which thus gives a finite subcover of $\{A_\alpha\}_{\alpha\in\Lambda}$. □

Explanation. It's important to know why we might even think of such proofs, so let's summarize the key ideas. The goal is to find a finite subcover. We break things down by seeing what we *can* find a finite subcover for. We know we can do closed intervals, and it is natural to hope that if we can extend this to a product of two closed intervals, then we can extend it to three, four, and so on – there's the induction.

Then, assuming we can find finite subcovers for each $(d-1)$-dimensional "slice" of our d-dimensional box, we notice that this finite subcover should cover further than just a single slice, because it is open – it has wiggle room. If this is true, we have an open cover of the interval $[a_d, b_d]$ parameterizing the slices – and compactness of this interval then means we can join together finitely many of these finite subcovers.

The meat of the proof was in guaranteeing that wiggle room. The natural thing to do was to explore how it could possibly be false – it gave us a sequence of points in the box but in the complement of B_x, wherein the last coordinate converges to x. This looks like it should contradict openness of B_x, but since the first $d-1$ coordinates could be changing a lot, we can't say anything concrete – this is where we use Bolzano-Weierstrass to give ourselves that extra control.

Moving on from this, we can establish some each properties of compact sets:

Theorem 2.13. *Let (M, d) be a metric space and $C \subseteq M$.*

1. *If C is compact then C is **bounded** – for any x there exists r so that $C \subseteq B_r(x)$.*
2. *If C is compact then C is closed.*
3. *If M is compact and C is closed then C is compact.*

Proof.

1. Let $x \in M$. Every point $y \in C$ lies in some $B_r(x)$ (since $d(x, y)$ is finite!), and $B_r(x)$ is open, so $\{B_r(x)\}_{r>0}$ is an open cover of C. Thus there is a finite subcover, $B_{r_1}(x), \ldots, B_{r_n}(x)$.

   Set $r = \max_i r_i$. Then $\cup_{i=1}^n B_{r_i}(x) \subseteq B_r(x)$, so $C \subseteq B_r(x)$.

2. If C were not closed, there would be some sequence of points x_n in C convergent to some x not in C. Note that each set $M \setminus \overline{B}_r(x)$ is open, and $C \subseteq \cup_{r>0} M \setminus \overline{B}_r(x)$, since the the only point of M not contained in some $M \setminus \overline{B}_r(x)$, that is, the set of y with $d(x, y) \geq r > 0$, is x.

   By compactness of C, this should have some finite subcover $M \setminus \overline{B}_{r_1}(x), \ldots, M \setminus \overline{B}_{r_n}(x)$. Setting $r = \min_i r_i$, this gives $C \subseteq M \setminus \overline{B}_r(x)$, so $d(x, y) \geq r > 0$ for every $y \in C$. But this contradicts that $x_n \to x$.

3. Take an open cover of C. Since C is closed, $M \setminus C$ is open, and adding this to an open cover of C gives an open cover of M. By compactness of M, this has a finite subcover, which necessarily also covers C.

□

Corollary 2.14. *A set is compact in $\mathbb{R}^d$ (with respect to the Euclidean metric) if and only if it is closed and bounded.*

Proof. The first two points in the previous theorem show that being closed and bounded is necessary. To see that it is sufficient, note that any bounded set in $\mathbb{R}^d$ is contained in some closed box (product of closed intervals), which is a compact set by our earlier theorem. Thus the third point above shows that a closed set contained in this box is compact. □

This gives a nice characterization of compact sets in $\mathbb{R}^d$, but the situation in other settings can be much more complicated. For instance, what about the compact subsets of $C([a, b])$ with the supremum norm[40]?

We should also say some more substantial applications of compactness! This, we take up here. For starters, we have the following generalization of the EVT:

Theorem 2.15 (Extreme Value Theorem)**.** *Let $f : M \to \mathbb{R}$ be a continuous function on a compact metric space. Then f is bounded[41] and attains its bounds[42].*

Proof. We have by an earlier theorem that $f(M)$ is compact, thus it will be bounded in the absolute value metric. This means, in particular, that it is bounded above and below, so has both an infimum and a supremum. We

[40] An answer to this is given by the famous Arzelà-Ascoli Theorem, which is an excellent extra topic I'd encourage you to explore after this chapter!

[41] That is, the image of f is bounded.

[42] That is, the supremum and infimum of the image are themselves in the image!

show that the supremum, which we denote by c, is in fact in the image. A similar proof works for the infimum.

Suppose that c is not in $f(M)$. Then $f(M) \subseteq \cup_{x<c}(-\infty, x)$. By compactness of $f(M)$, this open cover has a finite subcover, that is, for some collection of $x_i < c$ we have $f(M) \subseteq \cup_{i=1}^{n}(-\infty, x_i) = (-\infty, \max_i x_i)$. Thus every $y \in f(M)$ has $y < \max_i x_i$, so $\max_i x_i$ is an upper bound for $f(M)$. But $\max_i x_i < c$, contradicting c as the supremum. $\square$

Another nice property of real-valued continuous functions on closed intervals was that they were always uniformly continuous. This generalizes too! We'll need a supplementary result first.

Definition 2.16. *Let (M, d) be a metric space, $C \subseteq M$ and $\{A_\alpha\}_{\alpha\in\Lambda}$ be an open cover of C. We say $\delta > 0$ is a Lebesgue number for this cover if for every $x \in C$, there is some A_α with $B_\delta(x) \subseteq A_\alpha$.*

Lemma 2.17. *Let (M, d) be a metric space with a compact subset C, and let $\{A_\alpha\}_{\alpha\in\Lambda}$ be an open cover of C. Then this cover has a Lebesgue number.*

Proof. Every $x \in C$ is in some A_α, and thus there is some radius $0 < r \leq 1$ so that $B_r(x) \subseteq A_\alpha$. Denote by $r(x)$ the supremum of all $0 < r \leq 1$ such that $B_r(x)$ is contained in some A_α. Then $0 < r(x) \leq 1$ for every $x \in C$.

I claim that for any $\varepsilon > 0$, the set $U(\varepsilon) := \{y \in C : r(x) > \varepsilon\}$ is an open subset of the metric space C. Indeed, let $y \in U(\varepsilon)$. Set $\delta = (r(y) - \varepsilon)/3$. By the definition of the supremum $r(y)$, that is some A_α containing a ball of radius $r(y) - \delta$ centred at y. Then, for z with $d(y, z) < \delta$, we have by the triangle inequality that $B_{\varepsilon+\delta}(z) \subseteq B_{\varepsilon+2\delta}(y) = B_{r(y)-\delta}(y) \subseteq A_\alpha$. Thus $r(z) \geq \varepsilon + \delta > \varepsilon$.

We conclude that $B_\delta(y)$ is in $U(\varepsilon)$, and thus $U(\varepsilon)$ is open. Thus $\{U(\varepsilon)\}_{\varepsilon>0}$ is an open cover of C, and hence has a finite subcover[43]. In particular, by taking the minimal ε in that subcover, we see that there is some particular $\varepsilon > 0$ so that $C \subseteq U(\varepsilon)$. Thus $r(x) > \varepsilon$ for every $x \in C$. In particular, ε is a Lebesgue number for the cover, since by definition of $r(x)$ there must be a radius r at least ε so that $B_r(x) \subseteq A_\alpha$ for some α. $\square$

Theorem 2.18. *Let (M, d_1) and (N, d_2) be metric spaces with M compact. Let $f : M \to N$ be continuous. Then f is uniformly continuous.*

[43] There is a subtle point here which we have avoided addressing directly, but we'll clarify it here. Is it true that compactness, as defined via a cover of C by sets open in M, is the same as compactness defined via a cover of C by sets open in C? Yes! This is because every open subset of a metric subspace can be realized as the intersection of the subspace with an open subset of M, and also because the intersection with C of any set open in M is open in C.

Proof. Fix $\varepsilon > 0$. For every $x \in M$, $V_x := f^{-1}(B_{\varepsilon/2}(f(x)))$ is open and contains x, so the V_x form a cover of M. Thus this cover has a Lebesgue number $\delta > 0$.

Now, given any $y \in M$, we have that if $d_1(y, z) < \delta$, then $z \in B_\delta(y)$, which for some x is a set contained in V_x. By definition of V_x, this means $d_2(f(y), f(x)) < \varepsilon/2$ and $d_2(f(x), f(z)) < \varepsilon/2$. By the triangle inequality, this yields $d_2(f(y), f(z)) < \varepsilon$. $\square$

Another helpful property of compact sets in metric spaces is a generalization of the Bolzano-Weierstrass theorem.

Definition 2.19. *Let M be a metric space with $C \subseteq M$. We say that C is* ***sequentially compact*** *if every sequence in C has a subsequence converging to some point of C.*

Theorem 2.20. *Let (M, d) be a metric space and $C \subseteq M$. Then C is compact if and only if it is sequentially compact.*

Proof. ($\Rightarrow$) First, assume C is compact and x_n is a sequence of points in C. Consider the open cover of C by open balls of radius 1 centred at the points of C. By compactness of C, this has a finite subcover, thus is contained in a finite union of open balls of radius 1.

One such ball must contain infinitely many terms of the sequence x_n, say some $B_1(y)$. Let us denote by $B^{(1)}$ the closed ball $\overline{B}_1(y)$, which contains $B_1(y)$ and thus also infinitely many points of the sequence x_n.

Denote by $C^{(1)}$ the intersection of C and $B^{(1)}$. Since C is compact, it is closed, hence the intersection of these closed sets is closed. It is thus a closed subset of the compact metric space C – and so is itself compact.

Iteratively, given $C^{(i-1)}$, we find by the same argument a closed ball $B^{(i)}$ of radius $1/i$ so that the closed compact set $C^{(i)} := B^{(i)} \cap C^{(i-1)}$ contains infinitely many points of the sequence x_n.

Note that $\cap_{i=1}^{\infty} C^{(i)}$ is closed. I claim it is non-empty, for if it were empty, its complement would contain C – thus $C \subseteq \cup_{i=1}^{\infty} M \setminus C^{(i)}$ gives an open cover of C. This must have a finite subcover, where since $C^{(j)} \subseteq C^{(i)}$ for $i \leq j$, we must have that C is contained in some particular $M \setminus C^{(i)}$. Thus for this i, $C^{(i)}$ must not intersect C - but this contradicts that each contains terms from the sequence x_n, which all lie in C.

Thus there exists some x in $\cap_{i=1}^{\infty} C^{(i)}$. Now we iteratively construct a subsequence of the x_n converging to x as follows. Pick x_{n_1} to be some term of the sequence that lies in $C^{(1)}$. Iteratively, after choosing $x_{n_1}, \ldots, x_{n_j}$, pick $x_{n_{j+1}}$ to be a term of the sequence for which $n_{j+1} > \max(n_1, \ldots, n_j)$ and lying in $C^{(j+1)}$.

For each j, we have $x_{n_j} \in C^{(j)} \subseteq B^{(j)}$. The same is true for x. So $x_{n_j}, x \in B^{(j)}$. For some y_j, we have $B^{(j)} = \overline{B}_{1/j}(y_j)$, so

$$d(x, x_{n_j}) \leq d(x, y_j) + d(y_j, x_{n_j}) \leq 2/j.$$

Thus $x_{n_j} \to x$ by the Sandwich rule.

($\Leftarrow$) Now assume C is sequentially compact, and we will show that it is compact. Let $\{A_\alpha\}_{\alpha \in \Lambda}$ be some open cover of C, and suppose that it does not have a finite subcover.

Now, for every $x \in C$, we can choose a number $s(x)$ with $0 < s(x) \leq 1$ such that $B_{s(x)}(x) \subseteq A_\alpha$ for some α (since x is in some A_α, and it is open).

We now use a "greedy algorithm". Set $C_1 = C$. Find x_1 with

$$s(x_1) > \frac{1}{2} \sup_{x \in C_1} s(x),$$

and choose A_{α_1} ($\alpha_1 \in \Lambda$) such that $B_{s(x_1)}(x_1) \subseteq A_{\alpha_1}$. Now, given $x_1, \ldots, x_j$, set

$$C_{j+1} = C \setminus \cup_{i=1}^{j} B_{s(x_i)}(x_i).$$

We see that C_j is always a closed[44] subset of C.

Crucially, C_{j+1} is non-empty, since if it were empty, this would mean C is contained in $\cup_{i=1}^{j} B(x_i, s(x_i))$, and in particular, $A_{\alpha_1}, \ldots, A_{\alpha_j}$ would be a finite subcover – a contradiction.

So we can pick x_{j+1} with

$$s(x_{j+1}) > \frac{1}{2} \sup_{x \in C_{j+1}} s(x),$$

and pick $A_{\alpha_{j+1}}$ containing $B_{s(x_{j+1})}(x_{j+1})$.

Iteratively, this constructs a sequence x_j, which must therefore have a convergent subsequence $x_{j_k} \to x$ for some $x \in C$. However, since clearly $C_m \subseteq C_n$ when $m > n$, we have that x_j, and thus x_{j_k}, is eventually (i.e. after finitely many terms) contained in any given C_i. Since each C_i is closed, this means $x \in C_i$.

In particular, that $x \in C_{i+1}$ implies x is not in $B_{s(x_i)}(x_i)$, thus $d(x, x_i) \geq s(x_i)$, but $x \in C_i$ implies $s(x_i) > s(x)/2$, so $d(x, x_i) > s(x)/2 > 0$. But this violates $d(x, x_{j_k}) \to 0$, a contradiction. □

[44] C is sequentially compact, hence closed (follows easily from convergence formulation of closedness). The union of open balls is open, and subtracting a set is really intersection with its complement, which is a closed set. Thus this is an intersection of closed sets.

2.3 Equivalence of metrics and norms

Definition 2.21. *Let (M, d_1) and (N, d_2) be metric spaces and let $f : M \to N$ be a continuous bijection with continuous inverse. We say f is a* ***homeomorphism****, and (M, d_1) and (N, d_2) are* ***homeomorphic****. If $M = N$ and $f(x) = x$, we say d_1 and d_2 are* ***topologically equivalent****.*

Thus, homeomorphic metric spaces are those which are morally the same from the topological perspective[45]. Topological equivalence of metrics is a special case, in which the underlying set M is literally the same, and we aren't transforming it in any way, we are just considering different ways of measuring distances within it.

Homeomorphisms, and thus topological equivalences, preserve many features of interest. For example:

Lemma 2.22. *Let x_n be a sequence in M, let $x \in M$, and let d_1 and d_2 be topologically equivalent metrics on M. Then $x_n \to x$ with respect to d_1 if and only if $x_n \to x$ with respect to d_2.*

Proof. Apply sequential continuity to $f(x) = x$. □

We say a property of a metric space is a topological property if it is preserved by homeomorphisms. While homemorphisms send convergent sequences to convergent sequences (same proof as above!), interestingly, they do not send Cauchy sequences to Cauchy sequences!

Example: Consider $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ where the domain and codomain are both given the absolute value metric. This is a homeomorphism. However, $x_n = \pi/2 - 1/n$ is Cauchy in $(-\pi/2, \pi/2)$, but $\tan(x_n) \to \infty$, and thus $\tan(x_n)$ is not Cauchy in $\mathbb{R}$. Moreover, $(-\pi/2, \pi/2)$ is clearly not complete, but $\mathbb{R}$ is. So **completeness is not a topological property!**

Indeed, for some purposes, we wish for something stronger than topological equivalence:

Definition 2.23. *Let $f : M \to M$ be given by $f(x) = x$, and suppose the domain is given the metric d_1, and the codomain is given the metric d_2. Suppose that f is Lipschitz continuous with Lipschitz continuous inverse. Then we say that the metrics d_1 and d_2 are* ***Lipschitz equivalent****. Note that this is equivalent to the existence of constants $C_1, C_2 > 0$ with*

$$C_1 d_1(x, y) \leq d_2(x, y) \leq C_2 d_1(x, y)$$

for all $x, y \in M$ (also with the roles of d_1 and d_2 reversed).

[45] Like a donut and a coffee mug!

It is easily seen that Lipschitz equivalent metrics have the same Cauchy sequences, since distances in one metric can be used to directly bound distances in the other. This means that if a Cauchy sequence has a limit with respect to one metric, then it has the same limit in the other, and so a Cauchy sequence is convergent with respect to one metric if and only if it is convergent with respect to the other. Thus:

Lemma 2.24. *Let d_1 and d_2 be two Lipschitz equivalent metrics on a set M. Then (M, d_1) is complete if and only if (M, d_2) is complete.*

Fortunately, metrics induced by norms are well-behaved:

Proposition 2.25. *Let X be a vector space with two norms $\|\cdot\|$ and $\|\cdot\|'$. Suppose that the induced metrics are topologically equivalent. Then they are Lipschitz equivalent.*

Proof. Topological equivalence implies that $f : X \to X$ given by $f(x) = x$ is a continuous function from $(X, \|\cdot\|)$ to $(X, \|\cdot\|')$. But also, f is linear, so this continuity is necessarily Lipschitz continuity. Likewise for f^{-1}. $\square$

Remark. By translation, two norms $\|\cdot\|$ and $\|\cdot\|'$ are equivalent if there are $C_1, C_2 > 0$ with

$$C_1\|x\|' \leq \|x\| \leq C_2\|x\|'$$

for all x.

While there are many inequivalent norms on infinite-dimensional vector spaces such as $C([a, b])$, as was hinted at earlier, fortunately, the choice of norm on a finite-dimensional space does not matter!

Theorem 2.26. *Let $\|\cdot\|$ and $\|\cdot\|'$ be norms on a finite-dimensional vector space X. Then they are Lipschitz equivalent.*

Proof. Suppose X is a vector space of dimension d. Fix a basis $e_1, \ldots, e_d$ of X. Then for $x = x_1e_1 + \ldots + x_de_d$, we may define a norm $\|x\|_\infty$ by

$$\|x\|_\infty := \max_i |x_i|.$$

We will show that any other norm $\|\cdot\|$ on X is equivalent to this one, from which it follows that any two norms are equivalent[46]

[46]For clarity, note that the identity map $f : X \to X$ given by $f(x) = x$ would be continuous from $\|\cdot\|$ to $\|\cdot\|_\infty$ and from $\|\cdot\|_\infty$ to $\|\cdot\|'$, and so by composition of continuous functions (the identity composed with itself is the identity), we would have f is continuous from $\|\cdot\|$ to $\|\cdot\|'$.

We have

$$\begin{aligned}\|x\| &= \|x_1e_1 + \ldots + x_de_d\| \le |x_1| \cdot \|e_1\| + \ldots + |x_d| \cdot \|e_d\| \\ &\le (\|e_1\| + \ldots + \|e_d\|)\|x\|_\infty\end{aligned}$$

which shows that $\|x\| \le C_2\|x\|_\infty$ for $C_2 = \|e_1\| + \ldots + \|e_d\|$. Now, note that $T : \mathbb{R}^d \to X$ defined by $T(x_1, \ldots, x_d) = x_1e_1 + \ldots + x_de_d$ is a linear map, hence continuous. Hence the image of the compact set $[-1, 1]^d \setminus (-1, 1)^d$ is compact. This is exactly the set A of $x \in X$ with $\|x\|_\infty = 1$.

We have that $\|\cdot\|$ is a (Lipschitz) continuous function on X with the norm $\|\cdot\|_\infty$, since[47]

$$|\|x\| - \|y\|| \le \|x - y\| \le C_2\|x - y\|_\infty$$

It is therefore continuous on $A \subseteq X$, so by the Extreme Value Theorem, there is C_1 so that $\|x\| \ge C_1$ for $x \in A$, and also for some $x \in A$, $\|x\| = C_1$. But 0 is not in A (since $\|\cdot\|_\infty$ is a norm), so $\|x\|$ is never 0, and so $C_1 \ne 0$.

Moreover, for $x \ne 0$, we have $x/\|x\|_\infty \in A$, and so $C_1 \le \|x\|/\|x\|_\infty$. Multiplying this inequality by $\|x\|_\infty$ gives the result.

□

Corollary 2.27. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and let $A \subseteq \mathbb{R}^d$ be given the metric d induced by the norm. Then all open sets of A are the same as those induced by the Euclidean norm, A is complete with respect to d if and only if it is complete with respect to the Euclidean metric, continuity with respect to d is the same as continuity with respect to the Euclidean metric, and convergence with respect to d is equivalent to convergence in the Euclidean metric.*

Thus, when talking about these things with regards to subsets of $\mathbb{R}^d$, we may use whichever norm is most convenient.

2.4 Uniform convergence

We close out this section with a discussion of uniform convergence. Now that we know a real-valued continuous function on any compact metric space is bounded, so we can generalize the supremum norm.

[47]The first inequality is called the "reverse triangle inequality, and follows from $\|x\| \le \|x - y\| + \|y\|$ and $\|y\| \le \|x - y\| + \|x\|$.

Definition 2.28. *Let (M, d) be a compact metric space and let $C(M)$ denote the set of continuous functions $f : M \to \mathbb{R}$ (using the absolute value metric on $\mathbb{R}$). Then $C(M)$ is a (real) vector space under pointwise addition and scaling, and we define for $f \in C(M)$ the* ***supremum norm****, or* ***uniform norm****, by*

$$\|f\|_\infty := \sup_{x \in M} |f(x)| = \max_{x \in M} |f(x)|.$$

It is straightforward to confirm that this is a norm. Convergence in this norm is often called **uniform convergence**.

Theorem 2.29. *The space $(C(M), \|\cdot\|_\infty)$ is a Banach space.*

Proof. Recall that a Banach space is a normed vector space which is complete as a metric space. Thus we take some sequence of functions $f_n \in C(M)$ which is Cauchy with respect to the metric induced by the norm, and show that it converges uniformly to some $f \in C(M)$.

Step 1. We first find a candidate for what f should be. This is easy: Since f_n is Cauchy in the uniform norm, and $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty$ for any $x \in M$, it follows immediately from the definition that for each $x \in M$, $f_n(x)$ is a Cauchy sequence of real numbers, and thus converges to some real number. Call this number $f(x)$ – doing this for each $x \in M$ thus defines a function $f : M \to \mathbb{R}$.

Step 2. We note that f is bounded. Taking $\varepsilon = 1$ in the definition of Cauchy, we have that there is N so that for $n, m \geq N$, $\|f_n - f_m\|_\infty < 1$. Thus, taking $m = N$, we have $|f_n(x) - f_N(x)| < 1$ for every $x \in M$.

This is equivalent to $f_N(x) - 1 < f_n(x) < f_N(x) + 1$. Since f_N is continuous, it is bounded, so there are $a, b \in \mathbb{R}$ with $a \leq f_N(x) \leq b$, and thus $a - 1 \leq f_n(x) \leq b + 1$ for every $x \in M$. Taking the limit as $n \to \infty$ gives $a - 1 \leq f(x) \leq b + 1$.

Step 3. Since $f_n - f$ is bounded, it makes sense to take its supremum norm. We show f_n converges to f uniformly, that is, for every $\varepsilon > 0$ there is N so that $n \geq N$ implies $\|f_n - f\|_\infty < \varepsilon$. As in the previous step, there is N so that $n, m \geq N$ implies $\|f_n - f_m\|_\infty < \varepsilon/2$, which implies $f_n(x) - \varepsilon/2 < f_m(x) < f_n(x) + \varepsilon/2$, and then taking the limit in m gives us $f_n(x) - \varepsilon/2 \leq f(x) \leq f_n(x) + \varepsilon/2$. This implies $\|f_n - f\|_\infty \leq \varepsilon/2 < \varepsilon$, as desired.

Step 4. We must now confirm that in fact $f \in C(M)$. Fix $c \in M$, we will show f is continuous at c. Fix $\varepsilon > 0$. Take n to be such that $\|f_n - f\|_\infty < \varepsilon/3$. For this n, use the continuity of f_n to choose $\delta > 0$ so that $d(x, c) < \delta$ implies $|f_n(x) - f_n(c)| < \varepsilon/3$. Then $d(x, c) < \delta$ implies

$$|f(x) - f(c)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)|$$

$$\leq \|f - f_n\|_\infty + |f_n(x) - f_n(c)| + \|f_n - f\|_\infty$$
$$< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$$

as required. □

3 Part 3: Derivatives

The derivative in one real variable is the rate of change of a function. The "change" is a difference, taken in a linear space, occurring as the result of the input variable being modified by a linear increment. The natural setting to generalize to is functions between vector spaces.

The familiar expression is

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

This does not make sense in a general vector space, because we cannot divide by a vector. However, we can rearrange to an equivalent form,

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

where we are dividing by an absolute value – a size of h. This makes sense when our vectors have size, that is, in a normed vector space. Naturally, if $f: X\to Y$ is a mapping between normed vector spaces, we should replace the absolute values with norms.

We also now need to make sense of the expression $f'(x)h$. The derivative at x is to be a fixed object, acting on h and returning something like $f'(x)h$. So the derivative at x should be some function from X to Y. It should come as no surprise that the natural way we should expect a sufficiently well-behaved function between vector spaces to change as its input changes in a linear way is in fact linear[49].

We arrive at the definition of the Fréchet derivative.

3.1 The Fréchet derivative

Definition 3.1. *Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be Banach spaces, $U\subseteq X$ open, and let $f: U\to Y$ be a function. We say that f is (Fréchet)* ***differentiable*** *at $x\in U$ if there is a continuous linear map $Df_x: X\to Y$ with*

$$\frac{\|f(x+h)-f(x)-Df_xh\|_Y}{\|h\|_X}\to 0$$

as $h\to 0$. We call Df_x the ***derivative*** *of f at x.*

Lots of remarks!

[49] You *could* define derivatives in a way that allows them to be non-linear, but it would not lead to a very interesting or helpful theory

- Why Banach spaces? When limits and convergence are involved, it will lead to a far more robust theory if our spaces are complete.

- Why should U be open? Because then for every $x \in U$, some open ball $B_r(x)$ is contained in U, so that at least up to a distance r, we can move from x in any direction we like. Thus the derivative can capture a rate of change in *all* possible directions.

- Why should the derivative be continuous? Suppose we didn't require this. Consider, for instance, if f were a itself a linear map. Then we could take $Df_x = f$, since $f(x+h) - f(x) - Df_x h = f(h) - Df_x h = f(h) - f(h) = 0$. If f were a *discontinuous* linear map, we would have a function which is differentiable, but not continuous! That's not a very good definition...

- However, don't forget – every linear map between finite-dimensional vector spaces is continuous, so in $\mathbb{R}^d$ we'll never have to worry! Moreover, by the Lipschitz equivalence of norms on $\mathbb{R}^d$, the derivative is independent of the choice of norm for functions $f : \mathbb{R}^d \to \mathbb{R}^{d'}$.

- Note that this limit is over *all* directions at once! The quantity

  $$\frac{\|f(x+h) - f(x) - Df_x h\|_Y}{\|h\|_X}$$

  is assumed to be uniformly less than a given $\varepsilon > 0$, as long as we make $\|h\|_X$ sufficiently small – regardless of the direction of h. We could instead have tried to "take a derivative in each direction", but this again leads to lots of weird phenomena if we're not careful – if you're interested to see why, look into the "Gâteaux derivative".

- Returning to $\mathbb{R}$, not that all linear maps from $\mathbb{R}$ to $\mathbb{R}$ are just scaling by a fixed number, so $Df_x h = ch$ for some $c \in \mathbb{R}$. This c is of course the familiar derivative $f'(x) \in \mathbb{R}$.

- The definition being given in general Banach spaces is valuable in many applications, such as in the Calculus of Variations, where we might wish to optimize[50], for instance, some functional $T : X \to \mathbb{R}$, where X is a Banach space of functions. For instance, $T(f)$ might describe an "energy" of a physical system with data f.

[50]Some results on optimization will be given in the exercises.

We state the definition in this generality for this benefit, and prove things in this generality where it poses no extra challenge (and may even elucidate the key ideas of the proofs in ways working in coordinates might not!), but there are some results we would need a little more theory to prove in the general setting. We shall mention these as relevant.

- **A boring, but important warning:** The above definition shows the notation I'll be using for derivatives. For our purposes, this is more or less the most sensible one, but you should be aware that there are many others (there's a huge Wikipedia page on this, and it's far from complete!). Please try to stick to this one for this course, to avoid any confusion!

With the definition ready to work with, we can establish some of the standard results of multivariable calculus. Throughout this section, X, Y and Z will be used to denote Banach spaces and $\|\cdot\|_X$, $\|\cdot\|_Y$ and $\|\cdot\|_Z$ their norms.

Lemma 3.2. *Let $U \subseteq X$ be open and $f : U \to Y$ be differentiable at $x \in U$. Then f is continuous at x.*

Proof. By definition, this means that for any $\varepsilon > 0$ there is $\delta > 0$ so that $\|h\|_X < \delta$ implies

$$\frac{\|f(x+h) - f(x) - Df_x h\|_Y}{\|h\|_X} < \varepsilon.$$

Thus if $\|y - x\|_X < \delta$ we have

$$\begin{aligned}\|f(y) - f(x)\|_Y &\leq \|f(y) - f(x) - Df_x(y-x)\|_Y + \|Df_x(y-x)\|_X \\ &< \varepsilon\|y - x\|_Y + \|Df_X\|_{X\to Y}\|y - x\|_X.\end{aligned}$$

It follows that $f(y) \to f(x)$ as $y \to x$, as required. $\square$

Note the appearance of the operator norm in the proof! This important quantity is how we measure the size of the derivative itself – and will be important for taking repeated derivatives, as it will allow us to compare the derivative map at two nearby points.

Most of the "derivative rules" from single-variable calculus have trivial adaptations (possibly after some modifications, e.g. in the product rule the functions must be scalar-valued), with exceptions being the chain rule and the inverse function theorem. The latter we postpone for now, and the chain rule we discuss here.

Theorem 3.3 (Chain rule)**.** *Let $U \subseteq X$ and $V \subseteq Y$ be open and let $f : U \to V$ be differentiable at x and $g : V \to Z$ be differentiable at $f(x)$. Then $g \circ f$ is differentiable at x with*

$$D(g \circ f)_x = Dg_{f(x)} \circ Df_x.$$

That is to say, the derivative of a composition is a composition of the derivatives.

Proof. As $Dg_{f(x)} \circ Df_x$ is a composition of continuous linear maps, it is itself a continuous linear map. It remains to show that

$$\lim_{h \to 0} \frac{\|(g \circ f)(x+h) - (g \circ f)(x) - (Dg_{f(x)} \circ Df_x)h\|_Z}{\|h\|_X} = 0.$$

We introduce

$$\phi_f(h) = \frac{f(x+h) - f(x) - Df_x h}{\|h\|_X}$$

for $h \neq 0$ and $\phi_f(0) = 0$. By openness of U, this is defined in an open ball around x, and by definition of the derivative is continuous at 0. Likewise, we define

$$\phi_g(h) = \frac{g(f(x)+h) - g(f(x)) - Dg_{f(x)}h}{\|h\|_Y}$$

for $h \neq 0$ and $\phi_g(0) = 0$, which is continuous at 0. We then have

$$\begin{aligned}
&\|(g \circ f)(x+h) - (g \circ f)(x) - (Dg_{f(x)} \circ Df_x)h\|_Z \\
=&\|g(f(x) + (f(x+h) - f(x))) - g(f(x)) - Dg_{f(x)}(Df_x h)\|_Z \\
=&\|\|f(x+h) - f(x)\|_Y \phi_g(f(x+h) - f(x)) \\
&+ Dg_{f(x)}[f(x+h) - f(x) - Df_x h]\|_Z \\
=&\|\|\|h\|_X \phi_f(h) + Df_x h\|_Y \phi_g(f(x+h) - f(x)) + Dg_{f(x)}[\|h\|_X \phi_f(h)]\|_Z \\
\leq&\|\|h\|_X \phi_f(h) + Df_x h\|_Y \|\phi_g(f(x+h) - f(x))\|_Z \\
&+ \|h\|_X \|Dg_{f(x)}\|_{Y \to Z} \|\phi_f(h)\|_Y \\
\leq&(\|h\|_X \|\phi_f(h)\|_Y + \|h\|_X \|Df_x\|_{X \to Y}) \|\phi_g(f(x+h) - f(x))\|_Z \\
&+ \|h\|_X \|Dg_{f(x)}\|_{Y \to Z} \|\phi_f(h)\|_Y
\end{aligned}$$

Dividing by $\|h\|_X$ and using the continuity of $\phi_g(f(x+h) - f(x))$ and $\phi_f(h)$ at $h = 0$ shows that

$$\lim_{h \to 0} \frac{\|(g \circ f)(x+h) - (g \circ f)(x) - (Dg_{f(x)} \circ Df_x)h\|_Z}{\|h\|_X} = 0,$$

as required. □

Now, in the finite-dimensional case, linear maps can be represented by matrices. We concern ourselves here with this matrix representation. For notational simplicity, we assume $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. Points in these spaces can be represented as vectors with respect to the standard bases, and the action of the derivative Df_x can be described via a matrix with respect to the standard bases, which we call the Jacobian Jf_x. Assuming that Df_x exists, we can easily derive Jf_x:

Since all norms on finite-dimensional vector spaces are Lipschitz equivalent, it does not matter which norms we use in the definition of derivative. In particular, if we write the coordinates of f in the standard basis as $(f_1, \ldots, f_m)$, by choosing the ∞-norm of X and Y and restricting the limit over those h to those h which are a multiple of a standard basis vector e_j, it is clear that for each i we have

$$\frac{|f_i(x + he_j) - f_i(x) - (Df_x[he_j])_i|}{h} \to 0$$

as $h \to 0$ in $\mathbb{R}$. Rearranging this gives the coordinate expression, that

$$(Df_x e_j)_i = \lim_{h\to 0} \frac{f_i(x + he_j) - f_i(x)}{h},$$

where $(Df_x e_j)_i$ represents the i^{th} coordinate of $Df_x e_j$. This then is exactly the $(i,j)^{\text{th}}$ entry of the Jacobian, $(Jf_x)_{i,j}$.

If x is represented in standard coordinates as $x = (x_1, \ldots, x_d)$, this says

$$(Jf_x)_{i,j} = \lim_{h\to 0} \frac{f_i(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_d) - f_i(x_1, \ldots, x_d)}{h}.$$

We can think of this as a single-variable derivative of f_i in which all coordinates of the input except x_j are fixed, and we take the single-variable derivative of f_i with respect to the input x_j. We call this expression the **partial derivative** of f_i with respect to x_j, typically denoted[51]

$$\frac{\partial f_i}{\partial x_j}(x) := \lim_{h\to 0} \frac{f_i(x + he_j) - f_i(x)}{h}.$$

In summary:

Definition 3.4. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$. Denote by f_i the i^{th} coordinate of f in the standard basis of $\mathbb{R}^m$. If the limit*

$$\frac{\partial f_i}{\partial x_j} := \lim_{h\to 0} \frac{f_i(x + he_j) - f_i(x)}{h}$$

[51] And also commonly denoted a couple dozen other ways! I personally use $\partial_j f_i$ the most, but I'll stick to the most common notation here!

*exists, we call it the **partial derivative** of f_i at x with respect to x_j. If the partial derivatives of f exist at x for every i and j, we define the **Jacobian matrix** Jf_x to be the $m \times n$ matrix with*

$$(Jf_x)_{i,j} = \frac{\partial f_i}{\partial x_j}(x).$$

We already demonstrated the following:

Theorem 3.5. *Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$ be differentiable at x. Then Jf_x exists and is the matrix of Df_x with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.*

Clearly, the Jacobian is a lot easier to compute in practice, because we have lots of computational and symbolic manipulation tools for finding single-variable derivatives. We would love to be able to say that the existence of the Jacobian implies the existence of the derivative. However, this is *false*!
Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} 1 & \text{if } x = 0 \text{ or } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

It is very easy to see that the partial derivatives exist and are equal to 0 at $(x, y) = (0, 0)$, however, f is clearly discontinuous there and thus cannot be differentiable.

All hope is not lost, however – we just need to impose some additional hypotheses.

Theorem 3.6. *Let $U \subseteq \mathbb{R}^n$ be open and suppose all partial derivatives of $f : U \to \mathbb{R}^m$ exist at each $x \in U$. Suppose that each of the maps*

$$x \mapsto \frac{\partial f_i}{\partial x_j}(x)$$

is continuous as a function of $x \in U$. Then f is differentiable at each $x \in U$.

Proof. For convenience, consider the 1-norm on $\mathbb{R}^n$ and the ∞-norm on $\mathbb{R}^m$. Fix $\varepsilon > 0$. By continuity, there is $\delta_{i,j} > 0$ so that $\|h\|_1 < \delta_{i,j}$ implies

$$\left| \frac{\partial f_i}{\partial x_j}(x + h) - \frac{\partial f_i}{\partial x_j}(x) \right| < \varepsilon.$$

Define $\delta = \min_{i,j} \delta_{i,j}$ so that this is true for all i and j whenever $\|h\|_1 < \delta$. We write $f_i(x+h) - f_i(x)$ as a telescoping sum of axis-parallel sum over axis parallel increments of h, that is,

$$f_i(x+h) - f_i(x) = \sum_{j=1}^{n} \left(f_i \left(x + \sum_{k=1}^{j} h_k e_k \right) - f_i \left(x + \sum_{k=1}^{j-1} h_k e_k \right) \right).$$

Note that for each j, $\xi \mapsto f\left(x + \sum_{k=1}^{j-1} h_k e_k + \xi e_j\right)$ is a continuous function of $\xi \in [0, h_j]$ (or $[h_j, 0]$ if $h_j < 0$) which is differentiable on $(0, h_j)$ with derivative

$$\frac{\partial f_i}{\partial x_j} \left(x + \sum_{k=1}^{j-1} h_k e_k + \xi e_j \right).$$

Thus by the mean value theorem there is ξ between 0 and h_j with

$$f_i \left(x + \sum_{k=1}^{j} h_k e_k \right) - f_i \left(x + \sum_{k=1}^{j-1} h_k e_k \right) = \frac{\partial f_i}{\partial x_j} \left(x + \sum_{k=1}^{j-1} h_k e_k + \xi e_j \right) h_j.$$

If $\|h\|_1 < \delta$, then clearly $\| \sum_{k=1}^{j-1} h_k e_k + \xi e_j \|_1 < \delta$, so

$$\left| f_i \left(x + \sum_{k=1}^{j} h_k e_k \right) - f_i \left(x + \sum_{k=1}^{j-1} h_k e_k \right) - \frac{\partial f_i}{\partial x_j}(x) h_j \right| < \varepsilon |h_j|.$$

Note that

$$\frac{\partial f_i}{\partial x_j}(x) h_j = (J f_x(h_j e_j))_i.$$

It now follows that

$$\begin{aligned}
&|f_i(x+h) - f_i(x) - (J f_x h)_i| \\
&\le \sum_{j=1}^{n} \left| f_i \left(x + \sum_{k=1}^{j} h_k e_k \right) - f_i \left(x + \sum_{k=1}^{j-1} h_k e_k \right) - (J f_x(h_j e_j))_i \right| \\
&< \sum_{i=1}^{n} \varepsilon |h_j| = \varepsilon \|h\|_1.
\end{aligned}$$

It now follows that

$$\frac{\|f(x+h) - f(x) - J f_x h\|_\infty}{\|h\|_1} < \varepsilon$$

whenever $\|h\|_1 < \delta$. This proves the limit, and thus that f is differentiable at x. $\square$

3.2 Higher-order derivatives

Definition 3.7. *Let $U \subseteq X$ be open and suppose $f : U \to Y$ is differentiable at each $x \in U$. Then we say f is differentiable and call $Df : U \to \mathcal{L}(X;Y)$, defined by $Df(x) = Df_x$, the derivative.*

Iteratively, if f is $k-1$ times differentiable on an open set containing x, and the $(k-1)^{th}$ derivative is differentiable at x, we call this derivative the k^{th} derivative of f at x, denoted $D^k f_x$. If the k^{th} derivative of f exists on U and is continuous, we say f is k-times continuously differentiable on U and write $f \in C^k(U;Y)$.

If f is k-times differentiable on U, we write $f \in C^\infty(U;Y)$ and say f is infinitely differentiable, or smooth.

If you take a moment to think, this is quite scary: The second derivative is a map from U to the space of linear maps from X to the space of linear maps from X to Y! It will only get worse with more derivatives. However, we can identify this with a space of bilinear maps – maps with two inputs which are linear in each – or for higher orders, multilinear maps. We demonstrate for the second derivative.

A second derivative is a linear map T from X to the space of linear maps from X to Y. So for every $x \in X$, $T(x)$ is a linear map. Define $A(x,y) = [T(x)](y)$. This is clearly a bilinear map from $X \times X \to Y$. Actually, one can verify that the correspondence of T to A defines a linear homeomorphism from the space $\mathcal{L}(X;\mathcal{L}(X;Y))$ to the space $\mathcal{L}^2(X;Y)$ of "bounded bilinear maps", with norm

$$\|A\| := \sup_{\|x\|_X = \|y\|_X = 1} \|A(x,y)\|,$$

which makes it into a Banach space when Y is Banach. Note that when X is finite-dimensional, all bilinear maps are bounded[52].

We will abuse notation and use e.g. $D^2 f_x$ to refer to this bilinear map. Precisely, we are saying that

$$D^2 f_x(v,w) := [D(Df)_x v]w.$$

In the finite-dimensional setting, the derivative has coordinates given by the partial derivatives. The matrix coordinates of the second derivative will be the coordinates of this, things like

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f_i}{\partial x_j}\right).$$

[52] All of these assertions are modifications of earlier results and we shall not prove them.

We denote this by

$$\frac{\partial^2 f_i}{\partial x_k \partial x_j},$$

which we call a second order partial derivative. Note that continuity of all second order partial derivatives is the same as being in $C^2(U)$ by an earlier theorem. Similar notations are used for higher-order derivatives.

In multivariable calculus, you will surely have noticed when computing mixed derivatives that it doesn't matter which order you do them in. For instance, take $f(x, y) = x^2 y$. If I differentiate this with respect to x, then y, I get $2xy$ and then $2x$. If I differentiate with respect to y then x, I get x^2 then $2x$. For these easy symbolic calculations, you may see the patterns and conclude that this is a general fact. The substantial result of this section is that in a fairly general setting, this remains true.

Note that

$$\frac{\partial^2 f_i}{\partial x_j \partial x_k}(x) = [(D(Df)_x e_j)e_k]_i = [D^2 f_x(e_j, e_k)]_i.$$

It suffices to consider when $D^2 f_x$ is a symmetric bilinear map, that is, $D^2 f_x(v, w) = D^2 f_x(w, v)$ for all v and w. This is expressed by Clairaut's Theorem.

Theorem 3.8 (Clairaut)**.** *Let X be Banach, $U \subseteq X$ be open and $f \in C^2(U; \mathbb{R}^d)$. Then for $x \in U$, $D^2 f_x$ is a symmetric bilinear map.*

Remark: Note that this requires something **more** than just being twice differentiable – the second derivatives need to be continuous!

Proof. Note that the result is trivial if either v or w is 0, since in this case $D^2 f_x(v, w) = 0 = D^2 f_x(w, v)$. Note that by scaling, it suffices to prove the result when $\|v\|_X = \|w\|_X = 1$, so assume this is the case. Furthermore, writing f in components $f = (f_1, \ldots, f_d)$, it is clear that $[D^2 f_x(v, w)]_i = D^2(f_i)_x(v, w)$, so it suffices to prove the claim for each i. In particular, we need only prove the case $d = 1$, so assume this.

By continuity of $D^2 f_x$ we have that $D(Df)_x v$ is continuous, since

$$\|D(Df)_x v - D(Df)_y v\|_{X \to \mathbb{R}} \leq \|D(Df)_x - D(Df)_y\|_{X \to (X \to \mathbb{R})} \|v\|_X.$$

In turn, $[D(Df)_x v](w) = D^2 f_x(v, w)$ is continuous. Consequently, for $v, w \neq 0$ and $\varepsilon > 0$, there is $\delta > 0$ so that $\|h\|_X < \delta$ implies

$$|D^2 f_{x+h}(v, w) - D^2 f_x(v, w)| < \frac{\varepsilon}{2}$$

and also that
$$|D^2 f_{x+h}(w,v) - D^2 f_x(w,v)| < \frac{\varepsilon}{2}.$$

Next, set $v' = \frac{\delta}{4}v$, $w' = \frac{\delta}{4}w$. Then $x + sv' + tw' \in B_\delta(x)$ for $0 \le s, t \le 1$. In particular,

$$\begin{aligned}
&|D^2 f_{x+sv'+tw'}(v',w') - D^2 f_x(v',w')| \\
&= \left(\frac{\delta}{4}\right)^2 |D^2 f_{x+sv'+tw'}(v,w) - D^2 f_x(v,w)| < \frac{\varepsilon}{2}\left(\frac{\delta}{4}\right)^2.
\end{aligned}$$

Now define

$$g(x) = f(x + v' + w') - f(x + w') - f(x + v') + f(x).$$

Consider $\tilde{g}(x) = f(x + v') - f(x)$, so $\tilde{g}(x + w') - \tilde{g}(x) = g(x)$. Applying the fundamental theorem of calculus to $\tilde{g}(x + tw')$ for $t \in [0, 1]$, we have by the chain rule that

$$\begin{aligned}
g(x) = \tilde{g}(x + w') - \tilde{g}(x) &= \int_0^1 D\tilde{g}_{x+tw'} w' \, dt \\
&= \int_0^1 Df_{x+v'+tw'} w' - Df_{x+tw'} w' \, dt.
\end{aligned}$$

Applying a similar argument to $Df_{x+sv'+tw'}w' - Df_{x+tw'}w'$, we obtain

$$g(x) = \int_0^1 \int_0^1 D^2 f_{x+sv'+tw'}(v', w') \, ds \, dt.$$

It follows that

$$\begin{aligned}
|g(x) - D^2 f_x(v', w')| &= \left| \int_0^1 \int_0^1 D^2 f_{x+sv'+tw'}(v', w') - D^2 f_x(v', w') \, ds \, dt \right| \\
&\le \int_0^1 \int_0^1 |D^2 f_{x+sv'+tw'}(v', w') - D^2 f_x(v', w')| \, ds \, dt \\
&< \frac{\varepsilon}{2}\left(\frac{\delta}{4}\right)^2.
\end{aligned}$$

We also have by similar reasoning that

$$|g(x) - D^2 f_x(w', v')| < \frac{\varepsilon}{2}\left(\frac{\delta}{4}\right)^2.$$

By the triangle inequality, it now follows that

$$\begin{aligned}|D^2 f_x(v', w') - D^2 f_x(w', v')| &\leq |D^2 f_x(v', w') - g(x)| + |g(x) - D^2 f_x(w', v')| \\ &< \varepsilon \left(\frac{\delta}{4}\right)^2,\end{aligned}$$

and thus by scaling that $|D^2 f_x(v, w) - D^2 f_x(w, v)| < \varepsilon$. Since this holds for all $\varepsilon > 0$, we must in fact have $D^2 f_x(v, w) = D^2 f_x(w, v)$. $\square$

Remark (non-examinable): Can we replace $\mathbb{R}^d$ by a general Banach space? Yes! Using a vector-valued theory of integrals on intervals, instead of working in coordinates, the proof adapts. However, we would need a fundamental theorem of calculus for vector-valued functions. This is not so straightforward, as there is no obvious way for us to conclude that two functions $f, g : [a, b] \to Y$ which are continuous, and differentiable on (a, b), with equal derivatives, differ by a constant. It is however true, but makes use of a powerful result called the Hahn-Banach theorem.

Remark (non-examinable): An important application of higher order derivatives is Taylor's Theorem. However, it is not especially interesting for us to address here, since the general vector-valued version again requires a Banach space fundamental theorem of calculus, and the scalar valued version essentially consists in applying the single variable Taylor's theorem along lines and deriving the relevant algebraic expressions. It will be given as an exercise for the interested reader.

3.3 The Implicit and Inverse Function Theorems

We now arrive at two highly related and highly important theorems. The implicit function theorem, in brief, says that given a C^k function g of $x \in X$ and $y \in Y$, the solutions to

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

can, under certain conditions, be written as $(x, \phi(x))$ for some C^k function ϕ taking values in Y. This proves incredibly helpful in many areas of mathematics, including the study of manifolds (the general dimensional analogue or curves and surfaces, which show up just about everywhere) and in optimization, where $g(x, y) = c$ may represent a constraint.

As a consequence of this, we will prove the inverse function theorem, which gives conditions for the local invertibility of a function.