mx^m\\
\vdots \qquad\qquad &\qquad\qquad \vdots\\
b^n_1y^n + b^n_2y^n + \cdots + b^n_ny^n &= c^n - a^n_1x^1 - a^n_2x^2 - \cdots - a^n_mx^m
\end{aligned}$$

which is now a linear system of n equations in the n unknowns $y^1, \ldots, y^n$. This will have a unique solution when the matrix

$$A := \begin{bmatrix} a^1_1 & a^1_2 & \cdots & a^1_n\\ a^2_1 & a^1_2 & \cdots & a^1_n\\ \vdots & \vdots & \ddots & \vdots\\ a^n_1 & a^n_2 & \cdots & a^n_n \end{bmatrix}$$

is invertible. In this linear case note

$$a^j_i = \frac{\partial g^j}{\partial y^i}$$

and therefore in this linear case the condition for solubility is that this matrix of partial derivatives is invertible.

Theorem 2.30. *Let m and n positive integers and $W \subseteq \mathbb{R}^m \times \mathbb{R}^n$ be an open set. Let $g\colon W \to \mathbb{R}^m$ be a C^1 function and assume that at the point $(a, b) \in W$ that the matrix*

$$\begin{bmatrix}
\dfrac{\partial g^1}{\partial y^1}(a,b) & \dfrac{\partial g^1}{\partial y^2}(a,b) & \cdots & \dfrac{\partial g^1}{\partial y^n}(a,b)\\[2ex]
\dfrac{\partial g^2}{\partial y^1}(a,b) & \dfrac{\partial g^2}{\partial y^2}(a,b) & \cdots & \dfrac{\partial g^2}{\partial y^n}(a,b)\\
\vdots & \vdots & \ddots & \vdots\\
\dfrac{\partial g^n}{\partial y^1}(a,b) & \dfrac{\partial g^n}{\partial y^2}(a,b) & \cdots & \dfrac{\partial g^n}{\partial y^n}(a,b)
\end{bmatrix}$$

is invertible. Then there is a neighborhood U of a in $\mathbb{R}^m$ and a neighborhood V of b in $\mathbb{R}^n$ and a C^1 function $f\colon U \to V$ with $f(a) = b$ and

$$g(x, f(x)) = g(a, b)$$

for all $x \in U$.

Problem 2.34. Prove this. *Hint:* Define a function $F \to W \to \mathbb{R}^m \times \mathbb{R}^n$ by

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ g(x,y) \end{bmatrix}.$$

Show the derivative of this is

$$F'\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} I & 0 \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{bmatrix}$$

where I is the $m \times m$ identity matrix, 0 is the $m \times n$ zero matrix, $\frac{\partial g}{\partial x}$ is the $n \times m$ matrix

$$\frac{\partial g}{\partial x} = \begin{bmatrix} \dfrac{\partial g^1}{\partial x^1} & \dfrac{\partial g^1}{\partial x^2} & \cdots & \dfrac{\partial g^1}{\partial x^m} \\ \dfrac{\partial g^2}{\partial x^1} & \dfrac{\partial g^2}{\partial x^2} & \cdots & \dfrac{\partial g^2}{\partial x^m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g^n}{\partial x^1} & \dfrac{\partial g^n}{\partial x^2} & \cdots & \dfrac{\partial g^n}{\partial x^m} \end{bmatrix},$$

and $\frac{\partial g}{\partial y}$ is the $n \times n$ matrix

$$\frac{\partial g}{\partial y} = \begin{bmatrix} \dfrac{\partial g^1}{\partial y^1} & \dfrac{\partial g^1}{\partial y^2} & \cdots & \dfrac{\partial g^1}{\partial y^n} \\ \dfrac{\partial g^2}{\partial y^1} & \dfrac{\partial g^2}{\partial y^2} & \cdots & \dfrac{\partial g^2}{\partial y^n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g^n}{\partial y^1} & \dfrac{\partial g^n}{\partial y^2} & \cdots & \dfrac{\partial g^n}{\partial y^n} \end{bmatrix}$$

and all these partial derivatives are evaluate at (x, y).

Show that the matrix $F'(a, b)$ is invertible and now look at Problems 2.21, 2.25, and 2.28 for hints on how to finish. □

2.9. **Higher Derivatives for Inverse and Implicit Functions.** In the inverse function theorem we assumed that the map we are trying to invert was C^1 and found a C^1 inverse. Let W be an open set in $\mathbb{R}^m$ and $f\colon W \to \mathbb{R}$. Then for any positive integer k the function f is a C^k ***function*** if and only if all the partial derivatives of f up to order k exist and are continuous. If f is C^k for all k we say that f is C^∞. More generally if $f\colon W \to \mathbb{R}^n$, the f is a C^k function if and only if all the component functions of f are C^k.

Proposition 2.31. *The sums, products, quotients (where the denominator does not vanish) and compositions of C^k functions are C^k.*

Problem 2.35. Prove this. □

Problem 2.36. Let f be a function such that all the first partial derivatives of f are C^k. Show that f is then C^{k+1}. $\square$

It is both natural and useful to ask if f is C^k then can we expect the inverse of f to also be C^k. This is true and turns out to be relatively straightforward to show because there is an explicit formula for the inverse of a matrix. We first look at some low dimensional examples.

Problem 2.37. Let $f\colon (a,b) \to \mathbb{R}$ be a C^k for some $k \geq 1$ function assume assume that $f[(a,b)] = (\alpha, \beta)$ and that $g\colon (\alpha,\beta) \to (a,b)$ is a C^1 inverse of f. Show that g is also C^k. *Hint:* We already know from the inverse function theorem that g is C^1 and that the derivative of g is given by

$$g'(y) = \frac{1}{f'(g(y))}. \tag{3}$$

Assume that f is C^2. Then the function f' is C^1 and therefore

$$y \mapsto \frac{1}{f'(g(y))}$$

is a composition of C^1 functions and therefore (3) shows that the derivative g' is a C^1 function, which implies that g is C^2.

Assume that f is C^3. Then f' is a C^2 function and we have just seem that g is a C^2 function. Therefore $y \mapsto \frac{1}{f'(g(y))}$ is a composition of C^2 function and therefore (3) shows that the derivative g' is C^2 which shows that g is C^3.

Now it show be easy to see that there is an induction that proves the general case. $\square$

Problem 2.38. Let U be an open subset of $\mathbb{R}^2$ and $f\colon U \to V \subseteq \mathbb{R}^2$ a C^k map such that there is a C^1 inverse $g\colon V \to U$ to f. Then show g is also C^k. *Hint:* Let f be given by

$$f(x^1,x^2) = \begin{bmatrix} f^1(x^1,x^2) \\ f^2(x^1,x^2) \end{bmatrix}$$

Then derivative of f is

$$f'(x^1,x^2) = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^1} & \dfrac{\partial f^1}{\partial x^2} \\[2ex] \dfrac{\partial f^2}{\partial x^1} & \dfrac{\partial f^2}{\partial x^2} \end{bmatrix}.$$

The inverse of this is

$$f'(x^1,x^2)^{-1} = \frac{1}{f^1_{x^1} f^2_{x^2} - f^1_{x^2} f^2_{x^1}} \begin{bmatrix} f^2_{x^2} & -f^1_{x^2} \\[1ex] -f^2_{x^1} & f^1_{x^1} \end{bmatrix}$$

Write g as

$$g(y^1,y^2) = \begin{bmatrix} g^1(y^1,y^2) \\ g^2(y^1,y^2) \end{bmatrix}$$

Then the derivative of g is

$$g'(y^1, y^2) = \begin{bmatrix} \dfrac{\partial g^1}{\partial y^1} & \dfrac{\partial g^1}{\partial y^2} \\ \\ \dfrac{\partial g^2}{\partial y^1} & \dfrac{\partial g^2}{\partial y^2} \end{bmatrix}$$

From the Inverse Function Theorem we know

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

Writing this in matrix form and equating the components gives the four equations

$$\begin{aligned}
g^1_{y^1}(y^1, y^2) &= \frac{f^2_{x^2}(g(y^1, y^2))}{f^1_{x^1}(g(y^1, y^2)f^2_{x^2}(g(y^1, y^2)) - f^1_{x^2}(g(y^1, y^2))f^2_{x^1}(g(y^1, y^2))} \\
g^2_{y_1}(y^1, y^2) &= \frac{-f^1_{x^2}(g(y^1, y^2))}{f^1_{x^1}(g(y^1, y^2)f^2_{x^2}(g(y^1, y^2)) - f^1_{x^2}(g(y^1, y^2))f^2_{x^1}(g(y^1, y^2))} \\
g^1_{y^2}(y^1, y^2) &= \frac{-f^2_{x^1}(g(y^1, y^2))}{f^1_{x^1}(g(y^1, y^2)f^2_{x^2}(g(y^1, y^2)) - f^1_{x^2}(g(y^1, y^2))f^2_{x^1}(g(y^1, y^2))} \\
g^2_{y^2}(y^1, y^2) &= \frac{f^1_{x^1}(g(y^1, y^2))}{f^1_{x^1}(g(y^1, y^2)f^2_{x^2}(g(y^1, y^2)) - f^1_{x^2}(g(y^1, y^2))f^2_{x^1}(g(y^1, y^2))}
\end{aligned}$$

Assume that f is C^2. Then all the first partial derivatives of f are C^1 and by the Inverse function theorem g is C^1. The formulas just given then show that all the first partial derivatives of g are C^1. Thus g is C^2.

If f is C^2 all the first partial derivatives of f are C^2 and we have just shown that g is C^2. Therefore our formulas for the first partial derivatives of g show that the first partials of g are compositions of C^2 functions and thus these first partials are C^2. Thus g is C^3.

Now do an induction to complete the proof. $\square$

Based on this arguments we should be able to prove the result in $\mathbb{R}^n$ as long as we have a formula for the inverse of a matrix analogous to the one we have in two dimensions. There is such a formula which we now give. Let

$$A = [a^j_i] = \begin{bmatrix} a^1_1 & a^1_2 & a^1_3 & \cdots & a^1_n \\ a^2_1 & a^2_2 & a^2_3 & \cdots & a^2_n \\ a^3_1 & a^3_2 & a^3_3 & \cdots & a^3_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a^n_1 & a^n_2 & a^n_3 & \cdots & a^n_n \end{bmatrix}$$

be an $n \times n$ real matrix. (Here we with the convention that in a matrix $A = [a^i_j]$ that the upper index is row and lower index is the column of the entry.) For $i, j \in \{1, 2, \ldots, n\}$ let $A[i/j]$ be the $(n-1)\times(n-1)$ matrix obtained by crossing on the i-th row and the j-th column. This $(n-1)\times(n-1)$ matrix is called the ij-th ***minor*** of A. If

$$A = \begin{bmatrix} a^1_1 & a^1_2 & a^1_3 \\ a^2_1 & a^2_2 & a^2_3 \\ a^3_1 & a^3_2 & a^3_3 \end{bmatrix}$$

then, using the notation $\cancel{a}^\ell_k$ for indicating that we are deleting the element a^ℓ_k, we have:

$$A[1/1] = \begin{bmatrix} \cancel{a}^1_1 & \cancel{a}^1_2 & \cancel{a}^1_3 \\ \cancel{a}^2_1 & a^2_2 & a^2_3 \\ \cancel{a}^3_1 & a^3_2 & a^3_3 \end{bmatrix} = \begin{bmatrix} a^2_2 & a^2_3 \\ a^3_2 & a^3_3 \end{bmatrix},$$

$$A[3/2] = \begin{bmatrix} a^1_1 & \cancel{a}^1_2 & a^3_1 \\ a^2_1 & \cancel{a}^2_2 & a^3_2 \\ \cancel{a}^3_1 & \cancel{a}^3_2 & \cancel{a}^3_3 \end{bmatrix} = \begin{bmatrix} a^1_1 & a^1_3 \\ a^2_1 & a^2_3 \end{bmatrix}$$

and if

$$A = \begin{bmatrix} a^1_1 & a^1_2 & a^1_3 & a^1_4 \\ a^2_1 & a^2_2 & a^2_3 & a^2_4 \\ a^3_1 & a^3_2 & a^3_3 & a^3_4 \\ a^4_1 & a^4_2 & a^4_3 & a^4_4 \end{bmatrix}$$

then

$$A[2/3] = \begin{bmatrix} a^1_1 & a^1_2 & \cancel{a}^1_3 & a^1_4 \\ \cancel{a}^2_1 & \cancel{a}^2_2 & \cancel{a}^2_3 & \cancel{a}^2_4 \\ a^3_1 & a^3_2 & \cancel{a}^3_3 & a^3_4 \\ a^4_1 & a^4_2 & \cancel{a}^4_3 & a^4_4 \end{bmatrix} = \begin{bmatrix} a^1_1 & a^1_2 & a^1_4 \\ a^3_1 & a^3_2 & a^3_4 \\ a^4_1 & a^4_2 & a^4_4 \end{bmatrix}.$$

If $A = [a^i_j]$ is $n \times n$ the ***c***lassical adjoint is the $n \times n$ matrix $\operatorname{adj}(A)$ with elements

$$\operatorname{adj}(A)^i_j = (-1)^{i+j} \det(A[j/i]).$$

Note the interchange of order of i and j so that this is the transpose of the matrix $[(-1)^{i+j}\det_{n-1}(A[i/j])]$. In less compact notation if

$$A = \begin{bmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1^n & a_2^n & \cdots & a_n^n \end{bmatrix}$$

then

$$\operatorname{adj}(A) = \begin{bmatrix} +\det(A[1/1]) & -\det(A[2/1]) & +\det(A[3/1]) & -\det(A[4/1]) & \cdots \\ -\det(A[1/2]) & +\det(A[2/2]) & -\det(A[3/2]) & +\det(A[4/2]) & \cdots \\ +\det(A[1/3]) & -\det(A[2/3]) & +\det(A[3/3]) & -\det(A[4/3]) & \cdots \\ -\det(A[1/4]) & +\det(A[2/4]) & -\det(A[3/4]) & +\det(A[4/4]) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The classical adjoint is related to finding the inverse of a matrix by the following.

Theorem 2.32. *Then for any $n \times n$ matrix A we have*

$$\operatorname{adj}(A)A = A\operatorname{adj}(A) = \det(A)I$$

where I is the identity matrix. Thus if $\det(A) \neq 0$ the inverse of A is given by

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A).$$

Proof. Letting $A = [a_j^1]$, the entries of $A\operatorname{adj}(A)$ are

$$\begin{aligned}(A\operatorname{adj}(A))_k^i &= \sum_{j=1}^n a_j^i \operatorname{adj}(A)_k^j \\ &= \sum_{j=1}^n (-1)^{j+k} a_j^i \det(A[k/j]).\end{aligned}$$

If we let $k = i$ in this and use then sum is just the expansion for $\det(A)$ along the i-th row and therefore

$$(A\operatorname{adj}(A))_i^i = \sum_{j=1}^n (-1)^{j+i} a_j^i \det(A[i/j]) = \det(A).$$

If $k \neq i$ then let $B = [b_j^i]$ have all its rows the same as the rows of A, except that the k-th row is replaced by the i-the row of A (thus A and B only differ along the k-the row). Then B has two rows the same and therefore

$\det(B) = 0$. Now for all j that $B[k/j] = A[k/j]$ as A and only differ in the k-th row and $A[k/j]$ and $B[k/j]$ only involve elements of A and B not on the k-row. Also from the definition of B we have $b^k_j = a^i_j$ (as the k-th row of B is the same as the i-row of A). Therefore we can compute $\det_n(B)$ by expanding along the k row

$$\begin{aligned} 0 = \det(B) &= \sum_{j=1}^{n} (-1)^{j+k} b^k_j \det(B[k/j]) \\ &= \sum_{j=1}^{n} (-1)^{j+k} a^i_j \det(A[k/j]) \\ &= (A \operatorname{adj}(A))^i_k. \end{aligned}$$

These calculations can be summarized as

$$(A \operatorname{adj}(A))^i_j = \det(A)\delta^i_k$$

where δ^i_k is given by

$$\delta^i_k = \begin{cases} 1, & i = k; \\ 0, & i \neq k. \end{cases}$$

But δ^i_k are the elements of the identity matrix. Thus this implies $A \operatorname{adj}(A) = \det_n(A)I$.

A similar computation (but working with columns rather than rows) implies that $\operatorname{adj}(A)A = \det_n(A)I_n$. □

Theorem 2.33. *Let W be an open subset of $\mathbb{R}^n$ and $f\colon W \to \mathbb{R}^n$ be a C^k function where $k \geq 1$. Assume that for some $x_0 \in W$ that $f'(x_0)$ is invertible. Let U, V, and g be as in the Inverse Function Theorem 2.15. Then the local inverse, g, to f is also C^k.*

Proof. Form Theorem 2.15 we have that for y in the domain of g that

$$g'(y) = f'(g(y))^{-1} = \frac{1}{\det f'(g(y))} \operatorname{adj}(f'(g(y))).$$

In terms of components this is

$$\frac{\partial g^i}{\partial y^j} = \frac{1}{\det f'(g(y))} \operatorname{adj}(f'(g(y)))^i_j \tag{4}$$

As $\frac{1}{\det f'(g(y))} \operatorname{adj}(f'(g(y)))^i_j$ is a rational function of the elements of $f'(g(y))$ this function has a many continuous partial derivatives as the components of $f'(g(y))$ have.

Now assume that f is C^2. Then all the components of $f'(x)$ are C^1. Also g is C^1. The functions $y \mapsto \frac{1}{\det f'(g(y))} \operatorname{adj}(f'(g(y)))^i_j$ are compositions of C^1 functions and therefore themselves C^1 functions. Now (4) shows that all the first partial derivatives of g^i are C^1. But if all the first partials of g^i are C^2, then g^i is C^2. This works for all the components g^i of g and therefore g is C^2.

Now assume that f is C^3. Then we have just seen that g is C^2. But then (4) shows that all the first partial derivatives of g^i are C^2 and therefore g is C^3.

Continuing in this manner, or more formally using induction, we see that if f is C^k, then so is g. □

Theorem 2.34. *With the set up the Implicit Function Theorem 2.30 if the function g is C^k for some $k \geq 1$, then the function f in the that theorem is also C^k.*

Problem 2.39. Prove this. *Hint:* Just trace through the proof and show this follows from the regularity result, 2.33, for inverse functions. □

Problem 2.40 (Curvature of level sets in the plane.). Recall that if C is a C^2 curve in the plane then at points where is is locally a graph over the x axis, say $y = f(x)$, then the ***curvature*** of C (up to a plus or minus sign) is

$$\kappa = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}}$$

with a similar at the points where C is locally a graph over the y axis. One way to give a C^2 curve is as the level set of a C^1 function. Let W be an open set in the plane, $g\colon W \to \mathbb{R}$ a C^2 function and for some constant c let $C := \{(x, y) : g(x, y) = c\}$ be a level set of g such that for all $(x, y) \in C$ we have $\nabla g(x, y) \neq (0, 0)$. Then by the implicit function theorem at each of its points C is locally a C^2 graph over either the x-axis or the y-axis. Show that, up to a sign, that the curvature of C is given by

$$\kappa = \frac{g_{yy}g_x^2 - 2g_{xy}g_xg_y + g_{xx}g_y^2}{(g_x^2 + g_y^2)^{3/2}}.$$

Hint: Only worry about the points where C is locally a graph over the x-axis, say it is locally the graph of $y = f(x)$. Then

$$g(x, f(x)) = c$$

for all x in the domain of f. Take the derivative of this to show (as we did in Proposition 2.32) to show that

$$f(x) = -\frac{g_x(x, f(x))}{g_y(x, f(x))}.$$

Take the derivative of this to find that

$$f''(x) = -\frac{g_{yy}g_x^2 - 2g_{xy}g_xg_y + g_{xx}g_y^2}{g_y^3}$$

where these are all evaluated at $(x, f(x))$. □