30. Calculus of Several Variables II

30.6 Rolle's Theorem

We've already covered Weierstrass's Theorem in section 30.2. We use Weierstrass's Theorem to prove Rolle's Theorem.[1]

Rolle's Theorem. *Suppose* $f\colon [a,b] \to \mathbb{R}$ *is continuous on* $[a,b]$ *and continuously differentiable on* (a,b). *If* $f(a) = f(b) = 0$, *there is a point* $c \in (a,b)$ *with* $f'(c) = 0$.

Rolle's Theorem Illustrated

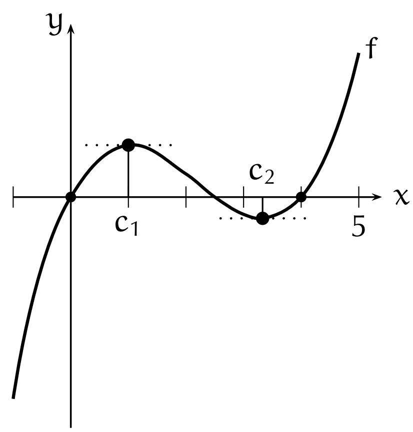


Figure 30.6.1: This function goes both above and below the axes, so there are at least two points in the closed interval $[0,4]$ that are extrema, obeying $f'(c) = 0$. In this case there are exactly two, labeled c_1 and c_2.

[1] Michel Rolle (1652–1719) was a French mathematician (number theory). He was the first to publish a description of Gaussian elimination, although the idea was known to Newton, and some cases had already been known to Chinese mathematicians by 179 AD.

Rolle proved a version of this theorem in 1691 that only applied to polynomial functions. Cauchy derived the general version from the Mean Value Theorem. Our proof is based on Weierstrass's Theorem which was not available to either Cauchy or Rolle. We follow the opposite approach from Cauchy and use Rolle's Theorem to prove the Mean Value Theorem.

30.7 Proof of Rolle's Theorem

Rolle's Theorem. *Suppose* $f\colon [a,b] \to \mathbb{R}$ *is continuous on* $[a,b]$ *and continuously differentiable on* (a,b). *If* $f(a) = f(b) = 0$*, there is a point* $c \in (a,b)$ *with* $f'(c) = 0$.

Proof. If f is constant on $[a,b]$, then $f(x) = 0$ for all $x \in [a,b]$ and so $f'(x) = 0$ for all $x \in (a,b)$. Any $c \in (a,b)$ will do.

Otherwise, f is not constant on $[a,b]$. Either there is $d \in (a,b)$ with $f(d) > 0$ or a $d \in (a,b)$ with $f(d) < 0$. In the former case, f has a maximum at some $c \in (a,b)$ by Weierstrass's Theorem. In the latter case, f has a minimum at some $c \in (a,b)$. Whichever happens, the first order necessary condition for an interior optimum in $\mathbb{R}$ shows $f'(c) = 0$ (remember your calculus!), and we are done. ∎

30.8 Mean Value Theorem

The Mean Value Theorem generalizes Rolle's Theorem, which is the key to the proving the Mean Value Theorem.[2]

Mean Value Theorem. *Let* $f\colon I \to \mathbb{R}$ *be a* $\mathcal{C}^1$ *function on an interval* $I \subset \mathbb{R}$. *Then for any points* $a, b \in I$ *with* $a < b$ *there is a point* c, $a < c < b$ *with*

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

or equivalently, $f(b) - f(a) = f'(c)(b - a)$.

Mean Value Theorem Illustrated

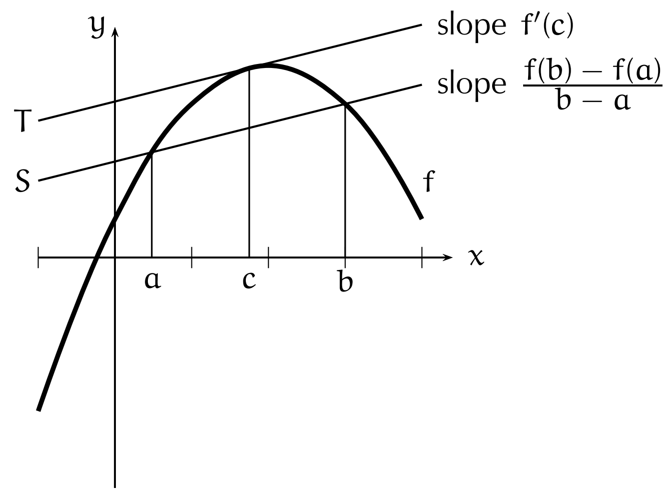


Figure 30.8.1: The Mean Value Theorem gives us a point $c \in (a, b)$ where the slope of the tangent T to f at $(c, f(c))$ is equal to the slope of the secant S through $(a, f(a))$ and $(b, f(b))$.

[2] The modern form of the Mean Value Theorem is due to Cauchy in 1821.

30.9 Proof of the Mean Value Theorem

Mean Value Theorem. *Let* $f\colon I \to \mathbb{R}$ *be a* $\mathcal{C}^1$ *function on an interval* $I \subset \mathbb{R}$. *Then for any points* $a, b \in I$ *with* $a < b$ *there is a point* c, $a < c < b$ *with*

$$f'(c) = \frac{f(b) - f(a)}{b - a}, \quad \textit{or equivalently,} \quad f(b) - f(a) = f'(c)(b - a).$$

Proof. We just need to apply Rolle's Theorem to the proper function. Define the function g by

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

For every $x \in (a, b)$, $g(x)$ is the vertical distance between the secant line S and $f(x)$.

The distance is zero at both a and b:

$$g(a) = f(a) - f(a) - \frac{f(b) - f(a)}{b - a}(a - a) = 0$$

and

$$g(b) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = 0.$$

This shows the secant intersects the graph of f at both a and b.

By Rolle's Theorem, there is a $c \in (a, b)$ with

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

Then $f'(c)(b - a) = f(b) - f(a)$, proving the result. ∎

30.10 Mean Value Theorem on $\mathbb{R}^m$

There is a version for $\mathbb{R}^m$ that follows directly from the basic Mean Value Theorem.

Theorem 30.10.1. *Let $f\colon U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subset \mathbb{R}^m$. Suppose $\mathbf{a}, \mathbf{b} \in U$ with the line segment $\ell(\mathbf{a}, \mathbf{b}) \subset U$. Then there is a point $\mathbf{c}$ in the line segment $\ell(\mathbf{a}, \mathbf{b})$ such that*

$$f(\mathbf{b}) - f(\mathbf{a}) = Df|_{\mathbf{c}}(\mathbf{b} - \mathbf{a})$$

Proof. Define $g\colon [0,1] \to U$ by $g(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ and set $h(t) = f\big(g(t)\big)$. Then $h\colon [0,1] \to \mathbb{R}$. By the Mean Value Theorem, there is a $t^* \in (0,1)$ with

$$h(1) - h(0) = h'(t^*)(1 - 0) = h'(t^*).$$

Let $\mathbf{c} = g(t^*) = \mathbf{a} + t^*(\mathbf{b} - \mathbf{a})$. Then

$$\begin{aligned} f(\mathbf{b}) - f(\mathbf{a}) &= h(1) - h(0) \\ &= h'(t^*) \\ &= D_t f\big(g(t^*)\big) \\ &= D_{\mathbf{x}} f\big(g(t^*)\big) g'(t^*) \\ &= Df|_{\mathbf{c}}(\mathbf{b} - \mathbf{a}). \end{aligned}$$

∎

This tells us that $f(\mathbf{b}) = f(\mathbf{a}) + Df(\mathbf{c})(\mathbf{b} - \mathbf{a})$ for some $\mathbf{c}$ on the line segment $\ell(\mathbf{b}, \mathbf{a})$, which is useful for approximating f at a point $\mathbf{b}$ based on its value at a point $\mathbf{a}$.

30.11 Derivatives and Factorials

There is a generalization of the Mean Value Theorem using Taylor Polynomials. Let $f^{(k)}$ denote the k^{th} derivative of f:[3]

$$f^{(k)}(x) = \frac{d^k f}{dx^k}(x)$$

where $f^{(0)}(a) = f(a)$.

Recall that $k!$ denotes k *factorial* which is defined inductively for non-negative integers by $0! = 1$ and $k! = k(k-1)!$ for any positive integer k. Thus

$$k! = 1 \cdot 2 \cdot 3 \cdots (k-1) \cdot k.$$

The gamma function extends the factorial function to the complex numbers, excepting the non-positive integers. They are related by $k! = \Gamma(k+1)$ whenever k is a non-negative integer. When $z \in \mathbb{C}$ has a positive real part,

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx.$$

[3] This is an alternate way of of writing Lagrange's notation, f', f'', f''', ... that works better with summation.

30.12 Taylor Polynomials

Taylor's formula will allow us to approximate a function f by polynomials—the Taylor polynomials.

Taylor Polynomial. The k^{th} order Taylor polynomial is

$$\begin{aligned}P_k(x;a) &= f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots + \frac{1}{k!}f^{(k)}(a)(x-a)^k \\ &= \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(a)(x-a)^n.\end{aligned}$$

The fact that $P_k(a;a) = f(a)$ will be useful when we prove various forms of Taylor's Formula.

Here are the first several Taylor polynomials:

$$\begin{aligned}P_0(x;a) &= f(x) \\ P_1(x;a) &= f(x) + f'(a)(x-a) \\ P_2(x;a) &= f(x) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 \\ P_3(x;a) &= f(x) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{6}f^{(3)}(a)(x-a)^3\end{aligned}$$

30.13 First Order Taylor's Formula in $\mathbb{R}$

To see how the proof of Taylor's formula works, we start with the first order Taylor's formula. (The Mean Value Theorem is the zeroth order Taylor's formula.) The proof is an exercise in using Rolle's Theorem.

Theorem 30.13.1. *Let* $f\colon I \to \mathbb{R}$ *be a* $\mathcal{C}^2$ *function defined on an interval in* $\mathbb{R}$*. If* $a, b \in I$ *there exists a* $c \in (a, b)$ *such that*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \frac{1}{2}f''(c)(b-a)^2 \\ &= P_1(b; a) + \frac{1}{2}f''(c)(b-a)^2. \end{aligned} \tag{30.13.1}$$

30.14 Proof of First Order Taylor's Formula in $\mathbb{R}$

Proof. Define

$$\begin{aligned} g(x) &= f(x) - P_1(x; a) - M(x-a)^2 \\ &= f(x) - \big[f(a) + f'(a)(x-a)\big] - M(x-a)^2. \end{aligned} \tag{30.14.2}$$

By definition, $g(a) = f(a) - P_1(a; a) = 0$. Now choose M so that $g(b) = 0$. Then

$$M = \frac{1}{(b-a)^2}\big[f(b) - f(a) - f'(a)(b-a)\big]$$

By Rolle's Theorem, there is a $c_1 \in (a, b)$ with $g'(c_1) = 0$. Now

$$g'(x) = f'(x) - f'(a) - 2M(x-a)$$

so $g'(a) = f'(a) - f'(a) = 0$. Both $g'(a) = g'(c_1) = 0$, so we apply Rolle's Theorem again, this time to g' on the interval (a, c_1).

From Rolle's Theorem we obtain a $c \in (a, c_1) \subset (a, b)$ with $g''(c) = 0$. Now

$$g''(x) = f''(x) - 2M$$

so $f''(c) = 2M$. In equation (14.43.6), substitute $M = f''(c)/2$ and set $x = b$ to obtain equation (30.13.1). ∎

30.15 k^{th} Order Taylor's Formula in $\mathbb{R}$

We now consider the k^{th} order Taylor's formula.

Taylor's Formula in $\mathbb{R}$**.** *Let* $f\colon I \to \mathbb{R}$ *be a* $\mathcal{C}^{k+1}$ *function defined on an interval* I *in* $\mathbb{R}$*. If* $a, b \in I$ *there exists a* $c \in (a, b)$ *such that*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \frac{1}{2!}f''(a)(b-a)^2 + \cdots + \frac{1}{k!}f^{(k)}(a)(b-a)^k \\ &\qquad + \frac{1}{(k+1)!}f^{(k+1)}(c)(b-a)^{k+1} \qquad (30.15.3) \\ &= P_k(b;a) + \frac{1}{(k+1)!}f^{(k+1)}(c)(b-a)^{k+1} \end{aligned}$$

Proof. Define

$$\begin{aligned} g(x) &= f(x) - P_k(x;a) - M(x-a)^{k+1} \\ &= f(x) - f(a) - f'(a)(x-a) - \frac{1}{2!}f''(a)(x-a)^2 \\ &\qquad - \cdots - \frac{1}{k!}f^{(k)}(a)(x-a)^k + M(x-a)^{k+1} \end{aligned} \tag{30.15.4}$$

where

$$M = \frac{1}{(b-a)^{k+1}}\left[f(b) - P_k(b;a)\right]$$

Proof continues . . .

30.16 Taylor's Formula Proof Part II

Part II of Proof. As before, $g(a) = f(a) - P_k(a; a) = f(a) - f(a) = 0$ and M has been chosen so that $g(b) = f(b) - P_k(b; a) - \left[f(b) - P_k(b; a)\right] = 0$.

By Rolle's Theorem, there is a $c_1 \in (a, b)$ with $g'(c_1) = 0$. Now

$$g'(x) = f'(x) - f'(a) - f''(a)(x - a) - \cdots - \frac{1}{(k-1)!} f^{(k)}(a)(x - a)^{k-1} \\ + (k + 1)M(x - a)^k.$$

Then $g'(a) = f'(a) - f'(a) = 0$. A second application of Rolle's Theorem, now to g', yields a $c_2 \in (a, c_1) \subset (a, b)$ with $g''(c_2) = 0$.

Computing g'' we obtain

$$g''(x) = f''(x) - f''(a) - \cdots - \frac{1}{(k-2)!} f^{(k)}(a)(x - a)^{k-2} \\ - (k + 1)kM(x - a)^{k-1}.$$

It follows that $g''(a) = f''(a) - f''(a) = 0$. A third application of Rolle's Theorem, now to g''', yields a $c_3 \in (a, c_2) \subset (a, b)$ with $g'''(c_3) = 0$.

Proof continues once more ...

30.17 Taylor's Formula Proof Part III

Remainder of Proof. We continue applying Rolle's Theorem to successive derivatives until we eventually get to $g^{(k)}(x)$ with $g^{(k)}(c_k) = 0$ and

$$g^{(k)}(x) = f^{(k)}(x) - f^{(k)}(a) - (k+1)!\,M(x-a).$$

It follows that $g^{(k)}(a) = f^{(k)}(a) - f^{(k)}(a) = 0$, so we apply Rolle's Theorem one last time to find a $c \in (a, c_k) \subset (a, b)$ with $g^{(k+1)}(c_k) = 0$.

We compute

$$g^{(k+1)}(x) = f^{(k+1)}(x) - (k+1)!\,M.$$

Then $f^{(k+1)}(c) = (k+1)!\,M$, so $M = f^{(k+1)}(c)/(k+1)!$.

In equation (30.15.4), substitute $M = f^{(k+1)}(c)/(k+1)!$ and set $x = b$ to obtain equation (30.15.3). ∎

30.18 Big O

Big O and *little o* are notations used to describe the asymptotic behavior of functions. They can be applied at infinity, or at any finite point (often 0). The "O" stands for order, and indicates that one function is the same order as the other. They are useful for describing the precision of estimates.[4]

Thus $f(x) = O\big(g(x)\big)$ as $x \to \infty$ means there is an $M > 0$ such that

$$\frac{|f(x)|}{g(x)} \leq M$$

for x large enough.

Similarly $f(x) = O(x^3)$ at 0 means there is an $M > 0$ with

$$\frac{|f(x)|}{x^3} \leq M$$

for x near 0.

Thus

$$10 - \frac{1}{x} = O(1) \quad \text{as } x \to \infty.$$

[4] This type of asymptotic notation is known as the Bachman-Landau notation. Big O was introduced by Paul Bachman (1837–1920) in 1894, Edmund Landau (1877–1938) proposed the little o notation in 1909. There are other variants that are less commonly used, such as the Ω notation of G.H. Hardy (1877–1947) and J.E. Littlewood (1885–1977).

30.19 Little o

Big O is used when the ratio of two functions is bounded. We use *little o* when the ratio converges to zero. So $f(x) = o\big(g(x)\big)$ at infinity means that for every $\varepsilon > 0$, there is a K such that

$$\frac{|f(x)|}{g(x)} < \varepsilon$$

for all $x > K$. The definition at any finite point is similar.

When we say $R(x) = o\big(|x|^k\big)$ at $x = 0$, it means that for every $\varepsilon > 0$, there is a $\delta > 0$ with

$$|R(x)| < \varepsilon |x|^k$$

for $|x| < \delta$. In this case f converges to zero enough faster than $|x|^k$ that the ratio also converges to zero.

30.20 Limsup and Liminf

Two useful limit concepts are the *limit superior* or *limsup*, and the *limit inferior* or *liminf*.

Limit Superior. The *limit superior* or lim sup is defined by

$$\limsup_{x \to a} f(x) = \lim_{\varepsilon \to 0} \Big(\sup\{f(x) : x \in B_\varepsilon(a)\} \Big)$$

when a is finite, and

$$\limsup_{x \to \infty} f(x) = \lim_{k \to \infty} \Big(\sup\{f(x) : x > k\} \Big)$$

when $a = +\infty$.

The definition is a little simpler for sequences. When we have a sequence $\{a_n\}$,

$$\limsup_{n \to \infty} = \lim_{n \to \infty} \Big(\sup\{a_m : m \geq n\} \Big).$$

Limit Inferior. The *limit inferior*, lim inf, is defined analogously. E.g.,

$$\liminf_{n \to \infty} a_n = \lim_{n \to \infty} \Big(\inf\{a_m : m \geq n\} \Big).$$

30.21 Limits, Big O, and little o

The definitions of big O and little o can also be stated in terms of limits.

Big O. We write $f(x) = O\big(g(x)\big)$ as $x \to a$ to mean

$$\limsup_{x\to a} \frac{|f(x)|}{g(x)} < \infty.$$

The case $a = \pm\infty$ is allowed.

As for little o, we use the same quotient, but here the limit superior is 0. In that case, it is enough to use the limit itself rather than the limit superior.

Little o. We write $f(x) = o\big(g(x)\big)$ as $x \to a$ when

$$\lim_{x\to a} \frac{|f(x)|}{g(x)} = 0.$$

The case $a = \pm\infty$ is allowed.

30.22 The Remainder in Taylor's Formula

We can describe the behavior of the remainder from Taylor's formula using little o. Define the k^{th} remainder term by

$$R_k(x; a) = f(x) - P_k(x; a).$$

Use Taylor's formula to write

$$\frac{R_k(x; a)}{(x-a)^k} = \frac{1}{k!}\frac{f^{(k+1)}(c)(x-a)^{k+1}}{(x-a)^k} = \frac{1}{k!}f^{(k+1)}(c)(x-a).$$

As $x \to a$, $c \to a$, so the limit of the remainder is

$$\lim_{x\to a} R_k(x; a) = \frac{1}{k!}f^{(k+1)}(a)(a-a) = 0,$$

or in the little o notation,

$$R_k(x; a) = o\big(|x-a|^k\big).$$

This shows that Taylor's formula is a good approximation of f for x near a, and the approximation is better the closer you are to a.

30.23 Example: A Power Series

▶ Example 30.23.1: Power Series. In some cases, the approximation is perfect as $k \to \infty$, and we get a convergent power series. Consider $f(x) = \sin x$ and set $a = 0$. Then $f^{(2k)}(0) = 0$, $f^{(4k+1)}(0) = +1$ and $f^{(4k+3)}(0) = -1$, so the Taylor expansion for $\sin x$ is

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

Applying the ratio test, we find that this series converges absolutely for any $x \in \mathbb{R}$ because

$$\left|\frac{x^{2n+3}}{(2n+3)!}\right| \cdot \left|\frac{(2n+1)}{x^{2n+1}}\right| = \left|\frac{x^2}{(2n+2)(2n+3)}\right| \to 0$$

as $n \to \infty$ for every real x.

The remainder obeys

$$|R_{2k+1}| = \frac{|x|^{2k+1}}{k!} \to 0 \quad \text{as } k \to \infty,$$

so the power series converges to $\sin x$. The convergence is uniform on any compact interval. As a result, the limit is continuous. In fact, the limits of the derivatives also converge uniformly to the derivatives of $\sin x$ on compact intervals, yielding a $\mathcal{C}^\infty$ function. Finally, since the power series converges, $\sin x$ is not only $\mathcal{C}^\infty$, but also is an analytic function, as is its complex version. ◀

30.24 Analytic vs. $\mathcal{C}^\infty$ Functions

Analytic (*holomorphic*) functions are very special. This is especially true in complex function theory where any function that is complex differentiable is also analytic, something that is not even remotely true for functions on the real line. In fact, even $\mathcal{C}^\infty$ functions need not be close to being real analytic.

We demonstrate this in Example 30.25.1. It shows a $\mathcal{C}^\infty$ function that is zero on half of the real line. It's power series at zero is identically zero. In other words, **the remainder term** in the Taylor series **is the function**, and the Taylor series contributes nothing. In contrast, with analytic functions, the remainder will converge to zero and the power series is everything.

30.25 Example: Nothing but the Remainder

▶ Example 30.25.1: Infinitely Flat Function. Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} e^{-1/x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

This is a $\mathcal{C}^\infty$ function. All derivatives from the left at zero are zero because the function is always 0 on $(-\infty, 0]$. As for the function to the right, $f'(x) = (1/x^2)e^{-1/x}$. This has limit zero at $x = 0$ because the exponential converges to zero a lot faster than the polynomial $(1/x^2)$ explodes. In fact, all of the derivatives from the right involve a polynomial in $1/x$ times $e^{-1/x}$, and the exponential always wins.

Since the derivatives are zero, each $P_k(x, 0) = 0$. Each Taylor approximation is zero! That means the remainder when $x > 0$ is always $f(x)$. The function itself **is** the remainder. However, as f is very near zero when $x > 0$ is small, that does not prevent Taylor's formula from still giving us a good approximation of the function f near 0.

Yes, it is kind of weird. ◀

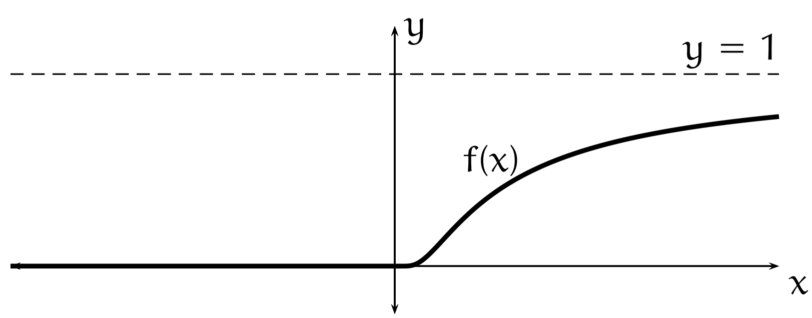


Figure 30.25.2: This $\mathcal{C}^\infty$ function f is infinitely flat at $x = 0$, meaning that all of its derivatives are zero there. It converges to 1 as $x \to +\infty$.

30.26 Another Infinitely Flat Function

An interesting variation is

$$g(x) = \begin{cases} e^{-1/(x-b)} & \text{for } x \geq b \\ 0 & \text{for } a \leq x \leq b. \\ e^{-1/(a-x)} & \text{for } x \leq a \end{cases}$$

This $\mathcal{C}^\infty$ function is zero exactly on the interval $[a, b]$ and strictly positive outside that interval. It is strictly decreasing on $(-\infty, a)$ and strictly increasing on $(b, +\infty)$.

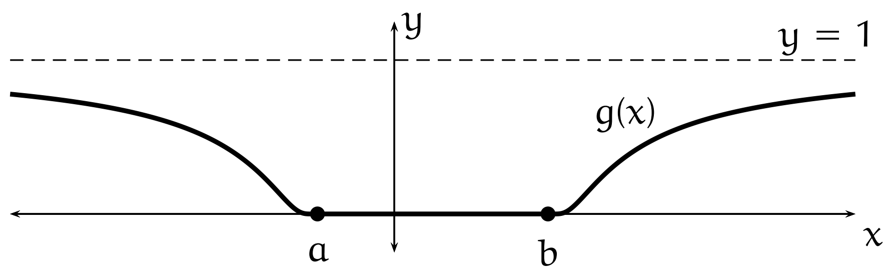


Figure 30.26.1: This $\mathcal{C}^\infty$ function f is infinitely flat at both $x = a$ and $x = b$, meaning that all of its derivatives are zero there. It is zero if and only if $x \in [a, b]$ and converges to 1 as $x \to +\infty$.

Although you can do this sort of thing with $\mathcal{C}^\infty$ functions, it doesn't work with analytic functions. This function has a power series representation on a neighborhood of every point **except** a and b.

30.27 Taylor's Formula in $\mathbb{R}^m$

Just as we did with the Mean Value Theorem, we can derive Taylor's Formula in $\mathbb{R}^m$ from the $\mathbb{R}^1$ version. We will use the shorthand notation $\left[D^k f_{\mathbf{a}}\right] \mathbf{h}^{\otimes k}$ to denote the k-tensor $\left[D^k f_{\mathbf{a}}\right](\mathbf{h}, \ldots, \mathbf{h})$ that is the k^{th} derivative applied to $\mathbf{h} \otimes \cdots \otimes \mathbf{h} \in (\mathbb{R}^m)^{\otimes k}$.

Taylor's Formula in $\mathbb{R}^m$. *Let $f : U \to \mathbb{R}$ be a $\mathcal{C}^{k+1}$ function defined on an open set in $\mathbb{R}^m$. Suppose that for every $\mathbf{a}, \mathbf{b} \in U$, $\ell(\mathbf{a}, \mathbf{b}) \subset U$. Then for all $\mathbf{a}, \mathbf{b} \in U$ there exists a $\mathbf{c} \in \ell(\mathbf{a}, \mathbf{b})$ such that*

$$\begin{aligned} f(\mathbf{b}) = f(\mathbf{a}) &+ \left[D f_{\mathbf{a}}\right](\mathbf{b} - \mathbf{a}) + \frac{1}{2!}\left[D^2 f_{\mathbf{a}}\right](\mathbf{b} - \mathbf{a})^{\otimes 2} + \cdots \\ &+ \frac{1}{k!}\left[D^k f_{\mathbf{a}}\right](\mathbf{b} - \mathbf{a})^{\otimes k} \\ &+ \frac{1}{(k+1)!}\left[D^{k+1} f(\mathbf{c})\right](\mathbf{b} - \mathbf{a})^{\otimes (k+1)} \end{aligned} \qquad (30.27.5)$$

30.28 Proof of Taylor's Formula in $\mathbb{R}^m$

Proof. We will piggyback off Taylor's Formula in $\mathbb{R}$. Define

$$\phi(t) = f\big((1-t)\mathbf{a} + t\mathbf{b}\big) = f\big(\mathbf{a} + t(\mathbf{b}-\mathbf{a})\big).$$

Since U is open, $\phi\colon I \to \mathbb{R}$ is a $C^{(k)}$ function defined on an open interval $I \supset [0,1]$. We can apply Taylor's Formula on the interval I between 0 and 1 to find

$$\phi(1) = \sum_{n=0}^{k} \frac{1}{n!}\big[\phi^{(n)}(0)\big]1^n + \frac{1}{(k+1)!}\big[\phi^{(k+1)}(t^*)\big]1^{k+1}.$$

for some $t^* \in (0,1)$. Then we apply the Chain Rule to find

$$\begin{aligned}\phi'(t) &= \big[Df\big(\mathbf{a} + t(\mathbf{b}-\mathbf{a})\big)\big](\mathbf{b}-\mathbf{a}),\\ \phi''(t) &= \big[D^2f\big(\mathbf{a} + t(\mathbf{b}-\mathbf{a})\big)\big](\mathbf{b}-\mathbf{a})^{\otimes 2},\\ \phi'''(t) &= \big[D^3f\big(\mathbf{a} + t(\mathbf{b}-\mathbf{a})\big)\big](\mathbf{b}-\mathbf{a})^{\otimes 3}\\ &\text{etc.}\end{aligned}$$

Setting $t = 0$ we obtain equation (30.27.5). ∎

30.29 The Remainder Term in $\mathbb{R}^m$

Again, the k^{th} remainder term is $R_k(\mathbf{x}; \mathbf{a}) = f(\mathbf{x}) - P_k(\mathbf{x}; \mathbf{a})$, so

$$R_k(\mathbf{x}; \mathbf{a}) = \frac{1}{k!}\left[D^{(k+1)} f_{\mathbf{c}}\right](\mathbf{x} - \mathbf{a})^{\otimes(k+1)}.$$

Dividing by $\|\mathbf{x} - \mathbf{a}\|^k$, we obtain

$$\frac{R_k(\mathbf{x}; \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|^k} = \frac{1}{k!}\frac{\left[D^{(k+1)} f_{\mathbf{c}}\right](\mathbf{x} - \mathbf{a})^{\otimes(k+1)}}{\|\mathbf{x} - \mathbf{a}\|^k}$$

Let $\mathbf{u}$ be the unit vector $(\mathbf{x} - \mathbf{a})/\|\mathbf{x} - \mathbf{a}\|$. Then

$$\frac{R_k(\mathbf{x}; \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|^k} = \frac{1}{k!}\left[D^{(k+1)} f_{\mathbf{c}}\right]\mathbf{u}^{\otimes(k+1)}\|\mathbf{x} - \mathbf{a}\|$$

As $\mathbf{x} \to \mathbf{a}$, $\mathbf{c} \to \mathbf{a}$, and we find

$$\frac{R_k(\mathbf{x}; \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|^k} \to 0$$

as $\mathbf{x} \to \mathbf{a}$. Alternatively

$$R_k(\mathbf{x}; \mathbf{a}) = o\left(\|\mathbf{x} - \mathbf{a}\|^k\right)$$

as $\|\mathbf{x} - \mathbf{a}\| \to \mathbf{a}$. This means that the remainder goes to zero as $\mathbf{x} \to \mathbf{a}$ enough faster than $\|\mathbf{x} - \mathbf{a}\|^k \to 0$ that their ratio converges to zero.

29.3 Connected Sets

10/25/22

Our next collection of important topological concepts relates to connected sets. Roughly speaking, a connected set is a set that is a single contiguous piece. Before defining connectedness, we need another definition.

29.48 Relative (Subspace) Topology

We start by defining the relative or subspace topology.

Relative (Subspace) Topology. Let $(X, \mathcal{T})$ be a topological space and S a subset of X. The *relative* or *subspace topology* on S is defined by $\mathcal{T}_S = \{S \cap U : U \in \mathcal{T}\}$ and $(S, \mathcal{T}_S)$ is called a *subspace of* X.

The relative topology is the weakest topology where the inclusion map $i_S : S \to X$, defined by $i_S(x) = x$, is continuous. In any topology where it is continuous, the sets $U \cap S = i^{-1}(U)$ must be open. The relative topology $\mathcal{T}_S$ demands exactly that—no more, no less. The relative topology on S makes S into a topological space (in this case a subspace).

When (X, d) is a metric space, the subspace topology is equivalent to metric topology (S, d). That's the point of the subspace topology. It's the topology a set gets by virtue of being a subset of a larger space.

Sets of the form $U \cap S$ with U open in X are referred to as *relatively open* and sets of the form $F \cap S = i^{-1}(F)$ with F closed in X are called *relatively closed*.

29.49 Embedding $\mathbb{R}$ in $\mathbb{R}^2$

An *embedding* of A into B is defined by a one-to-one continuous map $f\colon A \to B$ which has a continuous inverse on $f(A)$. The map is also referred to as an *embedding*.[1]

► Example 29.49.1: Embedding $\mathbb{R}$ in $\mathbb{R}^2$. When $\mathbb{R}^2$ has the usual topology, we can embed $\mathbb{R}$ into $\mathbb{R}^2$ as $R = \{\mathbf{x} \in \mathbb{R}^2 : x_2 = 0\}$. Here the embedding map is $f(x) = (x, 0)$ and $R = f(\mathbb{R})$. The relative topology on R is the same as the usual topology on $\mathbb{R}$. This is illustrated in Figure 29.49.2 which shows that the intersection of R with an open ball in $\mathbb{R}^2$ is an open interval (and vice-versa). This means the two topologies are the same.

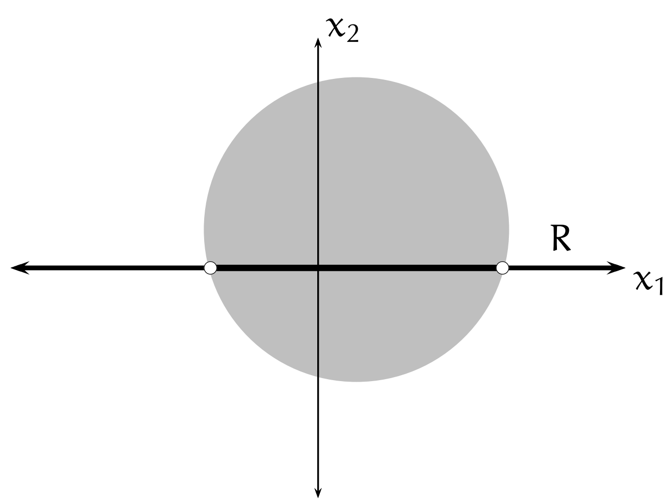


Figure 29.49.2: The intersection of an open ball in $\mathbb{R}^2$ and the line $R = \{\mathbf{x} : x_2 = 0\}$ is an open interval in R. Had we used a closed ball, we would have gotten a closed interval.

◄

[1] Such maps are called *homeomorphisms*. We will learn more about them in section 34.1.

29.50 Relative Complements

The *relative complement* of A in S, denoted $S \setminus A$, is defined by $S \setminus A = \{x \in S : x \notin A\} = S \cap A^c$.

It's easy to see that the relative complement of a relatively open set is relatively closed and vice-versa.

Theorem 29.50.1. *Let $S \subset X$ have the relative topology induced by $(X, \mathcal{T})$. Then A is the relative complement of an open set if and only if $A = S \cap F$ for some set F that is closed in X.*

Proof. Let $A = S \setminus U = S \cap U^c$ for some open set U. Then $A = S \cap F$ where $F = U^c$ is closed.

Conversely, if $A = S \cap F$ for some closed set F, then $A = S \setminus F^c$. Setting $U = F^c$ we get an open set, allowing us to write and $A = S \setminus U$. ∎

29.51 Disconnected Sets

To define a connected set, we first define disconnected sets.

Disconnected Sets. Let $(X, \mathcal{T})$ be a topological space. The space X is *disconnected* if there are two disjoint non-empty open sets U and V that obey $X = U \cup V$. Then (U, V) is called a *disconnection* of X.

Thus $[0, 1) \cup (1, 2]$ is disconnected. Just take $U = [0, 1)$ and $V = (1, 2]$ (both of these sets are relatively open).

The definition of disconnected sets can be recast in terms of closed sets. Theorem 29.51.1 applies regardless of whether we are using the base topology on X, or if we are considering a subset with the relative topology.

Theorem 29.51.1. *A space X is disconnected if and only if there are non-empty disjoint closed subsets A and B that cover X.*

Proof. The set X is disconnected if and only if there are non-empty open sets U and V with $U \cap V = \emptyset$ and $U \cup V = X$. Taking complements, and defining the closed sets $A = U^c$ and $B = V^c$, we find this holds if and only if $A \cup B = U^c \cup V^c = X$ and $A \cap B = \emptyset$. Also there are $u \in U$ and $v \in V$, if and only if $u \notin U^c = A$ and $v \notin V^c = B$. Since the sets cover X, that is equivalent to A and B being non-empty. ∎

29.52 A Disconnected Set

Disconnected sets have two or more pieces. Here's an example of an obviously disconnected set.

► **Example 29.52.1: A Disconnected Set.** A set such as

$$S = [0, 1) \cup (1, 2] \subset \mathbb{R}$$

is not connected. One disconnection is $A = (-1, 1)$ and $B = (1, 3)$. Both sets are open, both have non-empty intersection with S. Here $A \cap S = [0, 1) \neq \emptyset$ and $B \cap S = (1, 2] \neq \emptyset$. The sets A and B cover S as

$$\begin{aligned} A \cup B &= (-1, 1) \cup (1, 3) \\ &\supset [0, 1) \cup (1, 2] \\ &= S. \end{aligned}$$

Finally, the sets A and B are disjoint: $A \cap B = \emptyset$. ◄

29.53 Characterizing Relative Disconnections I SKIPPED

Let S have the relative topology and U, V be relatively open sets that disconnect S. What can we say about the open sets in X that generate them?

Theorem 29.53.1. *Let $(X, \mathcal{T})$ be a topological space and $S \subset X$. Suppose U and V are relatively open sets that disconnect S. Then there are open sets U' and V' in X obeying*

1. *$U' \cap S$ and $V' \cap S$ are both non-empty.*
2. *$S \subset U' \cup V'$.*
3. *$U' \cap V' \cap S = \emptyset$.*

Conversely, if U' and V' are open sets obeying (1)–(3), $U = U' \cap S$ and $V = V' \cap S$ disconnect S in the relative topology.

Proof. By the definition of the relative topology, there are open sets $U', V' \subset X$ with $U = U' \cap S$ and $V = V' \cap S$. (1) Now $U' \cap S$ and $V' \cap S$ are U and V, which are non-empty. (2) Because U and V are disjoint, $U' \cap V' \cap S = \emptyset$. (3) Finally, U and V cover S, so $(U' \cap S) \cup (V' \cap S) = S$. This follows if $S \subset U' \cup V'$.

We now prove the converse. By (1), $U = U' \cap S$ and $V = V' \cap S$ are are non-empty. By (2), U and V are disjoint. By (3), their union is all of S. Since U and V are also relatively open, they disconnect S. ∎

Notice that it need not be the case that $U' \cap V'$ is empty, only that it contain no points in S.

29.54 Characterizing Relative Disconnections II SKIPPED

There's a similar result for closed sets. The proof is quite similar, and has been omitted.

Theorem 29.54.1. *Let* $(X, \mathcal{T})$ *be a topological space and* $S \subset X$*. Suppose* A *and* B *are relatively closed sets that disconnect* S*. Then there are closed sets* A' *and* B' *in* X *obeying*

1. $A' \cap S$ *and* $B' \cap S$ *are both non-empty.*
2. $S \subset A' \cup B'$.
3. $A' \cap B' \cap S = \emptyset$.

Conversely, if A' *and* B' *are closed sets obeying (1)–(3),* $A = A' \cap S$ *and* $B = B' \cap S$ *disconnect* S *in the relative topology.*

29.55 Connected Sets

Intuitively, a set is connected if there are no breaks in the set, if it is all one piece. That's not how we define connected sets. We define them negatively. Rather than using any positive attributes of the set, connected sets are defined by what they don't have.

A connected set is a set that is not disconnected.

Connected Sets. Let $(X, \mathcal{T})$ be a topological space. The space X is *connected* if if cannot be disconnected.

A subset S of X is *connected* it if is connected in the relative topology.

29.56 Totally Disconnected Sets

How bad can disconnections get? Totally disconnected sets are the opposite extreme from connected sets. They contain no connected subsets bigger than a single point. A set is *totally disconnected* if its only connected subsets are the trivial ones—singletons and the empty set. Equivalently, a set is totally disconnected if you can disconnect any two distinct points in the set.

▶ **Example 29.56.1: The Rationals are Totally Disconnected.** The rational numbers $\mathbb{Q}$ provide an example of total disconnection. Suppose a is an irrational number (e.g., $\sqrt{2}$, π). Then $A = (-\infty, a] \cap \mathbb{Q}$ and $B = \mathbb{Q} \cap [a, +\infty)$ are relatively closed sets in $\mathbb{Q}$ that disconnect $\mathbb{Q}$. Moreover, you can disconnect any subset of $\mathbb{Q}$ that contains at least two distinct points in the same fashion. That means that the set $\mathbb{Q} \subset \mathbb{R}$ is totally disconnected. ◀

SKIPPED

Another totally disconnected set is the Cantor set of section 12.40.

▶ **Example 29.56.2: The Cantor Set is Totally Disconnected.** Suppose $x < y$ are distinct elements of the Cantor set. As we saw, both can be written as ternary numbers consisting entirely of 0's and 2's. Take the first digit that differs, call it digit k. Digit k is 0 in x, 2 in y. Let z have ternary expansion identical to x, except that the k^{th} digit is 1. Then $x < z$ and $y > z$. Let $A = \mathfrak{C} \cap (-\infty, z]$ and $B = \mathfrak{C} \cap [z, +\infty)$. Then A and B are closed sets that disconnect $\mathfrak{C}$. ◀

29.57 Connected Components of Disconnected Sets

Let $(X, \mathcal{T})$ be a topological space that is not connected. If a subspace S is connected, we can ask it if is there are any larger connected subspaces that include it. If there are not, then we call S a *connected component* or just a *component* of X.

Given a connected subset S, a maximal connected subset containing S can be constructed by taking the union of all connected subsets of X that contain S. The key step in showing this is that the union of two connected subsets containing S is also connected. In fact, we can show more than we need. If two connected subsets share even a single point, their union is connected.

Theorem 29.57.1. *Suppose* A *and* B *are connected subsets of* $(X, \mathcal{T})$ *that both contain the point* x*. Then* $A \cup B$ *is connected.*

Proof. **Suppose** that $A \cup B$ **can be disconnected** by relatively open sets U and V. Label the sets U and V so $x \in U$. Keep in mind that $x \in A \cap B$ and take $y \in V \cap (A \cup B)$. There are two possibilities: $y \in A$ or $y \in B$.

Suppose $y \in A$. But then both $A \cap U$ and $A \cap V$ are non-empty, so U and V disconnect A, **which is impossible**.

Otherwise, $y \in B$. Now $x \in U$ and $y \in V$, so both $B \cap U$ and $B \cap V$ are non-empty. Then U and V disconnect the connected set B, **which is also impossible**.

As both possibilities are impossible, we cannot disconnect $A \cup B$. ∎

Earlier we considered the space $[0, 1) \cup (1, 2]$. It has two connected components, $[0, 1)$ and $(1, 2]$. Both sets are connected because they are intervals, but because they share no points in common, their union is disconnected.

29.58 Intervals are Connected

In contrast to examples 29.52.1 and 29.56.1, any interval in the real line is a connected set.

Proposition 29.58.1. *Any interval* I *in* $\mathbb{R}$ *is connected.*

Proof. Let I be an interval, possibly infinite. We prove this by contradiction. **Suppose** I **is not connected**, then there are non-empty relatively closed sets A and B that disconnect I. Take $a \in A$ and $b \in B$. We label the sets so that $a < b$ and consider the interval $[a, b]$. Note that $[a, b] \subset I$ because I is an interval and $a, b \in I$.

Define $z = \sup([a, b] \cap A)$. The set $[a, b] \cap A$ is non-empty as $a \in A$. Since the set is bounded, the supremum will exist. By construction, $a \leq z \leq b$, so $z \in [a, b] \subset I$.

Because A and $[a, b]$ are relatively closed,

$$z \in \operatorname{cl}\left([a, b] \cap A\right) = [a, b] \cap A \subset A.$$

Of course $z \neq b \in B$ because $A \cap B$ is empty.

Since z is the upper bound of $A \cap [a, b]$, any $w \in (z, b]$ must be in B. In particular, $z + \frac{1}{n} \in B$ for n large enough.

Letting $n \to \infty$ and using the fact that B is relatively closed, we find $z \in B$. But $z \in A$ so it cannot also be in B. **This contradiction** means that there are **no sets** A **and** B **that disconnect** I. Therefore, I is connected. ∎

29.59 Continuity and Connectedness

The continuous image of a connected set is connected. This means that we can make connected sets from other connected sets by applying continuous functions to them.

Theorem 29.59.1. *Suppose* $f: X \to Y$ *is continuous and* $S \subset X$ *is connected. Then* $f(S)$ *is connected.*

Proof. **Suppose not**. Let A, B be relatively closed sets that disconnect $f(S)$. Now $f^{-1}(A)$ and $f^{-1}(B)$ are relatively closed because f is continuous. Moreover,

$$f^{-1}(A) \cap f^{-1}(B) = f^{-1}(A \cap B) = \emptyset$$

and

$$f^{-1}(A) \cup f^{-1}(B) = f^{-1}(A \cup B) \supset f^{-1}\big(f(S)\big) = S.$$

Thus $f^{-1}(A)$ and $f^{-1}(B)$ disconnect S. **As this is impossible**, $f(S)$ cannot be disconnected. ∎

29.60 The Intermediate Value Theorem

An important consequence of Theorem 29.59.1 is the Intermediate Value Theorem, which says that if a continuous function defined on an interval takes two values, it also takes all values in between.[2]

Intermediate Value Theorem. *If* $f\colon [a, b] \to \mathbb{R}$ *is continuous, and* y *is a number between* $f(a)$ *and* $f(b)$*, then there is at least one* $c \in [a, b]$ *with* $f(c) = y$.

Proof. We may assume $f(a) < f(b)$ without loss of generality. We proceed by contradiction. If no such c exists, $(-\infty, y]$ and $[y, +\infty)$ disconnect $f([a, b])$. This is impossible by Proposition 29.59.1, and so such a c must exist. ∎

[2] The first proof of a form of the Intermediate Value Theorem was by Bernard Bolzano in 1817. Cauchy proved something close to this version in 1821. Amazingly, the idea of the theorem was already used by the Greek mathematician and philosopher Bryson of Heraclea in the 5^{th} century BC. Bryson may have been a student of Socrates. He was from Heraclea Pontica, a Megaran colony on the southern Black Sea coast. The modern Turkish city of Karadeniz Ereğli (which means Heraclea Pontica) is at the same location. I'm not sure if the city has been continuously occupied since ancient times.

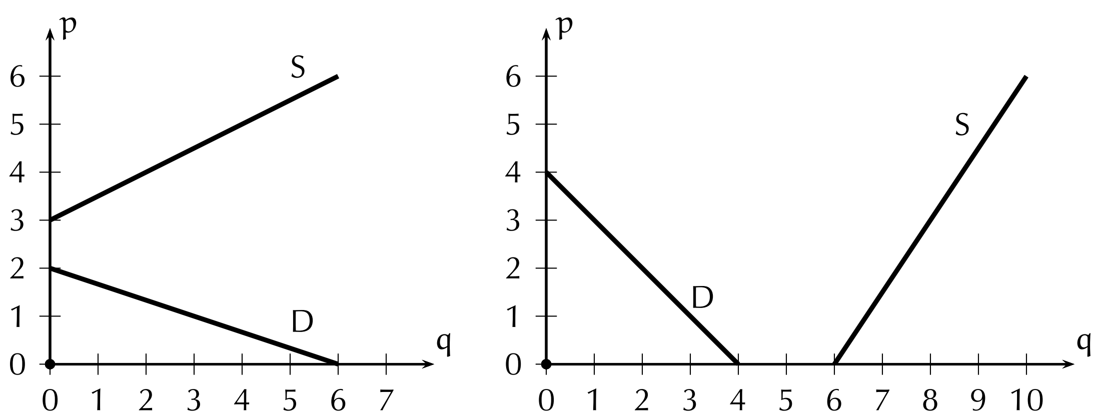
29.61 Market Equilibrium

The Intermediate Value Theorem can be used to show that a market equilibrium exists for a large class of supply and demand models.

Suppose that both supply $S(p)$ and demand $D(p)$ are continuous functions of price p and there are $p_0 < p_1$ with $S(p_0) - D(p_0) < 0$ (excess demand at a low price) and $S(p_1) - D(p_1) > 0$ (excess supply at a high price).

By the Intermediate Value Theorem, there is a p^* obeying $p_0 < p^* < p_1$ with $S(p^*) = D(p^*)$. In other words, a market equilibrium exists.

Here are two cases where there is no equilibrium. In neither are both the p_0 and p_1 conditions satisfied.



29.62 Path-Connected Sets

It is often easiest to show a set is connected by showing it is path-connected, rather than attempting to use the definition of connected.

A *path* from a to b in S is a continuous function $f\colon [0,1] \to S$ with $f(0) = a$ and $f(1) = b$. As the continuous image of a connected set, any path $f\big([0,1]\big)$ is itself connected (Theorem 29.59.1).

Path-connected. A set S is *path-connected* if for every $a, b \in S$ there is a path from a to b in S.

Paths include all sorts of curves as well as straight lines. Path-connected sets are connected. It is sometimes easier to show a set is path-connected rather than directly showing it is connected.

Proposition 29.62.1. *Any path-connected set is connected.*

Proof. Suppose S is path-connected and A, B are closed sets with $S \subset A \cup B$. Choose $a \in A \cap S$ and $b \in B \cap S$ and let f be a path between them in S.

Now consider $f^{-1}(A)$ and $f^{-1}(B)$. These are closed sets in $[0,1]$ with $0 \in f^{-1}(A)$ and $1 \in f^{-1}(B)$. Moreover, $f^{-1}(A) \cup f^{-1}(B) = f^{-1}(A \cup B) \supset [0,1]$. Since $[0,1]$ is connected, there is some $x \in f^{-1}(A) \cap f^{-1}(B)$. It follows that $f(x) \in A \cap B$ and that A and B cannot disconnect S. Thus S is connected. ∎

In some spaces the connected and path-connected sets are the same. This happens in $\mathbb{R}$, where the only connected sets are the intervals. This includes trivial intervals such as $[a, a]$ and infinite intervals like $(0, +\infty)$. Intervals are always path-connected, meaning that the set of connected subsets and the set of path-connected subsets are identical in $\mathbb{R}$.

29.63 Convex and Star-shaped Sets

A convex set is one that contains the line segment between any pair of its points.

Convex Set. Let V be a vector space. A set $S \subset V$ is *convex* if for every $\mathbf{x}, \mathbf{y} \in S$, $\ell(\mathbf{x}, \mathbf{y}) = \{(1-t)\mathbf{x} + t\mathbf{y} : 0 \leq t \leq 1\} \subset S$.

A star-shaped set is a set where all points can be connected to a special point, the star point, using a line segment.

Star-shaped Set. A set $S \subset \mathbb{R}^m$ is *star-shaped* if there is a point $\mathbf{x}_0 \in S$ so that for any $\mathbf{x}$, $\ell(\mathbf{x}_0, \mathbf{x}) \subset S$. We then say S is star-shaped with respect to $\mathbf{x}_0$. The point $\mathbf{x}_0$ is called a *star point*.

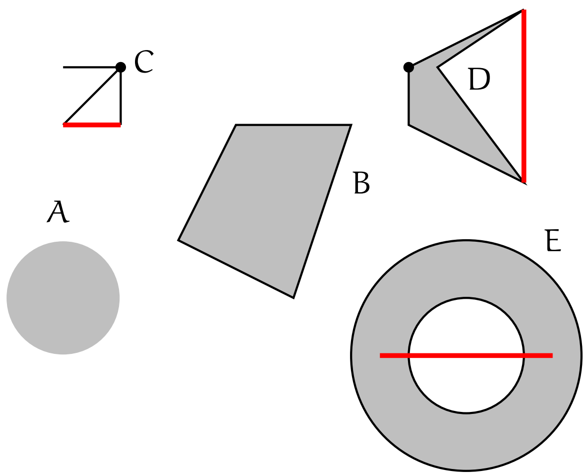


Figure 29.63.1: The sets A and B are convex. The sets C and D are not convex, as demonstrated by the red line segments that must leave the set to connect points. The sets C and D are star-shaped with respect to the heavy dots. The annulus E is neither convex nor star-shaped, although it is connected.

29.64 Convex and Star-shaped Sets are Connected

It's easy to show that convex sets are star-shaped.

Proposition 29.64.1. *Any convex set S is star-shaped and every point of it is a star point.*

Proof. Let x_0 be any point in S. Since S is convex, $\ell(x_0, y) \subset S$ for any $y \in S$, showing that S is star-shaped and that any point x_0 is a star point for S. ∎

An immediate corollary of the definitions together with Proposition 29.62.1 is that convex sets are connected.

Corollary 29.64.2. *Any convex set is connected.*

Proof. If S is convex and $x, y \in S$, the function $f(t) = (1 - t)x + ty$ is a path from x to y in S.

Star-shaped sets are also connected.

Corollary 29.64.3. *Any star-shaped set is connected.*

Proof. Suppose $x_0 \in S$ is a star point and $x, y \in S$. Since x_0 is a star point, there are paths $f_1 : [0, 1] \to S$ with $f_(0) = x_0$ and $f_1(1) = x$ and f_2 with $f_2(0) = x_0$ and $f_2(1) = y$.

Define the function f(t) by

$$f(t) = \begin{cases} f(t) = f_1(1 - 2t) & \text{for } 0 \le t \le \frac{1}{2} \\ f(t) = f_2(2t - 1) & \text{for } \frac{1}{2} \le t \le 1 \end{cases}$$

When $t = \frac{1}{2}$, both functions take the value x_0, so I can include in both intervals.

Now $f(0) = f_1(1) = x$ and $f(1) = f_2(1) = y$, so f is a path that runs between x and y by going through the star point x_0 when $t = \frac{1}{2}$. This shows S is path-connected.

29.65 Connected Sets Need Not Be Path-Connected

Although path-connected sets are connected, the converse is not always true. A set can be connected without being path-connected. Although such sets are all in one piece, parts of the set may be so far apart via the set, that you can't use a path to get there. After all, paths must be compact. This happens with the topologist's sine curve.

► Example 29.65.1: **Topologist's Sine Curve.** Let $S = \{(0, y) : -1 \le y \le 1\} \cup \{(x, \sin(1/x) : x > 0\}$. This is a variant of the topologist's sine curve of Example 13.18.1.

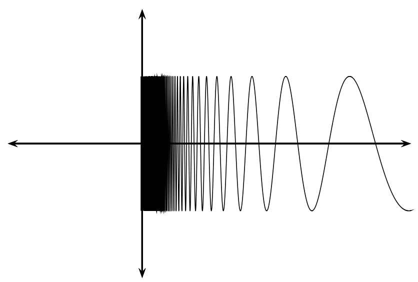


Figure 29.65.2: Topologist's Sine Curve

The set S is connected but not path-connected. **Suppose**, by way of contradiction, **that** S **is path-connected**. Then there is a continuous path in S, $f(t) = \big(x(t), y(t)\big)$, with $f(0) = (0, 0)$ and $f(1) = (1/\pi, 0)$. The components of f, x and y, inherit continuity from f.

By the Intermediate Value Theorem, there is a t_1, $0 < t_1 < 1$ with $x(t_1) = 2/3\pi$. Then there is a t_2, $0 < t_2 < t_1$ with $x(t_2) = 2/5\pi$. Continuing this process we obtain a decreasing sequence $\{t_n\}$ with $t_n \to 0$ and $x(t_n) = 2/(2n + 1)\pi$. Since t_n is a decreasing sequence that is bounded below, it converges to some t_0.

By continuity, $y(t_n) \to y(t_0)$. But this is impossible because $y(t_n) = +1$ when n is even and -1 when n is odd. **This contradiction shows there is no such path** f. The set S cannot be path connected. ◄

29.66 A Simple Fixed Point Theorem

A function $f\colon X \to X$ has a *fixed point* if there is an $x^* \in X$ with $f(x^*) = x^*$. The Intermediate Value Theorem can be used to show that any continuous function mapping the unit interval $[0, 1]$ to itself has a fixed point.

Theorem 29.66.1. *Let* $f\colon [0, 1] \to [0, 1]$ *be a continuous function. Then there is* $x^* \in [0, 1]$ *with* $f(x^*) = x^*$.

Proof. If $f(0) = 0$ or $f(1) = 1$, we are done.

Otherwise, define $g(x) = f(x) - x$. Then g is continuous with $g(0) > 0$ and $g(1) < 0$. By the Intermediate Value Theorem, there is a x^*, $0 < x^* < 1$, with $g(x^*) = 0$. But then $f(x^*) = x^*$ and we are done. ∎

29.67 Contraction Mappings

Connectedness is not the only way to prove a fixed point theorem. Banach's Contraction Mapping Theorem is based on completeness, not connectedness.

This powerful theorem can be used to prove the Inverse and Implicit Function Theorems of S&B Chapter 15. It can be used to show the existence of solutions to differential equations. It has economic applications, including finding solutions to Bellman's dynamic programming equation and to recursive utility for dynamic models.

Before we can state it, we must define contractions.

Contraction. Let (X, d) be a metric space. A function $f\colon X \to X$ is called a *contraction* if there is an $r < 1$ with $d\big(f(x), f(y)\big) \le r\, d(x, y)$ for every $x, y \in X$.

In other words, a mapping is a contraction if it the images of any two points are uniformly closer together than the original two points.

29.68 Contraction Mapping Theorem I

Contraction Mapping Theorem. *Let* f *be a contraction on a complete metric space* (X, d). *Then* f *has a unique fixed point* x^*. *Moreover, if we take any* $x_0 \in X$ *and define* $x_n = f(x_{n-1})$ *for* $n = 1, 2, \dots$. *Then* $x_n \to x^*$.

Proof. First we show uniqueness. Suppose there are two fixed points x^* and y^*. Then

$$d(x^*, y^*) = d\big(f(x^*), f(y^*)\big) \le r\, d(x^*, y^*).$$

As $r < 1$, this implies $d(x^*, y^*) = 0$, so $x^* = y^*$.

Consider the sequence $\{x_n\}$ given in the statement of the theorem. I claim that

$$d(x_{n+1}, x_n) \le r^n d(x_1, x_0). \qquad (29.68.1)$$

We prove equation (29.68.1) by induction. When $n = 1$ it follows because

$$d(x_2, x_1) = d\big(f(x_1), f(x_0)\big) \le r\, d(x_1, x_0).$$

Also, if equation (29.68.1) is true for n,

$$d(x_{n+2}, x_{n+1}) = d\big(f(x_{n+1}), f(x_n)\big) \le r\, d(x_{n+1}, x_n) \le r^{n+1} d(x_1, x_0)$$

showing equation (29.68.1) is true for $n + 1$. By induction it must be true for all $n = 1, 2, \dots$.

Proof continues . . .

29.69 Contraction Mapping Theorem II

Remainder of Proof. Suppose $m > n$. Then

$$\begin{aligned} d(x_m, x_n) &\le \sum_{i=0}^{m-n-1} d(x_{n+i+1}, x_{n+i}) \\ &\le \sum_{i=0}^{m-n-1} r^{n+i} d(x_1, x_0) \\ &= \frac{r^n}{1-r} d(x_1, x_0). \end{aligned}$$

Since the right hand side converges to zero as $n \to \infty$, we conclude $\{x_n\}$ is a Cauchy sequence.

By completeness of X the sequence has a limit x^*. Now f is continuous, so

$$\begin{aligned} f(x^*) &= f\big(\lim_n x_n\big) \\ &= \lim_n f(x_n) \\ &= \lim x_{n+1} = x^* \end{aligned}$$

showing that x^* is a fixed point of f. We established uniqueness at the beginning, so it is the unique fixed point of f. ∎

34. Topological Equivalence

34.1 Homeomorphisms

Homeomorphism is a key concept in topology. It's an isomorphism for topological spaces. It sets up an equivalence relation between topological spaces. A homeomorphism is a continuous map with a continuous inverse. More precisely,

Homeomorphism. Let X and Y be topological spaces. A bijective map $\varphi : X \to Y$ is a *homeomorphism* between X and Y if both φ and φ^{-1} are continuous. Where there is a homeomorphism between X and Y, we say X and Y are *homeomorphic*.

Homeomorphisms are one of the fundamental concepts of topology. In a sense, homeomorphisms are the defining concept. Homeomorphisms preserve topological properties such as openness, closedness, connectedness, compactness, and a number of others I haven't mentioned. If a property is not preserved by some homeomorphism, it isn't considered a topological property.

Homeomorphisms can be composed. If $f : X \to Y$ is a homeomorphism and $g : Y \to Z$ is a homeomorphism, then $g \circ f : X \to Z$ is also a homeomorphism. Together with the fact that the identity map $\mathrm{id}(x) = x$ is a homeomorphism from X to X, it shows that being homeomorphic is an equivalence relation.

That is, being homeomorphic, is reflexive (X is homeomorphic to itself by the identity map id), symmetric (by definition, X is homeomorphic to Y if and only if Y is homeomorphic to X), and transitive (if X is homeomorphic to Y and Y homeomorphic to Z, then X is homeomorphic to Z by composition).

34.2 Why Homeomorphisms?

Homeomorphisms have the important property that image of an open set is open.

Theorem 34.2.1. *Suppose* $\varphi : X \to Y$ *is a homeomorphism and* $U \subset X$ *is a open set. Then* $\varphi(U)$ *is also open.*

Proof. By hypothesis, $\psi = \varphi^{-1}$ is continuous. Now $\varphi(U) = \psi^{-1}(U)$, which is open as the continuous inverse image of an open set, U. ∎

As a consequence, a set is open in X if and only if its image is open in Y.

Corollary 34.2.2. *Suppose* $\varphi : X \to Y$ *is a homeomorphism. Then* $T \subset X$ *is open in* X *if and only if* $\varphi(T)$ *is open in* Y.

Proof. The "only if" part is Theorem 34.2.1. As for the "if" part, if $\varphi(T)$ is open in Y, then its inverse image, $T = \varphi^{-1}\big(\varphi(T)\big)$ is open in X. ∎

One consequence is that the Y-topology, the collection of open sets in Y, is the image of the X-topology. Homeomorphisms map topologies to topologies. That's the key to preserving topological properties.

We consider two topological spaces to be *equivalent* if they are homeomorphic.

34.3 Two Examples of Homeomorphism

Here are two homeomorphism. It is fairly easy to verify that all are bijective continuous mappings with continuous inverses.

► **Example 34.3.1: Stretching and Compressing $\mathbb{R}^m$.** A simple example of a homeomorphism between $\mathbb{R}^m$ and $\mathbb{R}^m$ is $f(\mathbf{x}) = 50\mathbf{x}$. This function stretches $\mathbb{R}^m$. The images of two points are always 50 times as far apart as the points themselves.

Its inverse is a contraction, compressing distances by a factor of 50. The inverse is defined by $f^{-1}(\mathbf{x}) = (1/50)\mathbf{x}$. Then

$$\left\|f^{-1}(\mathbf{x}) - f^{-1}(\mathbf{y})\right\|_2 = \frac{1}{50}\left\|\mathbf{x} - \mathbf{y}\right\|_2$$

Both f and f^{-1} map $\mathbb{R}^m$ to $\mathbb{R}^m$. ◄

► **Example 34.3.2: Graph of a Parabola.** Another homeomorphism maps $\mathbb{R}$ onto the parabola $S = \{(x, x^2) : x \in \mathbb{R}\}$, where the parabola S is given its relative topology in $\mathbb{R}^2$. The map $f\colon \mathbb{R} \to S$ is defined by $f(x) = (x, x^2)$ and its inverse $f^{-1}: S \to \mathbb{R}$ is given by projection onto the first coordinate: $f^{-1}(x, x^2) = x$. ◄

34.4 A Homeomorphism between $(-1, 1)$ and $\mathbb{R}$

The real line is homeomorphic to the open interval $(-1, 1)$.

► Example 34.4.1: Homeomorphism between a $(-1, 1)$ and the Real Line. This homeomorphism, *gnomonic projection*, maps the interval $(-1, +1)$ onto the real line. The term gnomonic refers to a sundial's gnomon.

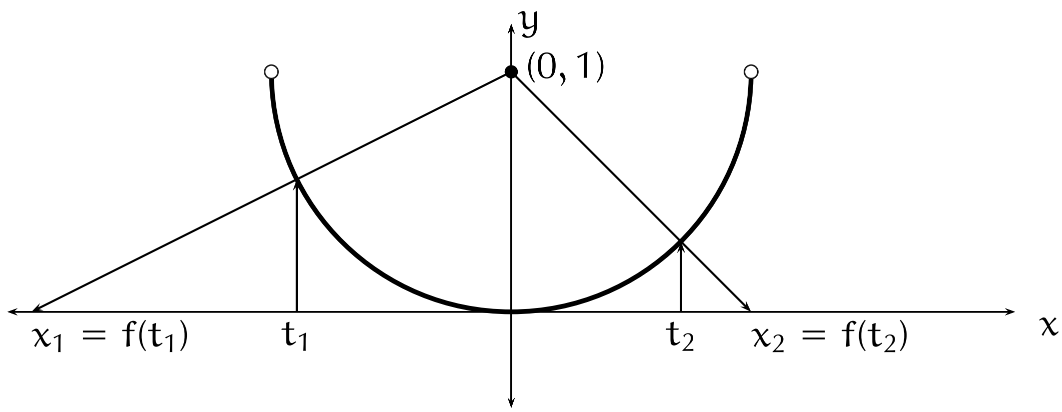


Figure 34.4.2: The diagram shows how the mapping f works. We start with a point $t \in (-1, 1)$, map it straight up to the semicircle of radius one centered at $(0, 1)$, then project the line though $(0, 1)$ and the resulting point back to the x-axis to obtain $x = f(t)$. Two samples are shown, from $t_1 = -2/5\sqrt{5}$ and $t_2 = +1/\sqrt{2}$.

The semicircle obeys $t^2 + (y - 1)^2 = 1$. Since we are taking the lower part, we must use the negative square root to compute y. Thus $y = 1 - \sqrt{1 - t^2}$. We use the point on the semicircle to compute the slope of the line through it and $(0, 1)$, which is $-\Delta y/\Delta t = (y - 1)/t$ for $t \neq 0$. The line is has equation $y' = 1 + (y - 1)x/t$ for $t \neq 0$. Its horizontal intercept is at $t/(1 - y) = t/\sqrt{1 - t^2}$, which is the value we are looking for. Notice that this formula works for $t = 0$ as well.

Summing up, the resulting function and its inverse are

$$f(t) = \frac{t}{\sqrt{1 - t^2}} \quad \text{and} \quad f^{-1}(x) = \frac{x}{\sqrt{1 + x^2}}.$$

The function f maps $(-1, +1)$ onto $\mathbb{R}$. The inverse maps $\mathbb{R}$ to $(-1, +1)$. ◄

34.5 Gnomonic Projection

Gnomonic projection also works if there are more dimensions. We can use it to show that the open unit disk in $\mathbb{R}^2$, $B(\mathbf{0}, 1)$, is homeomorphic to the lower open hemisphere in $\mathbb{R}^3$,[1]

$$S = \left\{(x, y, z) : x^2 + y^2 + (z-1)^2 = 1, z < 1\right\}.$$

Again, we take the line defined by $(0, 0, 1)$ and any point $(a, b, c) \in S$. Its intersection with the xy-plane has coordinates

$$\left(\frac{a}{1-c}, \frac{b}{1-c}\right), c = 1 - \sqrt{1 - a^2 - b^2}.$$

The inverse map is

$$(u, v) \mapsto \left(\frac{u}{\sqrt{1 + u^2 + v^2}}, \frac{v}{\sqrt{1 + u^2 + v^2}}, 1 - \frac{1}{\sqrt{1 + u^2 + v^2}}\right).$$

This generalizes to any number of dimensions. The formula is slightly different than in Example 34.4.1 because we only followed the x-coordinate there.

[1] We use the term "disk" to emphasize that this is in $\mathbb{R}^2$.

34.6 Characterizing Topological Spaces

Besides preserving the topologies, homeomorphisms preserve other topological features such as compactness and connectedness. The latter plays an important role in classifying topological spaces. For one, homeomorphic topological spaces must have the same number of (connected) components. In fact, counting the components is one of the first steps in classifying topological spaces.

Normally this is done using algebraic topology and concepts such as homotopy, homology, and cohomology. But a lot comes down to counting the connected components and seeing how they change if you change the space a bit.

34.7 Unit Interval and Unit Disk are Not Homeomorphic

11/01/22

The unit interval $(0, 1) \subset \mathbb{R}$ and unit disk in $\mathbb{R}^2$ are not homeomorphic. Suppose f is a homeomorphism between the open unit interval $I = (0, 1)$ in $\mathbb{R}$ and the unit disk in $\mathbb{R}^2$, $B = B(\mathbf{0}, 1)$. Then $f(I) = B$.

Take any a with $0 < a < 1$ and consider $I \setminus \{a\} = (0, a) \cup (a, 1)$. The image under f is all of B except the single point $\{f(a)\}$. In fact, the function f is still a bicontinuous bijection between $I \setminus \{a\}$ and $B(0, 1) \setminus \{f(a)\}$.

However, $f\big((0, 1) \setminus \{a\}\big) \subset B(0, 1)$ is the disk with a single point removed—a connected space. It cannot be homeomorphic to the **disconnected** space $(0, a) \cup (a, 1)$. This contradiction shows that unit interval and unit disk are not homeomorphic.

The same basic argument shows that the unit interval is not homeomorphic to any open ball in any $\mathbb{R}^m$ for $m > 1$. We can do the same thing with lines to show that open balls in $\mathbb{R}^2$ are not homeomorphic to open balls in $\mathbb{R}^3$, etc.

Figure 34.7.1: In the left panel, f maps $I = (0, 1)$ onto an open unit ball with diameter 1. In the right panel, we removed a single point $a = 0.3$ from the interval, which is now mapped to the disk minus a single point, $f(0.3)$. The sets in the right panel are **not** homeomorphic because we have disconnected the interval, but its image, the disk minus a point, remains connected.

34.8 Topological Dimension

Let's continue that line of thought.

So far, we've thought of dimension solely in algebraic terms, defining it by the size of the basis of a vector space. In that case, the number of elements in any basis defined the dimension. As we have just seen by considering the possibility of homeomorphisms between intervals and disks, dimension apparently also has a topological aspect. The disk remains connected when a single point is removed, while the interval is disconnected by the removal.

There is an important theorem in topology due to L.E.J. Brouwer (1881–1966) that relates topological dimension and homeomorphism. We state it without proof.[2]

Invariance of Domain. *Let* U *be an open set in* $\mathbb{R}^m$ *and* $f\colon U \to \mathbb{R}^m$ *be continuous and one-to-one. Then* $V = f(U)$ *is open in* $\mathbb{R}^m$ *and* f *is a homeomorphism between* U *and* V.

The theorem often concludes by merely saying that $f(U)$ is open, rather than f is a homeomorphism. Our version follows immediately from that one.

[2] L.E.J. Brouwer (1912), Beweis der Invarianz des n-dimensionalen Gebiets, *Mathematische Annalen* **71**, 305–315; see also **72**, 55–56. This can found in many textbooks. See the reference list in the Wikipedia page for Invariance of Domain, which includes a link to a blog post by Terrance Tao. Maurice Fréchet was also doing work on a similar concept at about the same time, although his approach ended up being less fruitful than Brouwer's.

L.E.J. Brouwer (1881–1966) was a Dutch mathematican and philosopher. Brouwer was one of the great mathematicians of the early 20^{th} century. He worked mainly in topology, set theory, measure theory and complex analysis. He is considered the founder of modern topology, and is particularly known for his Fixed Point Theorem and Invariance of Domain. He was also a proponent of mathematical intuitionism, a constructivist philosophy that considers mathematics a construct of the human mind rather than objective truth. In Brouwer's case, he insisted that mathematical concepts be based on sensory intuitions, with limited on the use of logical reasoning. In particular, he thought that the law of excluded middle cannot be "applied without reservation even in the mathematics of infinite systems". So much for proofs by contradiction.

34.9 Different $\mathbb{R}^m$'s are Not Homeomorphic

As a result, it's not possible to have a homeomorphism between $\mathbb{R}^k$ and $\mathbb{R}^m$ for $k \neq m$. Although we don't show it, it's not even possible to have homeomorphisms between open sets in $\mathbb{R}^k$ and $\mathbb{R}^m$ unless $k = m$.

Proposition 34.9.1. *If* $f\colon \mathbb{R}^k \to \mathbb{R}^m$ *is a homeomorphism, then* $k = m$.

Proof. **Suppose, by way of contradiction, that** $k > m$. We can regard $\mathbb{R}^m$ as the first m coordinates in $\mathbb{R}^k$. Define $\hat{f}$ by $\hat{f}(\mathbf{x}) = \big(f_1(\mathbf{x}), \dots, f_m(\mathbf{x}), 0, \dots, 0\big)$. Then $\hat{f}\colon \mathbb{R}^k \to \mathbb{R}^k$. Now $\hat{f}$ is one-to-one, so $\hat{f}(\mathbb{R}^m)$ must be open by Invariance of Domain. But it isn't because the last $k - m$ coordinates are zero. **This contradiction** shows $k \leq m$.

If $k < m$, consider its inverse and apply the above argument, which again **leads to contradiction**.

It follows that $k = m$. ∎

This shows that both topology and linear algebra agree that $\mathbb{R}^m$ is m-dimensional. The fact that topological and geometrical (vector) dimension agree is evidence of a deep relationship between them.

December 6, 2022